HAL
open science

Un nouveau regard cladistique sur l'anatomie comparée, la phylogénie, la systématique et la paléoécologie des rudistes (Bivalvia, Hippuritida)

Valentin Rineau

To cite this version:

Valentin Rineau. Un nouveau regard cladistique sur l'anatomie comparée, la phylogénie, la systématique et la paléoécologie des rudistes (Bivalvia, Hippuritida). Paléontologie. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066503 . tel-01814277

HAL Id: tel-01814277

https://theses.hal.science/tel-01814277

Submitted on 13 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

Ecole doctorale 398 – Géosciences et ressources naturelles

UMR 7207 - Centre de recherche sur la paléobiodiversité et les paléoenvironnements

Un nouveau regard cladistique sur l'anatomie comparée, la phylogénie, la systématique et la paléoécologie des rudistes (Bivalvia, Hippuritida)

Par Valentin Rineau

Thèse de doctorat de Paléontologie

Dirigée par Loïc Villier

Présentée et soutenue publiquement le 29 novembre 2017

Devant un jury composé de :

Mr. Jorge Cubo, professeur, Université Pierre et Marie Curie – Examinateur

Mr. Loïc Villier, professeur, Université Pierre et Marie Curie – Directeur de thèse

Mr. Pascal Neige, professeur, Université Bourgogne Franche-Comté – Rapporteur

Mr. David Williams, researcher, Natural History Museum, London – Rapporteur

Mr. Jean Philip, professeur émérite, Université d'Aix-Marseille – Examinateur

Mr. René Zaragüeta i Bagils, maître de conférences, Université Pierre et Marie Curie – Examinateur

Mr. Jean-Pierre Masse, directeur de recherche émérite, Université d'Aix-Marseille – Invité

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

À ma mère, mon grand-père, mon arrière-grand-mère

Remerciements

Dans quelques heures, je rendrai ce manuscrit, et trois ans recevront d'un coup d'un seul leur point final. C'est une sensation étrange que de voir se terminer une partie de sa vie aussi marquante qu'une thèse. On se rappelle des souvenirs du début avec l'arrivée au laboratoire, des moments marquants; bref, une rétrospective mentale s'enclenche, et les remerciements de tous ceux qui ont compté à un moment ou un autre arrivent…

Mes premiers remerciements vont à Loïc Villier, pour m'avoir fait confiance il y a trois ans, pour avoir été disponible à chaque fois, pour la somme de connaissances et la curiosité qui m'ont permis de découvrir de nouveaux domaines scientifiques et d'y prendre goût, et surtout pour le bon goût musical. Merci également à Jean-Pierre Masse pour avoir partagé ses connaissances sur les rudistes avec moi lors de mes voyages à Marseille, à Marc Floquet pour avoir pu travailler avec lui, pour sa rigueur, pour son aide et sa disponibilité, et pour l'étendue de ses connaissances dont on commence à prendre conscience au cours d'un trajet en voiture, ainsi qu'à René Zaragüeta pour m'avoir permis de rentrer dans le monde de la Cladistique si passionnant, et pour ce souci de la cohérence théorique qui fait de lui un grand scientifique. J'ai eu beaucoup de chance d'avoir pu côtoyer tous ces gens, et tous m'auront beaucoup apporté dans des domaines très divers.

À Paul Zaharias, pour être un des moteurs principaux de la génération de nouvelles idées et de nouveaux projets, pour le pouvoir heuristique du langage, pour la retraite scientifique dans les Landes et pour son amitié. À Stéphane Prin pour les discussions et les nombreuses corrections, pour l'exactitude, la rigueur et le talent. Au « cercle de Vienne » du bureau de René pour l'ensemble des discussions passionnantes. À toutes les personnes avec qui j'ai pu travailler : Michel Laurin, Estelle Bourdon, Mathieu Faure-Brac et bien d'autres.

À Jérémie Bardin, pour toutes ces émotions fortes durant trois ans, et pour la science.

À toutes les personnes qui ont participé au projet de cours de 3ia : Paul, Malcolm, Laetitia ainsi que Lou, Morgane, Artémis, Benjamin, Rémi, Charlène…

À Delphine Desmares pour l'identification des foraminifères planctoniques et à Marc Testé pour les pithonelles.

À tous les habitants du couloir entre les tours 46 et 56, pour l'ambiance de travail et la convivialité – en particulier celles et ceux qui, à la fin, étaient là le soir et les week-ends: Bérengère, Elise, Raquel, Isabelle et toutes et tous les autres…

À Didier Merle et Jean-Michel Pacaud qui m'ont permis d'accéder aux collections du MNHN, à Isabelle Rouget et Stéphane Jouve pour celles de l'UPMC, Axel Arnoux et Jean Philip pour celles de l'université Aix-Marseille, Emmanuel Robert pour l'Université Lyon 1, à ceux qui m'ont accueilli aux Etats-Unis pour mettre la main sur les rudistes du nouveau monde: Susan Butts à Yale, Bushra Hussaini au AMNH de New york, Mark Florence au Smithsonian NMNH de Washington et Ann Molyneux à l'Université d'Austin. À Nicolas Morel pour le prêt de nombreux spécimens Manceaux. À Frederic Delbes pour les lames minces. À tous ceux qui ont participé à l'illustration de la thèse : Sophie Fernandez, Justine Jacquot-Haméon et Charlène Letenneur ont réalisé les magnifiques dessins du chapitre 4, Alexandre Léthiers les planches photos, Philippe Loubry les photographies, mais également, par l'envoi de photographies : Javier Aguilar-Perez, Bogdan Jurkovšek, Fettouma Chikhi-Aouimeur, Carolina d'Arpa, Tony Costagliola, Peter Skelton, Simon Mitchell, Sacit Özer, Brian Pratt, ainsi qu'Alceo Tarlao.

À tous mes relecteurs, dont le premier est de loin Loïc Villier, mais également Isabelle Rouget, Stéphane Prin, Paul Zaharias, Valéry Zeitoun, Mathieu Faure-Brac, Isabelle Kruta, Lou Mary, ainsi que Raquel Sanchez-Pellicer pour la mise en page.

Aux faisans, Jérémy Tissier, Laetitia Carrive, Charles Ragusa, Marc Testé, ainsi qu'aux Francs-taxons et à tous mes amis.

À Lou pour m'avoir accompagné pendant ces derniers mois si intenses.

À ma famille pour être toujours là.

Sommaire

Remerciements vii

Sommaire x

Introduction générale 1

PARTIE 1. ANALYSE CLADISTIQUE 7

Chapitre 1. Cladistique et analyse à trois éléments 9

1. Introduction 11
2. Théorie et méthode cladistique 14
3. Reconstruction phylogénétique 30
4. Consensus trees: interpretation, informativity and efficiency 107
5. Conclusion du chapitre 131

Chapitre 2. Analyse à trois éléments et parcimonie : éléments de comparaison .. 135

1. Introduction 137
2. Méthode de parcimonie et 3ia : comparaison théorique et méthodologique 138
3. Experimental systematics: sensitivity of cladistic methods to polarization and character ordering schemes 161
4. Impact of errors on cladistic inference: simulation-based comparison between parsimony and three-taxon analysis 191
5. Conclusion du chapitre 215

PARTIE 2. LES RUDISTES: ANATOMIE COMPARÉE, PHYLOGÉNIE, SYSTÉMATIQUE, PALÉOÉCOLOGIE 219

Chapitre 3. Taxonomic revision of the genus *Ichthyosarcolites* Desmarest, 1812 (Bivalvia, Hippuritida): an example of high intraspecific variability 221

1. Introduction 223
2. Material and methods 225
3. Morphology 235
4. Systematic paleontology 239
5. Chapter conclusion 260

Chapitre 4. Homologies and the phylogeny of rudists (Bivalvia, Hippuritida): new insights from comparative anatomy 263

1. Introduction 265
2. Historical background 267
3. Material and methods 273
4. Results 317
5. Discussion 322
6. Chapter conclusion 345
7. Character list 347

Chapitre 5. Paléoécologie, paléoenvironnements et événements biosédimentaires enregistrés dans la succession de communautés à rudistes du Bassin Sud-Provençal au Cénomanien supérieur 459

1. Introduction 461
2. Cadre géologique 462
3. Matériel et méthodes 464
4. Résultats 470
5. Discussion 496
6. Conclusion du chapitre 504

Conclusion générale 507

Bibliographie 513

Lexique 553

Partie 1 : Analyse cladistique 553

Part 2 : Rudist morphoanatomy 559

Abbréviations et symboles - Abbreviations and symbols 563

Partie 1: Analyse cladistique 563

Part 2 : Rudist morphoanatomy 565

Annexes 567

Table des illustrations 573

Table des tableaux 613

Table des matières 617

Introduction générale

L'objectif de ce travail est d'apporter des éléments nouveaux à la compréhension de l'histoire des rudistes. Les rudistes (ordre Hippuritida) forment un groupe monophylétique au sein des bivalves, caractérisé notamment par une coquille inéquivalve ainsi que par une charnière dite pachyodonte, formée de dents épaisses et peu nombreuses (Skelton 2013a). Le groupe des rudistes est composé de deux sous-ordres, quinze familles, plus de 140 genres et près de 2000 espèces. Apparu il y a environ 160 millions d'années (Oxfordien), le groupe colonise progressivement les mers chaudes et peu profondes de l'ensemble du globe. Les rudistes disparaissent de manière brutale à la fin du Crétacé, marquant la fin de leur évolution (Swinburne 1990; Steuber *et al.* 2016). Leur évolution et leur diversification dépendent de l'effet conjoint de plusieurs facteurs. (i) Les *contraintes environnementales*. Au cours du Crétacé, la diversité et l'extension des rudistes suivent un rythme imposé par le développement cyclique de leurs biotopes, les plates-formes carbonatées. Deux évènements anoxiques EAO1 et 2 sont associés à des chutes importantes de leur diversité. Plus généralement, l'abondance et la diversité des rudistes dans les archives paléontologiques sont liées aux grands cycles de transgression/régression (Gili et Skelton 2000; Skelton 2003a; Skelton et Gili 2012). (ii) Les *contraintes fonctionnelles***.** Bien que les aspects fonctionnels de la morphologie des rudistes restent discutés et mal connus (Vogel 1975; Skelton 1976, 1978, 1979, Seilacher 1984, 1990), l'acquisition de certains caractères clés semble associée à des épisodes de diversifications et de radiations adaptatives majeures (Steuber et Löser 2000; Skelton 2003; Steuber *et al.* 2016). (iii) Les *contraintes écologiques*. La grande diversité morphologique et fonctionnelle des rudistes leur permet d'occuper une très large gamme d'environnements sur les plates-formes carbonatées, depuis l'intertidal au circalittoral (Philip 1981). Les espèces s'organisent en communautés qui se partagent l'espace écologique (Masse *et al.* 2003; Fenerci-Masse *et al.* 2004, 2005). La structuration et la spécialisation des communautés favorisent un

accroissement de la diversité locale et régionale (Fenerci-Masse *et al.* 2005) qui pourrait être un facteur de diversification globale.

A la base de la reconstruction de l'histoire d'un groupe se trouve la phylogénie. La première partie de la thèse a pour but de clarifier et de justifier du socle théorique et méthodologique employé pour résoudre les questions évolutives. Le cadre est celui de la théorie cladistique et de la méthode d'analyse cladistique qui permet la reconstruction phylogénétique. La théorie cladistique peut s'exprimer de manière claire et cohérente à partir de quelques principes simples dont découle un système de pensée complet de l'évolution et du changement. Nous montrerons que la théorie cladistique permet d'aboutir sans incohérence à une vision globale de l'évolution par hérédité. La théorie cladistique permet de justifier la méthode d'analyse à trois éléments, méthode permettant de reconstruire des phylogénies à partir d'hypothèses d'homologies. La théorie cladistique et la méthode d'analyse à trois éléments font l'objet du premier chapitre de la thèse. Une emphase particulière est portée sur la structure mathématique des objets cladistiques qui doit être nécessairement hiérarchique, et en quoi la méthode d'analyse à trois éléments permet d'appliquer la théorie. Le concept d'arbres consensus est discuté, ou comment choisir lorsqu'une méthode propose plusieurs hypothèses de phylogénie. L'analyse à trois éléments permet d'apporter une réponse non-ambigüe au problème du choix entre plusieurs phylogénies grâce au concept de triplet et à ses nombreuses ramifications. Les triplets et leurs relations de combinaisons nous permettent également de discuter de la pertinence des méthodes consensus. Cependant, l'analyse à trois éléments n'est pas la seule méthode se réclamant de la théorie cladistique. Dans un second chapitre, nous confronterons la méthode d'analyse à trois éléments avec la méthode la plus utilisée actuellement, la méthode d'analyse de parcimonie. Nous proposerons une confrontation de ces deux méthodes aux trois niveaux : théorique, méthodologique et empirique. Les arguments théoriques et méthodologiques feront appel au chapitre 1, et montreront que la parcimonie, au contraire de l'analyse à trois éléments, souffre de plusieurs éléments de méthode injustifiés par rapport au cadre de la théorie dans laquelle elle s'exprime. Nous proposerons également un test empirique des différences de comportement entre analyse à trois éléments et parcimonie, basé sur des simulations informatiques générées à partir de modèles d'évolution ainsi que sur quelques cas dont l'évolution est connue.

Dans la seconde partie de la thèse, les concepts et méthode cladistiques et d'analyse à trois éléments sont appliqués à la résolution de l'histoire phylogénétique, écologique et géologique des rudistes. L'ordre Hippuritida est composé de taxons aux morphologies extrêmement variées. Les premières descriptions d'espèces remontent au XVIIIème siècle (Picot de Lapeyrouse 1781) et, au fil du temps, de nombreux ajouts à la taxonomie par le biais de découvertes et d'émendations d'espèces se sont accumulés. Cependant, peu d'espèces ont fait l'objet de révisions taxonomiques utilisant des outils actuels de caractérisation de la morphologie et des différences entre taxons (Fenerci-Masse *et al.* 2011). Parmi les groupes de rudistes délaissés, le genre *Ichthyosarcolites* apparait comme un taxon clef dont l'arrangement des caractères morphologiques est crucial pour la reconstruction phylogénétique de l'ensemble des rudistes (Douvillé 1887; Skelton et Smith 2000; Skelton 2003b). Le chapitre 3 a pour objet la révision taxonomique du genre *Ichthyosarcolites* à partir de données nouvelles de terrain ainsi que de spécimens provenant de collections muséales.

Le chapitre 4 a pour but de proposer des hypothèses sur l'évolution des rudistes par la reconnaissance d'homologies. A partir d'un travail d'anatomie comparée sur les coquilles de 27 espèces de rudistes balayant l'ensemble des familles du sous-ordre Hippuritidina – les rudistes qui s'attachent par leur valve gauche –, nous proposons de nouvelles homologies et 40 caractères morpho-anatomiques. Les caractères sont construits dans le cadre de la théorie cladistique en hiérarchies. Ces caractères, traités par analyse à trois éléments, donnent deux arbres également optimaux dont le consensus permet de dégager une phylogénie non ambigüe retraçant l'histoire évolutionnaire des rudistes. Cette histoire est mise en relation avec l'histoire globale des plates-formes carbonatées et des crises biologiques du Crétacé afin de mettre en évidence deux grandes radiations évolutives des rudistes.

Dans le 5ème et dernier chapitre de la thèse, nous ferons le lien entre patrons évolutifs à l'échelle planétaire et évènements locaux par l'étude de deux localités du bassin sud-Provençal dont les couches datées du Cénomanien supérieur à terminal enregistrent les successions de communautés à rudistes. Ces successions nous permettent de proposer des hypothèses sur l'écologie des rudistes et leur implication dans la dynamique sédimentaire à une échelle locale, suivant les modalités d'études déjà mises en œuvre sur les rudistes du Crétacé de Provence (Philip 1967, 1970, 1980, 1998; Fenerci-Masse *et al.* 2005; Fenerci-

Masse 2006; Masse *et al.* 2009). L'étude des carrières de la Barre des Aiguilles et de la Bastide d'Orves (Var, France) nous permet également d'affiner le modèle de succession écologique des peuplements à rudistes depuis les organismes pionniers jusqu'au stade climacique, ainsi que le biotope de ces communautés. On terminera par une mise en relation des patrons à échelle globale et locale des rudistes, de leur évolution et de leur écologie.

PARTIE 1. ANALYSE CLADISTIQUE

Au bout d'un moment, j'entrais l'embrasser; Françoise faisait infuser son thé; ou, si ma tante se sentait agitée, elle demandait à la place sa tisane et c'était moi qui étais chargé de faire tomber du sac de pharmacie dans une assiette la quantité de tilleul qu'il fallait mettre ensuite dans l'eau bouillante. Le desséchement des tiges les avait incurvées en un capricieux treillage dans les entrelacs duquel s'ouvraient les fleurs pâles, comme si un peintre les eût arrangées, les eût fait poser de la façon la plus ornementale. Les feuilles, ayant perdu ou changé leur aspect, avaient l'air des choses les plus disparates, d'une aile transparente de mouche, de l'envers blanc d'une étiquette, d'un pétale de rose, mais qui eussent été empilées, concassées ou tressées comme dans la confection d'un nid. Mille petits détails inutiles — charmante prodigalité du pharmacien — qu'on eût supprimés dans une préparation factice, me donnaient, comme un livre où on s'émerveille de rencontrer le nom d'une personne de connaissance, le plaisir de comprendre que c'était bien des tiges de vrais tilleuls, comme ceux que je voyais avenue de la Gare, modifiées, justement parce que c'étaient non des doubles, mais elles-mêmes et qu'elles avaient vieilli. Et chaque caractère nouveau n'y étant que la métamorphose d'un caractère ancien, dans de petites boules grises je reconnaissais les boutons verts qui ne sont pas venus à terme; mais surtout l'éclat rose, lunaire et doux qui faisait se détacher les fleurs dans la forêt fragile des tiges où elles étaient suspendues comme de petites roses d'or — signe, comme la lueur qui révèle encore sur une muraille la place d'une fresque effacée, de la différence entre les parties de l'arbre qui avaient été « en couleur » et celles qui ne l'avaient pas été — me montrait que ces pétales étaient bien ceux qui avant de fleurir le sac de pharmacie avaient embaumé les soirs de printemps.

Du côté de chez Swann. Marcel Proust, 1913.

Chapitre 1. Cladistique et analyse à trois éléments

La partie 3.3 de ce chapitre a été écrite en collaboration avec M. Faure-Brac et R. Zaragüeta i Bagils, la partie 4 avec R. Zaragüeta i Bagils, J. Bardin et E. Bourdon.

1. Introduction 11
2. Théorie et méthode cladistique 14
3. Reconstruction phylogénétique 30
4. Consensus trees: interpretation, informativity and efficiency 107
5. Conclusion du chapitre 131

1. Introduction

La *reconstruction phylogénétique* (étymologie : « *phŷlon* » groupe, tribu, race, « *géneia* » qui engendre) repose sur trois postulats. Le premier est celui de l'existence de la vie comme un tout unifié ayant une origine commune. Le second est que la vie se montre diverse dans cette unité. Le troisième est qu'il est possible de faire des généralisations successives à partir des entités les plus particulières jusqu'aux entités les plus générales (Prin 2012 : 135-136). La reconstruction phylogénétique a pour but l'établissement de la phylogénie, c'est-à-dire de relations de parenté entre taxons. Un ensemble d'organismes se généralise en une classe appelée *taxon* (étymologie : « *taxis* » ordre). Lorsque l'on parle de *Felis catus*, on parle d'un concept qui regroupe tous les organismes possédant certaines propriétés et les généralise en un taxon. Une relation de parenté implique que plusieurs taxons soient reliés au sein d'une même classe. Or, un des postulats de base de la phylogénétique est l'acceptation de l'unité du vivant. Cette relation de parenté est donc vraie entre tous les organismes et il existe donc une relation de simple parenté (Prin 2012 : 34-37, 2016 : 441-443) entre toutes les formes de vie, actuelles comme fossiles. Ces liens de parenté pour être informatifs nécessitent l'ajout d'un troisième terme (Hennig 1965 : 97-98), définissant ainsi une relation ternaire de degré de parenté. Il ne s'agit donc pas de savoir si deux objets sont apparentés en soi (ce qui est trivial) mais s'ils le sont par rapport à un troisième. Cette relation, qui a déjà été identifiée sous diverses expressions telles que « relation phylogénétique » (« *phylogenetic relationship* »; Hennig 1965; Nelson 1994 : 104, 120), « *relationship of common ancestry* » (Nelson 1970 : 377, 1972) ou encore « *relative recency of common ancestry* » (Platnick 1979 : 545), est le fondement de la phylogénétique cladistique (Hennig 1965 : 97-98; Nelson et Platnick 1981 : 32; Nelson 1994 : 128; Zaragüeta i Bagils 2011 : 52-54; Prin 2012 : 136-137, 143). La théorie cladistique est la théorie des clades et des cladogrammes. La théorie cladistique est le cadre conceptuel par lequel nous avons accès à l'histoire des taxons, par la formulation de clades, groupes reflétant cette relation de degré de parenté. La phylogénétique cladistique implique également et nécessairement un système classificatoire hiérarchique inclusif (Nelson 1972, 1973; Colonius et Schulze 1981; McMorris et Powers 2003; Prin 2012 : 38-43, 2016 : 444-446). C'est pourquoi on parle de systématique phylogénétique.

L'analyse à trois éléments (3ia pour « *three-item analysis* »; Nelson et Ladiges 1991a, 1991b; Nelson et Platnick 1991; Zaragüeta i Bagils et Bourdon 2007; Cao 2008) est une méthode de la systématique phylogénétique et s'inscrit dans la théorie cladistique. Ce chapitre est consacré à une explication et une argumentation de la méthode de 3ia, initialement proposée par Nelson et Platnick en 1991 et qui a été privilégiée pour résoudre la phylogénie des rudistes, objet d'étude de la thèse. La 3ia y est présentée dans le cadre de la théorie cladistique (et plus particulièrement de Nelson et certains de ses continuateurs; voir notamment Nelson 1979, 1994; Platnick 1979; Brady 1982; Ebach *et al.* 2007, Zaragüeta i Bagils 2011, Prin 2012). La première partie concerne les fondements de la méthode cladistique, articulée autour des six principes définis par Prin (2012) nécessaires pour exprimer le déroulement général de la procédure cladistique. Il sera souligné comment la 3ia peut et doit être considérée comme une méthode cladistique, car elle en partage la structure logique. La seconde partie propose une explication détaillée de la manière d'appliquer la 3ia dans un cadre phylogénétique où les objets à classer sont des taxons. Cette partie est découpée en trois étapes : (i) la formulation des hypothèses cladistiques, (ii) la construction du cladogramme et (iii) son interprétation. Dans une troisième partie, nous proposerons une discussion sur les méthodes permettant de construire les arbres consensus – les arbres qui permettent de synthétiser plusieurs cladogrammes lorsqu'il n'est pas possible de choisir l'un de ces cladogrammes – et leur statut au regard de la méthode cladistique.

On détaillera tout d'abord la structure que doivent prendre ces objets à classer, les UT, et la manière de les définir dans l'ontologie cladistique. On détaillera ce qui sépare ces UT des TH (les clades) et l'importance fondamentale de cette distinction lors de l'analyse. Lors de la définition des hypothèses cladistiques, chaque UT forme un tout qui est décomposé en parties, ces parties étant rassemblées en homologues. Les structures ainsi construites s'appellent des homologies, et spécifient des relations entre parties de taxons. Au sein de ces structures, remplacer les parties par les taxons correspond au passage de l'homologie composée d'homologues au *caractère* (étymologie : « *kharakter* » signe, empreinte) composé d'*états de caractères*. On expliquera comment se constituent les arguments permettant la construction des hypothèses cladistiques et quelle forme doivent nécessairement prendre ces hypothèses. On expliquera comment la 3ia répond aux problèmes méthodologiques spécifiques à l'étape de construction des hypothèses

cladistiques, de la « polarisation des caractères », du polymorphisme, des données manquantes et des données non-applicables.

On expliquera ensuite comment passer des caractères au cladogramme en décomposant ces caractères en unités informatives minimales, qui correspondent à des hypothèses minimales et seront appelées par la suite *triplets*. La solution générale est le cladogramme le plus informatif, c'est-à-dire celui qui contient le plus de triplets. C'est donc celui qui retient le maximum des hypothèses formulées en amont, ce qui revient à rejeter le minimum d'hypothèses informatives. La décomposition des caractères nécessite l'application d'une *pondération fractionnaire* (Nelson et Ladiges 1992, 1994) permettant d'éviter la redondance entre les différentes hypothèses formulées. Nous expliquerons les raisons de cette dépendance (Prin, 2012 : 187-193, Wilkinson 1994b). On détaillera le lien entre les caractères et les triplets, ces unités minimales fournissant un outil de mesure de l'information portée par n'importe quelle hiérarchie cladistique. Cet outil permet de montrer pourquoi certains caractères sont plus informatifs que d'autres, pourquoi certains cladogrammes sont plus informatifs que d'autres. Ils permettent d'introduire la différence entre les clades paralogues (clade incluant immédiatement au moins deux sous-clades inclusifs) et les clades orthologues (clade incluant immédiatement un seul sous-clade inclusif) ou encore que les polytomies engendrent une incertitude sur le contenu en triplets. Une réflexion sera également apportée sur les algorithmes de construction de cladogrammes en 3ia et les problèmes liés à la combinaison de triplets. (iii) L'interprétation du cladogramme final découle des deux phases de formulation des caractères et de construction du cladogramme à partir de ces caractères. On précisera ce que veut dire un cladogramme, des relations cladistiques entre taxons, et comment ne pas le surinterpréter. L'interprétation *a posteriori* des caractères se fait grâce à l'Indice de Rétention (IR : Archie 1989; Farris 1989; Kitching 1998 : 179) qui mesure la part des triplets d'un caractère retenus lors de l'analyse. L'IR se généralise en un Indice de Rétention Inter-Arbres (IRIA) dès lors qu'il s'agit de comparer quantitativement deux hiérarchies cladistiques quelconques (Grand *et al.* 2013). A partir de l'IRIA, nous proposons une mesure du contenu informatif des nœuds d'un cladogramme grâce à l'Indice de Rétention Nodal (IRN). Seront spécifiés l'assignation des états de caractères aux nœuds d'un cladogramme, la procédure de test des hypothèses cladistiques, c'est-à-dire de savoir comment et quand elles sont rejetées, ainsi que l'interprétation d'une *homoplasie* (étymologie : « *homós* » semblable,

pareil, « *plásis* », façonner, modeler). Un état de caractère rejeté est qualifié d'*homoplasique*. L'assignation des états de caractères aux nœuds d'un cladogramme peut être généralisée à l'assignation de propriétés à un clade. Ce point permet d'introduire la notion de datation des clades (Zaragüeta i Bagils et Lelièvre 2001; Zaragüeta i Bagils *et al.* 2004; Lelièvre *et al.* 2008).

La théorie cladistique est donc une théorie qui peut s'appliquer à des champs disciplinaires variés. Bien qu'elle ait été historiquement fondée dans le but de construire des phylogénies, elle dépasse aujourd'hui largement ce cadre pour s'appliquer à tous types d'objets qui satisfont à certains prérequis : pouvoir se généraliser en clades par des propriétés intrinsèques, que ces clades forment une hiérarchie unique et qu'une théorie explicative sur l'évolution de ces objets permette de donner un sens aux cladogrammes produits. Le but principal de ce chapitre est de montrer les avancées liées à la 3ia depuis sa première formulation en 1991 jusqu'à aujourd'hui, de montrer comment la 3ia est définie aujourd'hui comme une méthode cladistique complète, ou chacune de ses étapes sont justifiées, et d'expliquer la méthode de 3ia, ses principes et comment l'utiliser.

2. Théorie et méthode cladistique

2.1. Théorie cladistique

Jardine (1969 : 37) délimite les trois grands questions de la classification biologique permettant de définir les grandes disciplines de ce domaine scientifique. (a) Le problème de la description d'organismes et de leur partitionnement en taxons appartient à la taxonomie. La taxonomie permet de formuler des taxons de base. C'est la science de la diversité des organismes. (b) Le problème de la construction d'une classification d'un ensemble de taxons de base construits en (a) en taxons plus inclusifs concerne la systématique, discipline de la construction des systèmes. (c) Le problème de la diagnose ou de l'identification émerge lorsqu'il s'agit de placer des individus dans une classification générée en (b). Bien que ce chapitre ne traite que de l'aspect classificatoire et donc de systématique, les trois classes de problèmes de la taxonomie, de la systématique et de l'identification sont intimement liées. Les objets de la taxonomie sont les organismes, et leur groupement produit des taxons. Les objets de la systématique sont les taxons, et leur

agencement produit des systèmes. La systématique peut cependant être nourrie par une théorie, cette théorie fournissant des interprétations causales au système classificatoire. La théorie de l'évolution biologique explique que du monde vivant émergent des variations qui sont ensuite soumises à la sélection naturelle, entraînant l'évolution des organismes. Il est nécessaire dans ce cadre de déterminer quelle forme doit prendre la classification des taxons selon le modèle explicatif de la théorie de l'évolution.

La classification traduit les relations entre des taxons. Dans ce cadre évolutionnaire, une relation implique une vision particulière du changement. Le changement permet de spécifier des relations de degré de parenté entre taxons et de dire que ces taxons sont « *les mêmes* » à un niveau de généralisation donné. Le changement est un problème qui agite les philosophes depuis la Grèce antique. Prenons l'exemple d'un objet a quelconque qui change et devient un objet b comme suit : $a \rightarrow b$ (Popper 1998; Zaragüeta i Bagils et Bourdon 2005). Héraclite est connu pour son aphorisme « *on ne se baigne jamais deux fois dans le même fleuve* » qui illustre sa pensée sur l'impermanence des choses. Selon lui, seul le changement existe et tout le reste ne serait qu'illusion. Observer un bébé, c'est observer un vieillard en devenir. a et b ne sont alors que des illusions et seul le changement $\rightarrow$ existe. Parménide, lui, part de la prémisse selon laquelle tout ce qui existe *est* quelque chose. Cela implique que ce qui n'est pas ne peut pas exister. Le vide étant défini par l'absence d'être, il ne peut pas exister. Or, si le monde est un, plein, et sans vide, tout changement est alors impossible. La solution devient $a = b$, et le changement $\rightarrow$ n'existe pas. Ces deux interprétations du monde sont en contradiction avec les deux postulats de la diversité (a et b) et du changement ($\rightarrow$) spécifiés plus haut. Le modèle de transformation $a \rightarrow b$ est en accord avec ces postulats, où a se transforme en b. Selon ce modèle, on affirme que a change pour devenir b. Or, lors de la transformation, a disparaît et b prend sa place ($a \cap b = \varnothing$). Si a disparaît, il ne peut pas se transformer en b. La transformation implique une classification de type partition ordonnée. Dans un cadre historique, la transformation de a vers b est incohérente. Une autre possibilité du changement existe : le modèle de différenciation $a \rightarrow b$. Dans ce modèle, a se différencie en b où $a \cap b = b$ car $a \supset b$. La différenciation implique cette fois une classification de type hiérarchique (Zaragüeta i Bagils et Bourdon 2005). La différenciation $a \rightarrow b$ est une autre manière que l'esprit humain à d'appréhender la même intuition du changement. En se différenciant a donne naissance à b, mais a ne disparaît pas pour autant. Nous n'avons pas l'habitude de réfléchir

au changement de cette manière, mais les conséquences sont importantes dans de nombreux domaines scientifiques, en particulier dans les sciences historiques. En observant une plante, on peut se demander où est passée la graine. La graine, et plus particulièrement l'embryon qui est contenu dedans, s'est différenciée en une plante. La plante est une graine différenciée qui n'a pas disparu mais qui a changé. De même, lorsque nous contemplons un enfant, nous ne voyons que lui; cependant, observer un vieillard permet également d'observer l'enfant qu'il était et sur lequel se sont appliqués des changements, un développement, tout au long de sa vie. Cette idée prend tout son sens avec la notion d'héritage. L'enfant porte en lui un héritage que l'on peut retrouver chez le vieillard (fig.

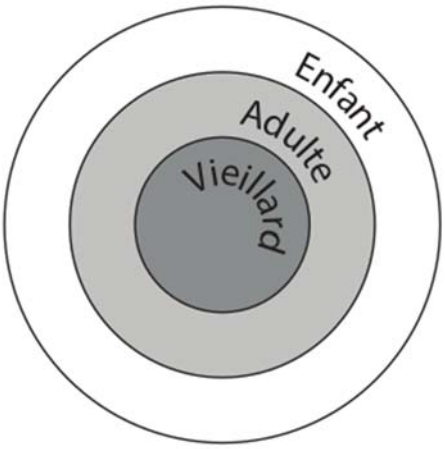


Figure 1.1. Représentation hiérarchique de la différenciation d'un être humain au cours du temps. Le vieillard se différencie à partir de l'adulte qui se différencie à partir de l'enfant. Il s'ensuit qu'un vieillard est un adulte car ce qui constitue l'adulte existe encore chez le vieillard, sous une forme ou une autre, et que l'adulte est un enfant.

1.1). Si l'enfant avait vraiment disparu, pourquoi retrouverait-on un code génétique identique, la même tâche de naissance ou la même cicatrice ? Ces attributs sont l'héritage porté par les objets et transmis aux entités en lesquelles ils se différencient. Selon cette vision, nous sommes bien dans l'époque contemporaine, mais nous sommes également encore au moyen-âge. Le moyen-âge s'est différencié en l'époque moderne puis contemporaine. Parcourir la France permet d'observer tous ces châteaux forts qui sont un héritage du moyen-âge porté jusqu'à nous. La langue française est également un ensemble structuré dont on peut faire l'étymologie pour remonter dans le passé. Et le présent n'est que la dernière différenciation du passé.

La phylogénétique nécessite la théorie de l'évolution comme cadre explicatif à la diversité biologique. La phylogénétique est une science historique, dont le but est l'étude du passé des taxons par leur histoire évolutionnaire. Le propre des sciences historiques est de ne s'appliquer qu'à des objets uniques ayant une extension de lieu et de temps définie, comme la voie lactée, le roi Louis XIV ou le taxon de la coquille Saint-Jacques *Pecten maximus*. Le seul accès au passé que l'on ait est l'héritage qui nous en parvient. D'après la théorie de l'évolution, tous les organismes sur Terre sont les descendants d'autres organismes jusqu'à remonter au premier organisme vivant. Ce monde vivant est unique mais également divers par l'apparition de modifications chez certains de ces organismes qui les transmettent à leurs descendants. C'est cet héritage qui est l'objet fondamental de la systématique phylogénétique. Ces modifications intrinsèques aux organismes sont la trace d'une parenté commune. Cet héritage nous permet de fonder des classifications historiques dont la forme est nécessairement hiérarchique. La reconstitution de ces hiérarchies est le but de la phylogénétique. La théorie de l'évolution, elle, sert de cadre explicatif permettant d'interpréter ces hiérarchies.

Ce sont donc des parties d'organismes qu'il faut chercher. Elles sont liées par une relation d'identité, et donc de parenté. Les relations de parenté entre parties d'organismes permettent d'inférer les relations de parenté entre les organismes eux-mêmes. On a dit qu'une relation de parenté entraîne une classification; une classe peut être fondée, qui généralise des parties à un degré de parenté donné. Une classe de parties d'organismes est appelée *homologues* (étymologie : « *homós* » semblable, pareil, « *lógos* » étude; Owen 1843 : 379, 1848 : 7). Par exemple, la présence d'une coquille bivalve constituée d'une valve gauche et d'une valve droite est une partie d'organisme. Celle-ci s'inclut dans un homologue « coquille bivalve ». Elle permet de dire que toutes les coquilles bivalves sont identiques. La généralisation de toutes ces coquilles individuelles en un même groupe homologue permet de construire un état de caractère qui définit le taxon Bivalvia par la propriété « avoir une coquille bivalve ». Un état de caractère qui caractérise un taxon est une *synapomorphie* (étymologie : « *sún* » commun, « *apo* » dérivé de, « *morphos* » forme). La proposition de la systématique phylogénétique est donc d'établir des groupes d'organismes à la fois fondés sur des généralisations (de parties) et interprétés comme le signe de l'existence d'une parenté exclusive aux constituants de ce groupe. En tant qu'*Homo sapiens*, nous sommes membres d'une multitude de taxons caractérisés par une

mosaïque d'homologues qui nous renseignent sur une histoire longue de plus de 3,5 milliards d'années (Schopf *et al.* 2007). Le fait que les différents homologues renseignent sur des niveaux de généralisation différents est décrit par Hennig (1966) selon le terme d'hétérobathmie. Nous, *Homo sapiens,* sommes également des *Primates* au même titre que les lémuriens du fait de nos pouces opposables; nous sommes des *Metazoa* au même titre que les araignées du fait de notre corps multicellulaire hétérotrophe composé de cellules eukaryotes; nous sommes également inclus dans le groupe qui englobe tous les organismes vivants que l'on peut nommer *Vie* du fait de la présence d'un génome ainsi que de gènes codant par exemple pour les protéines d'ADN polymérase, d'ARN polymérase, d'ARN ribosomial et d'ATPase[1].

Cet exemple permet de montrer la pertinence du modèle de la différentiation par rapport à celui de la transformation. Puisque le seul accès à l'histoire évolutive des taxons est l'héritage qui nous parvient, et que cet héritage se présente sous la forme de parties d'organismes qui donnent des indices sur une éventuelle parenté commune par la construction d'homologues, la différenciation $a \rightarrow b$ appliquée aux taxons implique une classification hiérarchique où la classe *a* inclut nécessairement *b*. Dans la transformation $a \rightarrow b$, des individus du taxon *a* ont donné naissance à des individus du taxon *b* qui ne sont plus *a*: *a* est un groupe incomplet en termes de parenté puisque les individus de *b* sont porteurs du même héritage que les individus de *a* (Zaragüeta i Bagils 2011). On définit, à la suite de Hennig (1965, 1966), un taxon complet en termes de parenté comme *monophylétique* (étymologie : « *mónos* », unique, « *phŷlon* » groupe, tribu, race; fig. 1.2, arbre de gauche). Tous les membres d'un taxon monophylétique sont plus proches parents entre eux que d'aucun autre. La différenciation est un modèle qui ne génère que des ensembles monophylétiques. Un exemple de taxon monophylétique est le taxon *Mammalia*, caractérisé par plusieurs propriétés homologues dont les poils et les glandes mammaires; tous les taxons membres partagent un degré d'apparentement exclusif. Un groupement incomplet en termes de parenté est un taxon non-monophylétique. Un taxon non-monophylétique peut être un taxon *paraphylétique* (étymologie : « *pará* », à côté de,

[1] On note que d'après cette définition, de nombreux groupes de virus se retrouvent sans ambigüité dans le taxon *Vie*. Les *Orthopoxvirus* par exemple possèdent une ARN polymérase structurellement et fonctionnellement proche d'*Archea*, *Bacteria* et *Eukaryota* (Broyles et Moss 1986; Patel et Pickup 1989).

« *phŷlon* » groupe, tribu, race; fig. 1.2, arbre de droite) ou *polyphylétique* (étymologie : « *polús* », beaucoup, « *phŷlon* » groupe, tribu, race). Un taxon paraphylétique implique que tous ses membres soient apparentés mais pas de manière exclusive. Ainsi, d'autres taxons en sont exclus alors qu'ils partagent le même degré de parenté que les constituants du taxon monophylétique. Le modèle $a \rightarrow b$ implique nécessairement la paraphylie de a, et ne peut donc pas être utilisé dans l'objectif de reconstituer une histoire évolutive à partir d'homologues. Un exemple de taxon paraphylétique est le taxon *Reptilia*, groupe tronqué auquel il manque les oiseaux. Un taxon polyphylétique en revanche implique que ses membres se répartissent dans plusieurs groupes monophylétiques (au moins deux) qui ne fassent pas partie les uns des autres. Les algues forment un groupe polyphylétique qui rassemblent de nombreux groupes monophylétiques n'ayant rien à voir entre eux, dont des Bacteria (ex : *Nostoc*), des Chromista (ex : *Fucus*) ou encore des Plantae (ex : *Acer*).

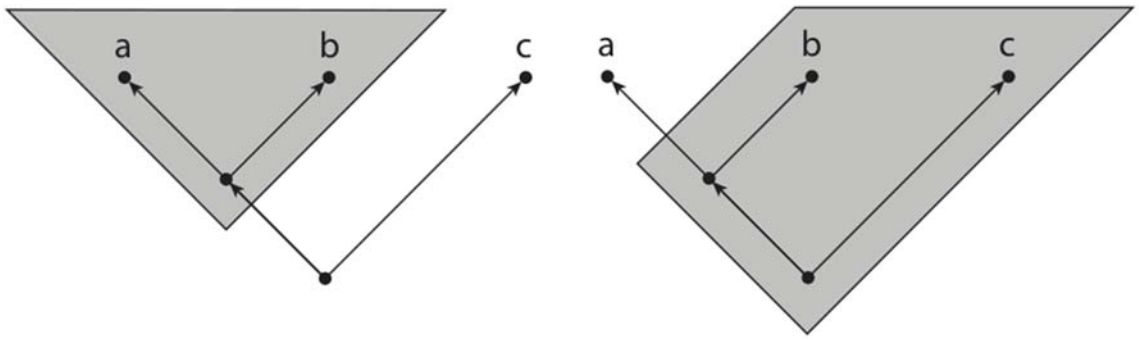


Figure 1.2. Représentation sur une même classification hiérarchique d'un groupe monophylétique et d'un groupe paraphylétique (en grisé). Le groupe de gauche est monophylétique car il est complet en termes de parenté : *a* et *b* partagent un degré de parenté exclusif. Le groupe de droite est paraphylétique car il est incomplet en termes de parenté : *b* et *c* partagent un degré de parenté non-exclusive, *a* possédant le même degré de parenté mais étant exclu du groupe.

L'utilisation du concept de différenciation peut ensuite se mêler à celui de l'unité et de la diversité du vivant. L'unité du vivant implique qu'au niveau de généralisation maximal, tous les taxons partagent un même degré de parenté. Il est nécessaire de différencier deux types de taxons : les *Unités Taxonomiques* (UT) et les *Taxons Hypothétiques* (TH). Une UT est un taxon défini *a priori* d'une analyse phylogénétique et considéré comme un objet de cette analyse. Dans un cadre cladistique, l'analyse phylogénétique a pour but de classer hiérarchiquement des UT. Un TH est un taxon résultant d'une analyse phylogénétique. Les taxons ainsi identifiés par les systématiciens sont des TH. Leurs classifications ont une structure hiérarchique, imposée par la

phylogénie. Regardons quelle classification minimale permet de dire quelque chose sur l'apparentement entre les UT. Soient les UT à classer *a* et *b*. La structure (*a, b*) est également nécessairement vraie car c'est la classe qui englobe tous les UT, ici {*a*, *b*}, et elle se justifie par le postulat de l'unité du vivant, qui peut être reformulé en postulat de la monophylie du vivant. La structure ((*a*) (*b*)) est elle aussi une relation de parenté. Elle est elle aussi triviale puisque {*a*} et {*b*} sont nécessairement vraies car implicitement définis comme monophylétiques. La relation minimale informative de la systématique phylogénétique nécessite l'ajout d'un troisième terme pour obtenir une *relation de degré de parenté* (*c,* (*a, b*)). Seule la classe {*a*, *b*} est non-triviale ici et donne une information supplémentaire autre que les prémisses de la monophylie du vivant et de la monophylie des UT. Le but de la systématique est donc de dire parmi trois objets lesquels deux sont « *les mêmes* » par rapport à un troisième. Établir une relation de degré de parenté ou d'*équivalence* revient mathématiquement à construire une classification hiérarchique entre ces taxons (Colonius et Schulze 1981; McMorris et Powers 2003; Prin 2012).

Il est important de préciser que l'idée transformationnelle peut aussi aboutir à une hiérarchie. Cependant, si la différenciation et la transformation peuvent se représenter toutes deux sous la forme d'une hiérarchie, leur contenu sémantique est radicalement différent. La figure 1.3 montre deux structures arborescentes, et les différences dans l'interprétation des objets qui les constituent (Prin, 2012 : 87-96, 111-115, 121-125). L'arbre de gauche présente un modèle transformationnel, ou *e* se transforme en *d* et *c* qui ne sont plus *e*, et *d* se transforme en *a* et *b* qui ne sont plus *d*. Cet arbre est une partition hiérarchiquement ordonnée. *a*, *b*, *c*, *d* et *e* sont des UT, et *d* et *e* sont des taxons ancestraux. L'identification d'UT ancestrales est cependant méthodologiquement impossible si nous n'avons accès qu'aux homologues. Si le seul accès épistémique à l'évolution que nous ayons sont les homologues, alors seuls les taxons monophylétiques doivent être pris en compte. Ici, le taxon *d* est paraphylétique. Il est incomplet car il n'intègre pas les taxons *a* et *b* qui partagent un même degré d'identité avec lui, car *a* est aussi proche de *d* que de *b* sur l'arbre (*a* et *b* sont parents au degré *d*). Le groupe incluant *a* et *b* inclut également *d*, mais un groupe monophylétique incluant *d* doit nécessairement inclure *a* et *b* qui lui sont directement apparentés. Le modèle transformationnel engendre des taxons ancestraux nécessairement paraphylétiques. L'arbre de droite au contraire ne présente que des entités

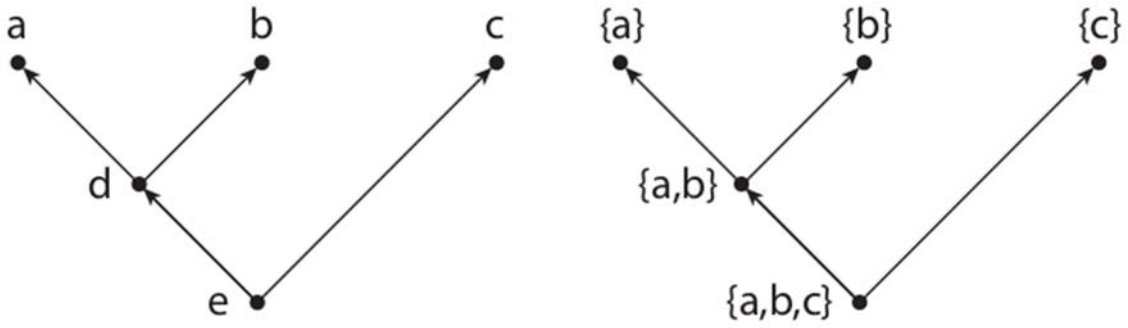


Figure 1.3. Deux classifications hiérarchiques construites sur le modèle de la transformation (à gauche) et de la différenciation (à droite). La hiérarchie de gauche montre la transformation de *e* en *d* et *c*, puis la transformation de *d* en *a* et *b*. Tous les taxons *a*, *b*, *c*, *d* et *e* sont mutuellement exclusifs, un organisme ne peut appartenir qu'à un seul de ces taxons. La hiérarchie de droite montre la différenciation de {*a*, *b*, *c*} en {*a*, *b*} et {*c*}, et la différenciation de {*a*, *b*} en {*a*} et {*b*}. Deux taxons sont soit mutuellement exclusifs ({*a*, *b*} et {*c*}), soit liés par une relation d'inclusion ({*a*} est inclus dans {*a*, *b*, *c*}). {*d*} et {*e*} sont des TH paraphylétiques, {*a*, *b*} et {*a*, *b*, *c*} sont des TH monophylétiques.

monophylétiques. Prendre le TH {*a*, *b*} implique nécessairement de prendre les UT *a* et *b* qui y sont incluses. Tous les nœuds de cet arbre désignent des groupes monophylétiques. De plus, cet arbre présentant des relations de degré de parenté entre *a*, *b* et *c* est également une classification qui procède au regroupement des UT par niveaux de généralisation. Il est également possible de représenter ces classifications sous la forme d'emboitements, appelés diagrammes d'Euler (fig. 1.4; Hennig, 1966). Les ancêtres sont ici complètement définis sans aucune ambiguïté : ce sont des taxons. L'ancêtre de *a* et *b* est le taxon {*a*, *b*}, tout comme le taxon *Mammalia* est le taxon ancêtre de tous les sous-taxons de mammifères. Tous les sous-taxons de mammifères descendent de *Mammalia*, mais sont également inclus dans *Mammalia*. Ainsi, les ancêtres sont des TH. Les TH sont les taxons que l'on reconstruit grâce aux homologues. L'ontologie de la systématique phylogénétique fait que le terme ancêtre ne peut être utilisé que pour définir des taxons, en aucun cas des individus, des populations ou des espèces comme cela est le cas dans le langage courant. La pertinence d'une telle représentation peut se comprendre en se posant la question de comment classer des objets qui sont constamment en évolution ? La hiérarchie de type différenciation est un ensemble de relations de parenté entre taxons monophylétiques qui s'applique par niveaux de généralisations successifs à tous types de taxons, quel que soit leur âge d'apparition. Faisons remonter le temps à un systématicien jusqu'au Protérozoïque il y a deux milliards d'années pour qu'il réalise une classification des organismes de l'époque. Imaginons

ensuite que le systématicien obtienne une classification cladistique qui partitionne l'ensemble du vivant en trois clades majeurs mutuellement disjoints : archées, bactéries et eucaryotes. Maintenant, posons-nous la question : cette classification est-elle encore vraie à notre époque, deux milliards d'années plus tard ? La réponse est oui, puisque non seulement les relations mises en place au cours du temps sont ensuite vraies dans la suite de l'histoire, mais également que tout taxon actuel ou futur peut être assigné à une de ces trois classes. Les taxons ne peuvent que se différencier, ils restent donc toujours vrais tout en se diversifiant. Cette expérience de pensée exprime l'adéquation de l'utilisation de ce type d'arbres avec l'étude de l'histoire évolutionnaire des taxons. La théorie cladistique est la théorie des clades et des cladogrammes; elle s'applique notamment à la systématique phylogénétique. Les arbres hiérarchiques qui reflètent ces hypothèses cladistiques sont appelés *cladogrammes*. Un cladogramme est hiérarchique parce qu'il est à la fois classificatoire, ordonné, et arborescent. Appliqué à la phylogénétique, un cladogramme est constitué des UT à classer et de la classification proprement dite, l'ensemble des TH, ou clades. Les cladogrammes de ce type sont appelés *phylogénies*.

La théorie cladistique est une théorie méthodologique. Le principal champ d'application de cette théorie est la systématique phylogénétique ou il s'agit d'établir des relations entre taxons, mais on verra que les problèmes de la biogéographie peuvent s'y appliquer où des relations sont établies entre aires biogéographiques. D'autres domaines peuvent également bénéficier d'une représentation de type cladogramme, comme la biogéographie, la linguistique (Warnow *et al.* 1995), ou en général tout objet qui possède une histoire cumulative par hérédité. Les principes spécifiques de la cladistique tiennent en quatre points (Prin, 2012 : 5-6). (i) Le premier point est d'ordre ontologique, c'est-à-dire qu'il porte sur la nature des objets de la théorie cladistique : les seuls groupes réels sont les clades car ils sont les seuls regroupements traduisant les phénomènes de parenté. Il en résulte que les groupes paraphylétiques, incomplets en termes de parenté, sont considérés comme artificiels et non-naturels dans l'ontologie cladistique. (ii) Épistémologiquement, les clades sont également les plus adaptés pour comprendre l'histoire et la structure de la

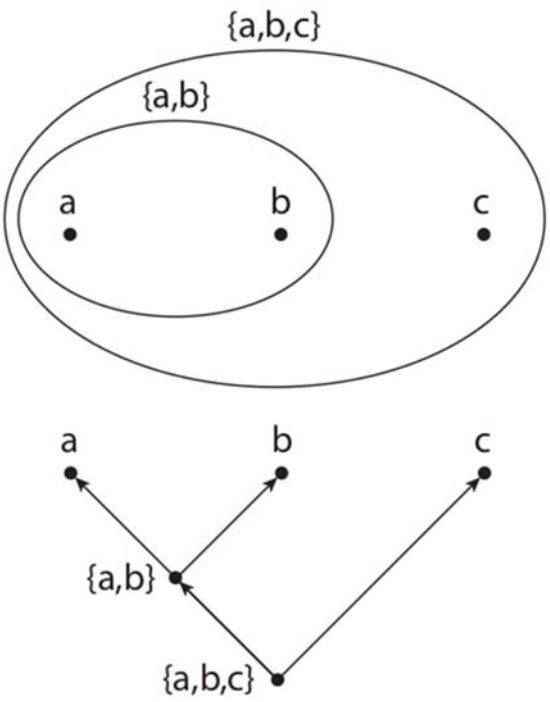


Figure 1.4. Deux représentations de la même information. En haut, un diagramme d'Euler, en bas, un arbre enraciné. Les deux représentations montrent une hiérarchie de type différenciation, avec trois objets à classer *a*, *b* et *c*, et une classification composée des deux classes {*a*, *b*, *c*} et {*a*, *b*}. La classification montre une relation de parenté entre a et b par rapport à c. Elle peut se représenter de manière parenthésée (*c,* (*a, b*)). On notera par la suite ce type de classe par une majuscule.

diversité du vivant. Ces regroupements sont en effet vrais d'un point de vue historique (un clade est une entité unique) et complets en termes de filiation lorsque l'on y ajoute une explication causale comme la théorie de l'évolution si le domaine d'étude est la phylogénie. Pour rendre complète la formulation de la théorie, il faut également statuer (iii) qu'il existe une hiérarchie unique des clades, et (iv) que de la théorie cladistique découle des méthodes dont l'objectif est la construction de cladogrammes.

2.2. Méthode cladistique

La méthode cladistique consiste en la reconstruction d'un cladogramme traduisant les relations de degré de parenté entre plusieurs (au moins trois) objets (distincts) comme des taxons ou des aires d'endémisme. La reconstruction de cette hiérarchie est fondée sur six principes (Prin, 2012 : 136-137) qui doivent être suffisamment précis pour identifier si une méthode est effectivement cladistique. Nous présentons ici notre interprétation de ces principes, dans un ordre différent de leur formulation initiale. Les deux premiers principes détaillent la structure générale de type *analyse/synthèse* cartésienne que doit prendre la méthode cladistique. L'analyse implique la décomposition d'un problème général complexe en problèmes partiels plus simples à résoudre. De son côté, la synthèse implique que les solutions partielles sont combinées, sommées, ou agrégées en une solution générale.

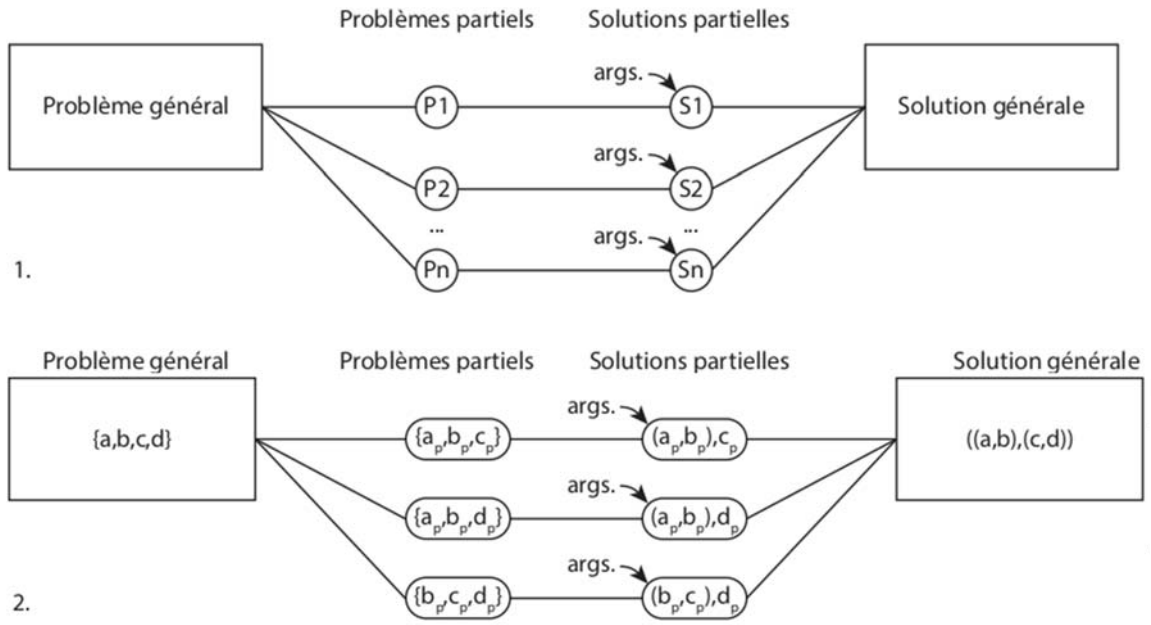


Figure 1.5 1 – Processus général d'analyse/synthèse. Le problème général complexe est décomposé en problèmes partiels plus simples. Chaque problème partiel est résolu par une solution partielle argumentée (args.). La synthèse de toutes les solutions partielles donne une solution générale. 2 – Processus d'analyse/synthèse dans le cadre de la méthode cladistique. Le problème général est de construire un cladogramme sur un ensemble d'objets, ici {*a*, *b*, *c*, *d*}. Ce problème général est décomposé en problèmes partiels consistant à rechercher les relations de parenté entre les parties des objets de l'ensemble (au moins trois). Chaque solution donne une relation de degré de parenté où deux parties sont plus apparentées entre elles que par rapport à une troisième. La combinaison de toutes ces solutions partielles donne la solution générale : le cladogramme.

Les deux principes suivants sont utiles pour définir comment obtenir des arguments pour la construction des solutions partielles aux problèmes cladistiques, et la forme que ces arguments doivent prendre. Les deux derniers principes décrivent comment synthétiser l'ensemble des solutions partielles en une solution générale. On montera comment il est possible de retrouver à partir de ces principes deux règles fondamentales décrites par Hennig que sont le principe auxiliaire et la règle de regroupement.

1. *Analyse du tout, synthèse des parties* (Hennig 1966; Nelson 1994; Cao *et al.* 2007; Zaragüeta i Bagils et Pécaud 2016)

La méthode cladistique est une méthode *d'analyse* qui consiste à décomposer un problème complexe en plusieurs problèmes partiels élémentaires plus simples à résoudre (figure 1.5-1). A chaque problème partiel correspond une solution partielle construite sur des arguments plus faciles à générer. Finalement, la méthode cladistique est une méthode *de synthèse* où tout ou partie des solutions partielles générées lors de l'analyse sont agrégées en une solution générale. Le problème général de la méthode cladistique est d'inférer, pour *n* objets, leur cladogramme (figure 1.5-2). L'analyse du problème des relations cladistiques entre ces *n* objets induit deux analyses complémentaires, c'est-à-dire deux types de décomposition.

Le premier est de décomposer le problème de la recherche des relations entre *n* objets en une série de problèmes demandant de rechercher les relations entre les parties de ces objets. On appelle *parties* ce en quoi peuvent être décomposés les objets de l'analyse et qui sont soumis aux mêmes contraintes ontologiques : les parties sont nécessairement des entités historiques; l'établissement de relations cladistiques entre ces parties doit être possible; les relations entre parties doivent également *in fine* être généralisables aux porteurs de ces parties. Dans le cadre de la méthode cladistique, les taxons se décomposent en parties de taxons et les aires d'endémismes se décomposent en taxons. Prenons un ensemble de trois taxons comprenant la souris *Mus musculus*, l'homme *Homo sapiens* et la rainette verte *Hyla arborea*. Les taxons peuvent se décomposer en homologues. Lors de l'analyse, on propose une homologie ((poils de *Mus musculus,* poils d'*Homo sapiens*) n'importe quelle partie de *Hyla arborea*) (la paire de parenthèses regroupant spécifiquement les poils de *Mus musculus* et d'*Homo sapiens* représente l'homologue « être un poil »), convertible en un caractère ((*Mus musculus, Homo sapiens*) *Hyla arborea*) (la

paire de parenthèse regroupant spécifiquement *Mus musculus* et *Homo sapiens* représente l'état de caractère « avoir des poils »).

Le second est de décomposer les relations. On a dit précédemment que la relation cladistique était une relation de degré de parenté dont l'unité d'information minimale est le triplet. Sachant un ensemble de quatre objets dont on cherche les relations cladistiques $\{a, b, c, d\}$ $\{a, b, c, d\}$ et représentant l'univers du discours, le problème se décompose en trois problèmes partiels consistant à rechercher les relations au sein de (1) $\{a, b, c\}$, (2) $\{a, b, d\}$ et (3) $\{b, c, d\}$. Cela n'interdit absolument pas de produire des arguments à plusieurs solutions partielles en même temps, c'est-à-dire de produire des arguments qui vont s'appliquer à plus de trois parties.

Dans le premier cas, les problèmes partiels sont de définir des relations cladistiques entre parties des n objets à classifier; dans le second cas, il s'agit de définir toutes les relations cladistiques entre trois objets parmi n. Les problèmes partiels de l'analyse cladistique sont donc de définir les relations entre trois parties (d'objets différents) parmi toutes les parties d'objets.

2. *Non circularité* (De Pinna 1991 : 372-375; Nelson 1994)

La génération des hypothèses trouve sa justification en dehors de la méthode ou de la théorie dans laquelle elles fonctionnent. Les arguments donnés pour la construction des solutions partielles doivent donc être indépendants de l'analyse cladistique. Dans le cas contraire ou des solutions partielles sont construites de manière dépendante de l'analyse, l'analyse devient circulaire. Cela se traduit notamment par l'indépendance entre les homologies : il n'est pas possible de tenir compte d'une hypothèse d'homologie pour la justification d'une autre homologie. Plus généralement, il doit y avoir une indépendance totale entre toutes les solutions partielles.

3. *Comparaison* (Prin 2012 : 137)

La comparaison est la source empirique de la connaissance des objets. Comparer des morphologies, des anatomies, des structures, des comportements ou des entités chimiques permet d'associer ou de dissocier deux objets par leurs ressemblances et leurs différences. La comparaison est donc la source fondamentale des arguments des solutions partielles. L'anatomie comparée par exemple est la source principale d'arguments pour la

construction d'homologies en paléontologie. Si les objets à classer sont incomparables car trop différents, ou s'ils sont si ressemblants qu'ils ne sont pas distinguables les uns des autres, aucun argument ne peut être proposé.

4. *Règle de regroupement* (Hennig, 1965 : 104; Nelson et Platnick, 1981 : 8; « *property-thinking* » pour Prin 2012 : 137)

La comparaison permet d'établir des similarités ainsi que des différences entre deux parties. Étendre cette connaissance empirique aux autres parties permet de faire des généralisations. Les généralisations de parties basées sur les similarités et différences s'appliquent donc à un ensemble de parties par rapport à d'autres. Cet ensemble de parties est défini par une propriété issue de la généralisation des comparaisons. La propriété minimale s'applique à deux parties par rapport à une troisième qui ne l'aurait pas, définissant l'homologue (a_p, b_p) et l'homologie ((a_p, b_p), c_p). Les propriétés basées sur la généralisation de comparaisons entre les parties sont les arguments des solutions partielles. La connaissance des relations entre objets ne peut donc s'établir que par la connaissance sur les propriétés intrinsèques, excluant l'usage de critères extrinsèques tels que l'âge de taxons. La pertinence de l'homologie pour reconstruire l'évolution repose sur l'hérédité des caractères. Une propriété intrinsèque à un taxon est soumise à l'hérédité et se transmet par filiation. Une propriété extrinsèque telle que l'âge d'un taxon n'est pas héritable. S'il n'est pas possible de faire des généralisations et d'établir des propriétés à partir de comparaisons, alors aucun argument ne peut être proposé.

5. *Maximisation de la congruence* (Hennig 1965 : 104, 1966; Patterson 1982 : 37-43; Farris 1983; De Pinna 1991 : 381-384; Nelson 1994)

Une fois les solutions partielles construites, celles-ci doivent être synthétisées en une solution générale, le cladogramme. Lors de l'analyse cladistique, l'ensemble des homologies sont standardisées en caractères. Cependant, les solutions partielles pour chaque caractère étant construites indépendamment les unes des autres, certaines peuvent se révéler être contradictoires entre-elles (incongruence). Il n'est pas possible dans ce cas-là d'agréger directement toutes les solutions partielles. Un critère est nécessaire pour choisir les solutions partielles à synthétiser. La solution générale de l'analyse cladistique est celle qui maximise la congruence ou minimise l'incongruence des solutions partielles

(*illumination réciproque;* Hennig, 1966). Partant du principe qu'une hypothèse ne peut être rejetée sans argument, et qu'une hypothèse ne peut être rejetée que par l'existence d'une autre hypothèse contradictoire, le cladogramme est la solution comportant le maximum possible des hypothèses argumentées. C'est également celle qui rejette le minimum d'hypothèses formulées. Le cladogramme est toujours la solution la plus informative possible. Plusieurs cladogrammes peuvent contenir le maximum de solutions partielles. Il peut donc y avoir plusieurs solutions générales équivalentes.

6. *Analyse logique* (Cao *et al.* 2007 : 7)

L'*analyse logique* est le dernier principe nécessaire à la construction de la méthode d'analyse cladistique en tant qu'ensemble autonome. Ce principe implique que la solution générale ne dit rien de plus que l'agrégation de toutes les solutions partielles. Il pose le fait qu'un cladogramme n'est rien d'autre que la somme des caractères mis en congruence et ne peut pas être interprété autrement. La connaissance des clades ne peut donc pas excéder la somme des connaissances sur les caractères. Il s'ensuit que ces deux structures doivent donner le même type d'information. Or, le cladogramme est une hiérarchie de clades. Tout caractère est une hiérarchie d'états où chaque état implique une hypothèse de clade. Comme tout caractère se déduit d'une homologie, l'homologie est une hiérarchie d'homologues et une relation de degré de parenté entre parties. Cladogramme, caractère et homologie sont donc des structures rassemblant et classant des entités par des relations de degré de parenté : telles et telles entités sont « *les mêmes* » à tel ou tel niveau de généralisation.

Deux sous-principes méthodologiques définis par Hennig peuvent être déduits de ces principes : la règle de regroupement et le principe auxiliaire (« *grouping rule* » et « *auxiliary principle* »; Hennig 1966 : 104; Wiley *et al.* 1991 : 14). La règle de regroupement peut être énoncée de la manière suivante : l'existence d'un clade ne peut être justifiée que par une synapomorphie, c'est-à-dire que ce clade ne peut exister que s'il y a un ou plusieurs homologues qui le justifient. Les homologies sont des hypothèses de relation de degré de parenté entre parties. Le remplacement des parties par leurs porteurs permet de passer de l'homologie composée d'homologues (solution partielle non-standardisée) au caractère composé d'états (solution partielle standardisée). En d'autres termes, une relation de degré de parenté entre des parties est toujours nécessaire pour justifier une relation de degré de parenté entre leurs porteurs. Le corolaire est qu'une

relation de degré de parenté entre des entités ne peut pas exister s'il n'y a pas au préalable de relation de degré de parenté entre leurs parties (Prin, 2012 : 149). L'homologie ((poils de *Mus musculus*, poils de *Homo sapiens*) n'importe quelle partie de *Hyla arborea*) se standardise en caractère ((*Mus musculus*, *Homo sapiens*) *Hyla arborea*). Une analyse cladistique inclut ce caractère en tant que solution partielle. Au sortir de l'analyse, le clade {*Mus musculus*, *Homo sapiens*} a été retenu dans le cladogramme. Ce clade est soutenu par l'état de caractère « avoir des poils » généré à partir de l'homologue « être un poil ». Cet état est donc une synapomorphie du clade {*Mus musculus*, *Homo sapiens*}. De son côté, le principe auxiliaire de Hennig énonce qu'une hypothèse de relation de degré de parenté entre porteurs doit toujours être considérée comme vraie à moins d'être contredite par d'autres. Pour résumer, la règle de regroupement implique qu'*une relation entre parties est nécessaire pour avoir une relation entre porteurs*, et le principe auxiliaire qu'*une relation entre parties est suffisante pour avoir une relation entre porteurs* (Prin, 2012 : 6). L'application de ces deux règles implique que le nombre minimal de relations doivent être rejetées pour construire la solution générale. Le corolaire est que le cladogramme est composé de l'ensemble maximal de relations congruentes.

Plusieurs méthodes d'analyse se réclament de la cladistique : la compatibilité (Wilson 1965; Le Quesne 1969, 1972; Estabrook *et al.* 1976a, 1976b), la parcimonie standard (Kluge et Farris 1969; Farris 1970; Fitch 1971; Farris 1977; Camin et Sokal 1965) et la 3ia (Nelson et Platnick 1991). Leurs différences proviennent de la représentation des solutions partielles ainsi que de la manière de les synthétiser en un cladogramme. Le problème général et les problèmes partiels sont en revanche communs à toutes les méthodes. Leur comparaison n'est cependant pas l'objet de ce chapitre. La partie suivante montre comment appliquer la 3ia lors d'une analyse de taxons. Les spécificités de la 3ia sont le traitement des caractères ainsi que la manière de construire et d'interpréter un cladogramme. On développera son usage dans le contexte phylogénétique et ses rapports avec les principes de la méthode cladistique.

3. Reconstruction phylogénétique

Le but de la 3ia en tant que méthode cladistique est la construction de relations de parenté. Nous développerons dans cette partie la 3ia dans le cadre de la recherche des relations de parenté entre UT, aboutissant à un cladogramme de taxons : la phylogénie. Nous présenterons le schéma standard d'une analyse cladistique utilisant la 3ia sur un nombre donné d'UT par étapes, selon le schéma présenté en figure 1.6. La première étape est la construction d'hypothèses de relations de parenté entre parties d'UT. Les relations de parenté entre parties d'UT forment une classification hiérarchique, l'homologie. La seconde est la standardisation des homologies. Les parties d'UT sont remplacées par les UT elles-mêmes. Les relations de parenté entre parties permettent de déduire les relations de parenté entre les porteurs de ces parties. Les homologies sont standardisées en caractères. Les caractères sont composés de relations de parenté entre UT, ils sont donc des hypothèses de clades d'UT. Un caractère implique au minimum trois UT pour former une relation de degré de parenté. La troisième étape est la décomposition des caractères en relations de degré de parenté minimales, les triplets. Un caractère peut contenir plusieurs relations de ce type, le triplet n'en contient toujours qu'une. Un triplet contient exactement

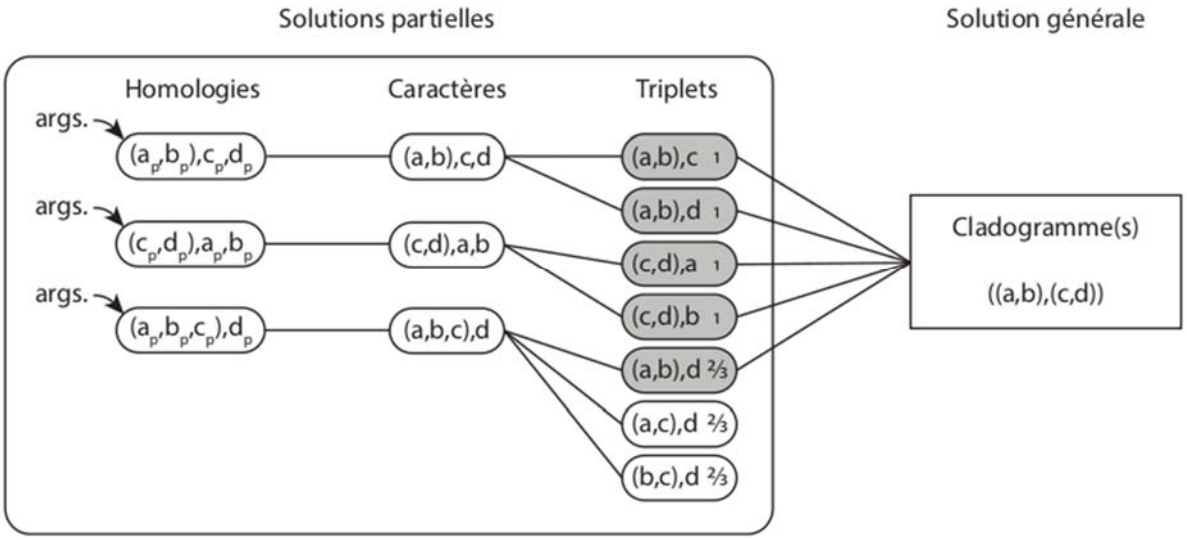


Figure 1.6. Schéma général de la méthode d'analyse cladistique en 3ia. Les homologies sont des structures établissant des relations entre parties d'UT. Les parties d'UT sont ensuite remplacées par les UT elles-mêmes lors de l'étape de standardisation pour former des caractères. Les caractères composés de relations entre UT sont décomposés en triplets, les relations minimales entre UT. Le cladogramme est la solution générale au problème des relations de parenté entre UT. Il est construit comme l'ensemble maximisant le nombre de triplets congruents.

trois UT, et spécifie que deux de ces UT sont plus proches parentes entre elles que de la troisième. L'indépendance entre les relations est maintenue par une pondération fractionnaire. La maximisation de la congruence est la quatrième étape. L'ensemble maximal de triplets est combiné pour former le cladogramme. Ce principe implique que la solution optimale est le cladogramme composé de l'ensemble maximal d'hypothèses, c'est-à-dire l'ensemble maximal de relations de parenté congruentes, ce qui est en accord avec la règle de regroupement et le principe auxiliaire. La cinquième étape est celle de l'attribution des états aux nœuds du cladogramme, et la sixième étape est celle de l'interprétation du cladogramme pour permettre par exemple la datation de ses clades.

3.1. Hypothèses

3.1.1. Unité taxonomique

La première tâche qu'une théorie doit accomplir est la définition des entités théoriques nécessaires à sa construction. La théorie cladistique manipule des entités qui peuvent se ranger en deux types : les objets et les concepts. Les objets sont les entités que manipule la théorie, en l'occurrence il s'agit des entités à classer. Les objets de la cladistique appliquée à la phylogénie sont les unités taxonomiques (UT). À opposer aux UT, les clades sont des TH, et forment la solution générale au problème de trouver les relations de parenté entre UT. Sur la figure 1.4, *a*, *b* et *c* sont des UT, et {*a*, *b*} et {*a*, *b*, *c*} sont des TH. Les UT considérées comme des objets cladistiques doivent remplir certains critères.

3.1.1.1. Constituants

Un *sémaphoronte* (étymologie : « *sêma* » signe, « *phore* » porteur, « *ontos* » ce qui est) est défini comme un porteur d'attributs. Il est une entité empirique sur laquelle il est possible de procéder à l'observation et à la comparaison (principe 3). Le sémaphoronte est défini par Hennig (1950, 1966) comme un organisme à un instant précis, ce qui permet d'observer ses attributs, comme la morphologie, l'anatomie, le comportement, le génome ou encore le métabolisme. Sur la figure 1.7, α_1, α_2 et α_3 sont des sémaphorontes. Plusieurs sémaphorontes peuvent se généraliser en un *holomorphe* (étymologie : « *hólos»* entier,

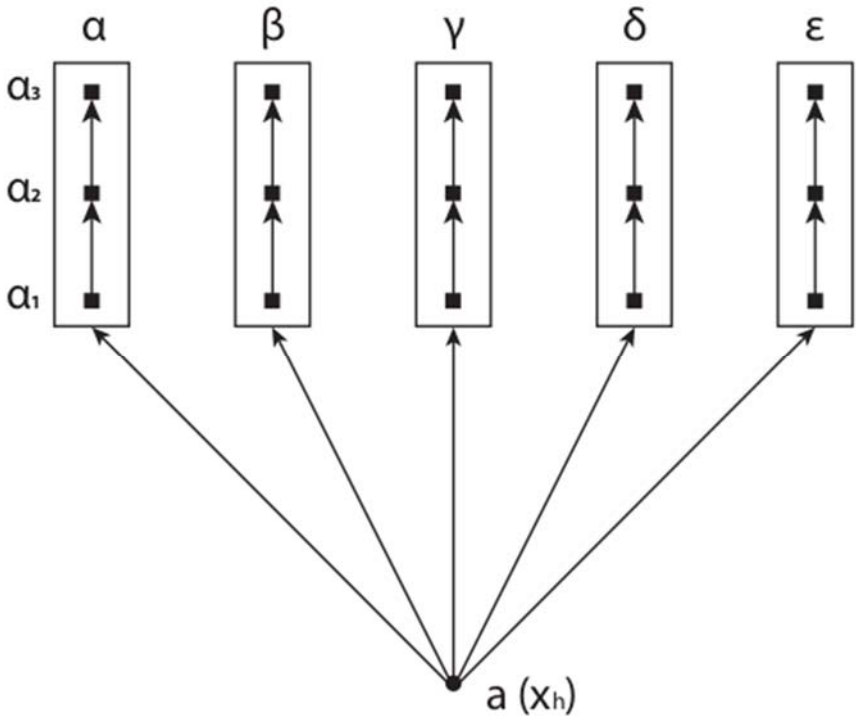


Figure 1.7. Composition d'une UTM. α_1 est un sémaphoronte. α_1, α_2 et α_3 sont les différents stades de vie de l'holomorphe α. Aucune relation de degré de parenté ne peut être établie entre α, β, γ, δ et ε car il n'y a aucun homologue que certains posséderaient et pas d'autres. a est une UTM qui rassemble α, β, γ, δ et ε par la présence de l'homologue X_h. α, β, γ, δ et ε sont donc plus proches parents entre eux que de tout autre holomorphe. Une UTM comme toute autre classe peut se définir en intension (X_h) ou en extension $\{\alpha, \beta, \gamma, \delta, \varepsilon\}$.

« *morphé* » forme). Les sémaphorontes caractérisent les différents stades d'un holomorphe (fig. 1.7, α, β, γ, δ, ε) qui définit le cycle de vie d'un organisme. C'est par le concept d'holomorphe que l'on peut par exemple relier une chenille à son papillon. Le concept d'holomorphe permet d'acquérir des connaissances sur l'embryologie ou le cycle de vie de l'organisme. Il peut ne pas être anodin de construire l'holomorphe d'un organisme ayant subi une métamorphose drastique. Un groupe d'holomorphes forme une UT. La formation d'UT à partir de sémaphorontes rassemblés en holomorphes peut comporter des difficultés. Une mauvaise distinction des holomorphes peut avoir pour conséquence de produire plusieurs UT distinctes correspondant chacune à un stade, par exemple une UT de larves et une UT d'imagos correspondant aux mêmes holomorphes. Un autre exemple est la mauvaise distinction du dimorphisme sexuel due à un manque de connaissances sur les holomorphes, qui peut amener à ériger une UT pour les femelles et une UT pour les mâles (e. g. Tintant 1977). Ces exemples entraînent la création d'UT paraphylétiques puisque les groupes sont incomplets. La théorie cladistique ne reconnaît que les entités

monophylétiques mais ne donne pas d'indications sur la manière d'argumenter la monophylie des objets qu'elle manipule. Les arguments sont à chercher en dehors, notamment en biologie des organismes. L'étude d'UT fossiles amène une incertitude intrinsèque sur leur statut monophylétique, puisqu'il manque les informations sur leur cycle de vie et la reproduction.

3.1.1.2. Structure

Une Unité Taxonomique Minimale (UTM) ou *Least-Inclusive Taxonomic Unit* (LITU; Pleijel et Rouse 2000) est une UT à l'intérieur de laquelle il est impossible de faire des regroupements entre individus (fig. 1.7 : a). Il est donc impossible par définition d'ériger un taxon qui serait inclus dans une UTM. En cladistique, cela implique également qu'il est impossible de définir des relations de degré de parenté à l'intérieur d'une UTM, et qu'il existe une relation de parenté exclusive entre tous ses constituants. L'UTM est donc l'objet minimal utilisable en méthode cladistique. La monophylie d'une UTM peut être supposée grâce à deux critères : (1) la présence exclusive d'un homologue (un homologue exclusif à une UTM est une autapomorphie (étymologie : « *autó»* propre, « *apo »* dérivé de, « *morphos* » forme; Rosen 1979; Nelson et Platnick 1981) notée x_h sur la figure 1.7, et (2) l'UTM peut être discriminée de tout autre UTM par une combinaison unique d'homologues (Wheeler et Platnick 2000). Une UTM est toujours relative à un état de connaissance à un moment donné.

La cohérence de la méthode cladistique appliquée aux taxons se juge à travers le prisme causal du processus de filiation. L'UTM peut être considérée comme un niveau séparant un domaine infra-UTM d'un domaine supra-UTM (Prin, 2012 : 31-32). Le domaine infra-UTM est caractérisable par son accessibilité empirique, c'est-à-dire que l'observation directe est possible, et par sa réversibilité (par exemple à travers les changements de fréquence allélique). S'il y a réversibilité, alors il n'y a pas d'historicité possible et il n'est pas possible de définir d'entité historique, dont les homologues. Ce niveau est soumis à des fluctuations sans structuration des populations par un homologue. L'UTM est donc le niveau historique élémentaire en dessous duquel la cladistique n'a pas prise car il n'y a par définition pas d'homologues. Si dans une population, il est possible de mettre en évidence un trait qui se fixe, alors une UTM se créé. Dans le niveau supra-UTM, les relations inter-UTM sont cladistiques car il existe des homologues permettant de

spécifier des relations de parenté. L'explication darwinienne de la diversité du vivant par mutation et héritabilité n'a cours qu'à partir du niveau de l'UTM. Le but de la phylogénie étant de rendre compte de la diversité du vivant, c'est donc à ce niveau que se passe l'analyse cladistique.

L'UTM n'est cependant pas le seul type de taxon pouvant être utilisé comme UT lors d'une analyse phylogénétique. Puisque le critère majeur est la monophylie, il est tout à fait possible de procéder à une analyse cladistique sur des UTM, et prendre un clade du cladogramme généré pour s'en servir comme UT d'une analyse subséquente. Les clades dans une analyse sont des TH, mais peuvent être utilisés pour d'autres analyses en tant qu'UT. Mais un organisme peut-il être utilisé en tant qu'UT (sémaphoronte ou holomorphe) ? Rien ne nous dit qu'un organisme puisse être considéré comme monophylétique (si on parle de la monophylie d'un individu par rapport à ses cellules pour un métazoaire, n'importe quel organisme devient paraphylétique dès lors qu'il a une descendance distincte de lui, en sachant de surcroît que la plupart des organismes perdent de leurs cellules en permanence). Dire d'un organisme qu'il est monophylétique implique sa réflexivité, c'est-à-dire qu'il soit plus proche parent de lui-même que de tout autre UT. Or, seule la présence d'un homologue peut confirmer la monophylie d'une entité. La découverte d'un homologue permet alors d'ériger une UTM contenant un unique organisme. Si l'holomorphe α de la figure 1.7 est indifférenciable en termes d'homologues des autres holomorphes β, γ, δ, ε, alors il n'est pas possible d'établir que α est plus proche parent de lui-même que de β, γ, δ, ε. α n'est donc pas une UTM car α seul est paraphylétique; l'UTM a doit inclure α, β, γ, δ, ε. La relation de degré de parenté minimale est « *a – et donc ses constituants – est plus proche parent de lui-même que d'aucun autre taxon* ». D'autres problèmes discutés précédemment comme le dimorphisme sexuel peuvent, en pratique, entraîner la création d'UT paraphylétiques.

3.1.2. Homologie

Le choix de l'ensemble d'UT à analyser impose la formalisation du problème général « *quelles sont les relations de parenté entre les UT de cet ensemble ?* ». Établir des relations entre UT étant un problème complexe, il est décomposé une première fois en un ensemble de problèmes partiels du type « *quelles sont les relations de parenté entre les parties d'UT de cet ensemble ?* ». La réponse à cette question correspond à la première

étape de l'analyse cladistique, la construction d'*hypothèses d'homologies*. La structure de ces hypothèses, vu la question posée, est nécessairement hiérarchique. Elles nécessitent en premier lieu de préciser les parties et les UT auxquelles elles se référent : « *poils d'Homo sapiens* », *« mitochondrie d'Helianthus annuus* » sont des exemples de parties d'UT. Une homologie est une classification dont les classes sont les homologues; les homologues spécifient une relation de degré de parenté comme suit : si *x*, *y*, et *z* sont trois parties quelconques, alors *x* et *y* sont plus proches entre eux que de *z* si et seulement s'il existe au moins un homologue contenant *x* et *y* mais pas *z*. Un homologue est donc un clade de parties. Si la théorie de l'évolution est vraie et que les taxons se différencient, et sachant que les taxons sont composés d'homologues, il s'ensuit que les homologues évoluent également et se différencient. Un homologue peut donc se différencier et donner naissance à de nouveaux homologues.

3.1.2.1. Structure

L'hypothèse d'homologie est une relation de degré de parenté entre parties d'UT : sa structure est donc hiérarchique au sens classificatoire du terme (Cao *et al.* 2007). Une classification hiérarchique est une structure mathématique vérifiant trois conditions (Margush et McMorris 1981; Barthélémy et Guénoche 1991) : (i) il y a une unique classe qui contient tous les éléments à classer; cette classe s'appelle la racine; elle se justifie par la monophylie du vivant; (ii) pour chaque élément à classer, il y a une classe qui contient exclusivement cet élément; ces classes s'appellent les singletons; elles se justifient par la monophylie a priori des taxons; une UT ou partie d'UT implique donc nécessairement son singleton associé, qui sera sous-entendu dans les parties suivantes; et (iii) l'intersection entre deux classes quelconques est soit vide, soit est l'une de ces deux classes. En d'autres termes, il n'y a jamais d'intersection entre deux classes, qui sont toujours ou disjointes, ou incluses l'une dans l'autre. Dans la pratique cladistique, on appelle feuille la classe qui ne contient qu'un seul objet à classer (UT, parties d'UT) et nœud une classe de la classification spécifiant une relation de degré de parenté (TH, clade, homologue, état de caractère).

C'est la *comparaison* des parties d'UT puis leur *généralisation en propriétés* qui permettent l'établissement des hypothèses d'homologie et des homologues. La comparaison permet par l'observation d'établir les ressemblances et les différences entre deux parties d'UT. Ces observations sont ensuite généralisées à l'ensemble des parties

(Frege 1971; Lebbe 1991; Vignes 1991). Prenons 4 UT {*a*, *b*, *c*, *d*} (fig. 1.8). Décomposons ces UT en parties et comparons leurs parties, respectivement a_p, b_p, c_p et d_p (fig. 1.8-1). Lors de la comparaison, on remarque une ressemblance entre a_p et b_p, entre a_p et c_p, entre b_p et c_p. Ces observations permettent de généraliser a_p, b_p et c_p en un homologue Y_h (fig. 1.8-2). L'homologue implique que a_p, b_p et c_p sont plus proches parents entre eux qu'aucun ne l'est de d_p. D'autres homologues peuvent être construits avec les contraintes suivantes : ils doivent être informatifs, c'est-à-dire qu'ils doivent inclure au moins deux parties et exclure au moins une partie pour établir un degré de parenté – les autapomorphies et les parties invariantes ne peuvent donc faire partie de l'analyse –, et ne doivent pas être redondants avec un autre homologue (inclut les mêmes parties et exclut les mêmes parties). Il doit nécessairement y avoir un homologue informatif. Lorsque tous les homologues ont été construits, il faut pour obtenir une hiérarchie complète construire la racine et les singletons (fig. 1.8-3). La hiérarchie prend alors le nom d'homologie. Par exemple, on remarque que *a*, *b* et *c* ont une ressemblance sur une de leurs parties : ils ont tous les trois une coquille composée de deux valves plus ou moins symétriques, et *d* possède une coquille univalve. La ressemblance est généralisée en une classe homologue nommée « *coquille bivalve* ». La propriété « *coquille bivalve* » (intension de l'homologue) implique une parenté entre les parties a_p, b_p et c_p (extension de l'homologue). Les classes de l'homologie sont les homologues. On considérera ici que la racine est également un homologue. Cependant, l'homologue racine revêt un statut un peu particulier. Il implique une relation de parenté entre toutes les parties et pas de degré de parenté, c'est donc une classe non-informative mais nécessaire à la structure de l'homologie qui spécifie l'univers du discours. Puisqu'elle est toujours présente, il est possible de ne pas spécifier son intension. Dans notre exemple, on peut soit laisser non-spécifiée l'intension de la racine, soit la spécifier, par exemple par la propriété « *coquille univalve* », la coquille bivalve se formant à partir d'une coquille univalve au cours du développement. On peut également nommer l'hypothèse d'homologie, par exemple « *structure de la coquille* ». Il faut également toujours avoir en tête le principe de *non-circularité* qui implique l'indépendance totale des homologies les unes vis-à-vis des autres.

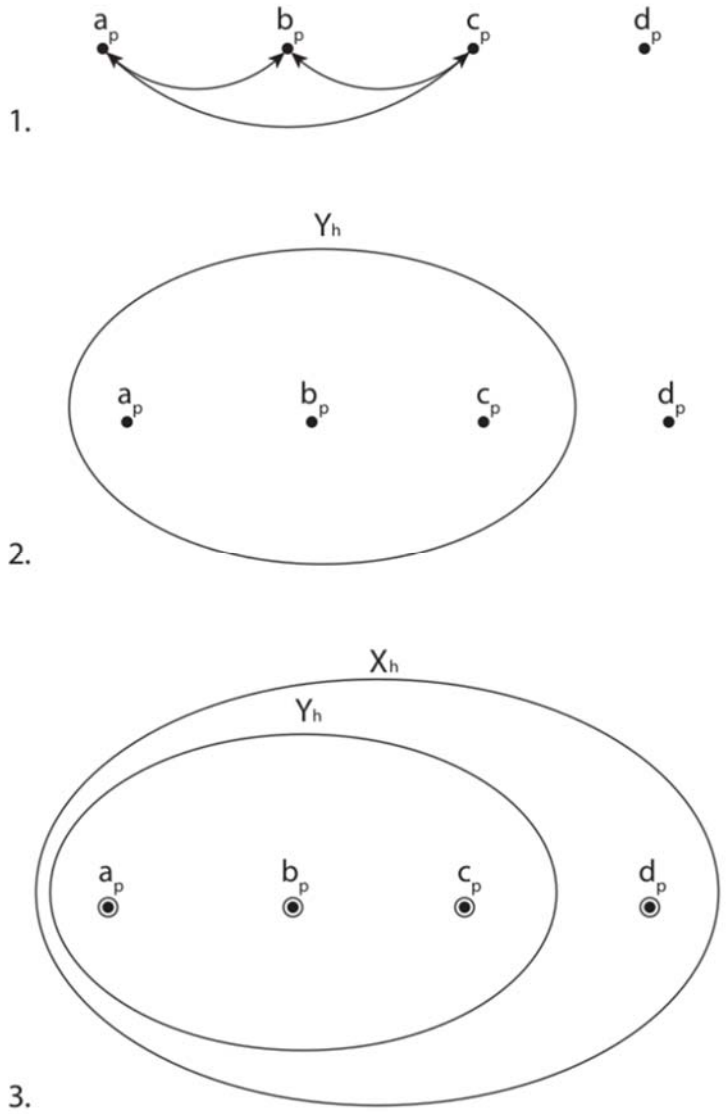


Figure 1.8. Etapes de construction d'une homologie à deux homologues. (1) On procède à la comparaison de quatre parties *p* de quatre UT *a*, *b*, *c* et *d*. Lors de la comparaison, on remarque des ressemblances entre a_p et b_p, entre b_p et c_p et entre a_p et c_p (flèches noires). (2) Ces ressemblances sont généralisées en un homologue Y_h incluant a_p, b_p et c_p. Les étapes (1) et (2) se répètent jusqu'à ce qu'on ne puisse plus générer d'homologue. (3) On ajoute un homologue racine X_h qui inclut toutes les parties et un singleton pour chacune des parties. L'ensemble des singletons de parties a_p, b_p, c_p et d_p et les homologues X_h et Y_h forment une homologie.

La figure 1.9 présente quatre homologies différentes pour un même ensemble de quatre parties a_p, b_p, c_p et d_p. Les homologies 1 et 2 varient par les parties qui les composent. L'homologie 1 propose sur l'ensemble un homologue U_h qui relie les parties a_p et b_p par rapport aux parties c_p et d_p. Sachant deux nœuds, dont l'un inclut l'autre, on dit que le nœud qui inclut est ancestral, et le nœud qui est inclus est dérivé (Hennig, 1965 : 102). L'homologie 2 présente une configuration identique à un point près : la partie d_p est une

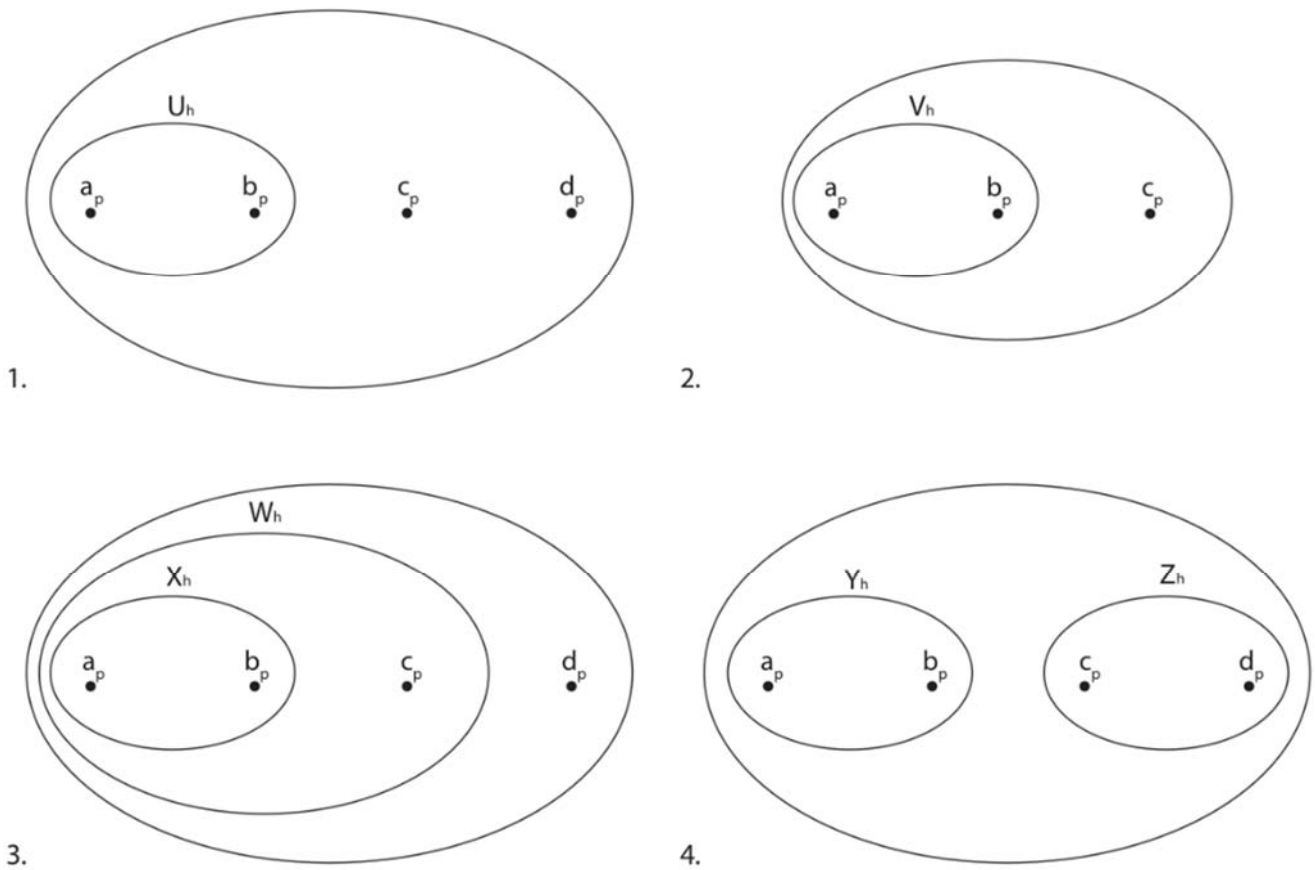


Figure 1.9. Quatre homologies différentes à partir du même ensemble a_p, b_p, c_p et d_p. L'homologie 1 présente un homologue U_h (a_p, b_p). L'homologie 2 présente également un homologue V_h (a_p, b_p). d_p est manquant et ne peut pas être observé. L'homologie 3 présente un homologue X_h (a_p, b_p) inclus dans un homologue W_h (a_p, b_p, c_d). X_h et Y_h incluent moins de deux nœuds internes, ils sont orthologues. L'homologie 4 présente un homologue Y_h (a_p, b_p) et un homologue Z_h (c_p, d_p) qui sont disjoints. Ici, la racine est paralogue : elle contient deux nœuds internes, Y_h et Z_h, qui sont indépendants.

donnée manquante. Ce cas de figure arrive lorsqu'il a été impossible d'observer d_p. Il est commun d'avoir des données manquantes en paléontologie. Par exemple, mettons que l'homologie 2 traite des canaux de la valve gauche chez des rudistes bivalves, et soit composée de l'homologue racine « *absence de valve gauche percée de canaux* » (l'homologue racine aurait également pu ne pas être spécifié) et de l'homologue V_h « *valve gauche percée de canaux* ». d_p est porté par un taxon fossile *d* dont on ne retrouve que des valves droites. d_p est dans ce cas une donnée manquante : il est impossible de savoir si la valve gauche manquante était ou non percée de canaux, et n'est assignable à aucun homologue. d_p ne participe donc tout simplement pas à l'homologie (Zaragüeta i Bagils et Bourdon 2007). Un autre cas de figure souvent confondu peut également émerger lors de la construction des hypothèses d'homologie. Reprenons une homologie composée d'un

homologue racine « *absence de valve gauche percée de canaux* » incluant un homologue « *valve gauche percée de canaux* ». Essayons maintenant d'incorporer le taxon *Escherichia coli* à cette homologie. Cette fois-ci, aucune partie ne manque puisque toutes les observations nécessaires peuvent être menées. Le placement de d_p dans l'homologie revient à se poser la question « *Escherichia coli possède-t-il une valve gauche percée de canaux ?* ». Ce problème est généralement connu comme un problème de *donnée non-applicable*. On peut considérer que la question de la structure de la valve gauche n'est pas applicable au taxon *Escherichia coli*. Que si *E. coli* avait une valve gauche, peut-être celle-ci serait percée de canaux. Eliminer une partie de taxon dans le cas des données manquantes est justifié : on perd de l'information car on n'a pas pu faire les observations sur cette partie. Dans le cas de *E. coli* par contre, toutes les observations nécessaires ont été faites. Il n'est donc pas souhaitable de perdre de l'information en supprimant cette partie de taxon de l'homologie. La réponse à la question « Escherichia coli *possède-t-il une valve gauche percée de canaux ?* » est évidemment non. Il n'est pas nécessaire de chercher une éventuelle structure homologue à la valve gauche des bivalves chez *Escherichia coli* : la comparaison de ce taxon par ses parties avec tous les autres montre qu'il ne possède pas cet homologue. *E. coli* fait en revanche *a priori* partie du monde vivant, et d'autres parties peuvent être comparées. Il est donc légitime d'inclure le taxon par ses parties dans la racine. En résumé, on sait que *E. coli* par ses parties peut être inclus dans la racine, mais qu'il ne peut être inclus dans l'homologue « *valve gauche percée de canaux* ». Il en résulte que le placement d'*E. coli* par ses parties à l'homologue racine n'est aucunement ambigu. Définir par des propriétés négatives est très souvent source d'erreur. En transposant à notre exemple le raisonnement par l'absurde de Platnick (Nelson et Platnick 1981; Platnick 2013), je peux dire que mon bureau non plus ne porte pas de valve gauche. Mais inclure mon bureau dans une analyse phylogénétique n'a aucun sens : l'intension est donc inexacte car trop permissive. Un quelconque homologue ne devrait ainsi jamais être appelé « *absence de x* » mais devrait être reformulé ou tout simplement être laissé sans intension dans le cas de l'homologue racine. En résumé, les données manquantes impliquent pour une partie donnée que *tous les homologues sont possibles* (Zaragüeta i Bagils et Bourdon 2007). La partie n'est alors pas incorporée à l'homologie. Les données non-applicables

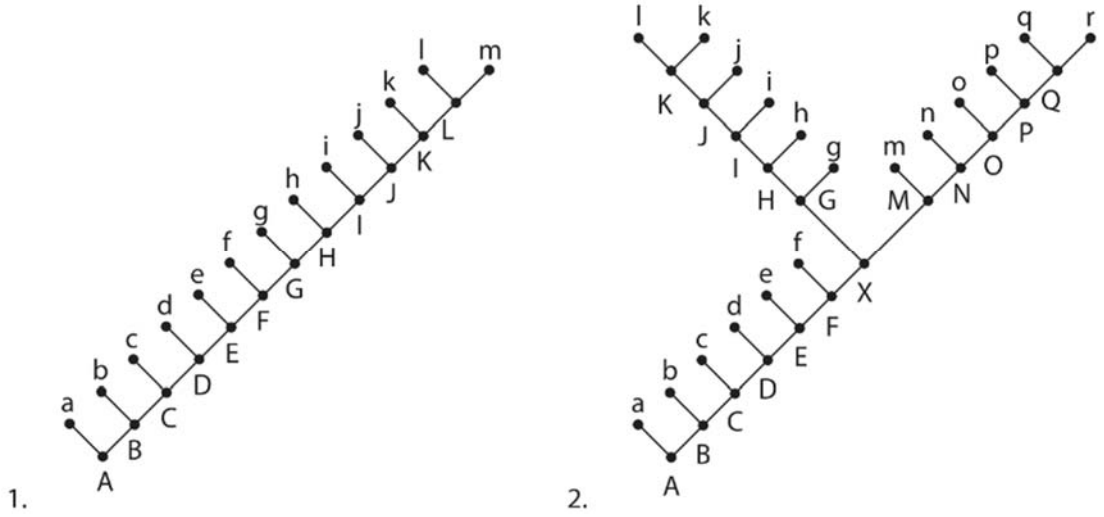


Figure 1.10. Illustration des nœuds orthologues et paralogues. (1) Orthologie, tous les nœuds internes sont en relation d'inclusion les uns par rapport aux autres. Aucune diversification ne peut être postulée. (2) Hiérarchie présentant un nœud paralogue *X* d'où partent deux lignées d'orthologues *G* et *M* indépendantes l'une de l'autre.

impliquent pour une partie donnée *qu'aucun homologue n'est possible*. Si la partie ne peut être incluse dans aucun homologue, il se branche directement à la racine qui par définition inclut tout.

Nous avons discuté des homologies 1 et 2 de la figure 9, qui sont des *homologies binaires* : elles ne contiennent que deux homologues dont la racine. Les homologies 3 et 4 contiennent elles trois homologues agencés de différentes façons. L'homologie 3 présente un homologue X_h (a_p, b_p, c_p) incluant l'homologue W_h (a_p, b_p). Ces homologues sont des *orthologues* (étymologie : « *orthós »* droit, *« lógos »* étude). Le terme orthologue décrit un nœud interne qui inclut directement au maximum un autre nœud interne (Cao 2008). Le terme *orthologie* décrit un ensemble d'orthologues inclus les uns dans les autres : l'homologie 3 de la figure 1.9 est donc une orthologie. Ce terme peut être rapporté à celui d'*anagenèse* qui implique l'évolution sans diversification d'une lignée. Un exemple est présenté en figure 1.10-1. Les feuilles font ici partie d'une orthologie, ce qui ne permet pas de postuler un épisode de diversification. C'est également la configuration la plus informative : à l'intérieur d'une orthologie, on peut prendre deux nœuds internes au hasard et il sera toujours possible de dire que l'un inclut l'autre. L'homologie 4 présente un homologue Y_h (a_p, b_p) et un homologue Z_h (c_p, d_p). L'homologue racine inclut deux homologues : c'est un *paralogue* (étymologie : « *pará »* à côté de, *« lógos »* étude : nœud

d'où partent plusieurs lignées parallèles). Le terme paralogue décrit un nœud interne qui inclut directement au minimum deux nœuds internes (Gray et Fitch 1983; Fitch 2000). En figure 1.10-2, à partir du nœud paralogue *X* (strictement équivalent à {*g*, *h*, *i*, *j*, *k*, *l*, *m*, *n*, *o*, *p*, *q*, *r*} divergent deux *lignées indépendantes G* et *M*. Ces lignées sont indépendantes puisque l'on peut prendre n'importe quel nœud de la première lignée et n'importe quel nœud de la seconde lignée et il sera impossible d'en retirer une information d'inclusion. Un arbre ne contenant aucun nœud paralogue est toujours plus informatif qu'un arbre aux feuilles identiques comprenant un paralogue. Dans le premier cas, chaque couple de nœuds est informatif (inclusion) tandis que le second présente des couples de nœuds dont on ne peut rien dire (disjonction). En phylogénie, on peut associer un paralogue à un épisode de *diversification*.

3.1.2.2. Arguments

Les homologues et les homologies nécessitent des arguments pour les construire. Si l'homologie est le pilier de la théorie cladistique, c'est également une notion ancienne dont on fait remonter l'ascendance à Pierre Belon (1555), même si des précurseurs bien plus lointains peuvent être trouvés jusqu'à Aristote (Tassy 2005). Le but de Pierre Belon était l'étude des oiseaux, et notamment de leurs os. Pour ce faire, il procéda à leur comparaison par rapport à ceux des humains pour « *faire apparaitre combien l'affinité est grande des uns aux autres* » (fig. 1.11). Cette première étude d'anatomie comparée lui permit d'assigner à un os d'oiseau et un os d'humain le même symbole (des lettres en l'occurrence). La lettre *k* par exemple désigne les vertèbres, de l'oiseau comme de l'humain. Cette « *affinité* » entre des parties de taxons est la relation de parenté explicitée tout au long de ce chapitre. Définir une affinité entre deux parties lui permit de les rassembler sous un concept, repris et détaillé sous divers noms par de nombreux auteurs, dont l'analogie d'Etienne Geoffroy Saint-Hilaire (1818, 1830) et l'homologie de Richard Owen (1848).

L'argument dont disposait Belon pour montrer l'affinité entre ces parties était la *similarité*. Le principe de *comparaison* permet d'étudier deux parties et de mettre en évidence une similarité sur leur forme, leur structure, etc. Elle nécessite de faire abstraction des différences qu'il peut y avoir entre différentes parties : on peut par exemple reconnaître la coquille bivalve chez le bénitier *Tridacna maxima* et la moule *Mytilus edulis* malgré leur

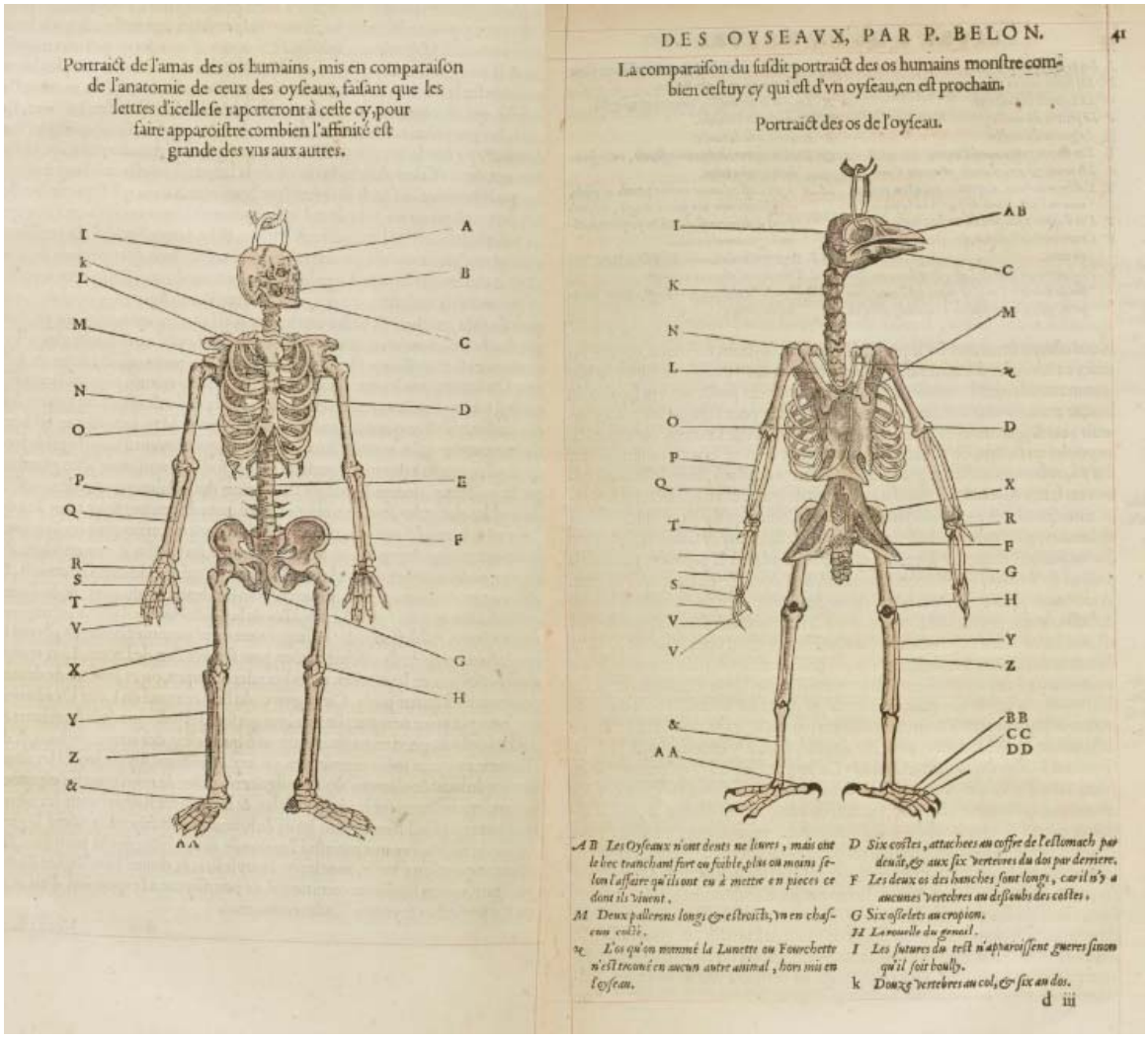


Figure 1.11. Illustration tirée de Belon, 1555 : *Histoire de la nature des oyseaux, avec leurs descriptions et naïfs portraicts retirez du naturel* (p. 40-41). L'affinité entre l'humain et l'oiseau est mise en évidence par l'utilisation du même symbole pour une partie de l'un et de l'autre. Le squelette de l'oiseau est modifié pour mettre en évidence l'identité des parties avec l'humain.

différence de taille et de forme. Cette coquille est composée des deux mêmes valves gauches et droites qui sont liées par une charnière, un ligament et sont disposées latéralement par rapport à l'animal. Ce dernier critère permet de les distinguer des coquilles des brachiopodes composées de deux valves qui sont, elles, insérées ventralement et dorsalement par rapport à l'animal : ce ne sont donc pas les *mêmes* coquilles, elles ne se rassemblent pas en un homologue. La similarité peut s'appliquer à n'importe quelle caractéristique des parties : morphologie, anatomie, biochimie, génétique, ontogénie, éthologie, etc. On remarque que ce critère implique la similarité structurale des parties; cela implique que la fonction de ces parties ne saurait être utilisée. De simples modifications

dans les proportions de parties peuvent en effet changer leur fonction, malgré l'homologie. La similarité peut également s'appliquer sous le concept de *principe des connexions* (Saint-Hilaire 1830). Ce concept permet d'identifier l'homologie par l'agencement des parties les unes par rapport aux autres. C'est ce principe qui permet, par exemple, de reconnaître les dents qui composent la charnière des mollusques bivalves en étudiant leur arrangement mutuel. Il permet également de reconnaître les homologies des os chez les vertébrés, ou encore les gènes homologues chez tous les êtres vivants. Se tromper dans une hypothèse d'homologie revient à se tromper dans l'hypothèse de parenté entre trois parties.

Un point de détail peut être apporté sur les cas de *perte* d'une propriété. On a dit précédemment que l'absence pouvait être évacuée de la définition des homologues racines. Cette « *absence* » peut également se retrouver comme hypothèse d'homologue dérivé. Il peut être légitime de se poser la question « *est-ce que ce que j'observe est une absence de preuve ou la preuve d'une absence ?* ». Il apparaît cependant légitime de coder une absence en tant qu'homologue dérivé, résultant par exemple d'une perte au cours de l'évolution, si l'on possède l'assurance que toutes les parties ont bien pu être observées et que l'on n'est pas face à un cas de donnée manquante.

La spécification des relations d'inclusion entre les homologues d'une homologie s'appelle la *polarisation* de cette homologie. L'ontogénie est un moyen efficace de polariser une homologie. La loi biogénétique de Nelson (1978) spécifie qu'étant donné une homologie et observant un homologue comme étant plus général et un homologue comme étant moins général, l'homologue général est ancestral et l'homologue particulier est dérivé. Elle fait suite aux deux premières lois de Von Baer (1828) : (1) « *les caractères* [homologues] *généraux apparaissent avant les caractères* [homologues] *spéciaux* » et (2) « *les caractères* [homologues] *généraux puis les spéciaux se forment à partir des plus généraux* ». Spécifiée ainsi, on constate même qu'il est possible d'utiliser la loi biogénétique pour des observations n'impliquant pas l'ontogénie. On peut par exemple uniquement par la comparaison d'adultes du taxon Mollusca proposer l'homologie (« *racine* » (« *coquille* » (« *coquille bivalve* »))), où « *coquille bivalve* » est un homologue dérivé qui se différencie à partir de l'homologue « *coquille* ». On note que les arguments ontogénétiques, s'ils sont très utiles pour l'établissement d'homologies, ne sont pas infaillibles. Les cas de paedomorphose comme la néoténie de l'Axolotl *Ambystoma*

mexicanum qui peut passer sa vie à l'état larvaire sans jamais se métamorphoser en adulte en sont un exemple classique.

Le problème de polarisation illustré en figure 1.12 peut également être résolu par le *critère extra-groupe* (« *analyse de groupes apparentés* » : Hennig, 1966; « *related groups* » : Maslin 1952; « *outgroup* » : Wiley 1976). Le critère extra-groupe implique une connaissance, *a priori* de l'analyse, de groupes apparentés au groupe analysé (intra-groupe) : que ces groupes apparentés porte l'homologue ancestral. Étant une connaissance *a priori*, ces extra-groupes ne doivent pas être inclus dans l'analyse, le but étant de polariser une homologie par une connaissance externe (*non-circularité*). Le critère de l'extra-groupe s'utilise homologie par homologie, certains taxons étant plus adaptés que d'autres pour la polarisation de telle homologie et pas de telle autre. Cette connaissance est basée sur le produit d'autres analyses cladistiques. On considère qu'observer un taxon, celui qui apparaît être le plus proche parent de l'intra-groupe, n'est pas suffisant, mais que ce critère nécessite que plusieurs taxons soient pris en compte (Farris 1982; Maddison *et al.* 1984). Le critère extra-groupe implique qu'il est impossible qu'il y ait eu une diversification entre l'intra-groupe et l'extra-groupe, c'est-à-dire qu'il y ait un nœud paralogue entre les deux : si tel est le cas, alors les deux homologues sont dérivés. Il faut donc s'assurer que les extra-groupes utilisés ne forment pas un clade soutenu par l'homologue impliqué dans l'homologie que l'on cherche à polariser, c'est-à-dire qu'il faut s'assurer que l'ensemble des extra-groupes observés plus l'intra-groupe forment une orthologie (par exemple, pour deux extra-groupes : (extra-groupe 1(extra-groupe 2(intra-groupe))). Le critère extra-groupe ne doit cependant pas être utilisé à la légère. Se tromper sur l'extra-groupe entraîne la paraphylie totale de l'homologue dérivé; par exemple, sur l'exemple de la figure 1.12, polariser par l'homologue X_h au lieu de Y_h entraîne la paraphylie de tout l'homologue qui devient entièrement faux. L'ensemble de ces arguments rend le critère extra-groupe faible. Il doit être utilisé avec une distribution non-ambigüe du plus grand nombre d'extra-groupes possibles. Pour illustrer cette notion, on cherche à polariser l'homologie (« *coquille bivalve maintenue par un ligament* » (« *coquille bivalve dépourvue de ligament* ») lors de l'analyse de trois taxons de rudistes : *Diceras arietinum* présentant un ligament et les deux autres en sont dépourvus, *Ichthyosarcolites triangularis et Titanosarcolites giganteus*. Les rudistes sont inclus dans *Bivalvia*. Grâce à cette connaissance *a priori*, le ligament est une synapomorphie de *Bivalvia* et l'absence de ligament chez *Ichthyosarcolites triangularis et*

Titanosarcolites giganteus est une perte secondaire. Cette polarité aurait pu être mal interprétée avec des comparaisons extra-groupes inappropriées, comme avec des brachiopodes par exemple qui ont une coquille bivalve, mais pas de ligament.

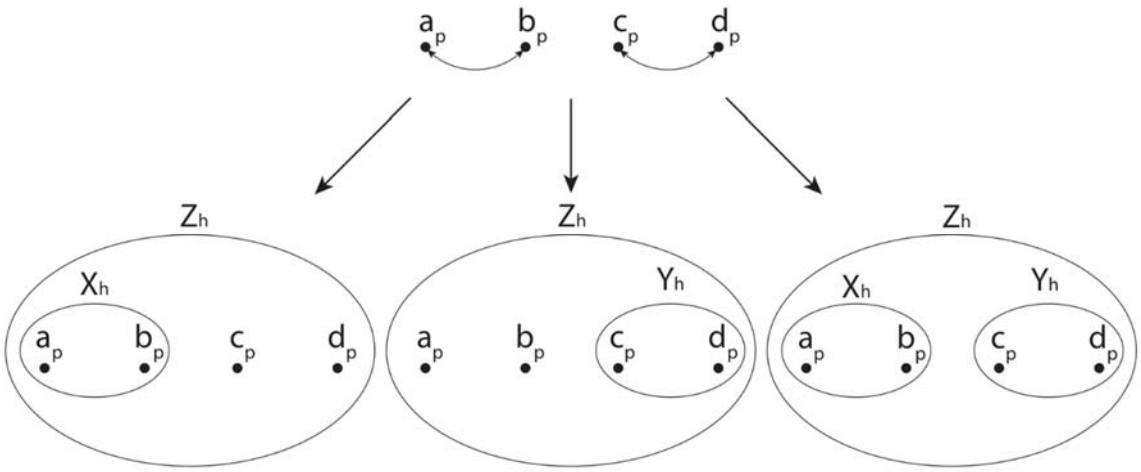


Figure 1.12. Illustration du problème de polarisation. La comparaison permet d'identifier deux ensembles $\{a_p, b_p\}$ et $\{c_p, d_p\}$. Il y a trois possibilités d'homologies. La première possibilité (gauche) est de considérer que X_h $\{a_p, b_p\}$ est le seul homologue informatif, et que des arguments ont été produits impliquant l'extension de la similarité de c_p, d_p à a_p et b_p. La seconde possibilité (centre) est de considérer cette fois que c'est Y_h $\{c_p, d_p\}$ qui est l'homologue dérivé, a_p et b_p faisant partie d'un homologue ancestral que possèdent finalement également c_p et d_p. La troisième solution est de construire deux homologues dérivés X_h $\{a_p, b_p\}$ et Y_h $\{c_p, d_p\}$ à partir de l'homologue racine Z_h.

D'autres critères ont été proposés pour résoudre les problèmes de polarisation. Le *critère paléontologique* (*précédence géologique*; Hennig 1966; Mayr 1986) est souvent cité comme le premier critère permettant d'identifier un homologue ancestral. Il s'énonce de la manière suivante : si dans un groupe monophylétique, un homologue est présent chez des UT plus anciennes et un autre homologue est présent chez des UT plus récentes, le premier est ancestral et le second dérivé (Darlu et Tassy 1993 : 57). Utilisé tel quel, son usage exclusif pour polariser les homologues reviendrait à dire que les Blattes actuelles portent des homologues tous plus dérivés de ceux des Mammouths disparus il y a 15 000 ans pour la simple raison de la présence des Mammouths avant celle des Blattes actuelles (Nelson et Platnick 1981; Darlu et Tassy 1993). L'utilisation de la position, comme celle de l'extra-groupe, implique donc qu'il n'y ait pas eu de diversification entre l'UT ancienne et l'UT plus récente. Ce qui rend absurde l'exemple des Blattes et des Mammouths est que l'on sait qu'il y a eu de la diversification entre ces deux taxons reliés par le nœud paralogue

Bilateria, conduisant aux deutérostomiens et aux protostomiens. On peut donc la soumettre aux mêmes critiques. Le critère paléontologique est encore plus faible que le critère extra-groupe, puisque la topologie ici peut être totalement inconnue. Cependant, on tire une certaine connaissance relative de la stratigraphie : on pense par exemple qu'il est impossible de retrouver des restes humains au Cambrien.

3.1.3. Caractère

3.1.3.1. Standardisation

Les hypothèses d'homologies sont des hypothèses sur les relations de parenté entre parties d'UT. Pour obtenir un cladogramme, il faut donc passer par une étape de remplacement des parties par leurs UT respectives. Cette étape s'appelle la *standardisation* (Prin, 2012 : 139) et permet l'établissement de relations de parenté entre UT. Ce sont les relations entre UT qui sont ensuite mises en congruence pour former le cladogramme. Une homologie est standardisée en caractère et un homologue est standardisé en état de caractère. La standardisation d'une homologie peut être complète, c'est-à-dire que chaque partie correspond exclusivement à une UT, et que chaque UT ne possède qu'une seule partie (fig. 1.13-1). Dans ce cas, la structure d'ordre entre les classes ne change pas, et l'homologie et le caractère ont le même nombre de feuilles. On peut alors remarquer que la classification ne change pas de l'homologie au caractère dans le sens ou à chaque classe du caractère correspond nécessairement une classe de l'homologie. Il n'y a alors pas de perte d'information. L'exemple 2 de la figure 1.13 montre un exemple de standardisation totalement non-informatif. Ce cas arrive lorsque toutes les parties de l'homologie appartiennent à la même UT. Ce cas s'appelle *homonomie* (Bronn 1858; Remane 1952), et correspond, dans une analyse donnée, à une diversification (paralogue) des homologues sans possibilité d'inférer une diversification (paralogue) de taxons. L'homologie (« *derme* d'*Homo sapiens* » (« *glandes mammaires d'Homo sapiens », « poils d'Homo sapiens »))* est un exemple d'homonomie et ne peut pas être convertie en caractère (dire qu'*Homo sapiens* est plus proche de lui-même que de lui-même est trivialement faux). Sachant qu'un

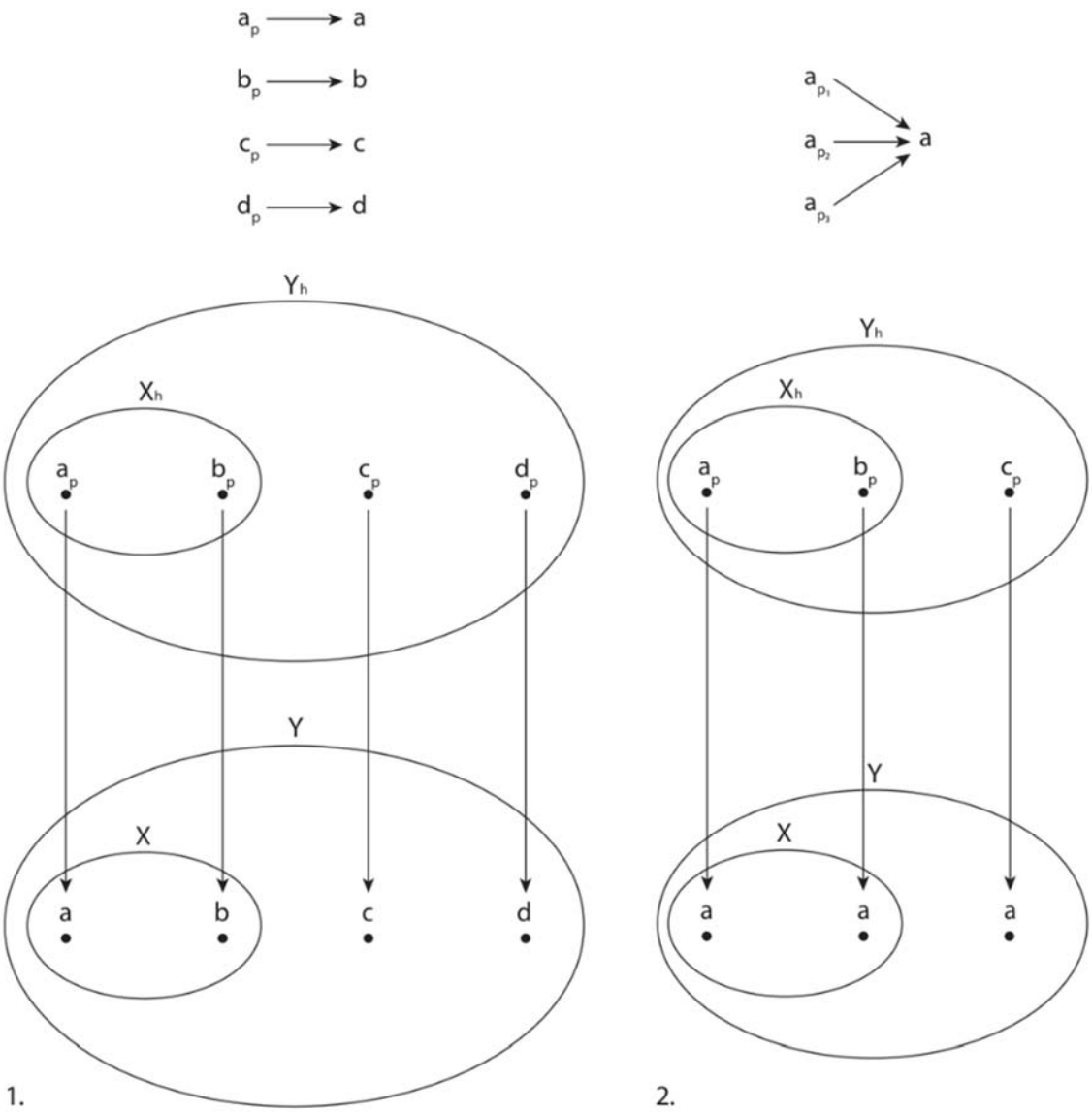


Figure 1.13. Standardisation d'homologies en caractères. En haut, correspondance entre parties d'UT a_p, b_p, c_p, d_p et UT a, b, c, d. En bas, les homologies sont composées de parties d'UT. Les parties sont remplacées par leurs UT correspondantes et forment les caractères. (1) Exemple de standardisation parfaite ou chaque partie correspond exclusivement à une UT. La standardisation ici ne fait perdre aucune information. (2) Exemple de standardisation non informative, ou homonomie. Toutes les parties correspondent à la même UT. Le caractère résultant de la standardisation n'apporte aucune information sur les relations de parenté entre UT.

caractère informatif est un caractère ayant au moins trois feuilles correspondant à trois UT différentes, on peut définir une homologie informative comme une homologie comprenant au moins trois parties d'UT appartenant à trois UT différentes. L'homonomie est une relation de parenté entre parties d'UT qui ne peut servir à l'établissement de relations de parenté entre UT.

3.1.3.2. Polymorphisme

La standardisation peut également être partielle en cas d'expression multiple des parties au sein d'une UT. Le caractère résultant de la standardisation est informatif, mais la structure obtenue affiche alors la même UT à plusieurs endroits de l'arbre. En phylogénie, on appelle ce phénomène de répétition de feuilles *polymorphisme*. La structure possédant des répétitions n'est pas encore un caractère : pour le devenir, il faut procéder à la suppression des répétitions. Le polymorphisme se résout en fonction du type de nœuds de la structure.

Le cas le plus simple de répétition de feuilles est dans une orthologie, comme exemplifié dans la figure 1.14. Ce cas part d'une homologie binaire (c_p, b_{p2}, (b_{p1}, a_p)) ou b_{p2} et b_{p1} sont deux parties portées par la même UT *b*. La structure standardisée à partir de cette homologie, que l'on appellera pseudo-caractère, est (*c*, *b*, (*b*, *a*)), composé du pseudo-état ancestral *Y* et du pseudo-état dérivé *X*. La structure hiérarchique permet de résoudre l'ambigüité liée à la répétition d'UT. *X* est un pseudo-état particulier du pseudo-état plus général *Y*. Il s'ensuit que tout ce qui est inclus dans *X* est également inclus dans *Y*. L'UT *b* incluse dans *X* est donc également incluse dans *Y*. Il en résulte que l'UT *b* incluse directement dans *Y* est redondante, puisque celui qui est dans *X* donne déjà cette information. On peut donc supprimer la feuille b incluse dans l'état Y et garder uniquement celle incluse dans l'état X. On obtient le caractère (*c*, (*b*, *a*)). Puisqu'entre le caractère et l'homologie, on a perdu une feuille, on a donc perdu de l'information. Cependant, on a pu conserver une information pertinente sur les relations entre UT. La méthode est de repérer toutes les instances de la même UT qui se répètent lors de l'étape de standardisation et de ne conserver qu'une de ces instances, placée dans le nœud orthologue le moins inclusif. On répète cette opération pour chaque UT répétée jusqu'à obtenir un caractère sans répétition de feuilles. Par exemple, la structure (*a*, (*a*, (*a*, (*a*, (*b*, (*a*, (*a*, (*a*, c)))))))) contient la classe {*a*, *c*}. Puisque *a* est présent dans cette classe, il est également présent dans toutes les autres classes qui incluent {*a*, *c*}. Expliqué d'une autre manière, on peut décomposer ce pseudo-caractère en toutes ces classes : {*a*, *a*, *a*, *a*, *b*, *a*, *a*, *a*, *a*, *c*}, {*a*, *a*, *a*, *b*, *a*, *a*, *a*, *a*, *c*}, {*a*, *a*, *b*, *a*, *a*, *a*, *a*, *c*}, {*a*, *b*, *a*, *a*, *a*, *a*, *c*}, {*b*, *a*, *a*, *a*, *a*, *c*}, {*a*, *a*, *a*, *a*, *c*}, {*a*, *a*, *a*, *c*}, {*a*, *a*, *c*}, {*a*, *c*}. La suppression de tous les taxons répétés donne cinq fois {*b*, *a*, *c*} et quatre fois {*a*, *c*}. Après traitement, on obtient le caractère (*b*, (*a*, *c*)). La structure (*a*, (*b*, (*a*, (*b*,

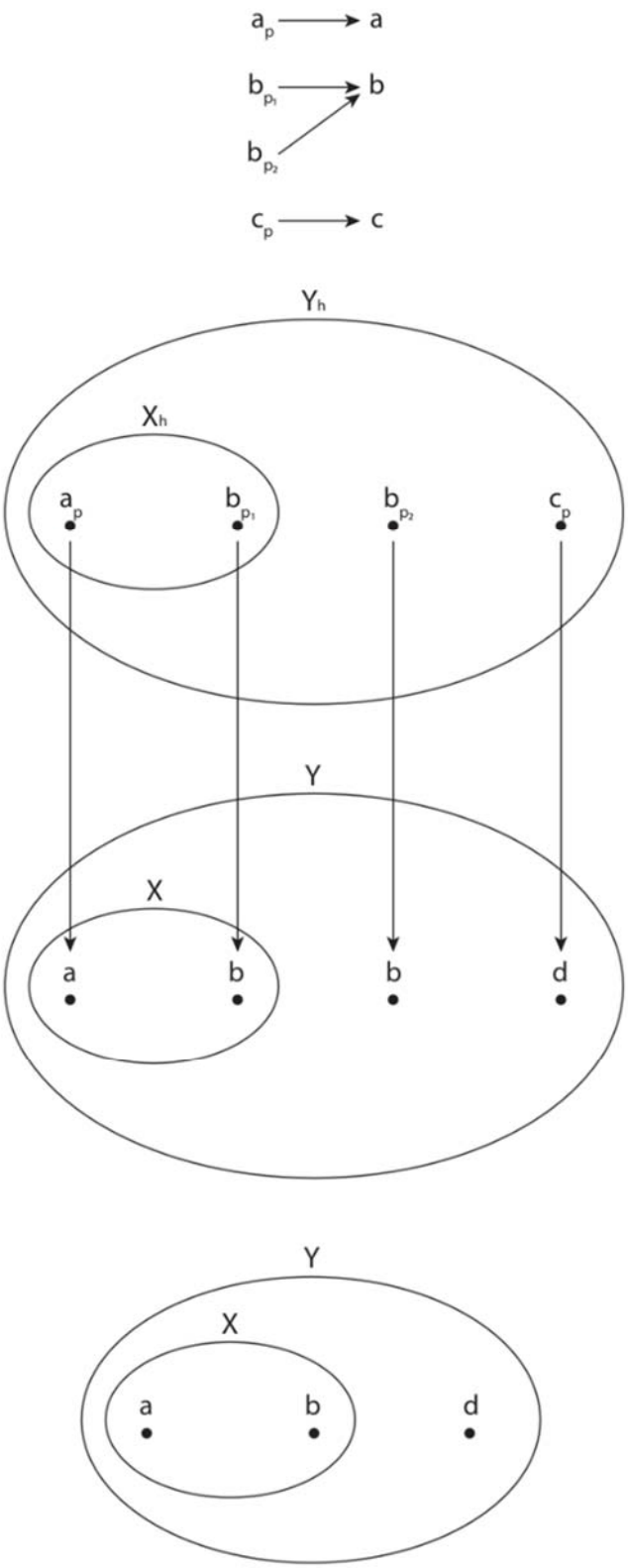


Figure 1.14. Standardisation d'homologies en caractères. En haut, correspondance entre parties d'UT (a_p, b_p, c_p, d_p) et UT (a, b, c,). En bas, l'homologie standardisée en caractère. L'UT b est polymorphique. Elle porte à la fois l'état X et Y. X étant inclus dans Y, toute feuille incluse dans X est également incluse dans Y. Dans ce cas, on peut supprimer la feuille b incluse dans Y, redondante avec l'autre instance de b. On obtient le caractère (d, (a, b)) (en bas).

(a, (b, (a, (b, (a, c))))))))) donne également le caractère (b, (a, c)). Plus simplement, parmi les taxons répétés, on ne garde que celui situé dans le taxon le moins inclusif. Cette méthode ne marche que lorsque l'on connaît les relations d'inclusion entre tous les nœuds concernés. Cette méthode ne fonctionne donc que lorsque les feuilles sont répétées dans une orthologie.

Le problème se complexifie lorsque des nœuds paralogues sont présents dans les caractères. Sur une structure ((a, b)(a, c)), trouver de manière non-ambigüe les relations de parenté entre UT semble beaucoup moins évident. Puisque les lignées dérivant d'un paralogue sont indépendantes les unes des autres, le principe consiste à « *couper* » le nœud paralogue. On obtient ainsi des sous-arbres dépourvus de nœuds paralogues (Nelson et Ladiges 1996). En figure 1.15, le nœud paralogue X est divisé en deux nœuds orthologues, chacun dans un sous-arbre, qu'il faut rassembler en une hiérarchie pour reconstruire le paralogue. Ces sous-arbres ont le même contenu en termes d'information que l'arbre dont ils sont issus. L'*analyse en sous-arbres libres de paralogie* (« *free-subtree paralogy analysis* »; Nelson et Ladiges, 1996) est basée sur ce principe. On définit un *sous-arbre* relativement à un autre arbre : c'est un arbre dont toute l'information est également contenue dans l'arbre dont il est issu et qui ne contient aucune information supplémentaire. Par exemple, (a, (b, c)) est un sous-arbre de (a, (b, (c, d))). Le cas d'analyse en sous-arbres le plus simple à comprendre est montré sur la figure 1.15 : il y a un nœud paralogue à couper (nœud X), et ce nœud est situé à la racine. On construit deux sous-arbres, chacun d'eux isolant une orthologie. Pour que l'analyse soit valide, on doit pouvoir reconstruire l'arbre source à partir des sous-arbres (pas de perte d'information). Cela explique pourquoi lors de cette analyse, on garde toutes les feuilles dans les deux sous-arbres (g, h, i, j, k, l, (a, (b, (c, (d, (e, f)))))) et (a, b, c, d, e, f, (g, (h, (i, (j, (k, l)))))). Si on ne garde pas ces feuilles, les sous-arbres sont (a, (b, (c, (d, (e, f))))) et (g, (h, (i, (j, (k, l))))). Or, ces deux sous-arbres ne sont pas suffisants pour retrouver l'arbre source. Ces deux sous-arbres autorisent par exemple l'arbre (a, (g, (b, (h, (c, (i, (d, (j, (e, (k, (f, l)))))))))). Lorsque l'on coupe un paralogue, les sous-arbres contiennent donc chacun toutes les feuilles incluses dans le nœud paralogue. Il faut découper le nœud paralogue en autant de sous-arbres qu'il y a d'orthologies qui partent du paralogue. L'arbre ((a, (b, c)), (d, (e, f)), (g, (h, i))) par

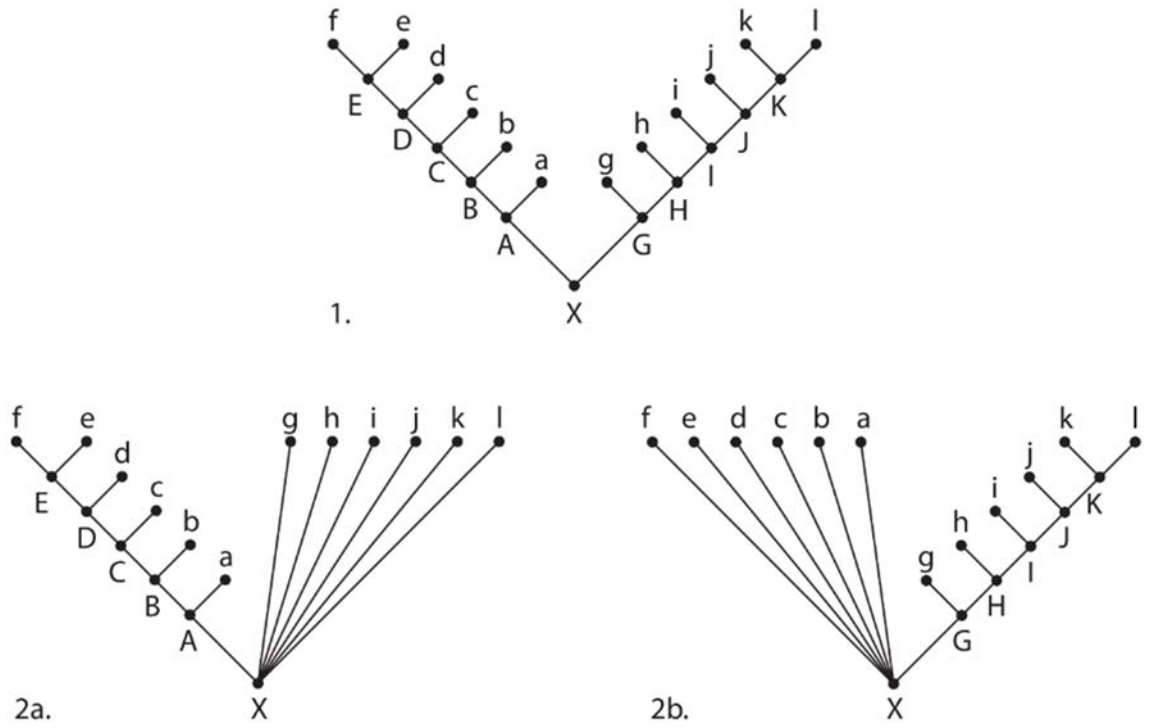


Figure 1.15. Analyse de l'arbre 1 en sous-arbres libres de paralogie. À partir du nœud paralogue *X* partent deux orthologies. L'analyse produit donc deux sous-arbres dépourvus de nœuds paralogues. Chaque sous-arbre contient également les feuilles appartenant à l'autre orthologie qui n'a pas été conservée. S'il y a du polymorphisme, celui-ci peut être résolu facilement à cause de la structure orthologique des sous-arbres.

exemple se découpe en trois sous-arbres (*d*, *e*, *f*, *g*, *h*, *i*, (*a*, (*b*, *c*)), (*a*, *b*, *c*, *g*, *h*, *i*, (*d*, (*e*, *f*)) et (*a*, *b*, *c*, *d*, *e*, *f*, (*g*, (*h*, *i*))) correspondant aux trois orthologies partant du nœud paralogue.

L'analyse devient un peu plus complexe lorsque le nœud paralogue n'est pas situé à la racine. Dans, ce cas, il sera nécessaire de générer un sous-arbre supplémentaire. Prenons l'exemple de la figure 1.16. Le caractère présente ici un nœud paralogue d'où partent deux orthologies. Si on suit le même processus, l'analyse de sous-arbres donne les deux arbres 2a et 2b. Or, ces deux sous-arbres ne sont pas suffisants pour reconstruire tout l'arbre : il manque une orthologie constituée des nœuds internes *A*, *B*, *C*, *D*, *E* et *F*. Dans ce cas, en plus des deux sous-arbres, on prend l'arbre source entier auquel on collapse tous les nœuds internes inclus dans le paralogue (nœud 7). Le troisième sous-arbre est le 2c, et l'ensemble des trois sous-arbres permet de reconstruire l'arbre source sans redondance d'information. Cet exemple nous permet de généraliser la procédure de suppression du

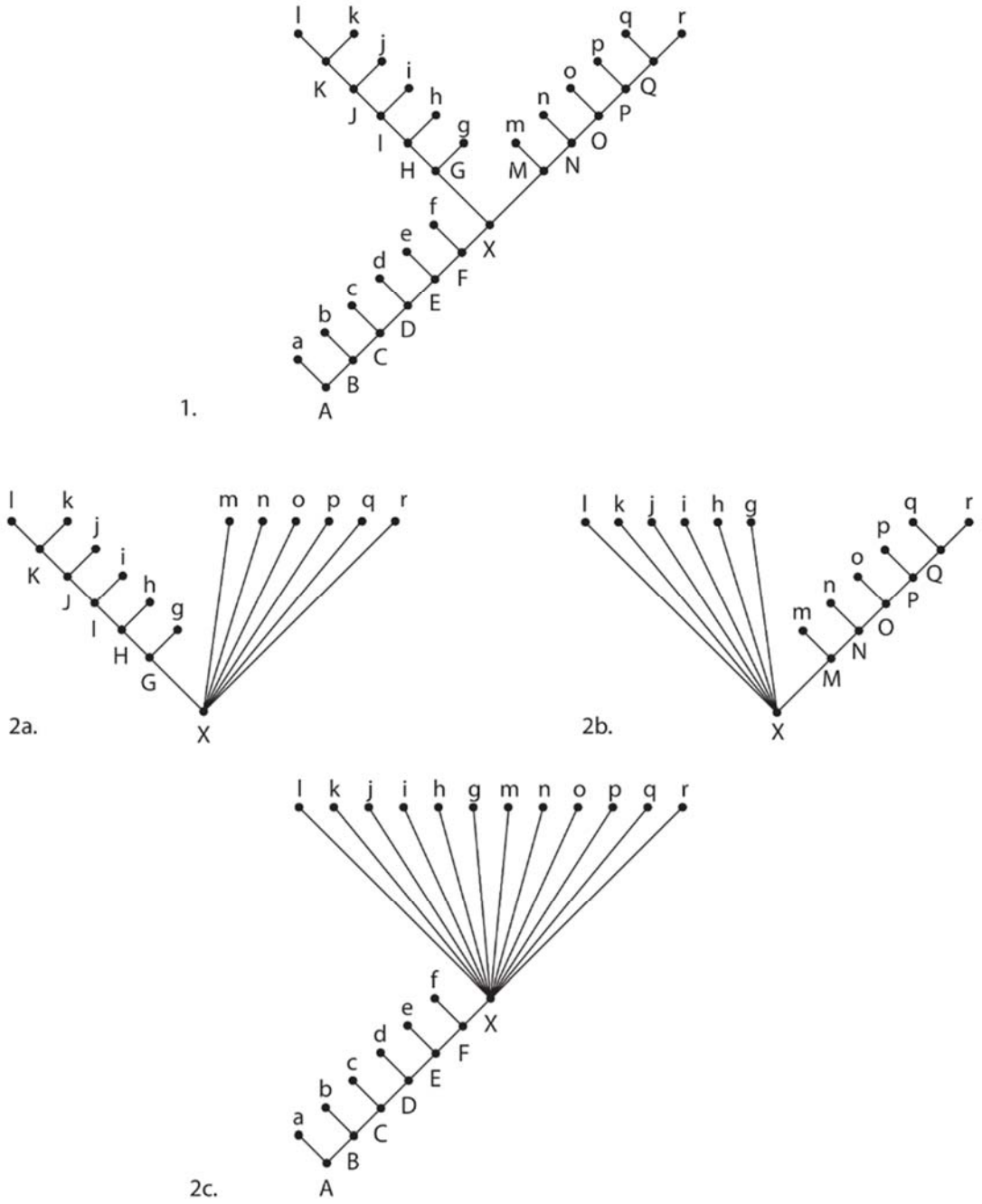


Figure 1.16. Analyse de l'arbre 1 en sous-arbres libres de paralogie. Trois sous-arbres 2a, 2b et 2c sont générés. 2a et b sont générés à partir des orthologies branchées au paralogue *X*. 2c est généré à partir de l'arbre source où tous les nœuds internes inclus dans le paralogue *X* (i à r) sont supprimés. Puisque tous les nœuds internes du nœud *X* sont supprimés, le nœud *X* devient orthologue.

polymorphisme et des paralogues. Soit un caractère c :

1. S'il existe au moins un nœud paralogue p_c dans c, aller à l'étape 2. S'il n'y a aucun paralogue, à l'étape 3.
2. Pour chaque nœud paralogue p_c, du plus inclusif au moins inclusif :
    a. Pour chaque nœud interne n_c directement inclus dans p_c :
        i. Construire un sous-arbre s_{pnc} comportant le nœud paralogue p_c, ainsi que toutes les feuilles incluses dans p_c, le nœud n_c et tous les nœuds internes inclus dans n_c.
    b. Construire un sous-arbre s_{pc} comportant toutes les feuilles de c et tous les nœuds internes excepté les nœuds inclus dans p_c. Remplacer le caractère c par les sous-arbres construits s_{pnc} et s_{pnc}.
3. Pour chaque arbre s_{pnc} et s_{pnc} :
    a. Pour chaque feuille f_c de c :
        i. vérifier le nombre d'occurrences de la feuille f_c. Si $f_c \geq 2$ et, supprimer toutes les occurrences de f_c excepté celle incluse dans le nœud interne le moins inclusif.

On obtient un ensemble de caractères dépourvus de paralogues et de répétition de terminaux.

Terminons avec un exemple empirique. Le terme de paralogie est né d'une étude de la protéine d'hémoglobine, constituée de deux chaînes α et de deux chaînes β (Fitch 1970, 2000; Gray et Fitch 1983). Considérons l'exemple classique de l'étude de trois UT, A, B et C à travers leurs gènes codant pour les protéines de globines (fig. 1.17). Les parties d'UT sont un gène de myoglobine porté par a (a_M), un gène d'hémoglobine α porté par b et c (b_α et c_α) et un gène d'hémoglobine β porté par b et c (b_β et c_β). Après standardisation, analyse en sous-arbres et suppression du polymorphisme, on obtient le caractère (a, (b, c)). On a perdu beaucoup d'information : l'arbre complexe avec un paralogue en (2a) donne un simple caractère à trois UT (2d). Cela correspond encore à une diversification des parties sans diversification du tout. Les instances d'homonomie (a_M, (b_α, b_β)) et (a_M, (c_α, c_β)) ne sont pas convertibles en relations de parenté entre UT, c'est pourquoi on obtient des caractères beaucoup moins riches que les homologies. Une fois standardisés, analysés et la

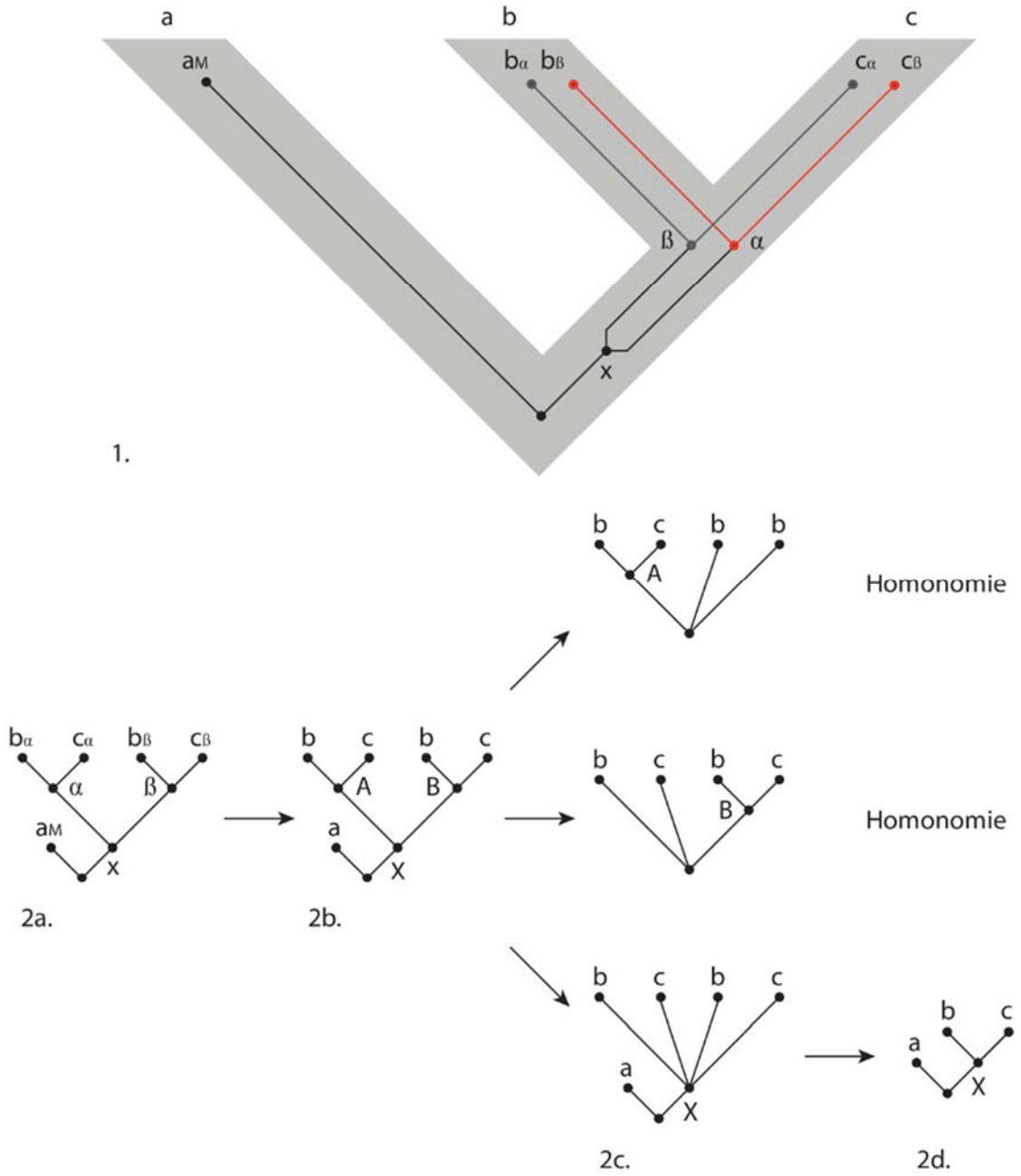


Figure 1.17. Illustration de Fitch 1970 et traitement par l'analyse en sous-arbres libres de paralogie. (1) L'arbre en noir et rouge correspond à l'homologie sur les parties a_M, b_α, c_α, b_β et c_β, l'arbre en gris correspond aux relations entre les UT *a*, *b* et *c*. Cette représentation met en évidence la diversification des parties (paralogue) sans diversification du tout. *X* représente l'homoglobine. (2) Représentation en 3ia : (2a) homologie, (2b) caractère, (2c) sous-caractères libres de paralogie, (2d) suppression des UT redondants. Les deux premiers arbres sont non-informatifs (seulement deux UT). Le résultat (*a*, (*b*, *c*))) correspond au schéma (1).

répétition supprimée, les caractères forment l'ensemble des hypothèses de parenté entre UT qui sont mises en congruence pour former le cladogramme.

3.2. Cladogramme

Un caractère est composé d'états de caractères et d'UT. En d'autres termes, un caractère est composé de relations de parenté entre UT. Le cladogramme est construit – selon le principe de *maximisation de la congruence* – comme la structure contenant le maximum de ces relations de parenté, c'est-à-dire que le cladogramme est la structure maximisant l'information concordante provenant des caractères. Le cladogramme est donc l'ensemble maximal de relations de parenté compatibles (on détaillera la notion de compatibilité dans la partie (3) recomposition). Or, on a défini la relation de degré de parenté minimale comme étant une assertion, ternaire, de type *a*, (*b*, *c*) : le triplet (Nelson et Platnick 1991). Dans une analyse, chaque nœud interne informatif correspond à un état, et donc à un homologue. Encore une fois, aucune information n'est perdue, et on peut retracer le parcours des homologues.

Ne pas décomposer les caractères équivaudrait à ne pas considérer les relations de parenté comme des entités indépendantes. Traiter les triplets de manière indépendante implique la possibilité, pour un caractère, de rejeter seulement une partie d'entre eux lors de la mise en congruence. La 3ia propose cette étape de décomposition des caractères en un ensemble de triplets correspondant à autant de solutions partielles indépendantes vis-à-vis de toutes les autres (*non circularité)*. Sachant deux triplets *a*, (*c*, *d*) et *b*, (*c*, *d*) provenant du même caractère (*a*, *b*, (*c*, *d*)), la 3ia permet de rejeter seulement un des deux triplets, et pas les deux qui seraient liés par une dépendance mutuelle. On remarque que l'utilisation des triplets permet que la maximisation de la congruence (accepter le maximum de triplets possibles dans le cladogramme final) soit strictement équivalente à minimiser l'incongruence (rejeter le minimum de triplets pour construire le cladogramme). Dans une première partie, on discute de l'utilisation de pondérations fractionnaires comme des poids attachés à chaque triplet permettant de corriger la redondance intrinsèque à l'étape de décomposition des caractères en triplets. On montre également la puissance du concept de triplets en tant qu'unité d'information. Dans la seconde, on discute des propriétés des triplets et de problèmes algorithmiques qui y sont liés.

3.2.1.1. Propriétés des triplets

On a dit que le cladogramme est l'ensemble maximal de relations de degré de parenté. C'est donc le cladogramme qui contient le plus de triplets. L'utilisation d'un critère de maximisation de la congruence implique que tous les triplets provenant de tous les caractères ne soient parfois pas inclus dans le même cladogramme. La raison est que certains triplets sont *compatibles* et d'autres non. Deux triplets sont compatibles lorsqu'ils peuvent être vrais en même temps. Que deux triplets ne soient pas compatibles implique que la vérité de l'un implique la fausseté de l'autre. Deux triplets sont compatibles lorsqu'ils sont identiques ou qu'ils possèdent au moins une feuille différente parmi les trois. *a*, (*b*, *c*) et *b*, (*a*, *c*) par exemple ne sont pas compatibles. La figure 1.18 montre quelques exemples de couples de triplets compatibles. Le principe de maximisation de la congruence peut alors être reformulé : dans ce cadre, il implique de chercher le plus grand ensemble de triplets indépendants compatibles.

Parmi les couples de triplets compatibles, certains possèdent des propriétés particulières qui doivent encore être détaillées pour comprendre la pondération fractionnaire. Deux triplets sont *combinables* lorsqu'ils peuvent être représentés au sein d'un arbre unique (fig. 1.18, combinabilité). Deux triplets combinables possèdent toujours exactement deux taxons en commun. Deux triplets combinables sont de plus *dépendants* si leur combinaison permet de déduire des triplets supplémentaires (fig. 1.18, dépendance). La combinaison de deux triplets se fait de la manière suivante. Soient les triplets compatibles t_x et t_y, avec $t_x \neq t_y$) :

1. Construire l'ensemble *n* des arbres z_n possibles contenant l'ensemble des feuilles de t_x et t_y et contenant t_x et t_y (t_x et t_y sont des sous-arbres de z_n).
2. Construire l'arbre *Z* qui contient les feuilles de t_x et t_y et qui contient les nœuds internes strictement présents dans l'ensemble des arbres z_n.
    a. Si *Z* ne contient aucun nœud interne, t_x et t_y ne sont pas combinables
    b. Si *Z* contient un ou deux nœuds internes, t_x et t_y sont *combinables*
    c. Si la décomposition de *Z* donne des triplets qui ne sont ni t_x ni t_y, t_x et t_y sont *dépendants*.

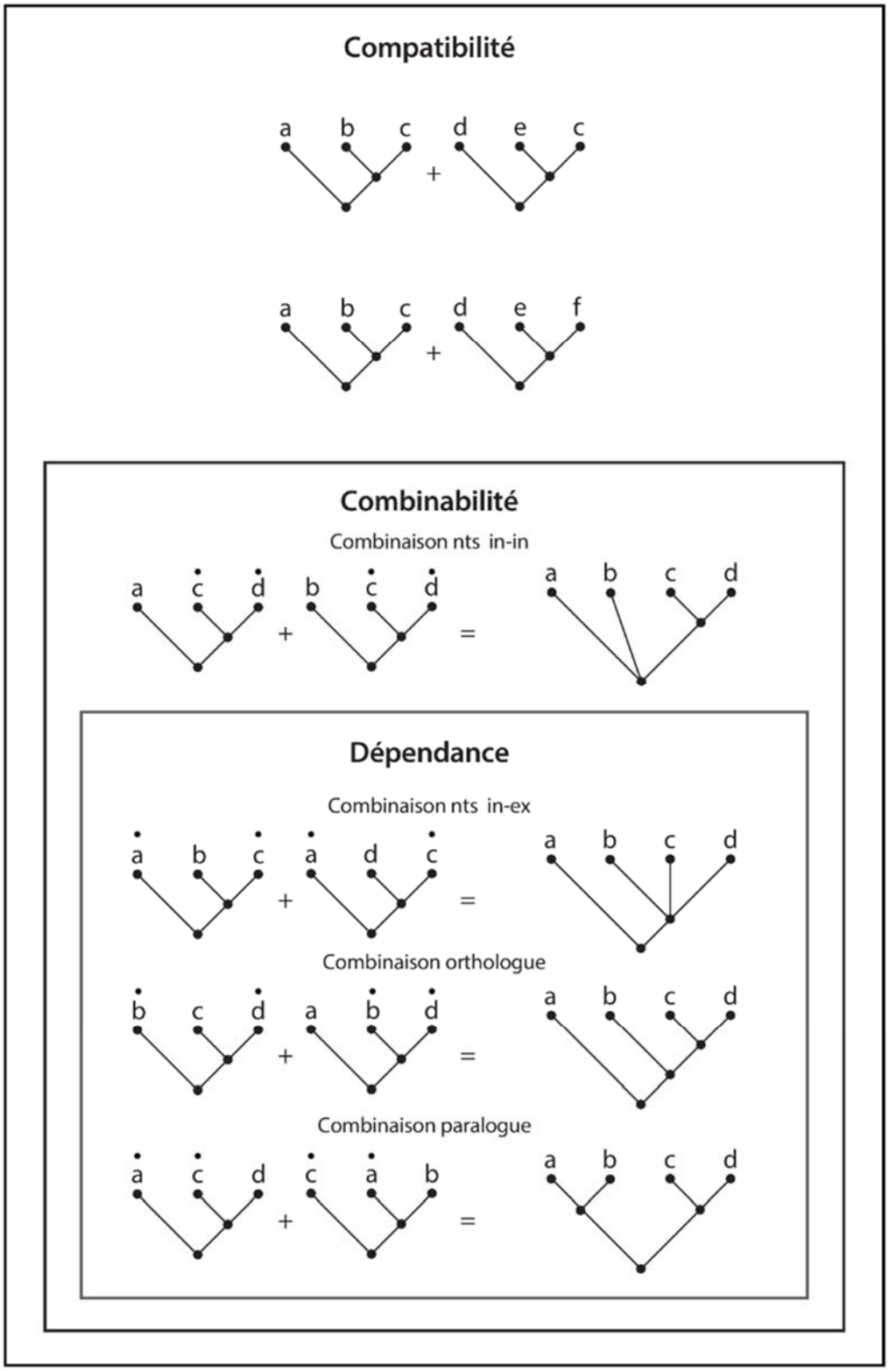


Figure 1.18. Représentation de différents types de relations entre deux triplets. La dépendance est un cas particulier de combinabilité qui est un cas particulier de compatibilité. Des exemples montrent la combinaison de deux triplets (sauf pour les relations simplement compatibles). Les quatre types de combinaison sont détaillés : la combinaison *in-in* et les trois combinaisons qui impliquent l'émergence de nouveaux triplets dépendants (dépendance), combinaison *in-ex*, *paralogue*, *orthologue*. Les points noirs signalent les feuilles communes au couple de triplets

La notion de dépendance entre triplets est reliée au contenu informatif de ces triplets. Soient les deux triplets *X* = a, (b, c) et *Y* = a, (d, c) (fig. 1.19). Si *X* et *Y* sont vrais en même temps, alors il existe nécessairement un troisième triplet *Z* = *a*, (*b*, *d*). On dira que *X* et *Y* sont des *triplets primaires*, et *Z* un *triplet secondaire*. Le triplet *Z* n'émerge pas de nulle part, son information est implicitement contenue dans l'existence simultanée de *X* et *Y*. Le triplet *Z* est dépendant de *X* et de *Y*. Il existe donc des combinaisons particulières de triplets qui contraignent l'existence d'autres triplets. La figure 1.19 résume les quatre types de combinaisons possibles et les triplets déductibles des couples de triplets combinables. Le type de combinaison varie en fonction de la position des deux feuilles communes au couple

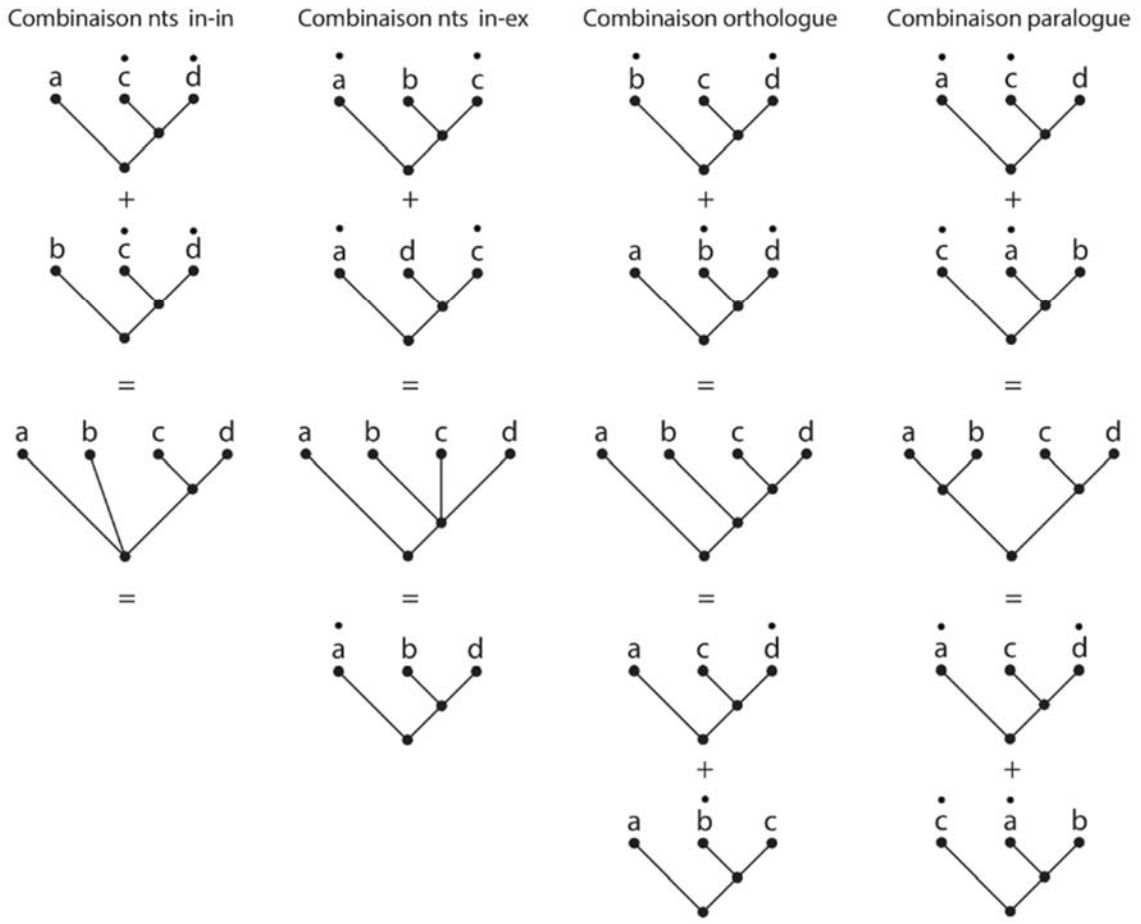


Figure 1.19. Quatre types de combinaison de couples de triplets en fonction de la position de leurs feuilles communes (notée avec un point noir). La disposition de haut en bas est : triplet primaire *x* + triplet primaire *y* = arbre *Z* = triplet secondaire *1* + triplet secondaire *2*. La combinaison *n*ts *in-in* ne produit aucun triplet. La combinaison *n*ts *in-ex* produit un triplet supplémentaire. La combinaison *in-in* et *in-ex* produisent un arbre *Z* avec un seul nœud interne. La combinaison paralogue produit un arbre *Z* dont la racine est paralogue. La combinaison orthologue produit un arbre *Z* qui est une orthologie. Les combinaisons paralogues et orthologues génèrent deux triplets supplémentaires.

de triplets. On notera *in* une feuille branchée à l'intérieur du nœud interne informatif d'un triplet (fig. 1.4, feuilles *a* et *b*) et *ex* une feuille branchée à l'extérieur du nœud interne informatif, c'est-à-dire branché directement à la racine (fig. 1.4, feuille *c*). La table 1.1 résume les quatre possibilités de combinaison en fonction de la position des deux feuilles communes. Un couple de triplets dépendants peut générer un ou deux triplets supplémentaires.

	triplet x		triplet y	
Combinaison	taxon a	taxon b	taxon a	taxon b
nts in-in	in	in	in	in
nts in-ex	in	ex	in	ex
orthologue	in	in	in	ex
paralogue	in	ex	ex	in

Table 1. 1. Quatre types de combinaison de couples de triplets *x* et *y* en fonction de la position de leurs feuilles communes *a* et *b*. *in* implique que la feuille concernée soit incluse dans le nœud informatif du triplet. *ex* implique que la feuille soit incluse à l'extérieur du nœud interne informatif, directement à la racine.

3.2.1.2. Décomposition

3.2.1.2.1. Caractères binaires

Les propriétés de la combinaison de triplets sont cruciales pour comprendre ensuite comment décomposer les caractères en triplets. La formule suivante permet de calculer le nombre de triplets contenus dans un caractère binaire (Nelson et Ladiges 1992), sachant le nombre total de feuilles t et le nombre de feuilles contenues dans le nœud informatif n :

$$\text{Nombre de triplets} = \frac{n(t-n)(n-1)}{2}$$

Suivant le principe de *non-circularité*, chaque triplet doit être indépendant de tous les autres, et donc chaque triplet doit pouvoir être réfuté indépendamment de tous les autres. Or, selon la structure du caractère, les triplets générés peuvent être liés par de la dépendance. Concrètement, cette dépendance amène de la redondance lorsqu'un triplet secondaire répète une information portée par un triplet primaire. Reprenons le caractère (*a*, (*b*, *c*, *d*)) (figs. 1.19, 1.20). Celui-ci se décompose en trois triplets : *a*, (b, c), *a*, (*b*, *d*) et *a*, (*c*, *d*). Ce caractère devrait donc en toute logique posséder un poids informationnel de trois

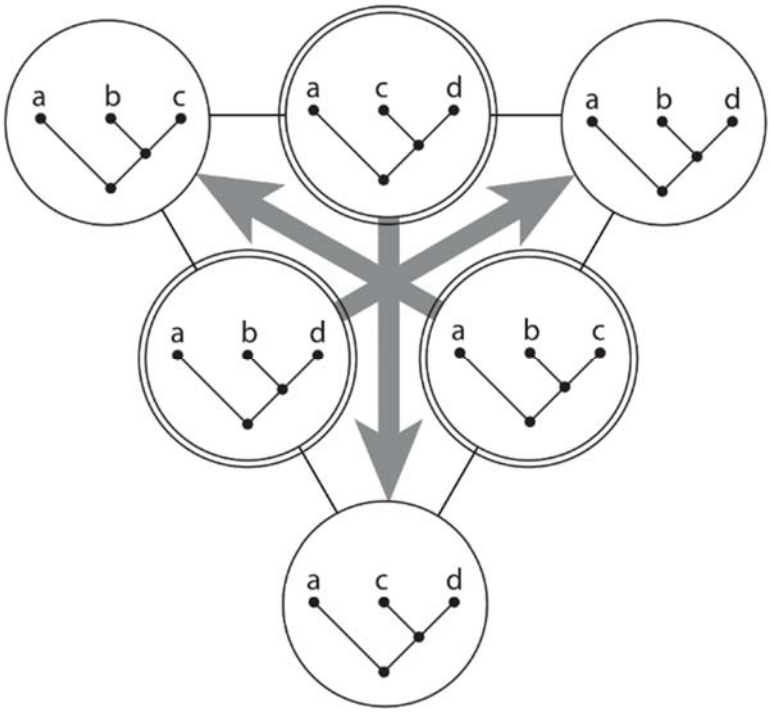


Figure 1.20. Trois triplets primaires (cercles simples) décomposés à partir du caractère (*a*, (*b*, *c*, *d*)) et les triplets secondaires déductibles (cercles doubles) de la combinaison de couples dépendants. Chaque triplet secondaire est relié par un trait noir à ses triplets parents. Les flèches grisent montrent la répétition du même triplet, entraînant la redondance. Il suffit de deux triplets primaires parmi les trois pour reconstruire le caractère initial.

triplets. Cependant, il existe trois relations de dépendance entre ces triplets générant des triplets secondaires qui entrent en répétition avec les triplets primaires (fig. 1.20). Si on sait que *b* est plus proche de *c* que de *a* et que *b* est plus proche de *d* que de *a*, alors *c* est plus proche de *d* que de *a*. Le troisième triplet est une nécessité logique. Rajouter le troisième triplet dans le poids du caractère apporte donc de la redondance. Le caractère porte en fait un poids informationnel de deux puisque seulement deux triplets sont nécessaires pour le reconstruire. La relation de dépendance implique que n'importe quelle paire de ces trois triplets est suffisante pour reconstruire le caractère. Ce « *deux parmi les trois* » correspond à la pondération fractionnaire (PF; fractionnal weighting, FW) associée aux triplets (Nelson et Ladiges 1992, 1994). Puisque le caractère possède en réalité deux triplets indépendants mais qu'il est impossible de choisir objectivement deux triplets parmi les trois, on pondère les trois triplets à 2/3 qui donnent en conséquence un total d'information de 3 triplets * 2/3 = 2 à ce caractère. Cet exemple illustre le contenu en information d'une hiérarchie. Le caractère ne contient que deux triplets indépendants sur les trois à cause de la redondance

provoquée par la dépendance entre les triplets. La formule suivante donne le nombre de triplets indépendants contenus dans un caractère binaire (Nelson et Ladiges 1992) :

$$\text{Nombre de triplets indépendants} = (t - n)(n - 1)$$

Cette formule correspond au nombre minimal de triplets nécessaires pour reconstruire le caractère. Il correspond également à l'ensemble maximal de triplets dépourvu de redondance. Pour un caractère à $n = 2$, il n'y a aucune dépendance entre triplets puisque tous sont liés par des relations de combinaison nts in-in. De plus, tous les triplets sont nécessaires pour reconstruire le caractère. Leur poids est logiquement toujours de 1. Cela inclut les caractères à trois feuilles qui contiennent un seul triplet : leur PF est de 1. Les caractères binaires à $n \geq 3$ contiennent systématiquement des relations *in-ex* qui nécessite l'ajustement par la PF. La valeur de chaque triplet par rapport à son caractère peut être calculée par la formule suivante.

$$\text{Valeur absolue d'un triplet} = \frac{2}{n}$$

Cette PF s'applique à tous les triplets puisque pour un caractère binaire, il est impossible de se dire que c'est tel ou tel triplet qui provoque la redondance.

3.2.1.2.2. Caractères à plus de deux états

Tous les caractères à plus de trois feuilles sont constitués de triplets qui forment des relations de combinabilité (fig. 1.18). Si le caractère comporte au moins un état informatif contenant plus de deux feuilles, des relations de dépendance *in-ex* apparaissent, gérées par la PF. Si le caractère porte plusieurs états informatifs, l'information sur l'inclusion entre états permet de corriger l'incertitude des PF (Prin, 2012 : 190-193). Les PF corrigent le fait que des triplets soient redondants; la correction pourrait se faire en supprimant les triplets qui entrent en redondance. Il est cependant impossible de décider lesquels supprimer de manière non-arbitraire pour un caractère binaire. L'information de l'inclusion entre états permet de résoudre cette ambiguïté. Voyons comment supprimer ces deux types de dépendance à partir d'un exemple.

Soit le caractère (a, (b, (c, d))) de la figure 1.21. Ce caractère est constitué de deux états informatifs $X = \{b, c, d\}$ et $Y = \{c, d\}$. La première étape est la décomposition des états en

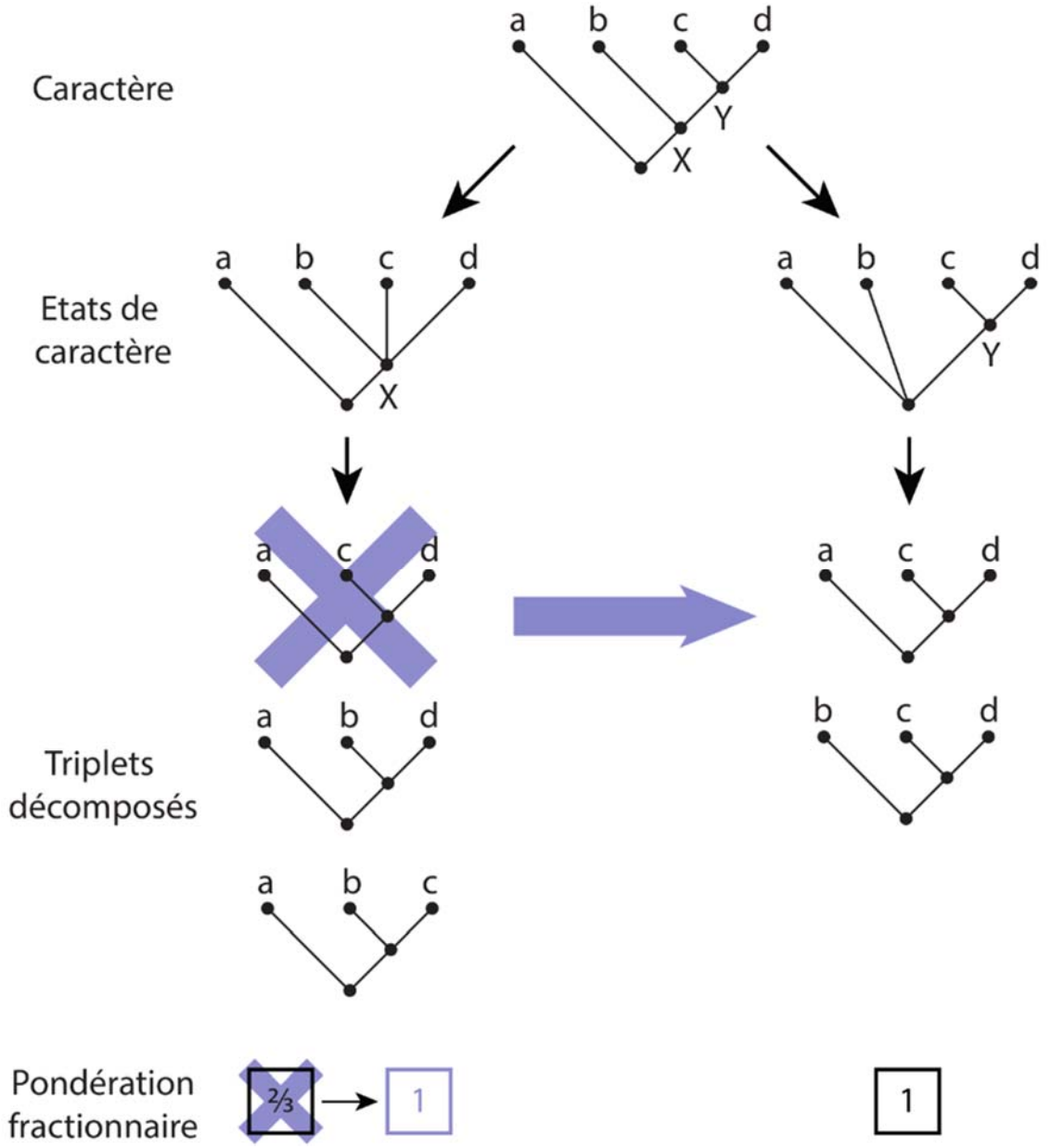


Figure 1.21. Calcul des PF pour le caractère (*a*, (*b*, (*c*, *d*))) en deux étapes. (1- redondance intra-état, noir) Chaque état de caractère (*X* et *Y*) est décomposé en triplets (3 et 2). La PF des triplets de l'état *X* vaut 2/3, celle des triplets de *Y* vaut 1. (2- redondance inter-états, violet) Le triplet (*a*, (*c*, *d*)) est présent dans les deux états. Sa suppression de *x* permet de lever l'ambigüité de la PF de 2/3 des triplets *X* (« *choisir deux parmi trois* »).

triplets. Il y a pour l'état *X* trois triplets *a*, (*c*, *d*); *a*, (*b*, *d*) et *a*, (*b*, *c*) pondérés à 2/3 puisque seuls deux parmi trois sont nécessaires pour reconstruire l'état. L'état *Y* contient *a*, (*c*, *d*) et *b*, (*c*, *d*) pondérés à 1 puisque les deux sont nécessaires. La seconde étape (fig. 1.21, en violet) intervient comme suit. Il faut garder deux triplets de l'état *X* et éliminer un troisième. Le triplet *a*, (*c*, *d*) de l'état *X* est aussi présent dans l'état *Y*. Or, on sait que dans l'état *Y*, ce

triplet est indispensable. La suppression de ce triplet de l'ensemble *X* est donc possible puisqu'il existe dans l'autre état *Y*. Supprimer ce triplet dans *X* rend obsolète la PF de 2/3 qui passe à 1. Les deux triplets restant de *X* sont tous deux indispensables pour reconstruire l'état. Notre connaissance sur l'inclusion entre les états nous a permis de lever l'ambigüité sur les triplets. Pour récapituler, la méthode de PF s'applique en deux étapes : (i) génération des triplets et calcul des PF état par état (suppression de la dépendance liée aux relations *in-ex*), (ii) ajustement des PF de l'état le moins inclusif à l'état le plus inclusif (résolution de la dépendance liée aux relations *in-ex* par l'information de l'inclusion entre états). La figure 1.22 montre un exemple avec un caractère-orthologie à cinq taxons. Soit un caractère *c* à *n* états informatifs du plus inclusif au moins inclusif e_1, e_2… e_n :

1. Décomposer chaque état en l'ensemble de ses triplets t_n.
2. Associer un poids de $2/f_n$ à chaque triplet de t_n (f_n est le nombre de feuilles que contient *n*).
3. Si $n \geq 2$, répéter l'opération suivante entre couples d'états directement inclus l'un dans l'autre du plus inclusif au moins inclusif (e_1 et e_2, e_2 et e_3… e_{n-1} et e_n) :
    a. Noter les *x* triplets identiques entre les deux états. Supprimer les triplets *x* de t_{n-1}. Additionner x/t_n au poids de chaque triplet restant de t_{n-1}.

Cette procédure ne fonctionne qu'avec des caractères sans paralogues, il faut donc d'abord supprimer tous les paralogues en utilisant la LP. Notons que l'on peut considérer qu'un vrai caractère cladistique ne comporte aucune instance d'indépendance, auquel cas la LP est une étape indispensable à la construction des caractères même sans répétition de terminaux.

3.2.1.2.3. Le triplet comme unité informationnelle

La quantité de triplets pondérés est un outil qui permet de juger du contenu informatif de n'importe quelle structure cladistique à partir de l'atome d'information qu'est le triplet. Dans la suite de ce chapitre, l'information cladistique ou juste information ou poids se réfère à une quantité de triplets indépendants. Revenons sur quelques notions abordées précédemment à la lumière de la notion d'information cladistique.

Considérons maintenant les quatre caractères binaires de la figure 1.23 afin de mieux comprendre comment les triplets permettent de quantifier l'information cladistique

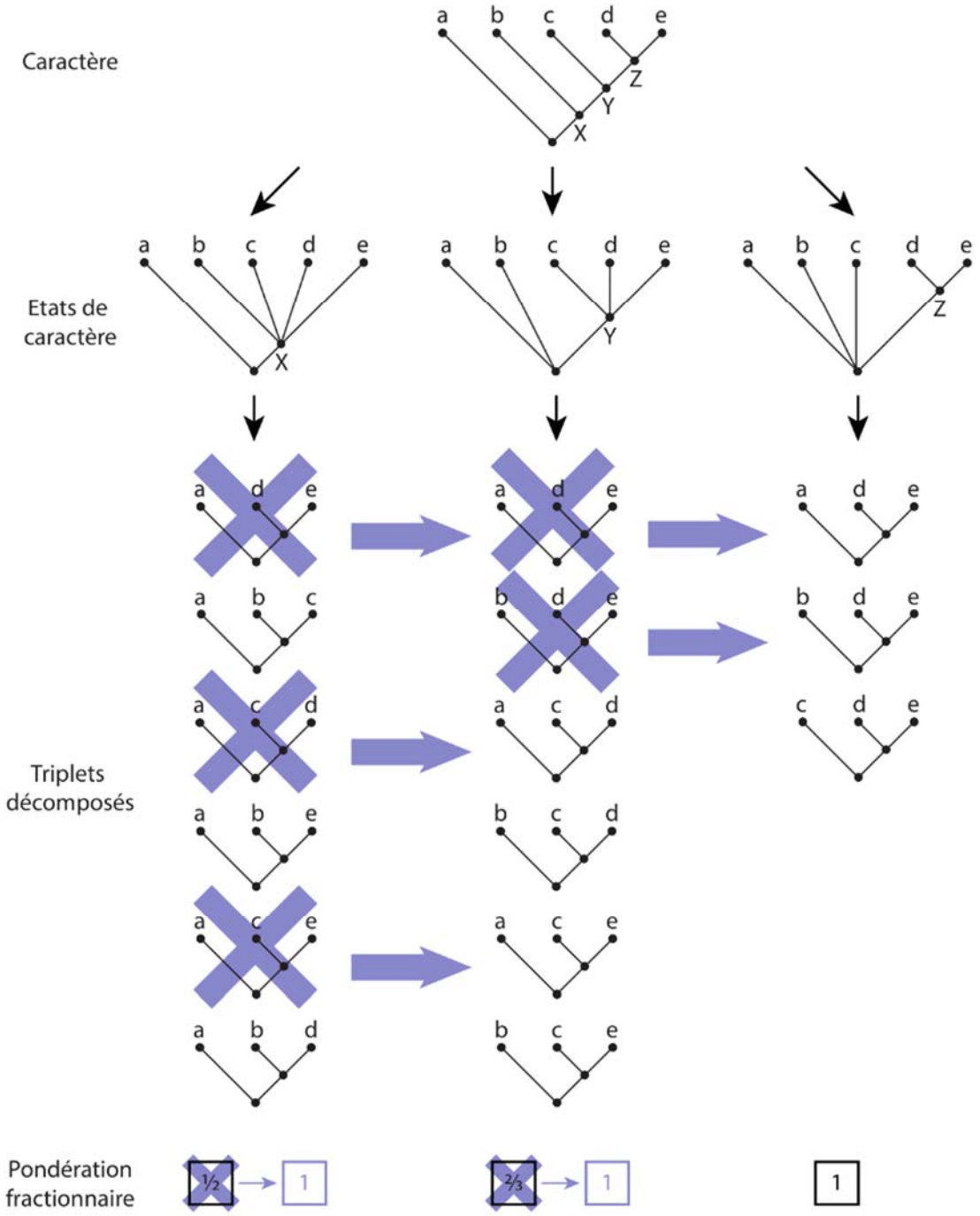


Figure 1.22. Calcul des PF pour le caractère multi-états (*a*, (*b*, (*c*, (*d*, *e*)))) en deux étapes. (1- redondance intra état, noir) Chaque état de caractère (*X*, *Y* et *Z*) est décomposé en triplets (6, 6 et 3). La PF des triplets de *X* vaut 1/2, celle des triplets de *y* vaut 2/3 et celle de *y* 1. (2- redondance inter-états, violet) On vérifie la redondance par couples d'états, des états les plus inclusifs aux moins inclusifs. On supprime les triplets redondants de *X* par rapport à *Y*, puis ceux de *Y* par rapport à *Z*. Le caractère étant une orthologie, toute ambiguïté est supprimée (toutes les PF=1).

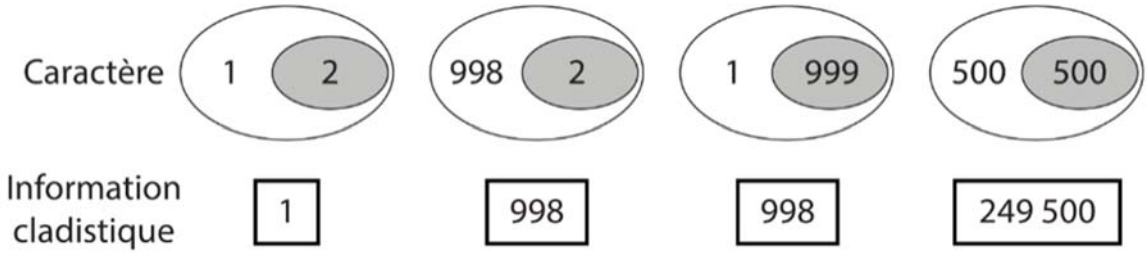


Figure 1.23. Illustration du concept d'information cladistique. Les caractères sont représentés par leurs nœuds internes, les chiffres correspondent aux feuilles qui s'insèrent à cet endroit. Le caractère le plus à gauche est l'entité informationnelle minimale : son poids est de 1. Les deux caractères du milieu ont un poids de 998 : il suffit de se tromper sur l'observation de 2 feuilles pour que le caractère soit faux. Le caractère de droite est le plus équilibré. Il faudrait se tromper sur l'observation de 500 feuilles pour que le caractère soit faux. Son poids de 249 500 souligne le degré de confiance que l'on peut lui accorder.

(Zaragüeta i Bagils 2002). Le premier d'entre eux porte sur trois feuilles et se résume donc à un seul triplet. Les trois suivants portent chacun sur 1000 feuilles et ne varient que dans la localisation des feuilles. Parmi eux, le premier contient 998 feuilles branchées à la racine et 2 feuilles sur l'état dérivé. Il porte un poids de 998. Si 1000 UT ont été observées et comparées, il est légitime d'accorder une confiance faible à ce caractère. Il aura suffi de 2 UT différant de toutes les autres pour fonder ce caractère. Se tromper sur une des deux UT est relativement facile et fait s'effondrer le caractère qui devient faux. Le second caractère possède à l'inverse 999 feuilles dans le nœud informatif et une unique feuille à la racine. Ce caractère porte cependant un poids identique de 998. Le problème est identique : il suffit de se tromper sur une seule UT pour que tout le caractère devienne faux. On peut donc comme pour le premier avoir une confiance faible sur ce caractère. Le dernier caractère est équilibré dans sa répartition des feuilles : 500 à l'intérieur et 500 à l'extérieur du nœud informatif. Son poids de 249 500 est 250 fois plus important que celui des caractères précédents. Réfléchissons de la même manière que précédemment : sur combien d'observations faudrait-il se tromper pour que le caractère soit faux ? Ici, il faut s'être trompé sur l'observation de la moitié de l'échantillonnage, c'est-à-dire sur 500 UT. Ce caractère apparaît donc bien plus robuste que les précédents, et c'est ce que traduit le poids élevé attaché à lui. Le poids d'un caractère traduit donc également un degré de confiance. On peut noter quelques faits intéressants supplémentaires :

1. Entre deux caractères possédant le même nombre de feuilles et le même nombre de nœuds internes, l'arbre qui présente plus de nœuds paralogues est toujours moins

informatif, par exemple (*a*, (*b*, (*c*, (*d*, *e*)))) (poids = 10) par rapport à (*a*, ((*b*, *c*), (*d*, *e*))) (poids = 9). Un arbre qui ne présente que des nœuds orthologues est donc la structure la plus informative puisqu'il y a une relation d'inclusion entre tous les nœuds internes. L'information des paralogues est moindre puisqu'ils provoquent de l'indépendance entre nœuds : il y a certains couples de nœuds (disjoints) pour lesquels on ne peut spécifier de relation. Les poids différents de ces deux caractères traduisent ce fait. La seule exception concerne la racine : ce sommet n'étant pas informatif, le fait qu'elle soit orthologue ou paralogue n'a pas d'influence sur le poids du caractère en triplets; par exemple (*a,* (*b*, (*c*, *d*))) et ((*a*, *b*), (*c*, *d*)) possèdent tous les deux le même poids global, à savoir 4. On considère alors que ((*a*, *b*), (*c*, *d*)) forme en fait deux caractères indépendants.

2. L'analyse de sous-arbres libres de paralogie corrige les problèmes de relations de dépendance des paralogues (fig. 1.18) en décomposant le caractère en sous-arbres sans nœud paralogue. L'analyse fait diminuer le poids du caractère qui serait sinon surévalué par la dépendance paralogue.
3. La gestion des PF sur les caractères multi-états est complète lorsque l'arbre est dichotomique. Tous les triplets ont alors un poids de 1 et toute l'ambigüité est résolue. L'orthologie dichotomique est donc la structure la plus informative possible.
4. On a expliqué plus tôt le traitement différentiel des données non-applicables et des données manquantes sur les caractères. Pour une donnée non-applicable, on place l'UT à la racine. Par exemple, pour un caractère (*b*, (*c*, *d*)) pour lequel l'état {*c*, *d*} est considéré non applicable à une UT *a*, on place *a* comme suit : (a, *b*, (*c*, *d*)) (poids = 2). On a une connaissance : *a* ne se place dans aucun état dérivé. *a* se place donc nécessairement à la racine qui représente l'univers du discours. Si au contraire *a* possède une partie manquante (donnée manquante) (i.e., qui ne peut pas être observée), alors on ne peut pas rajouter cette UT dans le caractère (*b*, (*c*, *d*)) (poids = 1). On a un déficit de connaissance, ce qui réduit le poids du caractère.
5. On a dit précédemment que la standardisation permettant de passer de l'homologie au caractère peut être complète ou incomplète. Si la standardisation est complète, le poids de l'homologie est identique au poids du caractère. Une standardisation

incomplète à cause de l'homonomie engendrant du polymorphisme fera toujours perdre du poids au caractère par rapport à l'homologie.

3.2.1.3. Recomposition

3.2.1.3.1. Méthode du cladogramme vers les triplets

A cette étape, on obtient une liste de tous les triplets générés ainsi que leur pondération. Pour construire le cladogramme à partir de cet ensemble, deux possibilités existent. On peut rechercher le meilleur arbre directement ou par triplet. La recherche d'arbres est la seule approche implémentée pour l'instant. C'est une recherche du haut vers le bas : on part des arbres et on calcule leur poids en fonction du nombre de triplets qu'ils contiennent. Lisbeth 1.2 (Zaragüeta i Bagils *et al.* 2012) est le seul logiciel actuellement en développement qui propose cette approche. La seconde approche, la recherche des triplets congruents n'a jamais été implémentée. C'est une recherche du bas vers le haut qui nécessite de trouver l'ensemble maximal de triplets compatibles, puis de construire le cladogramme à partir des triplets.

La recherche par le haut consiste à prendre un arbre et à calculer son contenu en information cladistique. Cet arbre contient un certain nombre de triplets; connaissant le poids des triplets, le poids de l'arbre est la somme des poids de tous les triplets compatibles avec cet arbre. Le cladogramme retenu est celui dont le poids en triplets est le plus grand, car c'est celui qui *maximise la congruence*. On choisit donc l'arbre qui tient pour vrai le maximum d'hypothèses formulées en amont. Plusieurs problèmes émergent de la recherche par le haut du cladogramme. Premièrement, les algorithmes utilisés actuellement contraignent la structure du cladogramme à être dichotomique. Or, rien ne justifie cette dichotomie. Pour une analyse de deux caractères *a*, (*c*, *d*) et *b*, (*c*, *d*), une des solutions est le cladogramme (*a*, *b*, (*c*, *d*)). La seconde difficulté vient du nombre d'arbres à vérifier lors d'une recherche exacte. La formule du nombre *T* d'arbres dichotomiques pour un ensemble de *n* UT à classer est (Cavalli-Sforza et Edwards 1967; Felsenstein 1978) :

$$T = \prod_{k=2}^{n}(2k-3) = \frac{(2n-3)!}{2^{n-2}(n-2)!}$$

On obtient très vite une quantité gigantesque d'arbres à vérifier (même pour les machines les plus récentes et les plus performantes). Par exemple, pour $n = 10$, $T =$

34 459 425, et pour $n = 15$, $T = 213\ 458\ 046\ 676\ 875$. La recherche exhaustive n'est le plus souvent possible que jusqu'à un maximum de 16 UT (selon les triplets et selon la puissance de la machine de calcul) avec un algorithme de séparation et évaluation (*branch and bound*). La recherche exhaustive de cladogrammes est un problème NP-complet; souvent, des heuristiques sont utilisées pour trouver la solution.

3.2.1.3.2. Méthode des triplets vers le cladogramme

Cette partie est d'ordre de la prospective. Son but est de montrer les voies possibles pour la recherche future d'algorithmes efficaces de 3ia.

La recherche par le bas n'a jamais été implémentée à cause d'une difficulté importante qui empêche de simplement choisir le plus grand ensemble de triplets compatibles (fig. 1.19), c'est-à-dire pouvant être vrais en même temps. Le *théorème de compatibilité par comparaison par paires* (*pairwise compatibility theorem*; Estabrook *et al.* 1976a, 1976b) énonce qu'étant donné un ensemble de caractères portant sur les mêmes UT (i.e. relatifs au même échantillonnage taxonomique), ces caractères sont tous mutuellement compatibles deux à deux si et seulement s'ils forment aussi une clique, c'est-à-dire si et seulement s'il existe au moins une structure classificatoire hiérarchique (ex : un cladogramme) avec laquelle ils sont tous compatibles (car « contenu » dans cette dernière). La conséquence principale de ce théorème est qu'au lieu de rechercher les cladogrammes maximisant les caractères compatibles (arbres de max-clique) parmi tous les cladogrammes possibles, on peut se contenter d'identifier les plus grands sous-ensembles de caractères mutuellement compatibles deux à deux, puis de construire les cladogrammes de max-clique directement à partir de ces sous-ensembles. L'intérêt est que même si la recherche des max-cliques est un problème NP-complet (Day et Sankoff, 1986), obtenir les plus grands sous-ensembles de caractères mutuellement compatibles deux à deux puis construire les cladogrammes de max-clique à partir de ces derniers est, dans la pratique, plus rapide que rechercher les cladogrammes les plus parcimonieux (Felsenstein, 2004; 60-61 et 91-92). On peut donc légitimement se demander si ce théorème peut également s'appliquer pour la compatibilité des triplets. Malheureusement, ce n'est pas le cas. En effet, des exemples montrent que parfois, trois triplets pourtant compatibles deux à deux ne peuvent pas être vrais en même temps (Wilkinson 1994b : 222-223; Prin 2012 : 193-194). Cette possibilité est illustrée par la figure 1.24. Soient les trois triplets d, (a, b), b, (a, c) et c, (a, d). Chaque

paire de triplets est non seulement compatible mais également dépendante (et donc combinable). Si on combine deux par deux les triplets, on obtient en effet :

$d, (a, b) + b, (a, c) = (d, (b, (a, c)))$

$d, (a, b) + c, (a, d) = (c, (d, (\mathrm{a}, b)))$

$b, (a, c) + c, (a, d) = (b, (c, (a, d)))$

Or les trois arbres générés sont tous incompatibles les uns avec les autres. On a choisi un ensemble de trois triplets compatibles deux à deux et pourtant, de l'incompatibilité apparaît lors de leur combinaison. Ce phénomène est lié à la dépendance entre triplets. On a dit plus haut que la dépendance générait des triplets secondaires. Ici, des triplets secondaires sont générés (fig. 1.24, cercles doubles) par la combinaison des triplets

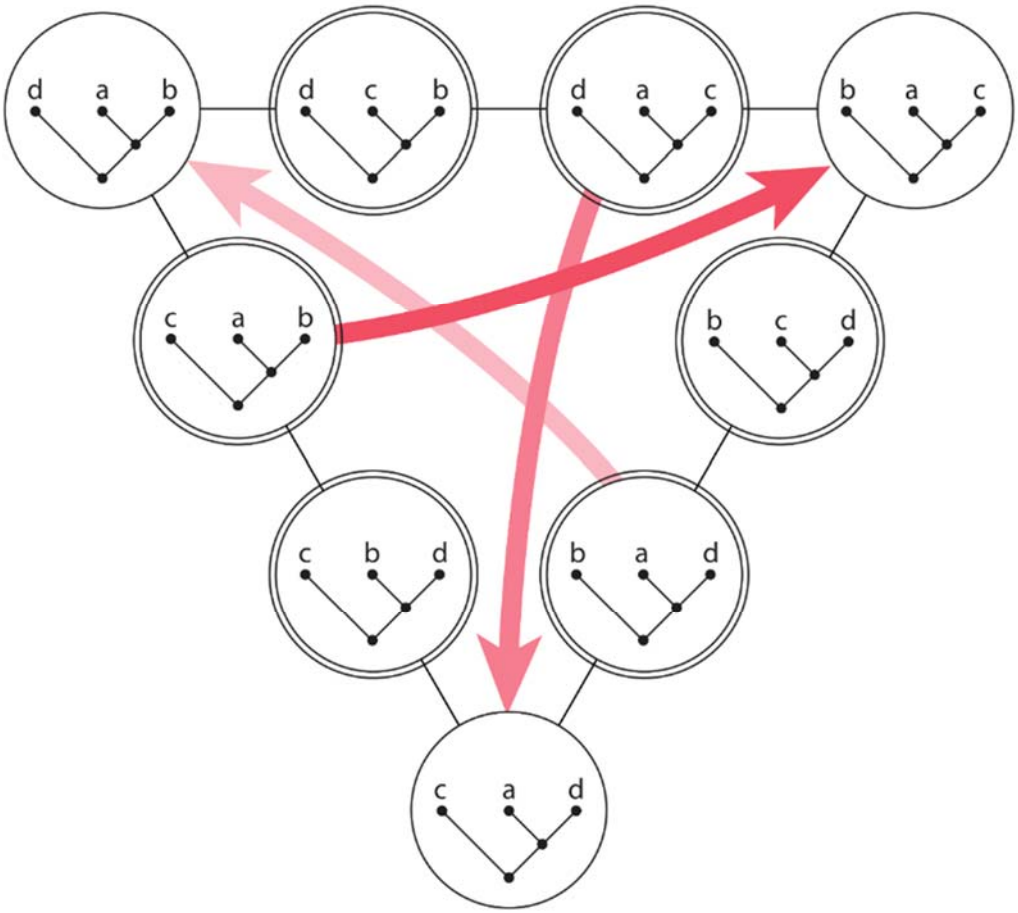


Figure 1.24. Illustration de l'incompatibilité secondaire entre triplets. Trois triplets primaires (cercles simples) sont compatibles deux à deux. Chaque couple est lié par de la dépendance et génèrent des triplets secondaires (cercles doubles). Certains triplets secondaires sont incompatibles (flèches rouges) avec les triplets primaires, rendant impossible la combinaison des trois triplets.

primaires. Certains triplets secondaires sont incompatibles avec les triplets primaires. C'est pour cette raison que la 3ia ne satisfait pas les conditions du théorème de compatibilité par comparaison par paires, qui, s'il était vrai pour les triplets, impliquerait que trois triplets distincts compatibles deux à deux peuvent toujours se représenter sous la forme d'une hiérarchie.

Il est possible cependant de contourner ce problème en « développant » la liste des triplets. La génération de tous les triplets secondaires, ternaires, etc., contraints par des relations de dépendance permet ensuite de détecter les relations de compatibilité et d'incompatibilité. Nous détaillons ici une proposition d'algorithme du bas vers le haut, basée sur ce principe. Son objectif est, sachant une liste contenant des triplets et leurs pondérations associées, de trouver l'ensemble maximal de triplets compatibles.

La figure 1.25 présente le fonctionnement de cet algorithme avec un exemple. Tout d'abord, les triplets primaires sont listés ainsi que leurs pondérations (1). Ensuite, on fait la combinaison des triplets deux à deux. Si une combinaison de type dépendance (fig. 1.19) est trouvée, on génère les nouveaux triplets et on place entre crochets les triplets primaires (2). Certaines combinaisons apportent de la redondance, il est dans ce cas inutile d'ajouter les triplets générés : c'est le cas de la combinaison d'un triplet père et d'un triplet fils (les triplets auront déjà été générés par la combinaison initiale entre les deux triplets pères) ainsi que la combinaison de deux triplets ayant un ou deux pères en commun. A partir de cette liste, on identifie toutes les relations d'incompatibilité. Ici, il existe deux relations d'incompatibilité : une entre *c*, (*d*, *e*) et *e*, (*c*, *d*) (violet), qui permet de construire I1 et de remplacer le triplet fils *e*, (*c*, *d*) par les deux triplets pères *a*, (*c*, *d*) et *e*, (*a*, *c*), et une entre *a*, (*d*, *e*) et *e*, (*a*, *d*) (rouge), qui permet de construire I2 et de remplacer le triplet fils *e*, (*a*, *d*) par les deux triplets pères *a*, (*c*, *d*) et *e*, (*a*, *c*). Le problème est le suivant : le cladogramme est généré à partir de l'ensemble de triplets de poids maximal ne contenant aucune relation d'incompatibilité. Cela revient donc à supprimer le minimum de triplets pour casser toutes les relations d'incompatibilité de la liste d'incompatibilités (3). L'arbre de décision (4) répond au problème de choisir le nombre minimal de triplets pour qu'il y en ait un dans chaque combinaison incompatible (I1, I2). A partir de la racine, on a le choix de supprimer un des trois objets de I1. Ensuite, pour chaque possibilité, on a le choix de supprimer un des triplets de I2 (et on continue l'arbre jusqu'à ce qu'il n'y ait plus de relations incompatibles). Un chemin de la racine vers une feuille de l'arbre est une solution (il faut

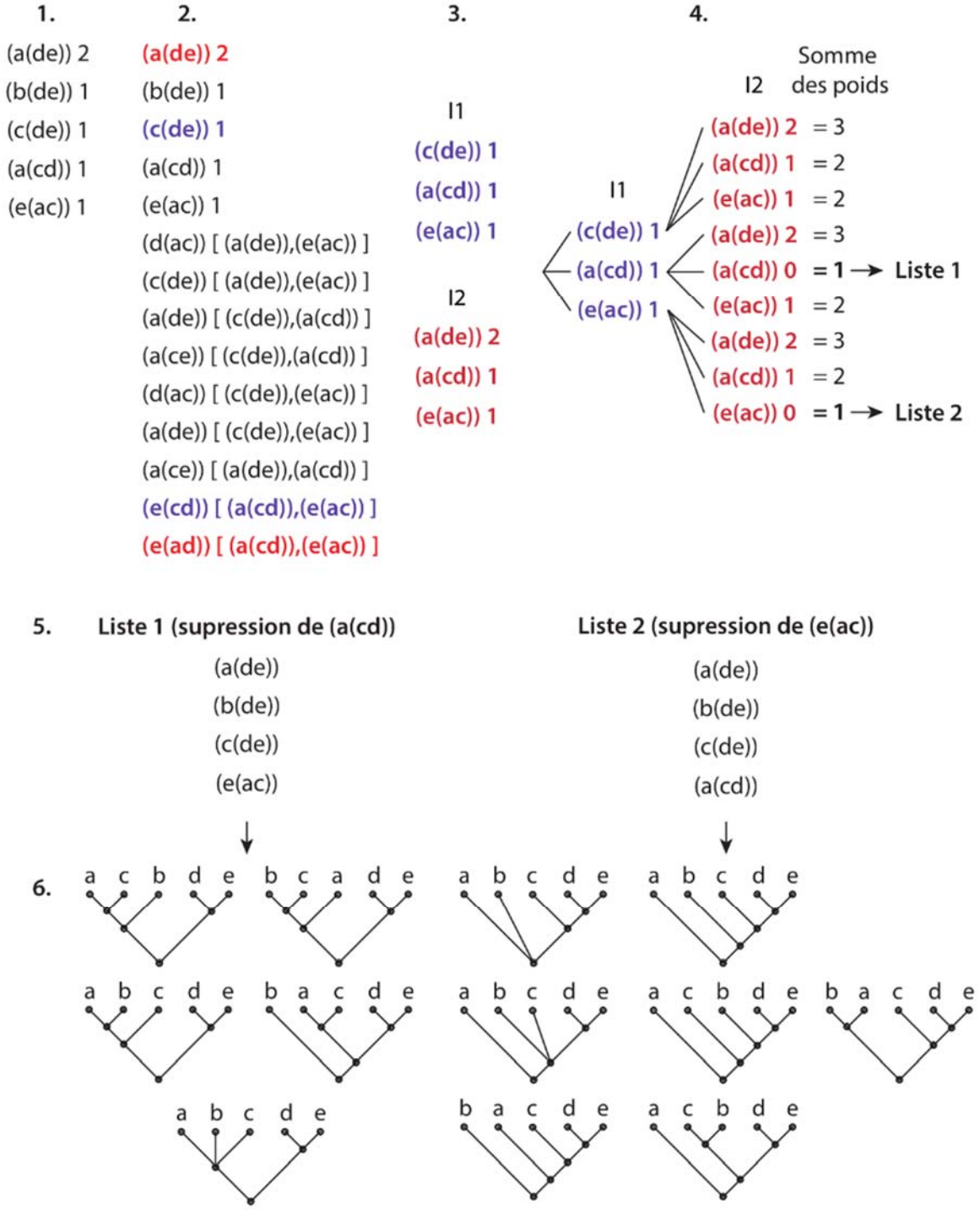


Figure 1.25. Exemple de construction des cladogrammes optimaux par la méthode du bas vers le haut pour l'ensemble de caractères {(*a*, *b*, *c*, (*d*, *e*)), (*a*, (*d*, *e*)), (*a*, (*c*, *d*)), (*e*, (*a*, *c*))}. (1) Liste des triplets primaires issus des caractères et leur PF associée. (2) déduction des triplets secondaires. (3) Liste des combinaisons de triplets incompatibles (si un triplet secondaire est présent dans une combinaison, il est remplacé par l'ensemble de ses triplets parents, entre crochets dans l'étape 2). (4) Arbre de décision de l'ensemble de triplets de poids minimal à supprimer pour faire disparaître toutes les relations d'incompatibilité. Pour une branche, on ne compte qu'une fois un même triplet. (5) deux solutions apparaissent : supprimer (*a*, (*c*, *d*)), et supprimer (*e*, (*a*, *c*)). Les deux listes de triplets optimales conduisent à 16 cladogrammes optimaux (6).

alors supprimer tous les triplets de la branche). La solution optimale est celle pour qui la somme des poids des triplets d'une branche est la plus faible. En effet, il faut supprimer le poids de triplets minimal pour obtenir un cladogramme ayant le poids informationnel le plus grand. À la fin de l'arbre de décision sont scorés les poids de chaque chemin. Dans un chemin, on ne compte pas plus d'une fois un triplet puisqu'un triplet ne peut pas être supprimé de la solution générale plusieurs fois (c'est pour ça que sont notés avec un poids de zéro *a*, (*c*, *d*) et *e*, (*a*, *c*), qui ont déjà été supprimés précédemment). Deux solutions émergent avec un poids minimal de 1 (5) : supprimer *a*, (*c*, *d*) (solution 1) et supprimer *e*, (*a*, *d*) (solution 2). Chaque solution permet de reconstruire tous les cladogrammes contenant ces triplets. On constate que cette méthode permet de générer des arbres polytomiques. Le résultat de l'analyse est un ensemble de 16 cladogrammes.

De manière générale, la méthode de bas vers le haut peut s'appliquer comme suivant. Soit une liste L_t contenant *x* triplets t_x de poids p_x provenant de *c* caractères :

1. Faire toutes les combinaisons de deux triplets t_{x1} et t_{x2} (ou $t_{x1} \neq t_{x2}$) dans L_t qui n'apparaissent pas dans le même caractère *c* (deux triplets issus d'un même caractère sont forcément compatibles).
2. Si t_{x1} et t_{x2} forment une combinaison *in-ex*, orthologue ou paralogue, générer le(s) triplet(s) fils et ajouter à la liste. Garder en mémoire les triplets primaires ayant servi à générer le(s) triplet(s) fils (liste développée de triplets). Répéter cette procédure sur toutes les combinaisons incluant des triplets générés à l'étape précédente, jusqu'à ce qu'aucun nouveau triplet ne puisse être généré. Les combinaisons (triplet père; triplet fils) et les combinaisons entre deux triplets ayant au moins un triplet père en commun n'ont pas besoin d'être vérifiées (*d'autres types de combinaisons sont certainement redondantes mais n'ont pas encore été identifiées*).
3. Construire une liste d'incompatibilités L_i. Faire toutes les combinaisons de deux triplets t_{x1} et t_{x2} (ou $t_{x1} \neq t_{x2}$) de la liste L_t développée. Pour chaque combinaison, s'il existe une relation d'incompatibilité, placer l'ensemble des t_x des deux objets dans un nouvel objet de L_i.
4. Remplacer dans L_i chaque triplet *n*-aire par l'ensemble des triplets primaires ayant servi à le générer.
5. Construire un arbre de décision pour savoir quels sont les ensembles de triplets de poids minimal pour qu'il y ait au minimum un triplet dans chacun des objets de L_i.

a. Pour chaque ensemble, créer une nouvelle liste L_c contenant tous les triplets de L_t moins les triplets de l'ensemble. Chaque ensemble L_c contient l'ensemble des triplets dont est constitué le cladogramme.

On note que s'il est possible de rechercher l'ensemble minimal de triplets à supprimer, il est possible de la même manière de construire un algorithme de recherche de l'ensemble maximal à conserver; cet algorithme sera plus rapide s'il y a un fort taux d'incompatibilité et que plus de la moitié des triplets doivent être supprimés.

3.2.1.3.3. Méthode mixte

Une alternative aux deux méthodes précédentes peut également être mise en place. Il s'agit d'une méthode alternative qui ne fonctionne ni complètement du bas vers le haut ni du haut vers le bas. Il s'agit de tirer parti des propriétés des triplets pour utiliser cette fois-ci les composantes comme structure intermédiaire. Le principe est de tester le poids de chaque composante possible. La formule du nombre C de composantes possibles pour un ensemble de n taxons est :

$$C = \sum_{k=2}^{n-1} \frac{n!}{k!\,(n-k)!}$$

Le graphe en figure 1.26 montre qu'entre vérifier le poids de tous les arbres possibles et vérifier le poids de toutes les composantes possibles, la dernière proposition semble avantageuse. La suite est une analyse de compatibilité de composantes. Les composantes présentent un avantage supplémentaire : puisque on va des composantes vers les triplets, on sait que chaque composante est possible et totalement dépourvue d'incompatibilité entre triplets n-aires. La combinaison de composantes satisfait donc le théorème de compatibilité par comparaison par paires. Il suffit alors de construire le cladogramme comme étant l'ensemble de composantes compatibles de poids maximal. Il faut cependant ajouter une étape de suppression de la redondance entre composantes de la même manière que les pondérations fractionnaires. On mesure un poids de 10 pour la composante (a, b, (c, d)) et un poids de 25 pour la composante (a, (b, c, d)). Or, le triplet a, (c, d) de poids 5 est présent

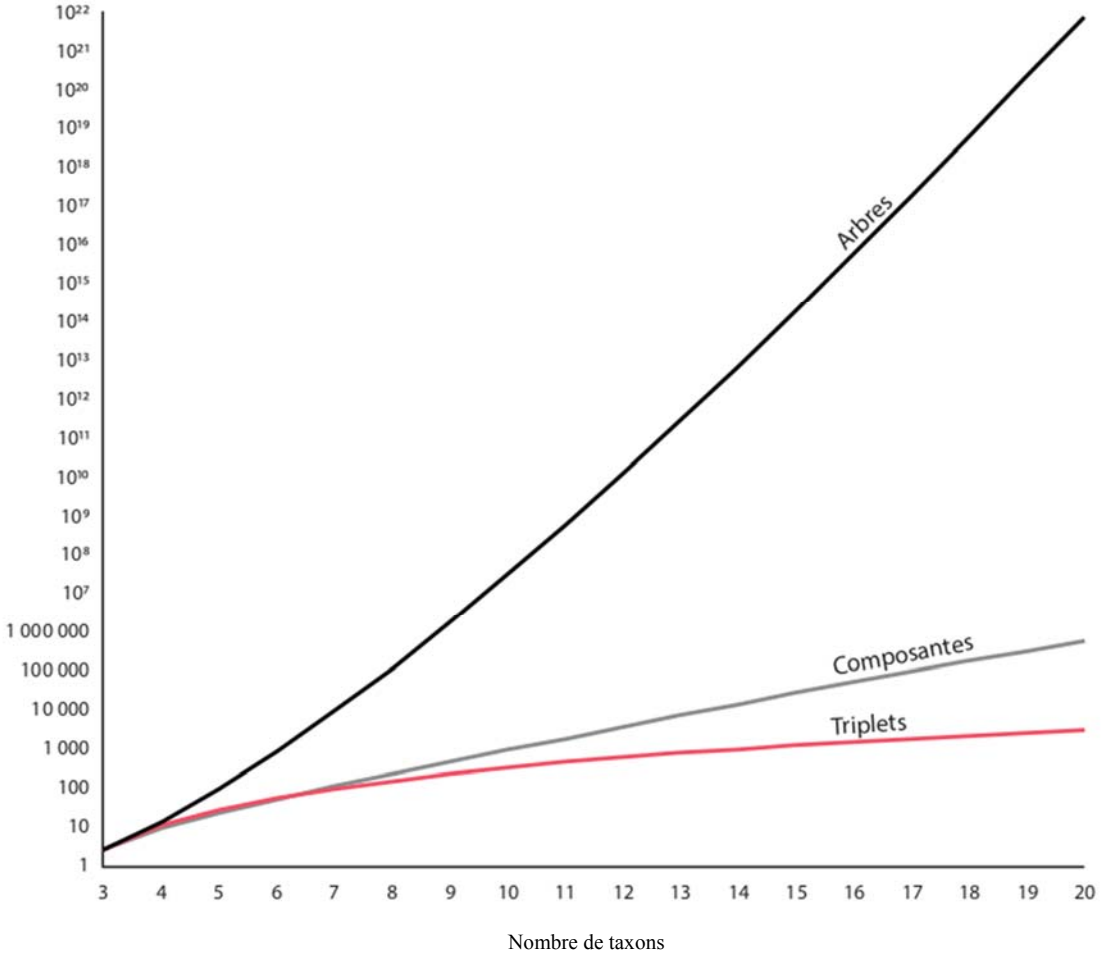


Figure 1.26. Graphe montrant le nombre d'arbres, de composantes et de triplets possibles (ordonnée, logarithmique) par rapport au nombre de taxons de l'analyse (abscisse, linéaire).

deux fois, une fois dans chaque composante. L'ensemble de composantes compatibles {(*a*, *b*, (*c*, *d*)), (*a*, (*b*, *c*, *d*))} vaut donc 30 (10 + 25 - 5). Il faut donc supprimer pour chaque couple de composantes la somme des poids des triplets communs aux deux composantes.

3.3. Caractères hiérarchiques et 3ia : application au problème de Pleijel (1995)

Auteurs: Mathieu Faure-Brac, Valentin Rineau, René Zaragüeta i Bagils.

Cette partie a pour but, après avoir exposé de manière théorique la méthode cladistique de 3ia pour la reconstruction des cladogrammes, de proposer un cas concret où le chercheur, face à une diversité d'objets avec des caractéristiques morphoanatomiques précises, a pour objectif de construire des caractères hiérarchiques. Nous avons retenu l'exemple donné par Pleijel en 1995 qui nous permet de faire appel à toutes les notions vues précédemment dans un exemple à la fois simple et complexe. Cet exemple s'inscrit dans un questionnement autour de la formulation de caractères complexes en cladistique, c'est-à-dire des caractères présentant plus de deux états ordonnés. L'exemple de Pleijel consiste en une caractéristique nommée *X* pouvant se manifester de quatre manières différentes, détaillées en figure 1.27. La caractéristique *X* peut être ronde et noire (fig. 1.27, morphologie 1), ronde et striée (fig. 1.27, morphologie 2), carrée et noire (fig. 1.27, morphologie 3) ou encore carrée et striée (fig. 1.27, morphologie 4). L'exemple ne présage pas ni de la fonction de *X*, ni d'aucune information autre que sa morphologie. L'interprétation en caractères hiérarchiques du problème de Pleijel et de sa caractéristique *X* est l'objet de cette partie.

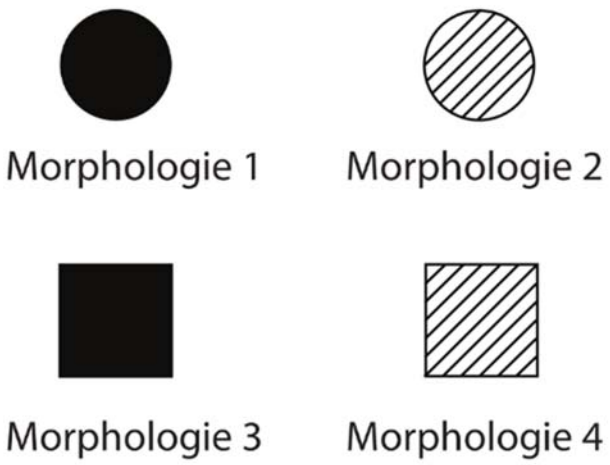


Figure 1.27. Quatre morphologies possibles de la caractéristique *X*. *X* peut être rond et noir (1), rond et strié (2), carré et noir (3)ou carré et strié (4).

Considérons dans cet exemple une analyse cladistique portant sur un ensemble de dix UT {*a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, *i*, *j*}. Parmi elles, les UT *a* et *b* ne portent pas la caractéristique *X*; les UT *c*, *d*, *e*, *f*, *g*, *h*, *i* et *j* au contraire portent la caractéristique *X*. Parmi ces dernières, les UT *c* et *d* possèdent la morphologie 1; *e* et *f* possèdent la morphologie 2; *g* et *h* possèdent la morphologie 3; et *i* et *j* possèdent la morphologie 4. A partir de ces différentes observations, il est possible de proposer différentes hypothèses d'homologies. Suivant l'exemple de Pleijel, nous proposons des rapprochements entre les UT postulés à partir du partage de formes et de pigmentations identiques, ainsi que sur le partage de la caractéristique *X*. Nous pouvons donc délimiter dans un premier temps l'existence de cinq classes homologues minimales : caractéristique *X;* forme carrée; forme ronde; pigmentation striée; pigmentation noire.

Pleijel propose quatre caractères possibles, dont deux sont détaillés à la suite. La première proposition, la plus simple et la plus intuitive, consiste à découper ces observations en plusieurs caractères minimaux considérés indépendants. Cette proposition consiste en un caractère construit pour chaque morphologie particulière (e.g. rond, carré, strié, noir; fig. 1.28-1, b-e) plus un caractère permettant de rassembler les UT sur le partage de la partie *X* par rapport à celles-qui ne l'auraient pas (fig. 1.28-1, a). Dans la seconde proposition, deux caractères relient les UT sur le partage de formes d'une part (carré, rond; fig. 1.28-2 b) et de pigmentations d'autre part (noir, strié; fig. 1.28-2 c), et un caractère rassemble sur le partage de *X* (fig. 1.28-2 a). Ces deux approches, réinterprétées en hiérarchies, ne sont néanmoins pas sans problèmes. Un des prérequis de l'analyse cladistique est l'indépendane entre tous les caractères. Or, ce n'est pas le cas pour ces deux propositions : en effet, le traitement de la pigmentation ou de la forme de la caractéristique *X*, dans quel cas que ce soit, nécessite la présence de cette caractéristique au préalable, présence qui est traitée par un autre caractère, ce qui induit une dépendance logique (i. e. dépendance due à la manière dont sont conceptualisés les caractères).

Les formes et pigmentations différentes de *X* nécessitent la préexistence de cette caractéristique *X*. Plus encore, ce sont des *morphologies particulières* de *X* : elles sont donc des différenciations de la caractéristique *X*. Une forme carrée est donc une forme carrée de *X*, et une pigmentation noire est une pigmentation noire de *X*. Forme et pigmentation sont donc des particularités de *X*. De fait, la pigmentation et la forme étant des différenciations particulières de la caractéristique *X*, elles ne sont pas indépendantes de *X*. Il est nécessaire

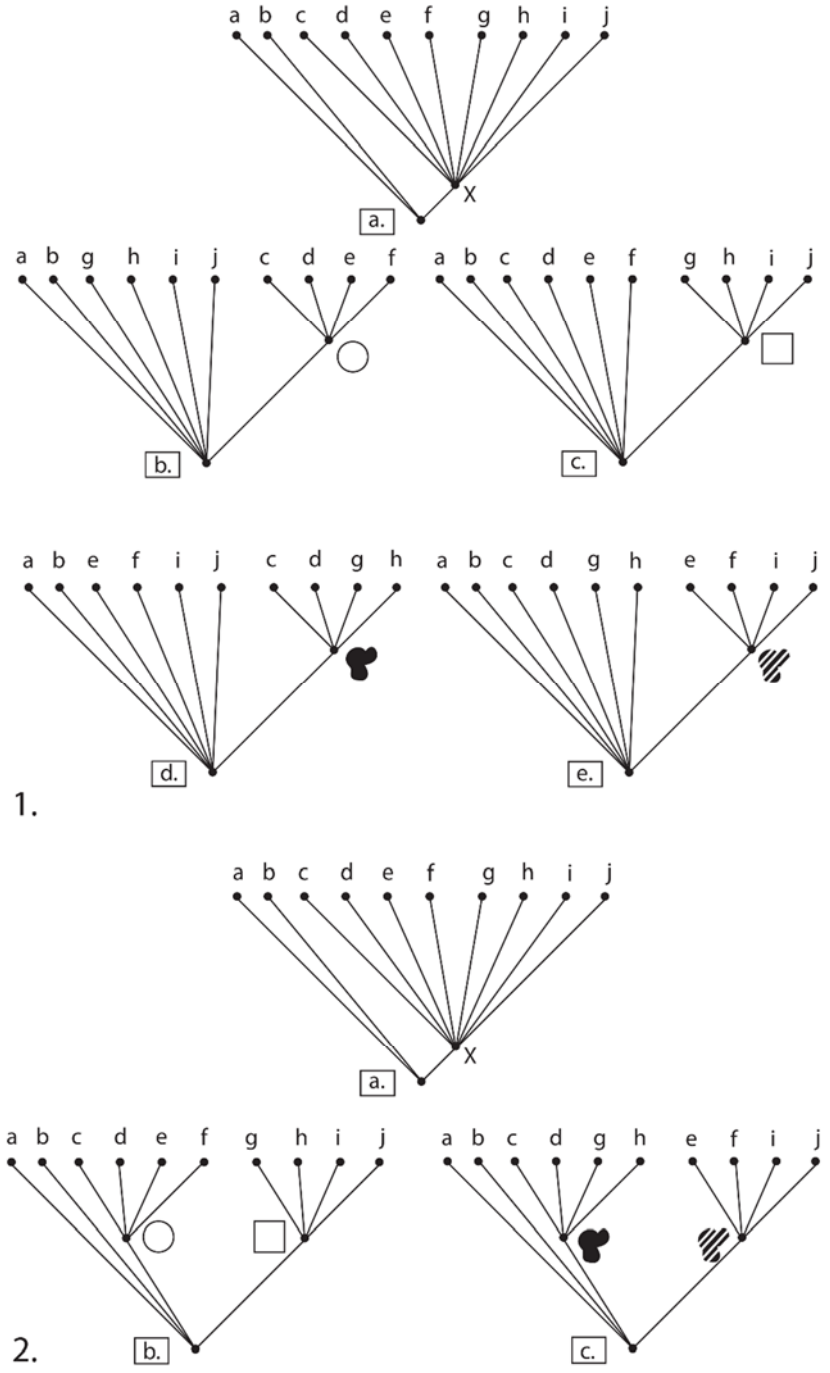


Figure 1.28. Propositions de Pleijel, réinterprétées en caractères hiérarchiques. La proposition 1 présente cinq caractères binaires (interprétation du codage D de Pleijel 1995 en hiérarchies) : le premier rassemble les UT par la présence de la caractéristique *X* (1.a) ; le second par la forme ronde (1.b) ; le troisième par la forme carrée (1.c) ; le quatrième par la pigmentation noire (1.d) ; et le dernier par la pigmentation striée (1.e). Cette proposition vaut 86 triplets pondérés. La proposition 2 présente trois caractères distincts (interprétation du codage A de Pleijel 1995 en hiérarchies) : le premier rassemble par la présence de la caractéristique *X* (2.a) ; le second rassemble selon la forme et réunit les UT sous deux états dérivés possibles, « *X rond* » et « *X carré* » (2.b) ; et le troisième rassemble selon les variations de pigmentation de *X*, distinguant les états dérivés « *X noir* » et « *X strié* » (2.c). Cette proposition vaut 62 triplets pondérés. Le *X* désigne l'apparition de la caractéristique *X* et les différents symboles renvoient aux formes et pigmentations associés.

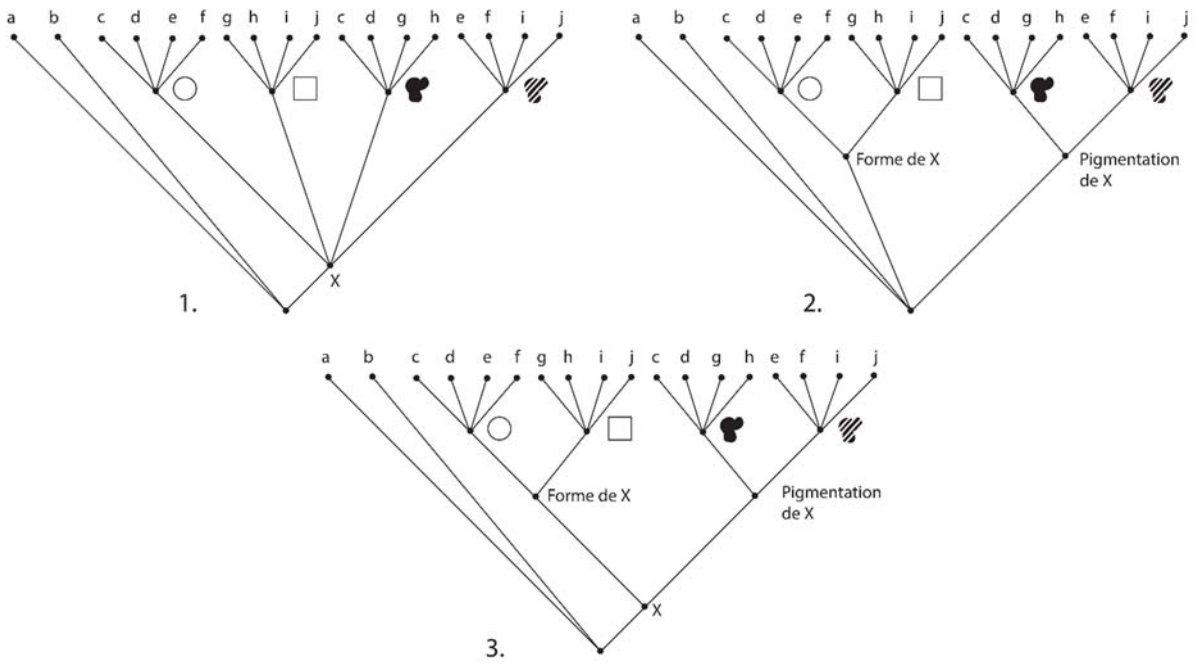


Figure 1.29. Nouvelles hypothèses hiérarchiques présentées ici. La proposition 1 correspond à l'hypothèse rassemblant les différentes formes et pigmentations sous un nœud *X* ; la proposition 2 rassemble ces formes et pigmentations sous deux nœuds distincts « *forme de X* » et « *pigmentation de X* » ; la proposition 3 cumule ces deux approches, rassemblant les nœuds « forme de X » et « *pigmentation de X* » sous le nœud *X*.

que chaque caractère étudié au sein d'une analyse cladistique soit logiquement indépendant des autres (Farris 1983) de manière à éviter la redondance d'information qui pourraient conduire à biaiser le résultat. Forme et pigmentation représentent donc deux homologues inclus dans l'homologue *X*. Etudier deux caractères *forme* et *pigmentation* distincts violerait le principe d'indépendance.

En comparaison des hypothèses de Pleijel, trois caractères peuvent être proposés respectant les principes de construction des hypothèses cladistiques énoncées dans ce chapitre. Ces hypothèses complexes sont liées à la structure hiérarchique de l'homologie et des caractères. Ces hypothèses partent sur une base commune : sont présents, à chaque fois, les homologues *rond*, *carré*, *noir* et *strié*. Dans une première proposition (fig. 1.29-1), les quatre homologues communs sont rassemblés dans un cinquième homologue, la caractéristique *X* elle-même. Dans une seconde proposition (fig. 1.29-2), cette fois les homologues ronds et carrés sont rassemblés sous un homologue *forme de X,* et les homologues noirs et striés sous un homologue *pigmentation de X*. Enfin, on peut faire une dernière proposition qui cumule les deux approches précédentes : les homologues forme et pigmentation sont rassemblés sous un homologue caractéristique *X* (fig. 1.29-3).

Les trois hypothèses d'homologies correspondant aux caractères détaillés précédemment rassemblent donc les caractéristiques *X* quelles que soient leurs morphologies sous un même ensemble (fig. 1.29-1), sous deux ensembles différents distinguant forme et pigmentation (fig. 1.29-2), voire encore en cumulant les deux approches (fig. 1.29-3). Elles sont en accord avec la manière d'appréhender la construction d'homologies en 3ia et mènent à autant de pseudo-caractères différents. Ce sont tous des pseudo-caracactères car ils sont tous concernés par du polymorphisme et, donc, par la répétition de terminaux révélée lors de la standardisation. Ainsi, pour tous les taxons (à l'exception des UT *a* et *b* qui ne portent pas *X*), on peut observer cette répétition. Ceci vient du fait que la caractéristique *X* possède toujours et une forme et une pigmentation. Pour l'UT *c*, la partie c_{pX} (la partie *X* portée par *c*) se retrouve dans les homologues rond et noir. Ce cas se retrouve pour l'UT *d* à l'identique (rond et noir), mais pas pour d'autres comme *e* qui apparait dans les états correspondant aux homologues rond et strié. Dans chacun des cas, les UT sont toutes répétées deux fois. Il faut alors avoir recours à l'analyse en sous arbres libres de paralogie pour passer du pseudo-caractère au caractère, dépourvu de répétition de terminaux. On constate alors que, même s'il s'agit d'hypothèses d'homologies différentes, l'analyse en sous arbre libre de paralogie aboutit au même résultat (fig. 1.30), et donc aux mêmes caractères finaux. Le contenu informatif en triplets pondérés est identique pour toutes ces propositions. Chacune de ces propositions donne 100 triplets pondérés et s'avèrent être plus informatives que les deux premières propositions de la Figure 2. Les propositions de Pleijel converties en hiérarchies et montrées en figure 1.28-1 et 1.28-2 apportent respectivement 86 et 62 triplets pondérés. Elles donnent donc une information moins riche que les trois nouvelles propositions (fig. 1.30). Qu'elle que soit l'hypothèse préférée dans ce contexte, le résultat reste le même; la 3ia a donc cela d'intéressant qu'un consensus dans le codage de ces caractères émerge de l'analyse. La raison en est que, bien que ces nœuds (*i.e.* les nœuds formes ou pigmentation et le nœud *X*) soient pertinents et non redondants, ils ne discriminent pas d'UT. Or, deux nœuds paralogues successifs qui ne discriminent pas d'UT vont juste collapser lors de l'analyse en sous arbre libres de paralogies, car celle-ci vise à retrouver des nœuds orthologues. Le résultat montré en figure 5 peut ensuite servir comme ensemble de caractères lors de l'analyse de l'ensemble des UT $\{a, b, c, d, e, f, g, h, i, j\}$. Ces caractères permettent de discriminer les UT en faisant des hypothèses de clades.

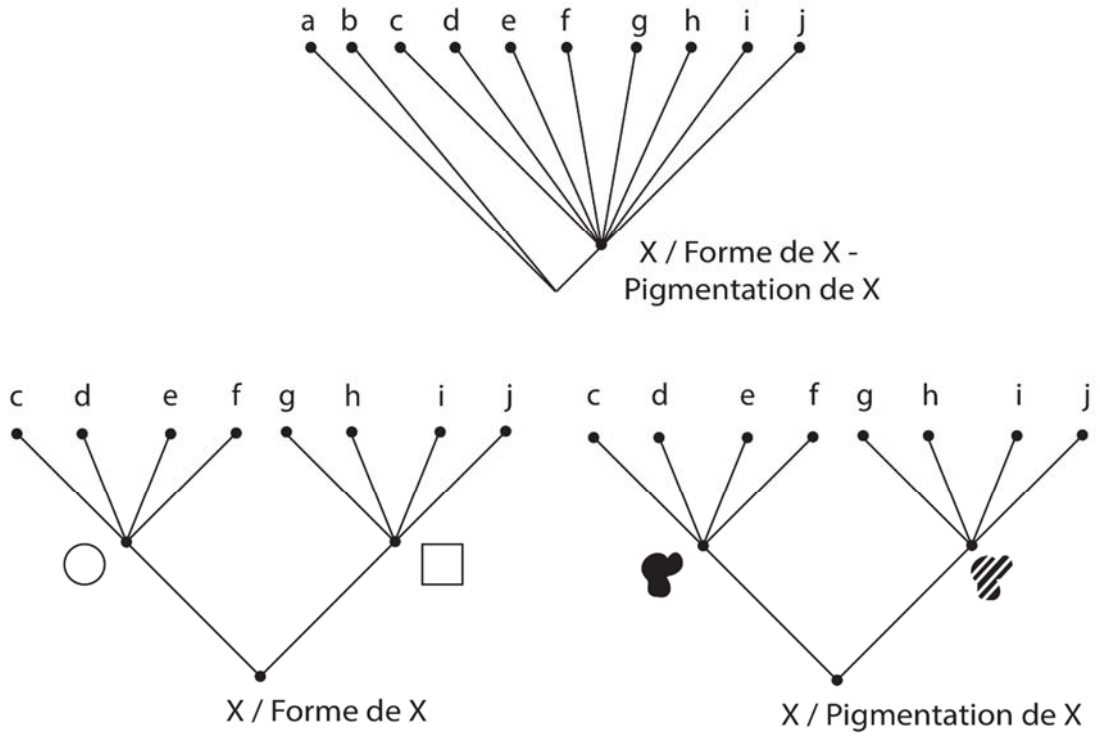


Figure 1.30. Résultat de l'analyse en sous arbres libre de paralogie. L'ensemble des trois propositions de la figure 1.29 donnent toutes le même résultat et valent toutes la même quantité d'information, soit 100 triplets pondérés. On note l'incertitude des noms apparaissant à certains nœuds : cela vient du fait que les nœuds collapsent. Même s'ils sont identiques en termes d'extension, ce n'est pas vrai pour leur intension, qui est ambigüe.

3.4. Logique cladistique et la méthode de 3ia

Le paragraphe suivant récapitule de manière synthétique et non-exhaustive les principes de la théorie cladistique, de l'analyse cladistique, de la méthode 3ia de la production d'hypothèses à la construction du cladogramme et de leurs connexions que nous avons argumenté précédemment.

1. Ontologiquement parlant, les seuls groupes réels sont les clades.
    a. Les clades sont des classes ou groupes monophylétiques d'objets au sens de Hennig (1965, 1966).
    b. Les objets à classer sont eux-mêmes monophylétiques, pour que les classes soient elles-aussi monophylétiques.
    c. Un groupe monophylétique est un groupe complet en termes de filiation.
    d. Les groupes incomplets en termes de parenté sont non-monophylétiques et se déclinent en groupes paraphylétiques et groupes polyphylétiques.
    e. Un groupe non-monophylétique est un groupe artificiel.
2. D'un point de vue épistémologique, les clades sont les taxons les plus pertinents pour comprendre l'évolution des entités qui les constituent.
3. Il existe une hiérarchie inclusive (c'est-à-dire classificatoire) de clades uniques.
    a. Cette hiérarchie peut se représenter par un arbre : le cladogramme.
    b. Cette hiérarchie classificatoire exprime fondamentalement deux relations (Prin 2012 : 89-92, 114, 123) :
        (i) Une relation de degré de parenté (mathématiquement une relation ternaire de degré d'équivalence) entre clades terminaux (et plus généralement entre clades non inclus les uns dans les autres) selon la règle suivante : pour n'importe quels clades terminaux *x*, *y* et *z*, *x* et *y* sont plus proches parents si et seulement s'il existe au moins un clade inclusif dans la hiérarchie qui contient *x* et *y* mais pas *z*;
        (ii) Une relation ancêtre-descendant entre clades (mathématiquement une relation binaire d'ordre inclusif), qui découle directement d'un processus de différentiation selon la règle suivante : *A* est l'ancêtre de *B* si et seulement si *B* s'est différencié à partir et au sein de *A*. Il s'ensuit que *A* est intégralement contenu dans *B*.

c. Cette hiérarchie est toujours composée d'une classe englobant tous les objets : la racine.
d. Cette hiérarchie est toujours composée d'une classe pour chaque objet qui englobe uniquement cet objet : le singleton.
e. Racine et singletons sont monophylétiques.
f. La plus petite unité d'information non-triviale en termes de relation de degré de parenté est le triplet, c'est-à-dire une assertion affirmant qu'étant donnés trois objets distincts, deux d'entre-deux sont apparentés par rapport au troisième.

4. Il est possible de reconstruire les clades grâce à la méthode cladistique.
   a. La méthode cladistique est une méthode permettant de spécifier des relations de degré de parenté entres objets.
   b. La racine et les singletons étant des connaissances préalables, la méthode cladistique nécessite un minimum de trois objets distincts pour spécifier des relations dont on n'avait pas la connaissance avant l'analyse. Ces relations sont des relations de degré de parenté.
   c. Cette méthode repose sur les points suivants
      i. Un objet se décompose en parties. Un objet est la somme de ses parties et l'ensemble des parties permet de reconstruire un objet.
      ii. Un objet soumis à l'analyse cladistique doit être monophylétique. Un objet se réduit en parties. Une partie est donc monophylétique.
      iii. Spécifier des relations de parenté entre parties permet de spécifier des relations de parenté entre objets.
      iv. Spécifier des relations de degré de parenté entre objets est un problème général complexe. Il est possible d'analyser ce problème en un ensemble de problèmes s'énonçant sous la forme de la spécification de relations de degré de parenté entre parties d'objets. La spécification de relations de degré de parenté entre parties est une solution partielle. La synthèse des solutions partielles permet de former la solution générale au problème général.
         1. La méthode cladistique est donc analytique.

2. Une relation de degré de parenté entre parties est une homologie. L'homologie est une hiérarchie d'homologues.
    a. La présence d'un homologue dans une homologie est une hypothèse de clade de parties.
3. Le remplacement des parties d'une homologie par les porteurs de ces parties est la standardisation. La standardisation permet le passage de l'homologie au caractère. La standardisation fait que les solutions partielles (relations entre parties d'objets) permettent de répondre à la solution générale (relations entre objets).
4. Un caractère est une hiérarchie d'états de caractères.
5. Toute homologie ne peut pas être standardisée. Un caractère portant nécessairement sur au moins trois objets distincts, il s'ensuit que toute homologie doit porter sur au moins trois parties de trois objets distincts pour être informative.
6. L'homologie entre parties d'un même objet (homonomie) n'est pas informative pour répondre au problème général.

d. La spécification d'homologues nécessite des arguments. Un argument doit être indépendant de la méthode d'analyse et de synthèse. Il s'agit du principe de *non circularité.* Tout argument doit donc être totalement indépendant de tout autre argument. La justification d'un homologue est donc indépendante de la justification de tout autre homologue.

e. La comparaison d'objets permet d'établir des relations de ressemblance et de différence entre deux objets. Ce point est résumé dans le principe de *comparaison.*

f. Les ressemblances entre objets peuvent être généralisées suivant un principe de *généralisation en propriétés.* La ressemblance deux à deux entre trois objets peut être généralisée en une ressemblance entre trois objets. La ressemblance permet de grouper des objets par rapport à ceux qui ne possèderaient pas la propriété.

g. Les homologies, et leur justification, étant indépendantes les unes des autres, certaines peuvent se révéler contradictoires. La solution générale ne

peut pas toujours inclure toutes les solutions partielles. La solution générale est la solution rassemblant le plus grand nombre de solutions partielles. C'est le principe de *maximisation de la congruence.* Le cladogramme est choisi comme étant la hiérarchie de clades incluant le maximum de relations de parenté provenant des caractères.

i. La décomposition des caractères en triplets permet de rendre indépendantes l'ensemble des relations de parenté d'un caractère.
ii. Dans un caractère, certaines relations peuvent être fausses, d'autres peuvent être vraies.

h. *Analyse logique.* La solution générale n'est que la somme des solutions partielles.
   i. Le cladogramme est donc l'ensemble maximal de triplets compatibles.
      1. Il peut y avoir plusieurs ensembles maximaux de triplets compatibles. Il existe alors plusieurs solutions générales équi-optimales.
      2. Les triplets incompatibles sont rejetés et ne participent pas au cladogramme.
      3. Si c, (a, b) est un triplet rejeté, alors tout taxon contenant a et b mais pas c sera paraphylétique.
   ii. Une relation entre parties est nécessaire et suffisante pour avoir une relation entre objets.

3.5. Interprétation

On construit des homologies, des hypothèses sur les relations de parenté entre parties d'UT. Les homologies sont standardisées en caractères qui sont des hypothèses de relations de parenté entre UT. Ceux-ci forment donc des hypothèses de clades. Les caractères sont décomposés en triplets. Un triplet est une hypothèse de relation minimale. Certaines hypothèses sont incompatibles par rapport à d'autres. Le cladogramme est la hiérarchie de clades maximisant les hypothèses émises par le chercheur (*maximisation de la congruence*). On interprète le cladogramme en revenant aux hypothèses de départ. Deux procédures de test permettre de montrer si ces hypothèses sont acceptées ou rejetées. Le

premier niveau de test des caractères se fait sur les triplets. Le second niveau se fait sur les états de caractères. Plus généralement, il est possible de tester toute hiérarchie cladistique par rapport à un cladogramme en cherchant si celle-ci est compatible à ce cladogramme. On montre que ce test permet l'assignation de propriétés de ces hiérarchies aux nœuds du cladogramme, comme des états de caractères ou du temps.

3.5.1. Indice de rétention

Dans le contexte de la méthode de 3ia, l'*indice de rétention* (IR; Archie 1989; Farris 1989) d'un caractère *r* pour un cladogramme donné *T* correspond au rapport entre le nombre de triplets déductibles de *r* qui sont retenus dans *T* et le nombre total de triplets déductibles de *r* (Nelson et Ladiges 1992). Si *C* est l'ensemble des caractères ayant permis de calculer *T* et *t* un triplet déductible d'au moins un caractère de *C*, alors *t* est retenu vis-à-vis de *T* si et seulement si *t* est compatible avec *T*, c'est-à-dire si et seulement si *t* est déductible de *T*; à ce titre, on peut considérer que *t* a passé le test de congruence au sens où en 3ia, la mise en congruence maximale des caractères consiste à obtenir *T* en maximisant la somme des poids des triplets déduis des caractères de *C* qui sont également compatibles avec *T*; un triplet rejeté par le test est donc simplement un triplet incompatible avec *T*, c'est-à-dire un triplet déductible d'au moins un caractère de *C* qui ne peut se déduire de *T*. La figure 1.31 illustre cette métrique avec un exemple. On calcule l'IR du caractère (*a, b, d,* (*c, e*)) sur le cladogramme (*a,* (*b,* (*c,* (*d, e*)))). Celui-ci contient trois triplets, dont un a été rejeté lors de l'analyse puisqu'il est contredit par le cladogramme. Le poids des triplets accepté est de deux, le poids total du caractère est de 3. L'IR est donc de 2/3.

Il est également possible de calculer l'indice de rétention de tous les caractères de *C* relativement à *T*. Il suffit pour cela de diviser (i) la somme des poids de tous les triplets compatibles avec *T* parmi tous ceux déductibles des caractères de *C* avec (ii) la somme des poids de tous les triplets déductibles des caractères de *C*. Cela peut s'écrire

$$IR = \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} p_i}$$

Où *n* correspond au nombre de caractères de *C*, s_i correspond à la somme des poids de tous les triplets du caractère r_i en accord avec *T*, et p_i la somme des poids de tous les triplets de r_i. Sachant que l'on choisit le cladogramme possédant le plus grand poids de triplets, le cladogramme retenu est toujours celui ayant l'IR maximal. Il n'existe donc pas

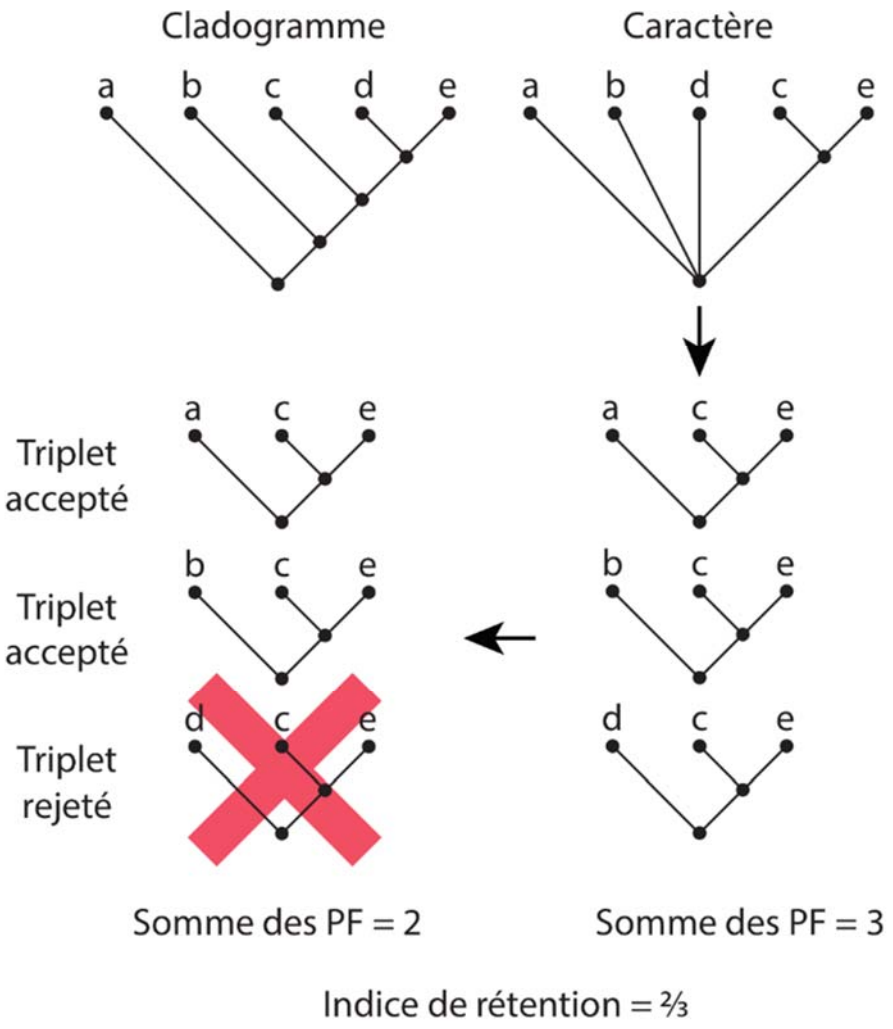


Figure 1.31. Calcul de l'indice de rétention du caractère (*a*, *b*, *c*, (*d*, *e*)) par rapport au cladogramme (*a*, (*b*, (*c*, (*d*, *e*)))). On retire du caractère trois triplets pondérés à 1. Parmi ces trois triplets, un est rejeté par le cladogramme. L'indice de rétention de 2/3 indique la proportion de l'information du caractère participant au cladogramme.

pour une analyse donnée d'arbre ayant un IR plus élevé que le cladogramme retenu. Un IR faible implique un grand nombre de triplets rejetés, et donc une forte incongruence dans les caractères. De nombreuses hypothèses entrent en conflit les unes avec les autres et engendrent de nombreuses relations d'incompatibilité. Un IR fort implique une congruence élevée des caractères dont l'information s'assemble de façon harmonieuse. Un IR = 1 implique qu'aucun triplet n'a été rejeté. Plus un IR est fort et plus on est porté à considérer comme élevée la qualité du cladogramme. Pour conclure, notons que L'IR peut être généralisé en Indice de Rétention Inter-Arbres (IRIA; Grand *et al.* 2013), qui calcule la proportion de triplets d'une hiérarchie contenus dans une autre.

Il est également intéressant d'obtenir des indices permettant de juger des nœuds d'un cladogramme les uns par rapport aux autres. Nous proposons ici une nouvelle métrique, l'*Indice de Rétention Nodal* (IRN_c), qui se calcule pour un clade *c* donné. L'IRN_c a pour utilité d'appréhender si les caractères structurent plus une partie de l'arbre qu'une autre (IRN_c hétérogène entre les clades de l'arbre, écart-type $\sigma(IRN_c)$ fort) ou si l'information des caractères se répartit de manière harmonieuse dans l'arbre. L'indice de rétention nœud par nœud est calculé comme un IR normal, avec la particularité que le cladogramme est remplacé par une unique composante *c* (on conserve du cladogramme uniquement le nœud à calculer). Soit *n* caractères comprenant *t* feuilles dont *p* à l'intérieur du nœud interne, et soit $p_{c\cup n}$ le nombre de feuilles incluses à la fois dans le nœud interne informatif de la composante et de l'état, et $t_{c\cup n}$ le nombre de feuilles à la fois en dehors du nœud interne informatif de la composante et de l'état.

$$\mathrm{IR}_{\text{par noeud}} = \frac{[\sum_{i=1}^{n}(p_{c\cup n}-1)(t_{c\cup n}-p_{c\cup n}))]}{[\sum_{i=1}^{n}(p_n-1)(t_n-p_n))]}$$

Cette seule métrique est cependant insuffisante. Sachant que le but est de comparer deux clades, on obtient un biais indésirable dû aux propriétés des triplets d'une hiérarchie d'être plus nombreux pour une composante équilibrée (fig. 23). La métrique donne donc toujours une valeur plus forte aux nœuds symétriques du cladogramme. Il faut mettre à l'échelle l'IR par nœud connaissant la structure des triplets. On représente la part due à la structure de la hiérarchie par le contenu informatif d'une composante *c* INF_c comme le rapport entre le poids d'une composante par rapport au poids total du cladogramme.

$$\mathrm{INF}_c = \frac{(p_c-1)(t_c-p_c)}{[\sum_{\mathrm{i}=1}^{\mathrm{c}}(p_c-1)(t_c-p_c))]}$$

Ce qui nous donne l'IRN_c, qui est l'IR par nœud mis à l'échelle en fonction de l'inclusivité du clade dans la hiérarchie (INF_c).

$$\mathrm{IRN}_c = \frac{\mathrm{IR}_{\text{par noeud}}}{\mathrm{INF}_c}$$

La formule développée est donc

$$\mathrm{IRN}_c = \frac{[\sum_{i=1}^{n}(p_{c\cup n}-1)(t_{c\cup n}-p_{c\cup n}))][\sum_{i=1}^{c}(p_c-1)(t_c-p_c))]}{[\sum_{i=1}^{n}(p_n-1)(t_n-p_n))]\,(p_c-1)(t_c-p_c)}$$

L'IRN_c est borné par]0; +∞[. Elle sert à comparer différentes composantes d'un même cladogramme. L'écart-type σ(IRN) de l'ensemble des valeurs de l'IRN_c du cladogramme peut être comparé entre différents cladogrammes de différentes analyses. Une application de l'IRN_c est donnée au chapitre 4 (p. 315). La figure 1.32 présente quelques exemples de l'IRN d'un caractère par rapport à un cladogramme, réalisés par un script R (R Core Team 2013) et disponible à l'adresse suivante : https://github.com/JeremieBardin/Valentin-Rineau-PhD-scripts (auteur: Jérémie Bardin). Ces exemples permettent de mieux comprendre à la fois la relation entre triplets et cladogramme, ainsi que le fonctionnement de l'IRN et comment celui-ci doit être lu. Dans l'arbre 1, on calcule les différents IRN du caractère (*a*, (*b*, *c*, *d*, *e, f, g, h, i, j, k, l, m, n, o*)). Le clade groupe frère de *a* est parfaitement congruent avec ce caractère. La totalité de l'information contenue dans le caractère, soit 35 triplets, se retrouve donc dans ce nœud. Cependant, on remarque que les autres nœuds sont également soutenus, bien que partiellement, par ce caractère. Le clade {*n*, *o*} par exemple est soutenu par 2,69 triplets*PF. Il s'agit en fait du triplet pondéré *a*, (*n*, *o*) qui est le seul triplet issu du caractère à être congruent avec ce clade. Le même raisonnement s'applique à tous les autres nœuds de l'arbre. Il faut donc retenir ici qu'un triplet soutient de nombreux clades au sein d'un cladogramme. L'arbre 2 présente exactement la même configuration avec un caractère (*a*, *b*, *c*, *d*, *e, f, g, h, i, j, k, l, m,* (*n, o*)) parfaitement congruent avec un clade, le groupe-frère de *m* en l'occurrence. Dans l'arbre 3, le caractère est cette fois-ci (*a*, *b*, *c*, *d*, *e, f, i, j, k, l, m, n, o,* (*g, h*)). On interprète l'état dérivé de ce caractère comme ayant été perdu chez *i*, *j*, *k*, *l*, *m*, *n* et *o*. Certains des triplets issus du caractère servent à soutenir le clade {*g, h, i, j, k, l, m, n, o*} (mais partiellement uniquement). Certains de ces triplets, comme *a*, (*g*, *h*), servent également à soutenir les clades incluant {*g, h, i, j, k, l, m, n, o*}. Cependant, aucun triplet issu du caractère ne peut servir à soutenir une quelconque relation au sein du clade {*g, h, i, j, k, l, m, n, o*}. On a ici une illustration du fait que les caractères en 3ia ne servent qu'à soutenir des clades selon les hypothèses effectuées en amont par le chercheur. Ici, la perte peut être invoquée pour illustrer la distribution de l'état sur l'arbre, mais aucun triplet ne peut soutenir la perte. On remarque que les valeurs de l'IRN sont nettement inférieures à celles des exemples 1 et 2. La perte, si elle ne joue pas sur le rejet ou non d'un état de caractère, fait perdre une quantité d'information très importante vis-à-vis du caractère initial et des triplets qui en découlent. Les deux derniers exemples 4 et 5 illustrent le

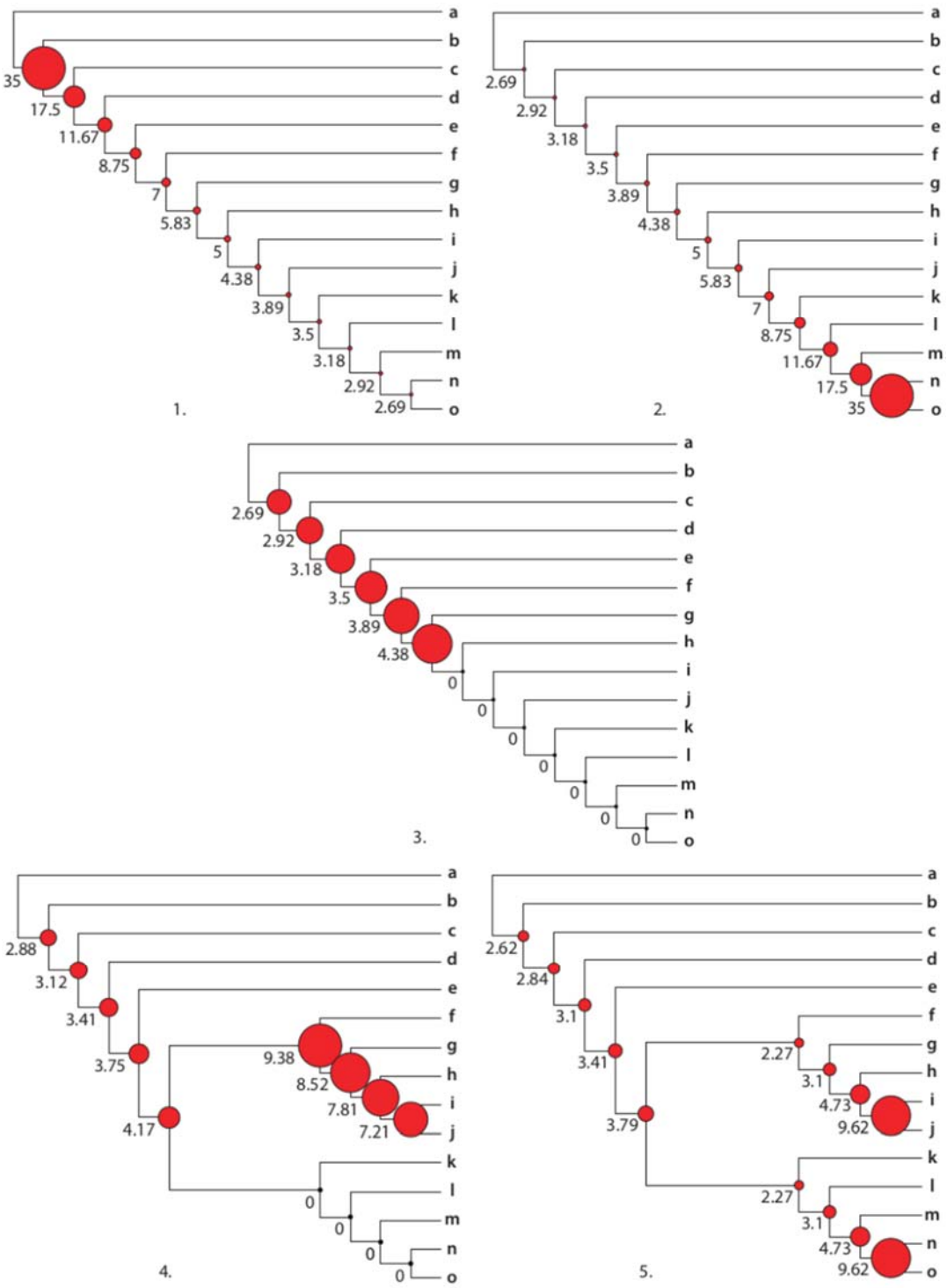


Figure 1.32. Exemples de répartition des valeurs de NRI sur un cladogramme pour un caractère donné. (1) Valeurs de NRI d'un arbre pectiné pour un caractère (*a*, (*b*, *c*, *d*, *e, f, g, h, i, j, k, l, m, n, o*)). L'état dérivé vaut 35 triplets pondérés qui se retrouvent tous dans le clade {*b*, *c*, *d*, *e, f, g, h, i, j, k, l, m, n, o*}. Cependant, certains triplets participent également à des nœuds moins inclusifs. (2) Valeurs de NRI d'un arbre pectiné pour un caractère (*a*, *b*, *c*, *d*, *e, f, g, h, i, j, k, l, m,* (*n, o*)). (3) Valeurs de NRI d'un arbre pectiné pour un caractère (*a*, *b*, *c*, *d*, *e, f, i, j, k, l, m, n, o,* (*g, h*)). (4) Valeurs de NRI d'un arbre avec un paralogue pour un caractère (*a*, *b*, *c*, *d*, *e, k, l, m, n, o, (f, g, h, i, j)*). Le nœud paralogue fonctionne comme un point de diversification à partir duquel deux lignées deviennent indépendantes l'une de l'autre. (5) Valeurs de NRI d'un arbre avec un paralogue pour un caractère rejeté (*a*, *b*, *c*, *d*, *e, f, g, h, k, l, m, (i, j, n, o*).

comportement de caractères vis-à-vis d'un arbre présentant un nœud paralogue. Dans l'exemple 4, le caractère (*a*, *b*, *c*, *d*, *e, k, l, m, n, o, (f, g, h, i, j)*) est parfaitement congruent avec le clade {*f*, *g*, *h*, *i*, *j*}; c'est donc à ce nœud qu'on trouvera la valeur maximale de IRN, qui est la valeur totale d'information portée par le caractère. Pour les nœuds plus inclusifs et moins inclusifs, la valeur de l'IRN décroit logiquement comme pour les exemples précédents. On a ici en plus un exemple clair des nœuds paralogues comme instances de diversification : à partir d'un nœud paralogue, deux lignées se mettent à diverger, des lignées qui deviennent indépendantes l'une de l'autre. L'état de caractère dérivé de cet exemple est synapomorphique d'un clade localisé dans l'une des deux lignées d'un paralogue. Aucun triplet ne peut soutenir de clade dans l'autre lignée du nœud paralogue, cette lignée étant totalement indépendante. De manière plus générale, un triplet peut soutenir deux clades liés par une relation d'inclusion, par exemple {*i*, *j*} et {*h*, *i*, *j*}. Cela est le cas de *a*, (*i*, *j*). Cependant, un triplet ne peut jamais soutenir deux clades disjoints, par exemple {*i*, *j*} et {*n*, *o*}. L'exemple 5 est encore un peu plus complexe. Cette fois-ci, le caractère (*a*, *b*, *c*, *d*, *e, f, g, h, k, l, m, (i, j, n, o*) soutient plusieurs nœuds disjoints. Il s'agit là du cas d'un état homoplasique. L'état, rejeté, sert à soutenir deux clades {i, j} et {n, o}; on s'est donc trompé sur l'hypothèse phylogénétique préalable. Le caractère se comporte *in fine* comme si ce n'est pas un mais deux états dérivés indépendants – car disjoints – qui constituent le caractère. Tous ces exemples permettent de mettre en évidence (i) la sensibilité des triplets à la dépendance ou à l'indépendance des clades les uns vis-à-vis des autres et (ii) la sensibilité des triplets à l'homoplasie et aux phénomènes de pertes.

Il peut également être intéressant de connaître la structuration du contenu informatif des nœuds. S'il est possible de faire des sous-ensembles de caractères, on peut connaître la proportion de chaque sous-ensemble pour un nœud donné. Par exemple, si un ensemble *n* de caractères peut être divisé en deux groupes, comme les caractères de la valve gauche et les caractères de la valve droite chez les bivalves, on peut calculer, pour un clade *c*, l'IRN_{c1} du premier sous-ensemble de caractères et l'IRN_{c2} du second sous-ensemble de caractères. On peut ensuite calculer IRN_{c1} / IRN_{c1} + IRN_{c1} comme la proportion de caractères de la valve gauche soutenant ce nœud. Entre autres exemples, on peut obtenir pour un clade la proportion de caractères de la valve gauche de bivalves, des caractères portant sur les dents de mammifères ou encore les caractères issus du CO1 par rapport aux autres gènes. Une application de l'IRN_c est également présentée au chapitre 4 (p. 317).

3.5.2. Test des états de caractères

L'acceptation ou le rejet des triplets et son calcul par l'IR constitue un premier niveau de test cladistique. Revenir aux hypothèses construites en première instance – les états de caractères – permet d'approcher une autre facette du cladogramme et de l'analyse phylogénétique. Le caractère est l'hypothèse fondamentale de l'analyse, une hypothèse de clade formulée à partir d'une homologie. On n'a du début à la fin de l'analyse jamais rompu la structure hiérarchique des homologies, des caractères et du cladogramme (fig. 1.33), puisque le cœur de l'information cladistique se représente sous la forme d'une hiérarchie. Lors de la standardisation d'une homologie *h* en un caractère *c*, on remplace les parties d'UT mise en relation dans *h* par les UT portant ces parties tout en respectant autant que possible la structure de *h*. Pour le cas spécifique ou deux parties ne proviennent jamais d'un même porteur (homologie spéciale au sens d'Owen 1848; orthologie au sens de Prin 2012) la structure hiérarchique de *c* est ainsi identique à celle de *h* (Prin 2012 : 147-151). Les homologies étant hiérarchiques, il s'ensuit de même pour les caractères ainsi que pour les composantes et les triplets déductibles de ces caractères. Enfin, recomposer des triplets combinables (et donc compatibles) implique, là encore, une structure hiérarchique.

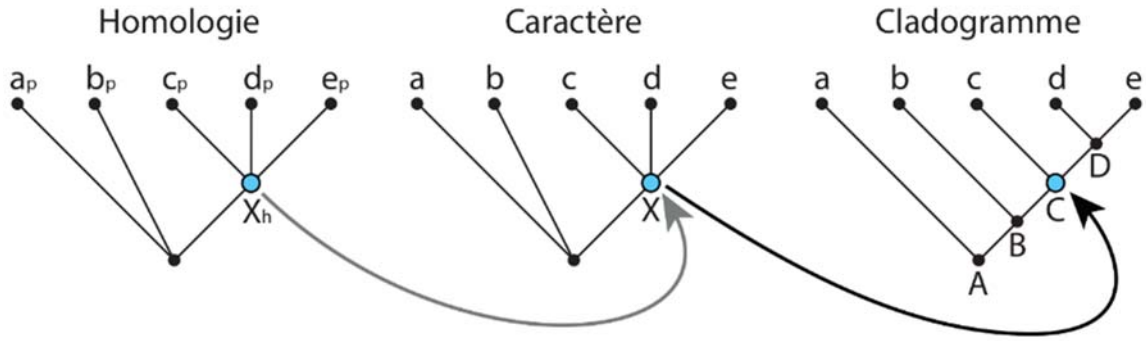


Figure 1.33. Correspondance entre un homologue X_h, hypothèse de clade sur des parties d'UT, et l'état de caractère correspondant (*C*) après standardisation, hypothèse de clade sur des UT (flèche grise). Assignation de l'état *X* au clade *C* (flèche noire). L'état *X* est la synapomorphie du clade *C*, que l'on note *C*(*X*). Le clade *C* regroupe les taxons *c*, *d* et *e* par la propriété « *avoir* X_h ».

Il est donc possible de comparer états et clades, ceux-ci provenant respectivement de caractères et de cladogrammes qui ont la même structure (une classification d'UT). Cette comparaison se fait lors du *test des états de caractère* qui constitue le second niveau de test cladistique après les triplets (Cao 2008 : 73-86; Prin 2012 : 194-197). Un état se teste par rapport à un cladogramme. Si un état de caractère peut être assigné sans ambigüité à un

seul clade du cladogramme, alors l'état de caractère est non-rejeté. Par exemple, dans la figure 1.33, l'état *X* est une hypothèse de clade $\{a, b, c\}$ que l'on peut retrouver dans le cladogramme et relier au clade $C = \{a, b, c\}$. Le clade *C* est – entres autres – l'état *X* qui a passé le test. S'il le test réussit, on assigne l'état à un clade; la synapomorphie est un état qui caractérise un clade. Par exemple, l'état de caractère « avoir une coquille bivalve » est une synapomorphie du clade Bivalvia. Dans notre exemple, l'état *X* est la synapomorphie définissant le clade *C*.

Un état de caractère, en tant qu'hypothèse de clade, postule le groupement de plusieurs UT sous un même degré d'identité. La relation de degré d'identité entre UT peut se traduire par l'assertion « *ces UT sont les mêmes par rapport aux autres* » et forment une classe de généralisation de ces UT par rapport à d'autres. On réfute un état de caractère lorsque l'on a montré qu'il est incongruent, c'est à dire lorsqu'on ne peut pas l'attribuer à un nœud unique du cladogramme, interdisant ainsi son interprétation en tant qu'héritage par ascendance commune. Dire que des UT ne sont pas les mêmes implique qu'il est impossible de grouper ces UT dans une classe unique : l'incongruence génère de la polyphilie. Un état qui est rejeté par le test ne peut pas caractériser un clade unique et devient homoplasique. La source de l'homoplasie est toujours l'incongruence entre le cladogramme et les caractères. L'homoplasie est donc au regard de la cladistique de l'erreur. Une infinité de raisons biologiques, taphonomiques, psychologiques ou sociologiques peuvent être recherchées a posteriori pour expliquer l'incongruence. Biologiquement par exemple, l'hérédité verticale (héritage par ascendance commune) ne peut générer que de la congruence; l'hérédité horizontale, la perte de copies paralogues et la convergence en revanche peuvent autant générer de la congruence que de l'incongruence. Par conséquent, il n'existe pas d'explication spécifique de l'incongruence. On doit donc maximiser la congruence et/ou minimiser l'incongruence, c'est-à-dire maximiser ce que l'on peut expliquer et/ou minimiser ce que l'on ne peut pas expliquer (Prin 2012 : 164-167). L'erreur dans la reconnaissance d'un homologue due à la convergence anatomique, dans la construction d'une homologie à cause de l'utilisation de mauvais arguments, dans l'identification d'une partie d'UT à un homologue ou encore dans le formatage d'un fichier destiné à l'analyse sont autant d'exemples générant de l'incongruence. L'étude de ces erreurs peut permettre de raffiner les arguments soutenant les homologies, et détecter une erreur permet de générer une homologie dont la monophylie sera potentiellement mieux

étayée : on appelle ce procédé le *retour aux caractères*. La justification des homoplasies en tant que telles sont hors du champ de la cladistique puisqu'elles sont, justement, non-monophylétiques.

Pour illustrer le test d'états, on a pour chaque exemple un état dérivé *X* et un cladogramme dont on note seulement les UT présentes dans *X*. L'exemple de la figure 1.34 présente un cladogramme (*a*, (*b*, (*c*, *d*))) et trois états de caractères. L'exemple 1 montre le cas le plus évident ou l'état X = {*c*, *d*} correspond parfaitement à un clade *C* = {*c*, *d*}. Le clade *C* est identifié à l'état *X* et acquiert ses propriétés. Le cas 2 apporte une différence dans la mesure où il n'existe pas de clade concordant parfaitement à l'état *X*. En termes phylogénétiques, un état est une hypothèse sur l'existence d'un TH ancêtre commun à des UT. On cherche la classe la moins inclusive contenant tous les UT de l'état dérivé. Dans ce cladogramme, seules trois options sont possibles pour placer un état *X*, puisqu'il n'existe

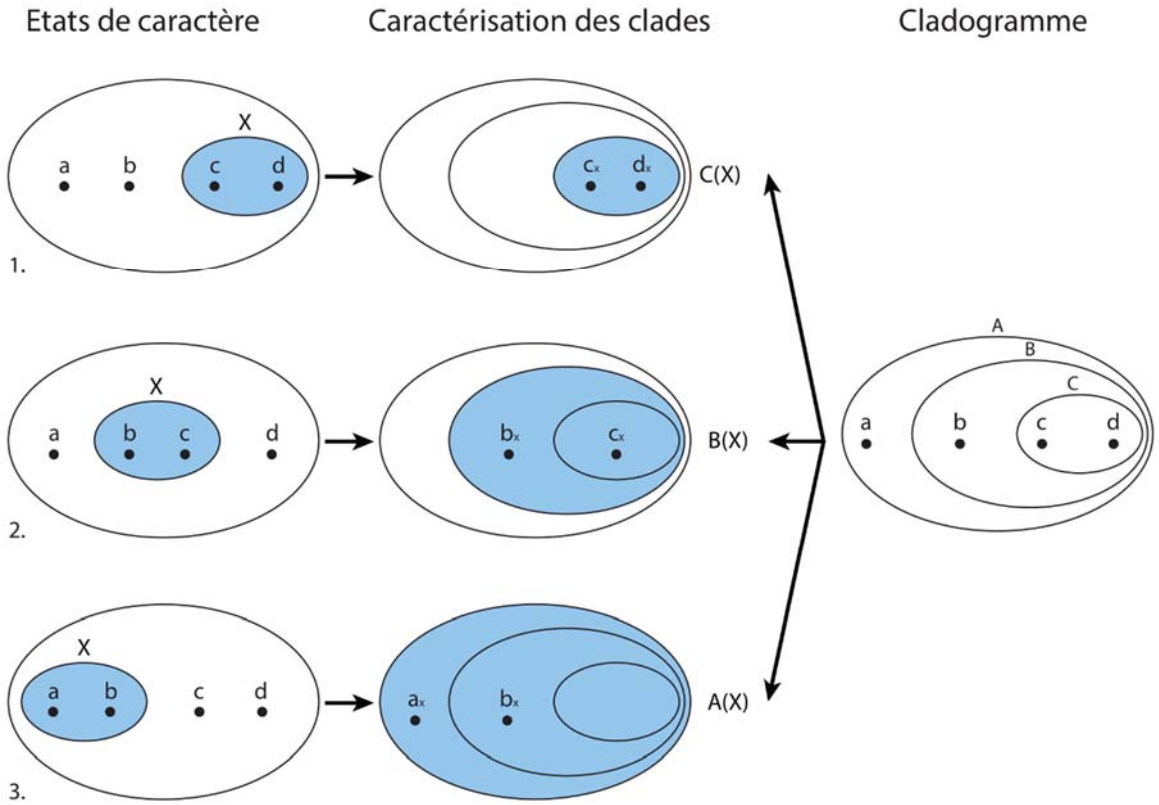


Figure 1.34. Caractérisation d'un état *X* à un clade du cladogramme (*a*, (*b*, (*c*, *d*))). A l'étape caractérisation des clades, on reprend le cladogramme sur lequel on n'affiche que les UT qui portent l'état *X*. Le cladogramme est composé de trois clades, il n'y a donc que trois possibilités. (1) L'état *X* correspond parfaitement au clade *C*. (2) L'état *X* est associé au clade *B*, *d* ayant potentiellement subi une réversion. (3) L'état *X* est associé à la racine *A*, *c* et *d* ayant potentiellement subi une réversion. Un état peut toujours être assigné à un clade si le cladogramme est une orthologie.

que trois clades : *A*, *B* et *C*. Pour qu'un état soit assigné à un clade, le clade doit logiquement contenir toutes les feuilles contenues dans l'état, sans quoi le clade ne serait pas commun à toutes les UT. Dans l'exemple 2, on élimine donc le nœud 1. Dans ces conditions, la classe la moins inclusive contenant tous les UT porteurs de l'état dérivé *X* est la classe *B*. On remarque que la classe *B* contient cependant l'UT d, qui ne porte pas l'état dérivé. Ce cas montre que l'assignation des états aux clades est une caractérisation plutôt qu'une définition (Prin 2012 : 203-214) : si l'état *e* (ex : « *avoir deux paires de membres chiridiens* ») caractérise le clade *c* (ex : « *être un tétrapode* »), alors posséder *e* est suffisant pour être dans *c* (et donc être dans *c* est nécessaire pour avoir *e*) mais pas l'inverse, c'est-à-dire qu'être dans *c* n'est pas forcément suffisant pour avoir *e* (et donc avoir *e* n'est pas forcément nécessaire pour être dans *c*). Il s'agit donc d'une relation entre état et clade plus faible (mais aussi plus souple) que la définition. En effet si *c* était défini par *e*, alors avoir *e* serait nécessaire et suffisant pour être dans *c*, ce qui interdit soit l'existence des pertes secondaires, soit la possibilité de les mettre en évidence. Le clade *c* peut donc contenir des UT qui n'instancient pas *e* tout en étant caractérisé par ce dernier. En d'autres termes, le fait que *c* soit caractérisé par *e* n'implique pas que toutes les UT de *c* aient des parties instanciant l'homologue sous-jacent à *e* (c'est-à-dire « *être un membre chiridien* » pour notre exemple). Dans un cadre phylogénétique, on interprète généralement l'absence d'une propriété chez certaines UT incluses dans un clade possédant cette propriété comme une réversion. La 3ia est donc une méthode qui fournit des cladogrammes pouvant être interprétés en termes évolutifs, et la réversion, c'est-à-dire la perte d'un état dérivé chez certaines UT, en fait partie. La présence de réversions n'empêche pas de dire que telles UT sont apparentées, ni d'assigner un état à un clade. Une réversion ne réfute pas non plus l'hypothèse d'une origine commune à plusieurs UT. Dans ce cas 2, les UT *b* et *c* sont bien apparentées par une origine commune, et le fait que cette origine commune ne soit pas exclusive ne change pas le fait qu'il n'y a qu'une origine, et donc bien un groupement unique. La réversion a en revanche des effets visibles sur les triplets : les réversions font baisser l'IR d'un caractère.

Il apparaît que quel que soit l'état défini en amont, il n'est pas possible de le réfuter sur ce cladogramme. La raison est que ce cladogramme est une orthologie. Cette structure ne contient aucun évènement de diversification. Sur une anagenèse, il est toujours possible d'assigner un état à *un clade unique* – éventuellement avec des réversions. La réfutation

d'un état, c'est-à-dire la mise en évidence que cet état n'a pas une origine unique, ne peut être montrée que si le cladogramme contient des lignées indépendantes. Alors, si on met en évidence que des UT que l'on croyait apparentées se retrouvent sur des lignées indépendantes, il est impossible de soutenir la thèse d'un ancêtre commun exclusif, et l'état est rejeté.

Le test des états ne peut donc se faire qu'en présence de nœuds paralogues. L'apparition d'UT de l'état dans plusieurs lignées indépendantes réfute cet état, puisqu'on ne peut alors plus dire que ces UT sont les mêmes, c'est dire qu'elles ont une origine commune. L'état réfuté est alors considéré homoplasique (car polyphylétique) et ne peut être assigné à un clade unique en tant que synapomorphie. La figure 1.35 montre un cladogramme contenant un nœud paralogue pour lesquels on essaye d'assigner un état. Dans l'exemple 2, toutes les instances de l'état *X* se situent sur une orthologie. Aucun paralogue n'est impliqué. La caractérisation de l'état se fait de la même manière que précédemment. L'exemple 3 montre exactement le même cas de figure, toutes les instances de *X* sont contenues dans une orthologie, et *X* peut être placé comme précédemment. On pourrait imaginer le même cas avec des réversions, ce qui ne changerait rien au placement de l'état. Le premier exemple à impliquer un paralogue est l'exemple 4. Dans ce cas, les orthologies doivent être traitées comme des lignées indépendantes. La première orthologie laisse penser que l'état se place au nœud *E*, rien ne permet de dire que l'état est plus inclusif sur cette orthologie. De la même manière sur l'autre lignée prise isolément, on place l'état sur le nœud *K*. L'état ici peut être place à deux endroits différents – indépendants – de l'arbre. On ne peut donc pas dire que les instances de *X* chez $\{f_x, e_x\}$ et $\{k_x, l_x\}$ sont les mêmes : l'état est rejeté. On a rejeté l'état *X* qui se retrouve dans deux lignées indépendantes reliées de l'arbre. L'exemple 5 montre une distribution plus importante de l'état *X*. Sur la première orthologie, on place *X* et on constate qu'il englobe l'ensemble des nœuds $\{A, B, C, D, E\}$, avec une possibilité de réversion. Dans la seconde orthologie, l'état *X* englobe également tout. On vient alors au nœud paralogue, et on constate que les deux lignées qui partent de lui portent l'état *X*. Il est donc légitime de considérer que le nœud paralogue lui-même porte l'état *X*. L'état *X* n'est pas rejeté. L'exemple 6 est un cas extrême ou de nombreuses réversions entrent en jeu. Cependant, les deux orthologies possèdent l'état *X*, aux nœuds *A* et *G*. Le paralogue est donc caractérisé par *X* puisque les orthologies qui partent du paralogues possèdent toutes l'état *X*.

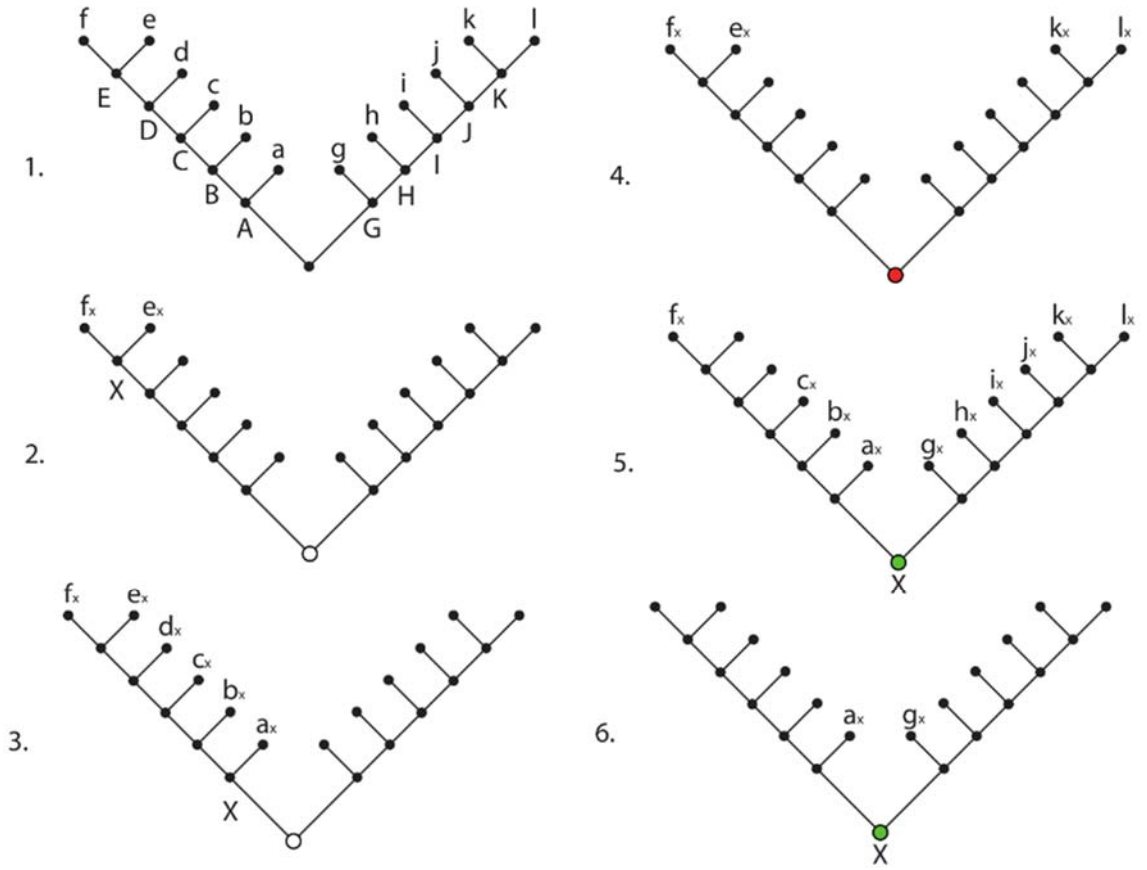


Figure 1.35. Test d'un état dérivé *X* sur un cladogramme (1) contenant un paralogue et illustration du nœud paralogue comme instance de test : un paralogue rejetant un état est noté en rouge, un paralogue acceptant un état en vert, et un paralogue non impliqué dans le test (les instances de *X* dans un seul nœud interne inclus dans le paralogue) en blanc. Les tests 2, 3, 4, 5 et 6 montrent le cladogramme et uniquement les UT qui sont également contenues dans l'état *X*. *X* est accepté partout et caractérise un clade sauf en (4) ou l'état est rejeté car apparaissant dans deux lignées indépendantes.

On peut maintenant faire une première généralisation du test des états de caractères. Pour un nœud paralogue, on regarde les clades qu'il contient (les clades formant tous des lignées indépendantes les unes par rapport aux autres). Pour qu'un nœud paralogue soit caractérisé par un état, cet état doit être présent dans l'ensemble des nœuds inclus dans le paralogue. Sur la figure 1.36, on note un nœud paralogue *Z* conduisant vers deux clades *X* et *Y*. Si *X* et *Y* possèdent l'état dérivé, *Z* le possède également. Si seulement *X* possède l'état dérivé, l'état caractérise *X* et uniquement lui. Enfin si *X* et *Y* ne sont pas caractérisés par l'état, alors *Z* non plus. Si des UT incluses dans *X* et *Y* possèdent l'état mais pas *X* et *Y* eux-mêmes, alors l'état est rejeté. Finalement, un nœud paralogue peut conduire vers plusieurs lignées, comme dans l'exemple 4 et 5. Dans ce cas, l'ensemble des lignées doit posséder l'état pour que le paralogue *Z* le possède. Dans l'exemple 4, l'état est donc rejeté puisqu'il

apparaît dans deux lignées. Dans l'exemple 5, l'état apparaît au niveau du paralogue, et est donc accepté. Le nœud paralogue joue donc le rôle d'instance de test d'un état : contrôler un paralogue permet de voir si un état est retrouvé dans des lignées divergentes, auquel cas l'hypothèse de clade est réfutée. Dans ce cas, l'état ne peut simplement pas caractériser de clade.

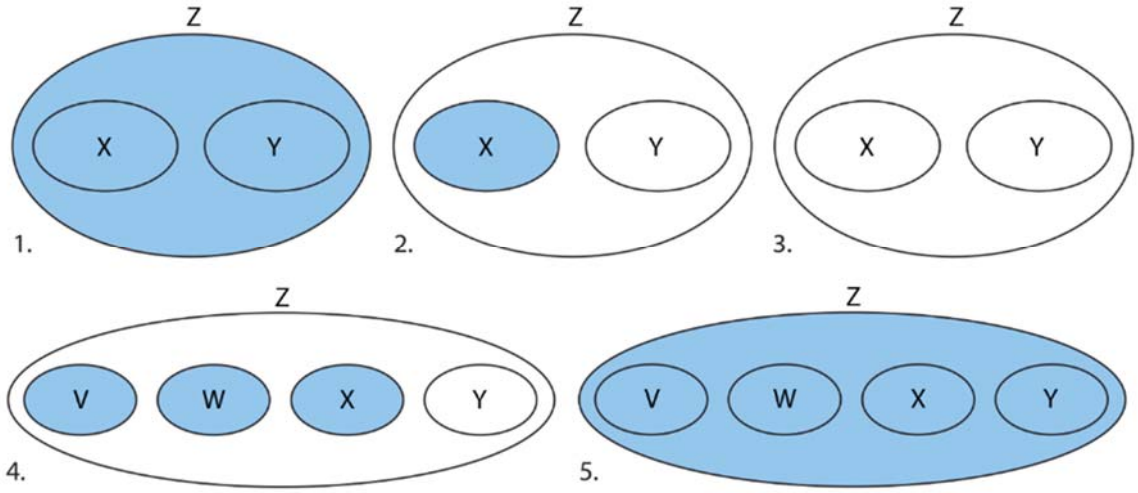


Figure 1.36. Schéma du test d'un état sur un nœud paralogue *Z* contenant les nœuds internes *V*, *W*, *X* et *Y*. Le clade caractérisé par *X* est indiqué en bleu. (1) Si *X* et *Y* sont caractérisés par l'état, alors *Z* aussi. Si seul *X* l'est, *Z* ne contient pas l'état dérivé. Si aucun nœud n'est caractérisé par l'état, *Z* non plus. L'état est rejeté si on trouve des instances dans différentes lignées d'un paralogue qui n'est pas lui-même caractérisé par l'état. (4) et (5) pour que le nœud paralogue soit caractérisé par l'état, toutes les lignées indépendantes doivent être caractérisées par l'état.

La figure 1.37 montre un dernier exemple permettant de comprendre le principe du test. L'exemple 1 montre un cas où toutes les instances de *X* sont contenues dans différentes lignées d'un paralogue. On vérifie le paralogue et on constate que chaque lignée (c'est-à-dire le clade le plus inclusif) possède l'état : le paralogue possède donc l'état. L'exemple 2 montre la même chose : on vérifie d'abord les lignées indépendantes, et on conclut que le paralogue possède *X*. On regarde ensuite les nœuds qui incluent le paralogue et on constate que l'état *X* se place plus profondément dans l'arbre et inclut *c*, *d*, *e* et *f*. Dans l'exemple 3, on constate que l'état *X* se trouve dans deux lignées divergentes, et que le paralogue ne peut pas être caractérisé par *X*. Le rejet se fait alors immédiatement, quels que soient les taxons (*c*, *d*, *e*, *f*) possédant *X* en dehors du paralogue. Le test se fait au niveau de l'instance d'indépendance du nœud paralogue, et confirme que l'état *X* ne peut caractériser un clade. S'il existe plusieurs paralogues dans le cladogramme, chaque paralogue est un niveau de

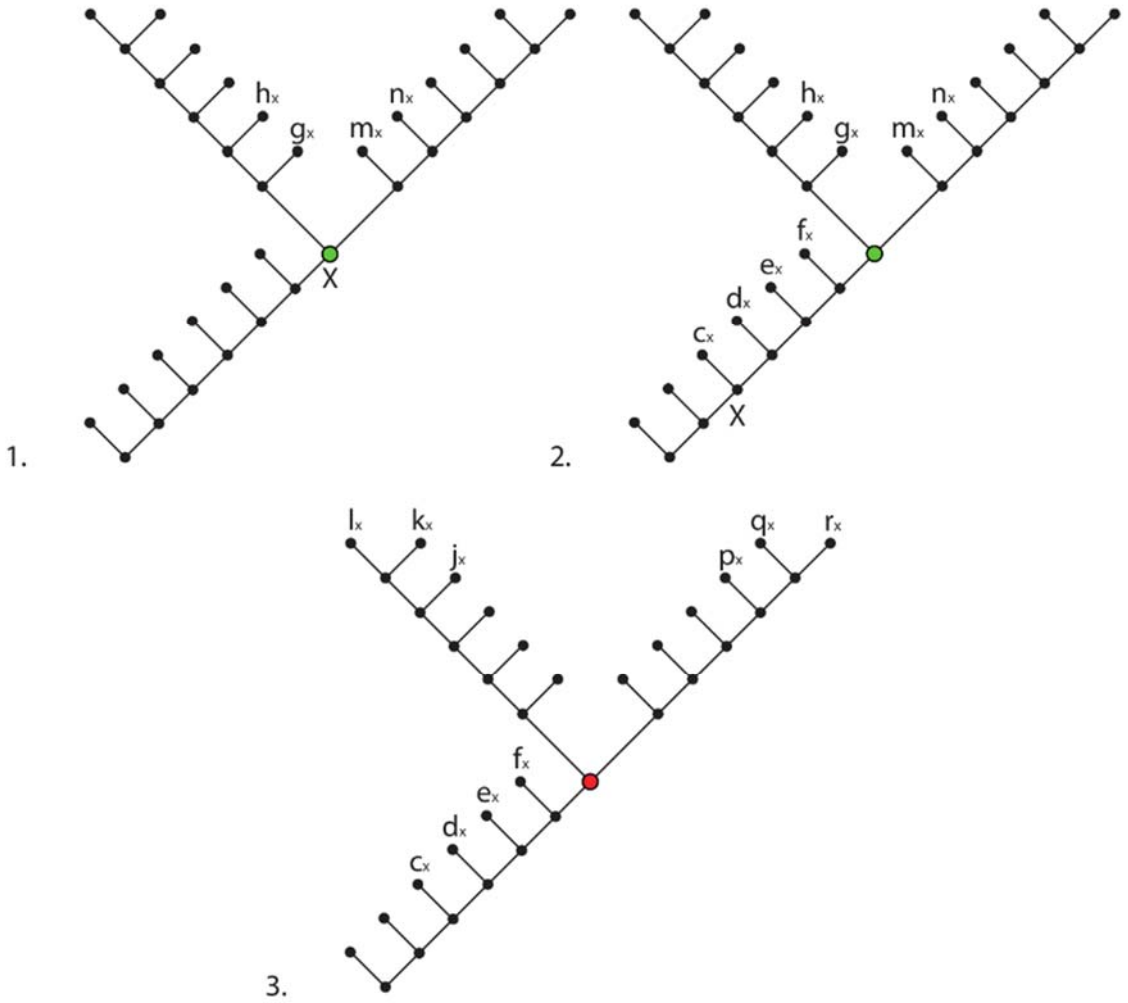


Figure 1.37. Test d'un état dérivé *X* sur un cladogramme contenant un paralogue lui-même inclus dans une orthologie. Un paralogue rejetant un état est noté en rouge, un paralogue acceptant un état en vert, et un paralogue non impliqué dans le test (les instances de *X* dans un seul nœud interne inclus dans le paralogue) en blanc. Seuls les UT inclus dans l'état dérivé sont affichés. (1) et (2) l'état *X* est accepté. (3) l'état est rejeté car apparaissant dans deux lignées indépendantes. L'état est rejeté car il ne passe pas le test du nœud paralogue rouge).

test. Si un paralogue est inclus dans un autre, on vérifie le paralogue le moins inclusif d'abord. On peut formaliser le test d'un état de caractère de la manière suivante. Soient un cladogramme *c* contenant *p* paralogues et un état dérivé *X*.

1. Si $p = 0$, on assigne *X* au clade le moins inclusif incluant toutes les UT contenues dans *X*. *X est accepté*.
2. Si $p \geq 1$, vérifier le nœud paralogue le moins inclusif non vérifié.
    a. Pour chaque nœud interne *n* non vérifié inclus dans le paralogue,

i. Assigner *X* au clade le moins inclusif incluant tous les UT contenus dans *X*. S'il n'y a pas d'instance de *X*, ne pas assigner *X*. n est noté vérifié.
    1. Si *X* n'est pas assigné, aller à l'étape 3.
    2. Si *X* est assigné à un seul clade, aller à l'étape 3.

b. Si *X* est assigné à chaque n du paralogue, assigner *X* au nœud paralogue.

c. Si le nœud paralogue est lui-même inclus dans une orthologie, assigner *X* au clade le moins inclusif de cette orthologie incluant tous les UT contenus dans *X*.

d. Si le nœud paralogue contient des instances de *X* dans au moins deux *n* différents mais que *X* n'est pas assigné au paralogue, *X est rejeté*.

3. Le paralogue est noté vérifié. S'il existe des paralogues non vérifiés, aller à l'étape 2.
4. Si *X* est assigné à un seul clade, *X* caractérise ce clade. *X est accepté*.

3.5.3. Datation

Le test des états de caractères est généralisable à d'autres types de hiérarchie cladistique. Pour assigner un classe d'une hiérarchie à un clade, il faut satisfaire plusieurs conditions : (i) la classe doit provenir d'une hiérarchie de type cladistique (fig. 1.4), (ii) la classe doit être un nœud interne qui n'est pas la racine et (iii) à chaque feuille de la hiérarchie doit correspondre une UT du cladogramme. Ainsi, diverses propriétés peuvent caractériser les clades.

Puisque l'analyse phylogénétique fonctionne dans un cadre historique, une des propriétés les plus intéressantes concerne le temps. Plus particulièrement, dater les nœuds permet d'apposer une chronologie au cladogramme d'UT et donne des informations sur l'âge des clades. Cependant, un problème important émerge dès lors qu'on essaye de dater un clade : le temps des taxons est représenté par l'échelle des temps géologiques, et cette échelle, comme son nom l'indique n'est pas une hiérarchie cladistique. Nous sommes en effet dans le cadre de la théorie cladistique. Cette dernière ne reconnaît que les entités monophylétiques. Or, dans sa définition, le Jurassique est une unité de temps paraphylétique, car le Crétacé descend du Jurassique. Le jurassique devrait alors inclure le Crétacé. Plus généralement, tout moment descend du moment le précédant. L'échelle n'est donc pas une structure mathématique pertinente. Cependant, Zaragüeta i Bagils *et al.*

(2004) ont montré qu'il était possible de convertir parfaitement et sans perte d'information l'échelle des temps géologique en une hiérarchie temporelle ou chaque unité de temps inclut l'unité qui descend d'elle. Les termes de l'échelle de temps géologique sont repris, en ajoutant un + après le nom. Par exemple, le Crétacé s'étend de 145 Ma à 66 Ma, mais est paraphylétique dans cette définition. Le *Crétacé*+ s'étend de 145 Ma à aujourd'hui, et inclut le *Barrémien*+ ou encore le *Paléogène*+. Selon cette définition, nous sommes encore dans le *Crétacé*+ puisque certains des évènements s'y étant déroulés, comme l'apparition des abeilles, se déroulent encore aujourd'hui (les abeilles existent encore). Il est cependant très important de noter que le temps n'est pas par essence partitionné ou hiérarchique. Le temps peut être envisagé de différentes manières, et l'échelle des temps géologiques est la représentation la plus pratique en stratigraphie. Le temps hiérarchique est par contre le seul à pouvoir être utilisé dans la théorie cladistique. La figure 1.38 montre une représentation de la hiérarchie du temps. On note que le temps par essence ne se diversifie pas. Une période cladistique (PC) est une période de temps monophylétique. Une PC commence mais ne s'arrête jamais. Chaque PC se différencie en exactement une autre PC. Le temps cladistique est une orthologie. La classification ainsi construite est composée de PC (par exemple, *Crétacé*+), les objets à classer étant des instances particulières de ces périodes (par exemple, *125 Ma*). Ce qu'on assigne à un clade, c'est une classe. Dater un cladogramme consiste donc à attribuer à des clades des PC, les instances de temps étant portées par les UT.

Si la hiérarchie temporelle est assimilable à n'importe quelle autre hiérarchie cladistique, comme l'homologie ou le cladogramme, elle possède néanmoins deux spécificités. Premièrement, on considère notre hiérarchie temporelle comme vraie *a priori*. Il n'y a donc pas de test des PC. Deuxièmement, on connaît le processus temporel : c'est un processus sans diversification, qui génère uniquement de l'orthologie. On sait également que le processus temporel est irréversible. On s'attend donc à trouver dans un clade daté au *Crétacé*+ uniquement des UT du *Crétacé*+. Il n'y a par définition pas de caractérisation des clades par le temps. On rappelle que l'assignation des états aux clades est une caractérisation plutôt qu'une définition car le fait de contenir toutes les UT de l'état dérivé est nécessaire et suffisant. Pour la datation, cela est nécessaire mais pas suffisant. L'assignation d'une PC à un clade est une définition puisque ce clade ne doit contenir que les UT de la PC et aucune UT qui n'appartient pas à cette PC. Si les clades sont datés de

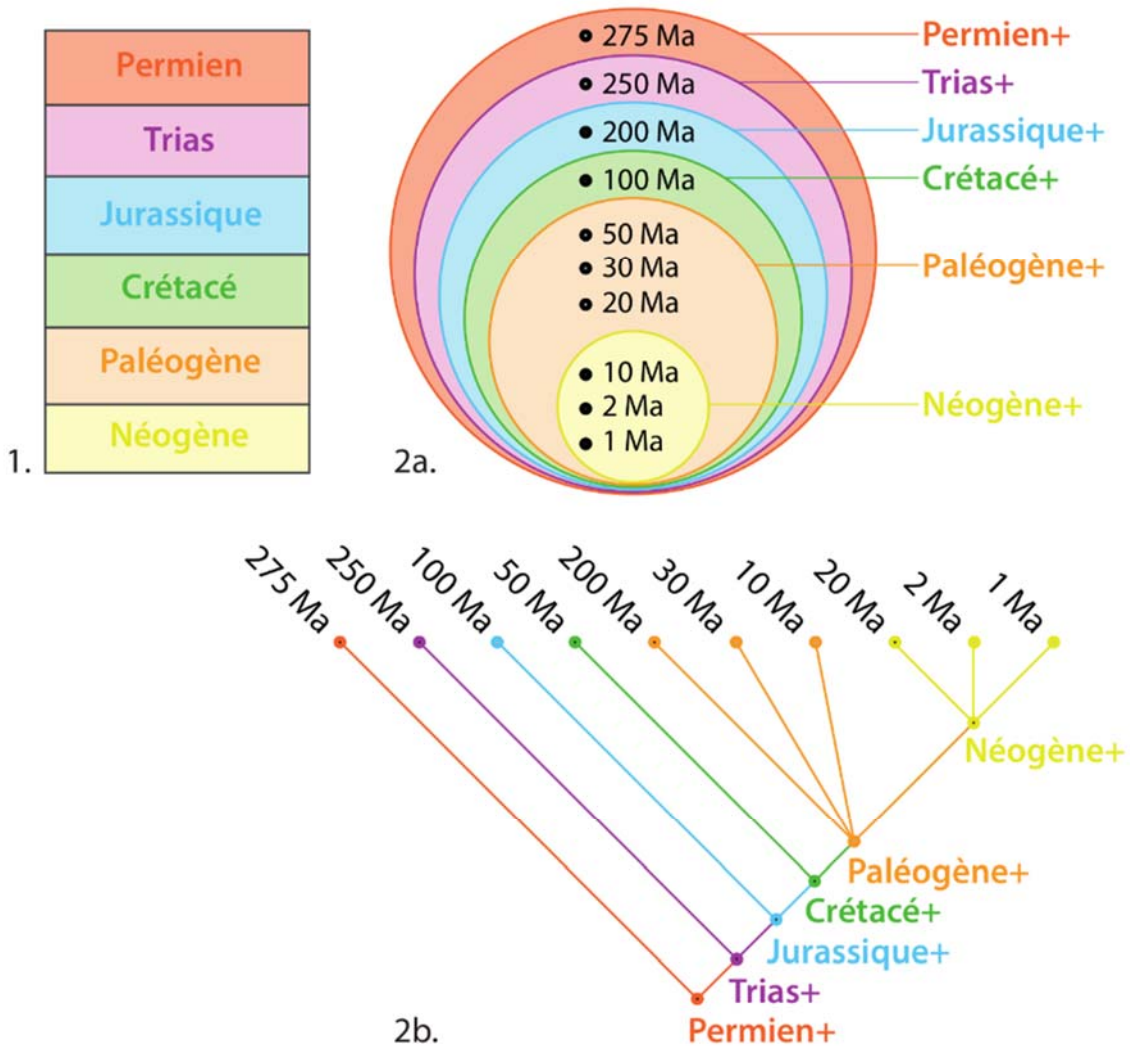


Figure 1.38. Représentations du temps. (1) Echelle des temps géologiques. Mathématiquement, une échelle est une partition. Une période commence lorsque l'autre s'arrête. (2) Hiérarchie temporelle représentée sous la forme d'un diagramme d'Euler. Une période commence à un moment donné mais ne s'arrête jamais. Une période se différencie en une autre. La hiérarchie temporelle est une classification : les classes sont des périodes, ou unités chronologiques (UC) et les feuilles sont des instances particulières de ces unités. (3) Même hiérarchie représentée sous forme d'arbre. La hiérarchie temporelle est une classification d'instants temporels (feuilles). Le temps hiérarchique est un temps cladistique. Les nœuds internes sont colorés selon la convention de l'échelle des temps géologique.

cette manière, il n'y a aucune réversion possible : le temps des cladogrammes s'écoule de manière irréversible.

La figure 1.39 montre trois exemples de datation, sur un cas parfaitement concordant avec la hiérarchie temporelle (a), un autre cas moins congruent (b) et un cas le moins concordant possible (c). Dans le premier cas, toutes les PC de la hiérarchie temporelle sont assignés à des clades concordant parfaitement. La chronologie des clades

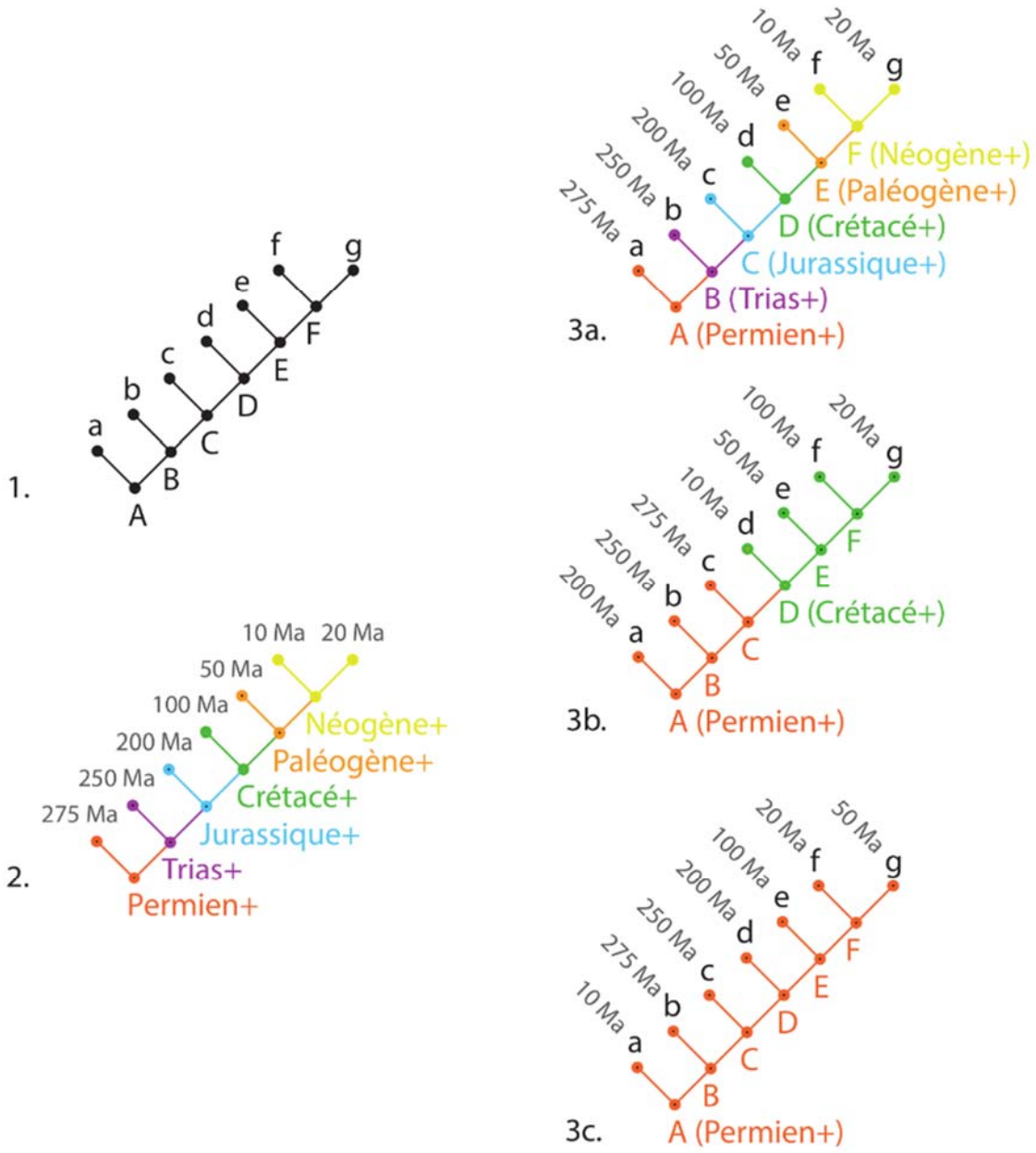


Figure 1.39. Datation d'un cladogramme-orthologie. (1) Cladogramme. (2) Hiérarchie temporelle composée de périodes (UC – nœuds internes) et de dates particulières (instances d'UC – feuilles). (3) Cladogrammes datés. On place une UC au clade le plus inclusif ne contenant que des instances de cette UC. Dans le cas (a), hiérarchie temporelle et cladogramme sont parfaitement congruents. Dans le cas (b), il y a de l'incongruence. Les nœuds *B*, *D* et *F* ne sont pas caractérisables. L'UC *Maastrichtien+* ne peut être placée sur un clade (un nœud interne). Dans le cas (c), l'incongruence est très forte : seul le nœud *A* est défini, par la période Permien+ bien que ne clade daté au Permien+ contienne des UT beaucoup plus récentes.

est parfaitement connue. Dans l'autre cas, le cladogramme ne suit pas exactement la hiérarchie temporelle. Puisque l'on sait que le temps est irréversible, il est interdit de placer une PC, par exemple *Trias+*, à un clade contenant des UT plus anciennes que *Trias+*. La méthode est donc de placer une PC au clade le plus inclusif ne contenant que des UT dont

l'âge est inclus dans cette PC. En d'autres termes, un clade est daté par le plus ancien taxon présent dans ce clade. Il en résulte qu'un clade daté au Trias+ ne peut contenir que des UT âgées de moins de 252 Ma, mais que des UT âgées de moins de 252 M peuvent également être situés en dehors de ce clade. C'est le cas dans l'exemple (b) qui montre de l'incongruence. Le clade *A* est défini par le *Permien*+ car il contient une UT *c* de *95 Ma*, même s'il contient également une UT *b* à *84 Ma*. La chronologie est incomplète (on ne connait que la chronologie entre *A* et *D*). L'exemple (c) montre un cas d'incongruence maximale ou seul le clade *A* peut être daté (*Permien*+). Dans ce cas de figure, on ne connaît rien de la chronologie du cladogramme.

La datation de cladogrammes incluant des nœuds paralogues dans leur structure fait émerger une propriété importante. On a dit que la hiérarchie temporelle est tenue pour vraie a priori; les nœuds paralogues ne servent pas de test. Nous allons réutiliser la procédure de LP (analyse de sous-arbres Libres de Paralogie) pour dater les nœuds sous-arbre par sous-arbre. On sait du fait du nœud paralogue que chaque sous-arbre est indépendant des autres. La figure 1.40 montre un exemple de cladogramme contenant un nœud paralogue *X* que l'on cherche à dater à l'aide de la hiérarchie temporelle de la figure 1.38. La LP produit trois sous-arbres, sur lesquels on applique la méthode précédemment décrite (on assigne une PC au clade le moins inclusif ne contenant pas d'UT plus ancienne). Le phénomène intéressant apparaît lorsque les sous-arbres sont recombinés : on s'aperçoit que des PC apparaissent dans deux lignées indépendantes. C'est le cas du *Paléogène*+ et du *Néogène*+, qui évidemment ne sont pas « réfutés » : le *Paléogène*+ du nœud *D* et du nœud *H* sont la même PC. Cela veut tout simplement dire que ce temps qui caractérise les clades est un temps biologique : à chaque évènement de diversification, chaque lignée suit une nouvelle hiérarchie temporelle indépendante de celle des autres (Zaragüeta i Bagils *et al.* 2004; Lelièvre *et al.* 2008). Ce phénomène de duplication temporelle génère artificiellement de nouvelles hiérarchies temporelles indépendantes les unes des autres. Les PC sont alors dupliquées, comme dans le cas de l'homonomie ou des UT sont dupliquées. On appelle ce phénomène *paralogie temporelle*. La paralogie temporelle, comme l'homonomie, fait perdre une quantité importante d'information. L'*indice de concordance hiérarchique* ou ICH (Hierarchical Fit index en anglais, HiFi) est un indice permettant de mesurer l'adéquation d'un cladogramme à une hiérarchie temporelle (Zaragüeta i Bagils 2002 : 249-

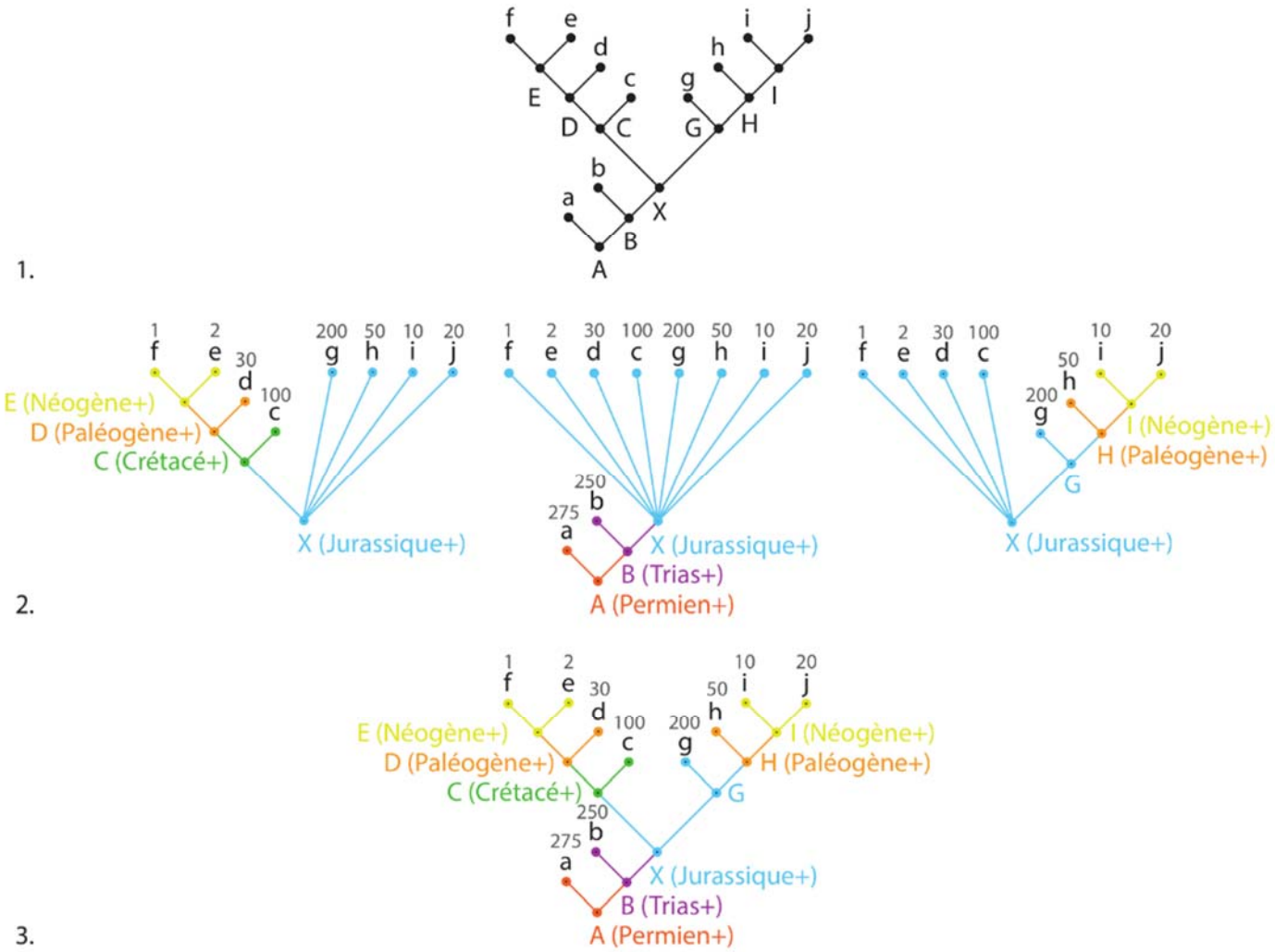


Figure 1.40. Datation d'un cladogramme contenant un paralogue *X*. (1) Cladogramme. (2) Décomposition du cladogramme en trois sous-arbres libres de paralogie et datation (en Ma). On place une UC au clade le plus inclusif ne contenant que des instances de cette UC. (3) La réunion des sous-arbres met en évidence un phénomène de paralogie temporelle : à partir du nœud paralogue, la hiérarchie temporelle se scinde en deux chronologies indépendantes l'une de l'autre.

299). L'ICH est un indice de rétention qui fonctionne uniquement sur les triplets déduits des nœuds orthologues, les nœuds paralogues étant considérés comme temporellement non-informatifs. On n'utilise pas non plus les triplets déduits des nœuds contenant exactement deux feuilles qui ne sont pas temporellement informatifs. Un triplet peut être temporellement congruent ou non, c'est-à-dire en accord avec la hiérarchie temporelle ou non. Le triplet *a*, (*b*, *c*) auquel on ajoute l'information temporelle donne en Ma *275*, (*250*, *100*). *a*, (*b*, *c*) est temporellement congruent puisque le premier nœud interne est plus ancien (*275*) que le nœud interne suivant (*250*). Le triplet c, (a, b) donne le triplet temporel faux

100, (*275*, *250*). *c*, (*a*, *b*) est donc temporellement incongruent avec la hiérarchie temporelle car on ne peut dater que le nœud racine, à *275*. L'ICH est donné par la formule suivante :

$$\mathrm{ICH} = \frac{\sum_{i=1}^{O} X_i}{\sum_{i=1}^{O} N_i}$$

Où *O* correspond au nombre de nœuds orthologues contenant plus de deux feuilles du cladogramme, *X* correspond à la somme des poids de tous les triplets déductibles de *O* temporellement congruents, et *N* correspond à la somme des poids de tous les triplets déductibles de *O*. Il manque cependant une information supplémentaire : on sait que les instances de paralogie temporelle sont non-informatives temporellement. Si on veut que l'ICH donne une mesure de l'information temporelle d'un cladogramme, on doit rajouter l'information de la proportion de paralogues dans l'arbre, ceux-ci devant faire diminuer l'indice. On obtient l'ICH rééchelonné $\mathrm{ICH_r}$:

$$\mathrm{ICH}_r = \mathrm{ICH} * \frac{\sum_{\mathrm{i=1}}^{\mathrm{O}} N_i}{\sum_{\mathrm{i=1}}^{\mathrm{O}} T_i}$$

Ou *T* vaut pour le poids total de triplets de l'arbre. On obtient alors :

$$\mathrm{ICH}_r = \frac{\sum_{i=1}^{O} X_i}{\sum_{i=1}^{O} T_i}$$

L'$\mathrm{ICH_r}$ vaut 1 lorsque le cladogramme est une orthologie et que tous les triplets sont temporellement congruents (fig. 1.40-3a). L'$\mathrm{ICH_r}$ diminue avec le nombre de paralogues. Il diminue également avec le nombre d'incongruences temporelles. Un $\mathrm{ICH_r}$ de 0 peut être atteint si tous les triplets sont incongruents temporellement (fig. 1.40-3c) et/ou si le cladogramme ne contient aucun nœud orthologue à plus de deux UT.

4. Consensus trees: interpretation, informativity and efficiency

Cette partie correspond au mansucrit d'un article qui sera soumis par la suite.

Auteurs: Valentin Rineau, René Zaragüeta i Bagils, Jérémie Bardin and Estelle Bourdon.

4.1. Abstract

Consensus trees aim to represent the phylogenetic information common to a set of optimal trees. In this paper we discuss the theoretical and interpretative aspects of the consensus tree building problem with a view to application for phylogenetic inference. We discuss two conflicting aims of consensus trees in the literature: the hard consensus aim (the consensus tree is a graphical summary of the information conveyed by a set of optimal trees) and the soft consensus aim (the consensus tree represents only the information shared by all the optimal trees). In the hard aim, all optimal trees are considered as potentially true, the consensus being a structure that does not represent any of the optimal trees as it is (without preliminary deduction procedure). It is a methodological tool that cannot be used for direct phylogenetic interpretations. In contrast, in the soft aim, there is one true tree within the whole set of optimal trees, all other optimal trees being false. In this interpretation, the consensus is a non-ambiguous representation of the information common to all optimal trees: it gathers the parts of optimal trees that are not contradicted. We show that the Consensus Efficiency (CE) is a metric adapted to the hard aim of consensus trees. We propose a new metric called Consensus Retention Index (CRI) for comparing consensus methods in agreement with the soft aim of consensus trees. We also show that the Reduced Cladistic Consensus (RCC) profile is the flawless solution to the soft aim of consensus tress. The RCC is the only consensus method that has always a maximal Consensus Retention Index. Therefore, the RCC must be regarded as the best consensus method for phylogenetic inference. An R script for constructing RCC based on Wilkinson's algorithm is proposed in this paper.

Keywords: Adams Consensus, compatibility, Consensus Efficiency, Consensus Retention Index, combinability, hierarchical tree, Majority-rule Consensus, phylogenetic information, Reduced Cladistic Consensus, Strict Consensus.

4.2. Introduction

The difficulty in representing the phylogenetic information that is common to a set of equally optimal trees is a classical problem in phylogenetics. Phylogenies are now widely used for testing evolutionary hypotheses since the dawn of phylogenetic comparative methods (Felsenstein, 1985). The concept of consensus tree applied to phylogenies was introduced by Adams (1972). Since then, a great variety of consensus methods have been proposed and discussed (see Bryant 2003 for a classification of the main methods). Despite the large amount of literature dealing with mathematical problems related to consensus tree building, the rationale for and interpretative aspect of consensus trees in evolutionary sciences has not been discussed enough. Bryant (2003) stated that: '[…] *consensus methods are a useful tool for phylogenetic inference, but only when used in conjunction with a model or paradigm*'. Our paradigm is phylogenetic analysis, which produces several equally optimal topologies. In this paper, we clarify the relations between different kinds of phylogenetic information. We discuss the coexistence of two incompatible interpretations of consensus methods, which has engendered confusion. Consensus trees are interpretable for evolutionary studies under the soft interpretation. In contrast, consensus trees are technical tools that gather mutually incompatible scenarios under the hard interpretation. A consensus tree must fulfill three properties under the soft interpretation: sensitivity, unambiguity (Wilkinson, 1994a) and strict retention of information (new notion). We show that the Consensus Efficiency Index (Wilkinson and Thorley, 2001) is inappropriate, and we propose a new metric called Consensus Retention Index. Finally, we examine the different consensus methods in light of these properties and metrics in order to advise biologists on the best soft-aimed consensus to use.

4.3. Phylogenetic Information

The concept of phylogenetic information has a dual nature. The information conveyed by a given tree is called positive phylogenetic information, while the information

not present in this tree is called negative phylogenetic information. According to Wilkinson (1994), two kinds of positive phylogenetic information are used in consensus methods: *n*-taxon statements (*n*ts) and nestings. *N*ts are characterized by their cardinality *n* (number of leaves). They consist of trees containing *n* leaves and a unique internal node (Cao 2008; Cao *et al*. 2009). For example, *AB(CD)* is a *n*ts of cardinality four. There are two specific kinds of *n*ts: triads and components. Triads (3-taxon statements) are the smallest *n*ts (i.e. *n* = 3) that convey phylogenetic information. They correspond to the statement that two taxa are more closely related to each other than either is to a third (e.g. *A(BC)*). Components are the largest *n*ts (i.e. *n* is maximal), and contain the exact same leaves as the optimal trees. For example, the optimal tree *A(B(C(DE)))* contains the component *ABC(DE)*. The strict consensus is based on components (Sokal et Rohlf 1981).

Adams (1986) described a second kind of positive phylogenetic information called nestings. In contrast to *n*ts, nestings represent a possibility of the existence of phylogenetic relationships (Wilkinson, 1994a): for example, the nesting *CD*{*AB*} implies the *n*ts *C*(*AB*) AND/OR the *n*ts *D*(*AB*). Because a nesting implies one hierarchy and/or another, it cannot be represented as a single hierarchy. Nestings are always less informative than *n*ts because the positive information they convey is ambiguous and uncertain. The main method based on nestings is the Adams II consensus (Adams, 1972).

Disqualifiers (Wilkinson 1994a) represent negative phylogenetic information, i.e. phylogenetic relationships absent from the set of optimal trees. The disqualifier not*C*(*AB*) states that taxon *A* and taxon *B* are not more closely related to each other than they are to taxon *C*. Disqualifiers do not represent independent phylogenetic information because there is an *orthogonal relationship* (Barbut, 1968) between them and *n*ts. The disqualifiers of a given *n*ts correspond to its orthogonal *n*ts. Consider the triad *A*(*BC*): the triads *B*(*AC*) and *C*(*AB*) are orthogonal to it. The triad *A*(*BC*) logically implies not*B*(*AC*) and not*C*(*AB*). For example, disqualifiers are used in the disqualifier faithful subtrees (Wilkinson 1994a; Wilkinson et Benton 1995). This method aims to restrict the number of trees permitted by a consensus by showing complementary disqualifiers and to summarize the original set of optimal trees more accurately.

4.4. Aim of consensus trees

The concept of consensus tree was first introduced by Adams (1972). He defined it as a '*method for computing a tree which represents only that information shared by all the rival trees*' (Adams 1972, p. 390; fig. 1.41). The information common to a set of optimal trees is combined into one single tree that must ideally represent only this information without gain or loss. Thorley, Wilkinson and Charleston (1998) and Wilkinson and Thorley (2001, p. 610) modified Adams' interpretation of consensus trees by asking '*how well do individual consensus trees represent the set of source trees for which they stand?*'. The consensus efficiency was defined as the '*relation between the trees permitted by the consensus tree and the optimal trees*' (Wilkinson and Thorley 2001, p. 610; fig. 1.41). For these authors, '*Consensus trees might deviate from the ideal in two ways. First, they might permit trees that are not source trees, and second they might fail to permit some of the source trees*' (Wilkinson and Thorley 2001, p. 610). We explain below how these two interpretations impact on the treatment of phylogenetic information.

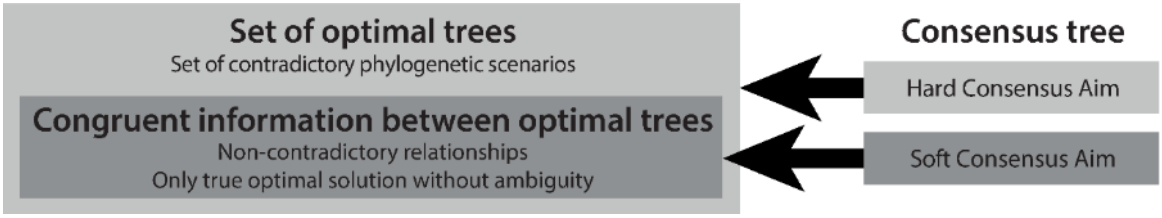


Figure 1.41. Diverging objectives of consensus tree methods. A consensus tree in agreement with the soft consensus aim must represent only that information shared by all the optimal trees. Its efficiency is calculated using the Consensus Retention Index (CRI). A consensus tree following the hard aim must summarize graphically all the information conveyed by a set of optimal trees. Its efficiency is calculated using the Consensus Efficiency (CE) index.

4.4.1. Treatment of phylogenetic information

Maddison (1989) defined two ways of interpreting polytomous trees: soft polytomies and hard polytomies. On a polytomous tree with three leaves *ABC*, each interpretation considers that there is no positive information. However, the two interpretations differ from each other in their use of disqualifiers. The soft interpretation implies that *ABC* contains no specific information, positive or negative, and does not forbid anything. The hard interpretation implies that ABC forbids every resolution, and contains

three disqualifiers not*A(BC)*, not*B(AC)* and not*C(AB)*. We can link these differences to the idea of orthogonal relationships. Under the soft interpretation, disqualifiers are deduced from the positive statements of a tree. Positive information logically implies negative one and no disqualifier exists if it is not linked to an orthogonal positive information. The hard interpretation considers that all possible disqualifiers are true by default. There are positive statements implied by the tree, and every other possible statement is a disqualifier that is contained in the tree. Under this interpretation, disqualifiers are pre-existing information and are not deduced from positive information. In figure 1.42, the tree *AB(CD)* is interpreted following soft and hard interpretations. The soft interpretation implies two minimal positive statements, which entail themselves four minimal orthogonal disqualifiers. The hard interpretation implies the two minimal positive statements plus every minimal disqualifier. Under this interpretation, six more disqualifiers are implied because they are considered as independent information. The objective which implies that negative information is derived from positive statements conveyed by a tree is defined here as the '*soft consensus aim'* for consensus methods. The '*hard consensus aim'* for consensus methods implies that positive and negative statements are independent. Soft and hard consensus aims extend Maddison's concepts of soft and hard polytomies to the interpretation of consensus trees.

Below we illustrate how sets of optimal trees are treated using soft and hard aims in the consensus tree problem. Our examples use minimal units of information, which consist of triads and disqualifiers (with n=3). Despite original algorithms are not based on triads, we use these for the sake of clarity. Figure 1.43 shows a set of two optimal trees *A(B(CD))* and *B(A(CD*)) and the information retained under the two interpretations of consensus trees: a consensus permits all the optimal trees and only these (fig. 1.43 – hard consensus aim), and a consensus represents only that information shared by all the optimal trees (fig. 1.43 – soft consensus aim). The first step is to decompose the two optimal trees into minimal statements in order to identify the points of agreement. A tree of four leaves yields 12 statements that are composed of four triads and eight minimal disqualifiers. Each tree agrees with the other on 2/4 of triads and 6/8 of minimal disqualifiers. Under the hard interpretation, a consensus tree represents the positive information *B(CD)* and *A(CD)*, which is combined into *AB(CD)*. Two optimal trees *A(B(CD))* and *B(A(CD*)) plus a third artefactual tree *(AB)(CD)* are deduced from *AB(CD)*. The hard consensus retains both

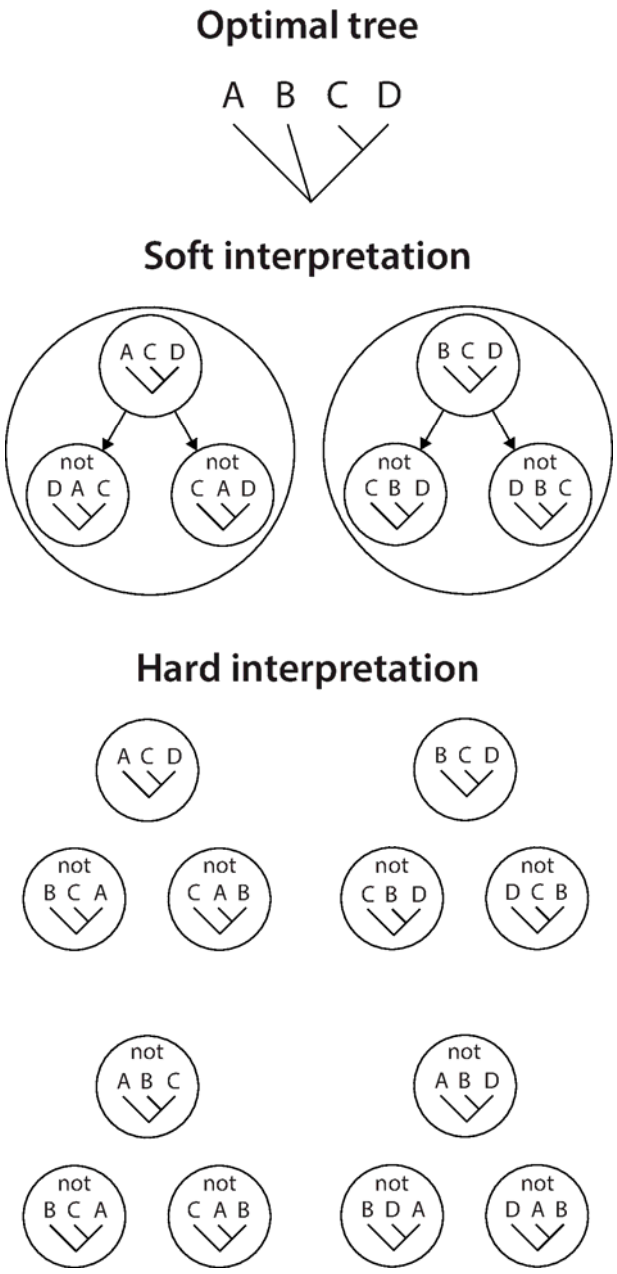


Figure 1.42. Diverging interpretations of the informative content of a hierarchical tree and the relations between phylogenetic statements. Soft interpretation of polytomies – the optimal tree *AB(CD)* shows a soft polytomy that expresses uncertain resolution. Under the soft interpretation, phylogenetic information consists of the positive statements *A(CD)* and *B(CD)* and their two orthogonal disqualifiers not*D(AC)* and not*C(AD)*, and not*C(BD)* and not*D(BC)*, respectively. Hard interpretation of polytomies – the optimal tree *AB(CD)* shows a hard polytomy that expresses a multiple speciation event. Its deduced statements include all disqualifiers (e.g. not*A(BC)*) and thus forbid any possible resolution of the polytomy. Because these disqualifiers are not linked to any positive information, the orthogonality rule is broken.

positive and negative information by breaking the orthogonality between triads and disqualifiers (fig. 1.43 – step of decomposition). The disqualifier not*C(AB)* is present in both trees: in the tree 1, not*C(AB)* is deduced from *A(BC)* whereas it is deduced from the

opposite triad *B(AC)* in the tree 2. However, in order to keep not*C(AB)* and not*D(AB)* that can be merged into not*CD(AB)* (fig. 1.43 – step of consensus), it is necessary to restrict the deductible trees from three to the exact two optimal trees *A(B(CD))* and *B(A(CD*)). *(AB)(CD)* is forbidden by the disqualifier not*CD(AB)*. In the soft consensus aim, the only information retained consists of the two triads *B(CD)* and *A(CD)* and their two respective disqualifiers (fig. 1.43 – step of decomposition: three orthogonal statements are surrounded by circle). However, only 4 disqualifiers (versus 8 for the hard aim) are retained as complementary statements logically deductible from triads. Under the soft interpretation, only positive statements are considered, and the retention of any supplementary disqualifiers is meaningless. The two triads *B(CD)* and *A(CD)* are combinable into the consensus *AB(CD)*, and the information deduced from it, *B(CD)* and *A(CD)*, corresponds exactly to the information shared by the two optimal trees.

4.4.2. Hard aim of consensus trees

According to Thorley *et al.* (1998) and Wilkinson and Thorley (2001), consensus trees are summary structures that aim to represent the entire set of optimal trees (fig. 1.41 – hard interpretation). This interpretation has some consequences from an epistemic point of view. In the hard interpretation, all the information conveyed by the optimal trees is considered as potentially true. The hard consensus needs to include all the information of the optimal trees, plus relationships that are not common to all optimal trees (i.e. deduced from disqualifiers; fig. 1.41). Under the hard interpretation, an ideal consensus tree retains all optimal trees regardless of their partial incongruences. Read like a classical cladogram, it is a false tree because it does not correspond to any of the summarized trees. The hard consensus tree generally cannot represent positive phylogenetic relationships non-ambiguously. Using the example in figure 1.43, two trees are deductible from the consensus *AB(CD)* and not*CD(AB)*: *A(B(CD))* and *B(A(CD))*. They show partially incompatible information because *A*(*BC*) comes from tree 1 and *B*(*AC*) comes from tree 2. A hard consensus tree is a summary structure that cannot be interpreted in a phylogenetic way, because it can show partially incompatible solutions.

Wilkinson and Thorley (2001) tried to solve the issue of the consensus efficiency under the hard interpretation by retaining two kinds of phylogenetic information: positive and negative. Maximizing the efficiency of the hard consensus requires the retention of

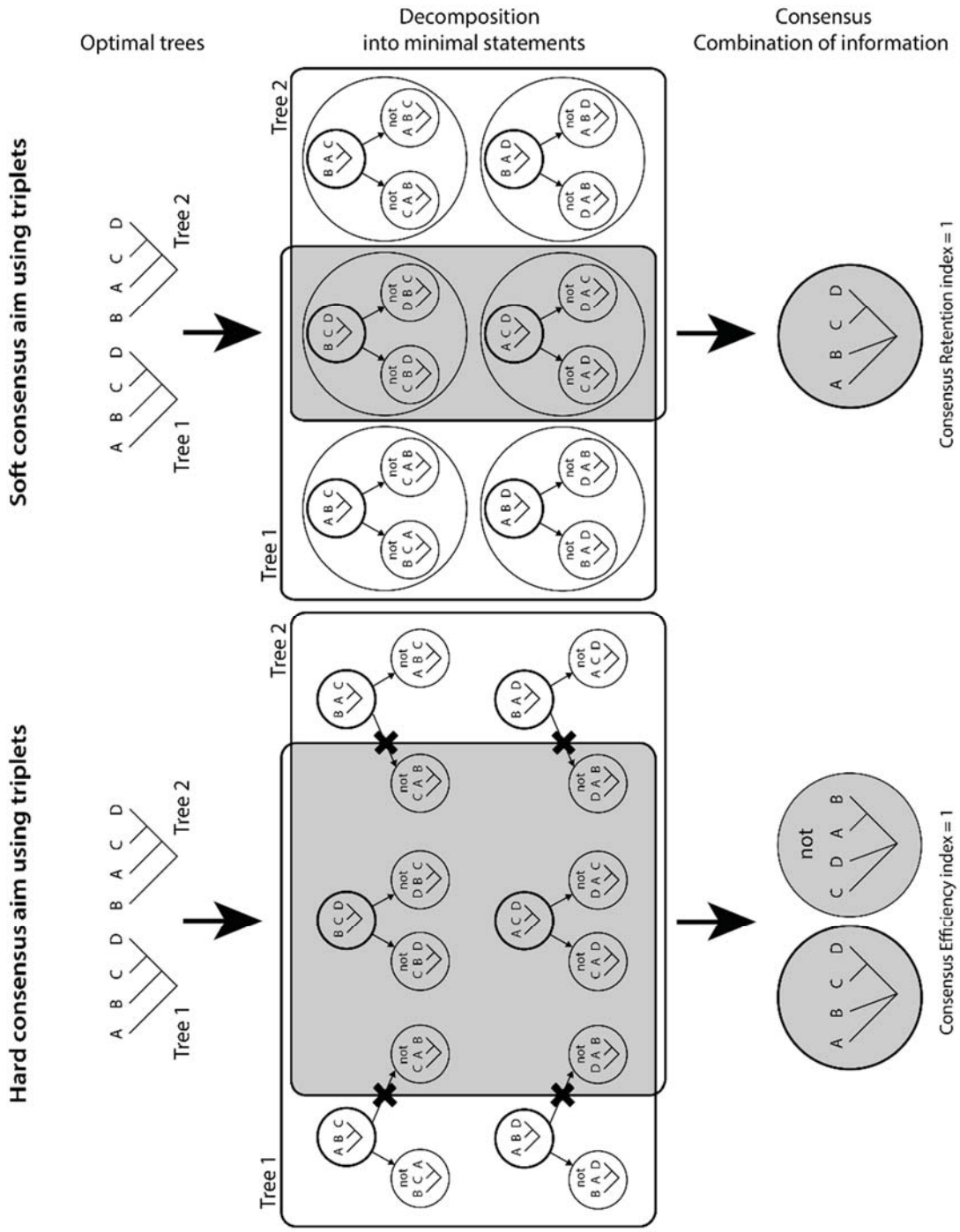


Figure 1.43. Two optimal cladograms (Tree 1 and Tree 2) and their differential treatment by applying the hard consensus aim (left – the consensus must represent the set of optimal trees) or the soft consensus aim (right – the consensus represents the congruent information between all optimal trees) on triads and their disqualifiers. The set of optimal trees is composed of the two trees *A(B(CD))* and *B(A(CD)*. Optimal trees are decomposed into triads and orthogonal disqualifiers. Hard aim – during the decomposition step, the orthogonality rule is broken (black crosses) in order to maximize consensus efficiency. 2/4 of triads and 6/8 of disqualifiers common to the set of optimal trees are retained (grey area). The recombination of these statements produces *AB(CD)* and not*CD(AB)* in order to forbid the tree *(AB)(CD)* that is not one of the two optimal trees. Trees permitted by the consensus are identical to the optimal trees, which leads to a maximal Consensus Efficiency index. Soft aim – during the decomposition step, the orthogonality is conserved (circles surround each mother triad and its two child disqualifiers). 2/4 of triads and 4/8 of disqualifiers common to the set of optimal trees are retained (grey area). The recombination of these statements produces *AB(CD)*, which is the only unambiguous true information. The Consensus Retention index is maximal since all the common information is displayed by the consensus.

both positive and negative information (figure 1.43). Wilkinson and Thorley considered the negative phylogenetic information as important as the positive one: '*consensus trees might deviate from the ideal in two ways. First, they might permit trees that are not optimal trees, and second they might fail to permit some of the optimal trees*' (Wilkinson and Thorley 2001, p. 610). In this approach, it is necessary to forbid the relationships compatible with the strict consensus that are not present among the optimal trees, using for example a disqualifier faithful tree.

An efficient summary tree is computed by retaining both positive and negative phylogenetic information (fig. 1.43). The method breaks the orthogonal relationship between *n*ts and their disqualifiers because it considers them as independent. In this example, the two disqualifiers not*C*(*AB*) and not*D*(*AB*) are retained because they are present in both trees. However, disqualifiers are deduced from different triads, which arise from different optimal trees. Disqualifiers are thus not really the same as they do not have the same origin.

The origin of disqualifiers is an important information that is systematically lost in the hard consensus trees. Figure 1.44 shows an example of limitation of consensus trees caused by this loss of information. As in figure 1.43, all triads plus all disqualifier triads common to the set of optimal trees *A(B(CD))* and *B(C(AD))* are retained (fig. 1.44 – optimal trees). In this example, only 1/4 of the triads for each tree is common to the set, and 5/8 of the disqualifiers (fig. 1.44 - decomposition step). The clustering of these statements produces *B(CD)* and not*D(C(AB))*. Three trees can be deduced from these statements including the two optimal trees plus an artefactual tree *B(A(CD))* that is not part of the original set of optimal trees (fig. 1.44 – consensus step). Why are artifactual trees permitted when retaining strictly all positive and negative phylogenetic information? This is because another kind of positive information is needed in this case but never defined. *A*(*BC*) AND *D*(*BC*) represent the *n*ts *AD*(*BC*); *A*(*BC*) AND/OR *D*(*BC*) represent the nesting *AD*{*BC*}; *A*(*BC*) OR *D*(*BC*) means that the triads are mutually exclusive, i.e., this kind of information tell us that only one triad is true. In our example, one must represent the following relation between triads: [*A*(*BC*) and *A*(*BD*) and *A*(*CD*)] OR [*B*(*AC*) and *B*(*AD*) and *C*(*AD*)] (fig. 1.44 - decomposition step: *B(CD)* is present in both trees, but the other triads of tree 1 are true only if the triads of tree 2 are false). Looking at the optimal trees, we need to convey the information that if the triads *A*(*BC*), *A*(*BD*) and *A*(*CD*) are true, the triads *B*(*AC*), *B*(*AD*)

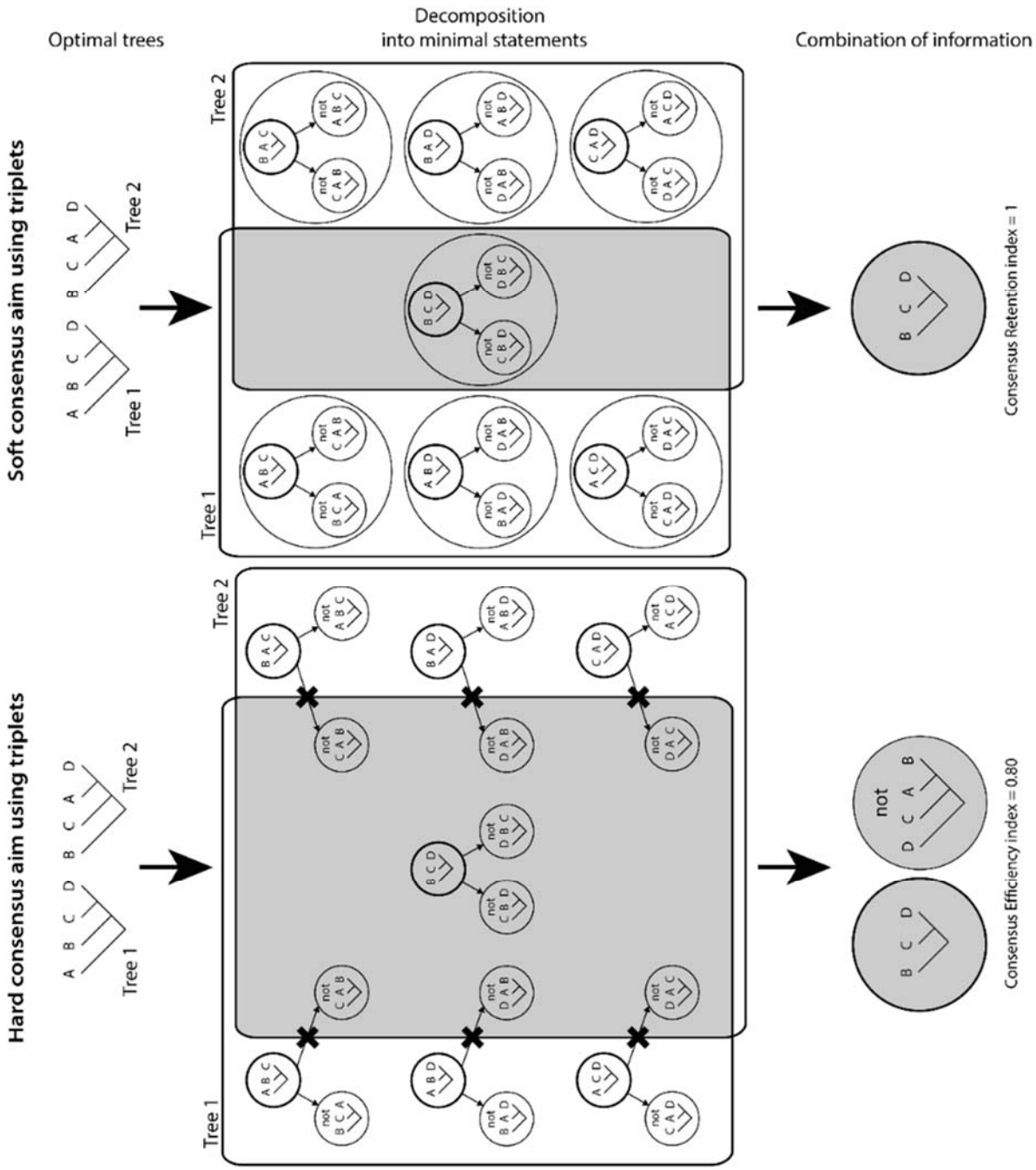


Figure 1.44. Two optimal cladograms *A(B(CD))* and *B(C(AD))* and their differential treatment by applying the hard or the soft consensus aims on triads and their disqualifiers. Hard aim – during the decomposition step, the orthogonality rule is broken (black crosses) in order to maximize consensus efficiency. 1/4 of triads and 5/8 of disqualifiers common to the set of optimal trees are retained (grey area). The recombination of these statements produces *B(CD)* and not*D(C(AB))*. The Consensus Efficiency index cannot be maximal because there is another kind of information that cannot be displayed on a hierarchical tree: the mutually exclusive statements [*A*(*BC*) and *A*(*BD*) and *A*(*CD*)] OR [*B*(*AC*) and *B*(*AD*) and *C*(*AD*)]. Breaking the orthogonality rule implies loss of information on the origin of disqualifiers. Thus, several triads are present in the set of optimal tree but can never be gathered into the same optimal tree. Soft aim – during the decomposition step, the orthogonality is conserved. 1/4 of triads and 2/8 of disqualifiers common to the set of optimal trees are retained (grey area). The recombination of these statements produces *B(CD)*, which is the only unambiguous true information. The Consensus Retention index is maximal since all the common information is displayed by the consensus.

and *C*(*AD*) cannot be true, and reciprocally. Artefactual trees can occur when statements that are not present in the same optimal tree are retained in the consensus tree. This methodological limitation forbids the existence of an ideal consensus method that would have a maximum efficiency in every case. Moreover, the nesting *A(BC)* AND/OR *D(BC)* or the mutually exclusive statement *A(BC)* OR *D(BC)* cannot be directly interpreted as a hierarchical tree or as a unique phylogeny. They must be interpreted beforehand as possibilities of trees. Consequently, consensus methods based on the hard interpretation and the use of nestings are of limited utility for evolutionary studies. The hard consensus methods that use phylogenetic information regardless of its origins (positive or negative) yield consensus tress that have no direct meaning in terms of evolution. Under the hard interpretation, the consensus tree is only a data compression format for synthesizing a set of trees.

4.4.3. Soft aim of consensus trees

The consensus tree has a different epistemic interpretation following Adams' rationale. The goal of a consensus tree under this interpretation is to convey the phylogenetic information common to a set of optimal trees. Each optimal tree is partially incongruent with all the other trees of the set. Thus, they cannot all be true at the same time. At most only one of them may be potentially completely true. Consequently, each optimal tree is regarded as partially false because it contradicts other optimal trees. In this context, the consensus tree is the only non-ambiguous true tree because it gathers parts of the optimal trees that are never contradicted (fig. 1.41), and each node of the consensus refers to a clade in the optimal trees. Under the soft interpretation, the consensus tree must be regarded as a non-ambiguous representation of the information common to all optimal trees. Because one cannot choose a particular optimal tree, one must retain the non-ambiguous common structure only, and nothing else. Moreover, because it is a non-ambiguous representation of the information common to all optimal trees, the consensus tree itself becomes an optimal subtree (this is not the case under the hard interpretation).

The consensus tree aims to represent phylogenetic relationships conveyed by all optimal trees, the negative phylogenetic information being secondary. Disqualifiers do not represent independent information, since they are deductible from their positive orthogonal *n*ts. In this context, the retention of disqualifiers along with the consensus tree becomes

meaningless, as shown in figures 1.43 and 1.44. The treatment of negative phylogenetic information highlights the contradictory aim of consensus trees under hard and soft interpretations (fig. 1.41). In the soft interpretation, orthogonal relationships between *n*ts and disqualifiers are always preserved (figures 1.43 and 1.44) whereas *n*ts and disqualifiers are regarded as independent under the hard interpretation.

In summary, the hard interpretation considers all optimal trees as potentially true. Because of this, all optimal trees must be retained in the consensus tree. In this context, the hard consensus must be understood as a methodological tool, a compressed structure that can be interpreted as a set of trees, but not as a single phylogeny. In contrast, the soft consensus considers that there is at most one true tree within the whole set of optimal trees, the other optimal trees being false because they partially contradict the true tree. The consensus becomes the only tree that can be chosen, because it shows the phylogenetic relationships that are never contradicted. Because the consensus is the only ascertained true tree produced by the analysis, it can be read as a phylogeny and used in evolutionary studies. We discuss below the efficiency of consensus methods under the soft interpretation.

4.5. Informativeness of consensus trees

Wilkinson and Thorley (2001) described a measure of consensus efficiency that allows us to say how well a given consensus tree represents the set of optimal trees for which it stands: the consensus efficiency (CE) index. This index calculates the proportion of binary trees permitted by the consensus tree compared to the number of binary optimal trees. It ranges between values of zero (minimum efficiency) and one (maximum efficiency). The CE is related to the measure of Phylogenetic Information Content (PIC) developed by Thorley *et al.* (1998). A maximally efficient consensus tree permits only the optimal trees that it represents (fig. 1.43). The CE is reduced if the consensus permits additional trees that are not optimal trees, or fails to permit some of the optimal trees. It is obvious that the CE index has been developed under the hard interpretation of consensus trees. Therefore, the CE index must be interpreted as a measure of summary efficiency, and cannot be used for comparing consensus tree methods under the soft interpretation.

Since a consensus is intended to represent the congruent information between all optimal trees instead of summarizing a set of optimal trees (fig. 1.41), we propose a measure in agreement with the soft interpretation of consensus trees. Phylogenetic information can be decomposed without loss into a set of triads, which are the smallest units of phylogenetic information. Here we use a new metric based on triads, which is a variant of the Inter-Tree Retention Index (ITRI; Grand, Corvez, Duque Velez, et Laurin, 2013). The ITRI is equivalent to a retention index (RI) calculated as a proportion of triads. Triads imply the use of fractional weighting for correcting their logical dependency (Mickevich et Platnick 1989; Nelson et Ladiges 1992). The ITRI aims to assess the degree of congruence between two trees (T_1 and T_2): ITRI{T_1,T_2} is the proportion of the fractional weights (FW) of the T_1 triads that T_1 shares with T_2 (Grand *et al.*, 2013).

$$ITRI\{T1, T2\} = \frac{\sum(FW\ of\ triads\ that\ T1\ shares\ with\ T2)}{\sum(FW\ of\ T1\ triads)}$$

Our new metric called Consensus Retention Index (CRI) is derived from the ITRI and aims to assess the efficiency of consensus trees under the soft interpretation. The CRI uses the same formula as the ITRI, where T_2 represents the consensus tree and T_1 is replaced by the set of triads common to all the optimal trees.

$$\text{CRI} = \frac{\sum(\text{FW of triads common to all optimal trees that occur in the consensus})}{\sum(\text{FW of triads common to the optimal trees})}$$

Two cases where all triads are strictly preserved (CRI = 1) using the soft interpretation are shown in figures 1.43 and 1.44. We use the CRI as a methodological tool for comparing the various consensus tree methods (see below).

4.6. Comparison of consensus methods

Here we provide a comprehensive review of various consensus methods, considering that an ideal consensus aims to represent the phylogenetic information common to a set of optimal trees (soft interpretation). We do not discuss other problems of representation such as bootstrap results, but only how to interpret efficiently the results of a phylogenetic analysis. A consensus with a perfect retention of information (CRI=1) must verify three properties (table 1.2):

1. Sensitivity (Wilkinson 1994a). A sensitive consensus method retains all available positive information (all possible *n*ts).
2. Non-ambiguity (Wilkinson 1994a). The hierarchical structure of a consensus makes it possible to assess unambiguously phylogenetic relationships supported by synapomorphies. Ambiguity occurs when nestings are conveyed by the consensus, because a nesting is an (and/or) structure that cannot be read as a hierarchy. Reading a nesting as a hierarchy is equivalent to misinterpreting a node of the consensus as a *n*ts while this node does not correspond to any node of the optimal trees.
3. Strict retention of information (new property). Strict retention methods require agreement among all optimal trees. Only the information common to all the optimal trees is unambiguous and necessarily true under the soft interpretation (fig. 1.41).

We follow Bryant's terminology for standard consensus methods (Bryant 2003), and Wilkinson's terminology for reduced consensus methods.

Properties	Strict retention of information	Sensitivity	Non-ambiguity
Nelson-Page consensus	.	.	y
Loose / semi-strict consensus	.	.	y
Matrix representation with parsimony	.	.	y
Asymmetric median consensus	.	.	y
Majority-rule consensus	.	.	y
Greedy consensus	.	.	y
R* (Q*) consensus	.	y	.
Buneman consensus	.	y	.
Adams-II consensus	y	y	.
Triad consensus	y	y	.
Strict consensus	y	.	y
Reduced Cladistic Consensus	y	y	y

Table 1. 2. Properties of consensus tree methods: strict retention of information, sensitivity and non-ambiguity. The letter 'y' means that the consensus method fulfills the property. Strict retention of information implies that only the information common to all the optimal trees is present in the consensus tree. Sensitivity implies retention of all possible *n*ts, and not only components. Non-ambiguity implies that the consensus can be read as a hierarchical tree, without nestings. The RCC is the only consensus method that verifies the three properties: it retains strictly all *n*ts present in all optimal trees, and keeps a non-ambiguous hierarchical representation.

4.6.1. Majority rule Methods

The majority rule consensus method (Margush et McMorris 1981) is among the most widely used consensus methods in phylogenetics, because it produces one single, fairly well-resolved consensus tree. It includes only those replicated components that occur in at least 50% of the optimal trees. Because it retains only components and does not consider *n*ts of lower cardinality, the majority rule consensus does not have the property of sensitivity (table 1.2). Moreover, this method retains replicated components that are not present in all the optimal trees. There are two ways of interpreting a majrule consensus tree. The first one is to read it as the *probable* optimal solution in virtue of a quantitative criterion (number of occurrences of a node). If the tree is read in such a way, there is a possibility for the tree to be false, because it displays the components of 99/100 trees while the true solution lies in the non-retained hundredth optimal tree. Indeed, if the optimal solution is not present in a majority of trees, it will not be represented in the consensus (see Sharkey et Leathers 2001 for biases due to retention of information). There is no reason to ignore one of the optimal trees that is supported by characters, by using a quantitative treatment.

In the second solution, the majrule consensus tree cannot be false, and the consensus suppresses all the implicit information conveyed by disqualifiers linked to *n*ts. The resulting consensus tree does not forbid any relation, because it allows all possible phylogenetic trees. Phylogenetic analysis is qualitative by nature and the CRI reflects this qualitative nature. The CRI calculates the amount of information present in all optimal trees which appears non-ambiguously in the consensus tree. Even with a retention index of 99%, the majrule method keeps a margin of 1% for other components, but one has no way to know which other components are involved. As a consequence, there is no way to be sure that the information displayed in the consensus is strictly present in all optimal trees. It cannot retain the information efficiently, and thus does not forbid any component: the majrule consensus has a null explanatory value. The CRI will always have a zero value for majority rule methods. Methods without strict retention of phylogenetic information are not adapted for clearly representing the information common to a set of optimal trees. The majrule method may prove useful in bootstrap measurements (Wilkinson 1995), but this is beyond the scope of this paper.

The greedy (or extended majority rule) consensus allows additional components to be included in the majrule consensus (Bryant 2003). It is created sequentially by including components one by one. A list of components appearing in the optimal trees is established ordered by frequency of occurrence. In each iteration, one includes the most frequent component that is compatible with all the included components. The process is repeated until no more component can be included. Like the majrule consensus, the greedy consensus does not possess the property of strict retention (CRI = 0).

The loose (or semistrict or combinable component) consensus (Bremer 1990) includes all components that are replicated in all the optimal trees. In addition, it includes components that appear only in one subset of the optimal trees if they are non-contradicted in those optimal trees that do not include them. As for the majrule, the loose consensus contains information that is not present in all the optimal trees. The loose consensus retains the information present in the most resolved trees and rejects the less resolved combinable trees. The loose consensus does not correspond to the inference that polytomies indicate that 'no relation is possible' (hard interpretation), or that 'all relations are possible' (including no relation - soft interpretation). It follows another interpretation in which only one resolution of the polytomy is possible but uncertain. Consider a set of two optimal trees *A(BC)* and *ABC*. The first tree produces one triad *A(BC)* and two disqualifiers not*C(AB)* and not*B(CA)*. In the soft interpretation, we consider that nothing is retained because the second tree *ABC* contains no information. In the hard interpretation, the tree ABC produces not*A(BC)*, not*B(AC)* and not*C(AB)*, and two disqualifiers are common to the set of trees: not*B(AC)* and not*C(AB).* The loose consensus approach seems close to the hard interpretation: not*B(AC)* and not*C(AB)* are retained. However, the method makes it possible to consider that if two disqualifiers are known, they imply the third orthogonal statement *A(BC)* as positive information. The loose consensus breaks the orthogonal links between statements and implies the false belief that there is necessarily a completely resolved solution. The loose consensus method does not prohibit any component and produces a consensus tree that displays more than the information shared by all optimal trees. Thus, it falls into the same problem as the majrule method. Like the majrule consensus, the loose consensus does not possess the property of strict retention (CRI = 0).

The Nelson-Page consensus (Nelson 1979; Page 1989) contains all the replicated components together with the remaining components that are compatible with all of these.

In this method, each component is assigned a weight equal to one less than the number of times this component appears in the set of optimal trees. The Nelson-Page consensus corresponds to the maximum weight compatible subset of the set of optimal trees (Bryant, 2003). It can therefore be assimilated to a majrule consensus with a maximum percentage. It suffers from the same defects as a majrule consensus, i.e. a CRI with a zero value. If the set of optimal trees contains unrooted trees and the splits are weighted by the number of optimal trees they appear in, then the asymmetric median tree (Phillips et Warnow 1996) is formed from a collection of compatible splits with maximum weight (Bryant, 2003). Like the majrule consensus, the Nelson-Page consensus does not possess the property of strict retention (CRI = 0).

The R* consensus is obtained by determining, for each subset of three taxa, which is the most frequent triad. The R* consensus is built based on these favored triads. The Q* consensus is analogous to the R* consensus, the triads being replaced by unrooted quartets (Berry et Gascuel 2000). The Buneman tree is analogous to the Q* consensus with a distance measure added (Buneman 1971). None of these consensus methods have the property of strict retention (CRI = 0).

The Matrix Representation with Parsimony (MRP) is a method that takes as input a set of optimal trees, recodes them as binary matrices, and returns a consensus tree using parsimony analysis (Baum 1992; Ragan 1992). Each component of each optimal tree is recoded as a binary character. The matrix includes all components from all the optimal trees. The parsimony analysis yields a strict consensus that is called the MRP consensus. This method has two flaws. First, the analysis retains the largest sets of components without taking into account the optimal trees from which they come. Second, the parsimony analysis creates homoplasy that distorts the initial information conveyed by components. Homoplasy is a concept of cladistic analysis that has no meaning in the combinatory problem of consensus methods. With MRP, initial components can be divided into several ones because of convergence. Statements can also appear because a component from an optimal tree transformed into a binary character can undergo a reversal. Finally, parsimony can create clades that are not supported by initial characters. The MRP method is highly ambiguous because it yields information that is not present in the optimal trees and because it does not have the property of strict retention of the information common to the set of optimal trees. The Average consensus method (Lapointe et Cucumel 1997) is closely

related to the MRP (Lapointe, Wilkinson, et Bryant, 2003), and has the same flaws. It takes as input a set of weighted trees (i.e., with branch lengths) and uses a matrix representation with distances (MRD) analysis to seek the weighted tree that is closest to the optimal trees using a least-squares criterion.

4.6.2. Strict retention Methods

4.6.2.1. Strict consensus

The strict consensus or strict component consensus (Sokal et Rohlf 1981) is one of the most widely used consensus method in phylogenetics. It includes only those components that are replicated in all optimal trees. As such, its interpretation is unambiguous, and can be understood as the intersection of all components from all optimal trees. The problem of the strict consensus lies in the fact that it retains only components, which are only a particular kind of *n*ts. Consider a set of optimal trees with a stable topology and one floating taxon (e.g. a taxon with a lot of missing data). There is no component that is present in all optimal trees. Consequently, the strict consensus is an unresolved bush. However, these optimal trees do share phylogenetic information, i.e. *n*ts of lower cardinality than components. Because the common phylogenetic information of the optimal trees is not component information, it is not represented in the strict consensus. Insensitivity of the strict consensus to some common phylogenetic information has been noted previously and has led to the view that this consensus is too strict (Wilkinson 1994a) in terms of retained information. The CRI value is comprised between 0 and 1, because the strict consensus is highly dependent on the information present in the optimal trees.

4.6.2.2. Adams consensus and triad consensus

The Adams consensus method (Adams 1972) has been the first method for computing a tree which represents only that information shared by all the optimal trees. The Adams consensus is based upon the intersections of all nestings of all optimal trees (see Barbut 1968). However, it is difficult to understand the significance of this operation from an empirical point of view. The Adams consensus includes all those nestings that are common to a set of optimal trees. Therefore, the Adams consensus is sensitive to common phylogenetic information of lower cardinality than replicated components. However, *n*ts of lower cardinality are constrained to get the same cardinality as components in the Adams

consensus because they are all displayed in a single structure. This operation causes major problems, which are detailed below.

The Adams consensus is very similar to a tree which represents all triads common to all the optimal trees (Adams 1986; Felsenstein 2004; Wilkinson, Cotton, Lapointe, et Pisani, 2007). Therefore, the problem is the same with the triad consensus (Adams 1972), local consensus (Aho, Sagiv, Szymanski, et Ullman, 1981; Kannan, Warnow, et Yooseph, 1998) and intersection tree (Cao *et al.* 2009): the goal is to show in a single hierarchy of cardinality *t* some information of lower cardinality than *t*. The triad consensus is based on triads, but the result is showed in the form of nestings because of the forcing of cardinality, and is very similar to the Adams consensus. The triad consensus illustrates how ambiguity can affect a consensus method. The triad consensus combines a set of triads common to all optimal trees. In this paper, we propose two new notions that are necessary to understand this procedure: compatibility and combinability. Two distinct triads are compatible if and only if there exists at least one tree containing them, i.e., if and only if they do not share the exact same OTUs (fig. 1.45, compatibility). *A*(*CD*) and *D*(*AC*) (fig. 1.42) are logically incompatible because they cannot be true together. Because the set of triads to retain is present in each optimal tree, all triads are compatible. The second property of triads is combinability. Two triads are combinable if and only if they are subtrees of the strict consensus of all possible trees containing the two triads, i.e., if and only if they have two OTUs in common (fig. 1.45, combinability). Combinability means that two triads can be represented into a single non-ambiguous hierarchy. If the triads are not combinable, the strict consensus is completely non-informative. Combinability logically implies compatibility, but the reverse is not always true. Operationally, compatible triads are combinable only if they have at least two taxa in common. The combination of two combinable triads produces a *n*-taxon statement (fig. 1.43; consensus tree), whereas the combination of two non-combinable triads is impossible, and can only produce nestings. Nestings should not be regarded as a kind of positive information, but as a limitation inherent to the way a consensus is constructed. Nestings are the outcome of the failure to read *n*ts in a hierarchical way, and imply ambiguity and loss of information. They appear only when the combinability rule is broken, i.e. when a consensus method enforces the merging of non-combinable information into a single structure.

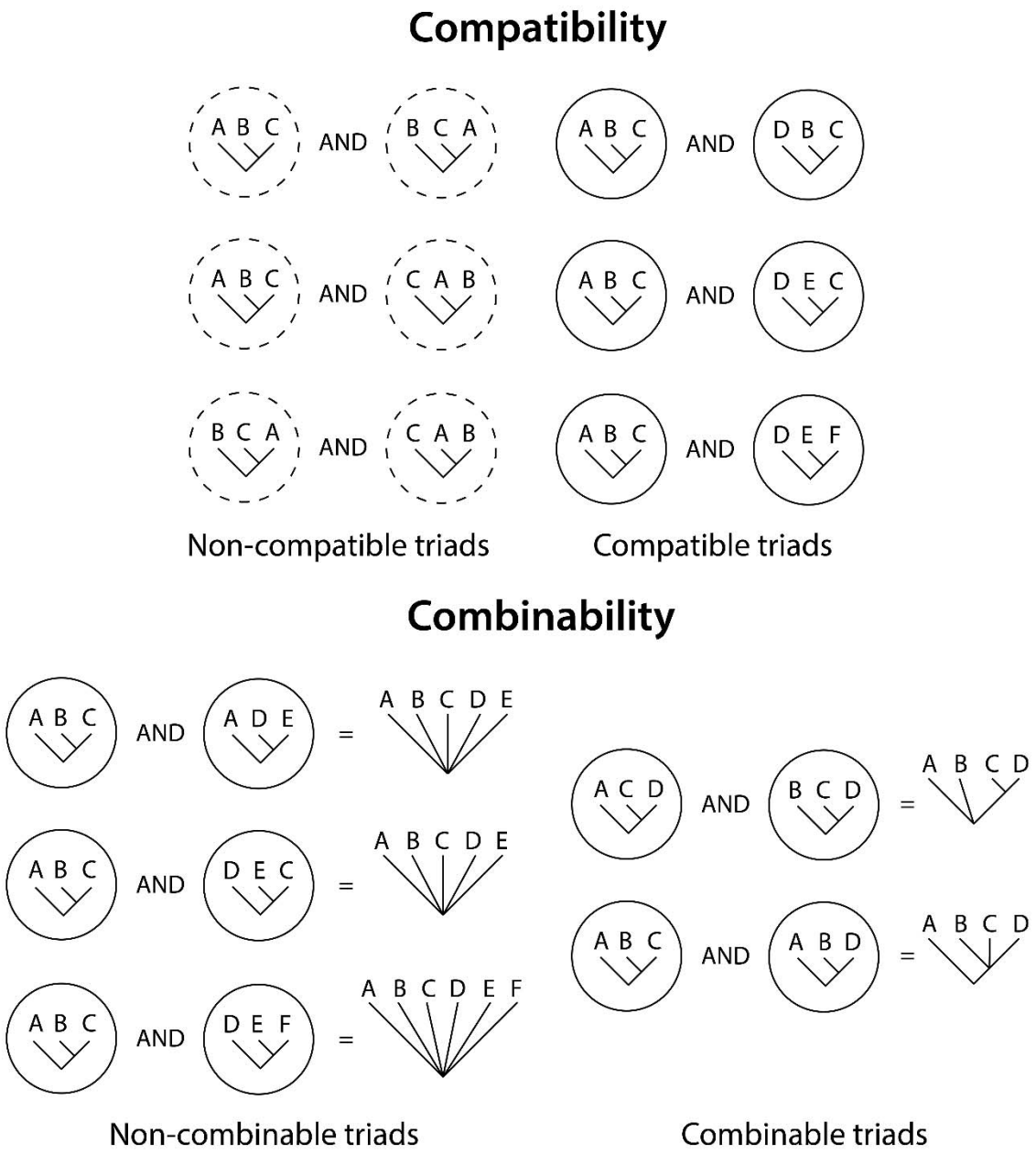


Figure 1.45. Illustration of the concepts of compatibility and combinability using triads. Compatibility – Left, examples of non-compatible triads. Triads within each couple cannot be true at the same time, because they necessarily contradict each other. Right, compatible triads that can be true at the same time. Triads are compatible if they are not orthogonal. Two triads are orthogonal if they have the same set of OTUs but different relationships between these OTUs. Combinability – Left, triads within each couple are non-combinable because they have less than two common OTUs. The strict consensus of all possible trees that contain two non-combinable triads is always non-informative: the triads cannot be combined into one single non-ambiguous hierarchy. Ten possible trees contain the triads of the first couple *A*(*BC*) - *A*(*DE*), e.g. ((*BC*)(*A*(*DE*))) and (*A(B*(*D*(*CE*)))). Right, triads within each couple are combinable because they share two OTUs. The strict consensus of all possible trees that contain two combinable triads always includes an informative internal node.

Importantly, the Adams consensus is not a hierarchy because of its nestings. *{AB}CD* cannot be represented as a tree *(AB)CD* because the node (*AB*) can contain *C* AND/OR *D* in the Adams consensus. In the example of two optimal trees *A(B(CD)* and *D(C(BA))*, the Adams consensus shows a node BC that excludes OTUs A and D. Most authors read Adams trees as classical hierarchies, and point out the fact that the Adams' method creates artificial phylogenetic information that is not present in the optimal trees. In our example, the consensus is read as *AD(BC)*. In this case, it is easy to show that the clade *AD(BC)* is not present in *A(B(CD)* or *D(C(BA))*, the Adams consensus being nothing else than a *false hierarchy* in the graph theory. Wilkinson (1994) pointed out that it is erroneous to consider the Adams consensus as a traditional tree, its status being more subtle. This author showed that Adams (1986) described a relation of subset nesting: '*If A and B share a node in common that is further from the root than the node including A, B, C and D, then A and B are a subset nesting of the larger group, symbolically {A, B}C,D .[…] The simplest cladistic interpretation of the nesting {AB}C,D is that A and B share a more recent common ancestor than do A, B, C and D. Nestings can also be interpreted in terms of relative closeness of relationships of non-overlapping sets of taxa (n-taxon statements) but only through ambiguous conjunctions of them. […] In other words, a given nesting does not unambiguously entail a specific n-taxon statement. For {A,B}C,D to be true of a particular cladogram, either (A,B)C,D could be true or just one of (A,B)C and (A,B)D: (A,B)C and/or (A,B)D*'. If a node must be understood in such a way, then *the Adams consensus is not a hierarchy* because a node does not necessarily imply that all its OTUs form a class (i.e. a clade). We name the structure produced by the Adams consensus method an '*Adams pseudo-hierarchy*' in order to avoid confusion. This structure should not be represented as isomorphic to hierarchies. A consensus must be a tree in a mathematical sense, i.e. a fully connected graph without cycle. The property of non-ambiguity of a consensus method allows keeping the hierarchical structure of the consensus as well as all the information used to construct it: the passage between the information common to optimal trees and the consensus tree must be possible in either way. Because the nestings always show possibilities, the Adams consensus does not have a unique CRI value, i.e., the CRI is comprised in an interval if the consensus shows polytomies. The CRI points out the flaws of the nesting consensus methods. If nestings show possibilities, the

choice of a particular structure is impossible, and makes it very difficult to assess the congruent information between all optimal trees.

4.6.2.3. Reduced cladistic consensus

The Reduced Cladistic Consensus (RCC) provides a solution to the problem of representing all *n*ts common to a set of optimal trees (Wilkinson, 1994a). It is intended to achieve a greater sensitivity to common phylogenetic information, as compared with the strict consensus, without the ambiguity that affects Adams consensus trees. This method allows consensus trees to represent relationships among a subset of the taxa included in the optimal trees. Consequently, there may not be a unique reduced consensus for a set of optimal trees. Appropriate selection of a set of reduced consensus trees (i.e. the RCC profile) provides a means of avoiding the ambiguity that results from the elimination of taxa (Wilkinson, 1994a).

The RCC profile is constructed by comparing components of the optimal cladograms (Wilkinson 1994a). The strict consensus is a member of the RCC profile if it is informative. A pairwise comparison of all the components of two optimal trees determines a set of common *n*ts. This set of *n*ts can then be compared with components of the next optimal tree, and so on. The result is a list of *n*ts of the highest possible cardinality, including all *n*ts common to the set of optimal trees. The trees of the RCC profile are built by combining statements from the list of *n*ts that have identical terminal taxa. The RCC algorithm iteratively prunes OTUs from the optimal trees in order to reconstruct a set of strict consensuses of lower cardinality called the RCC profile.

The RCC profile can also be constructed by retaining all triads common to the optimal trees (Wilkinson 1994a). *N*ts of higher cardinality can then be built up through the combination of these triads. However, this method differs from the triad consensus in combining sets of combinable triads only. We will highlight the properties of the RCC by constructing a RCC profile with triads.

Here we extend the compatibility and combinability of triads to all kind of *n*ts. Two *n*ts are compatible if they do not contradict each other. Incompatible *n*ts cannot be represented simultaneously in a strict consensus. Two triads are combinable if they share two taxa in their structure (combinability implies compatibility; fig. 1.45). Two *n*ts are combinable if they share an internal node (a node comprising the same taxa in both *n*ts),

including the root. Combinable *n*ts can be mixed together to form a hierarchical structure, without loss of information. *A*(*BCD*) and *B*(*CD*) yield *A*(*B*(*CD*)) because (*BCD*) is an internal node in both trees (an informative node in the first tree, the root in the second). This also explains why the strict consensus is hierarchical: in this case, the rule of combinability is respected because all components have a node in common: the root (i.e. the class containing all taxa). *N*ts can be combinable in other cases, but we only need this property for constructing a RCC profile with triads. The triad consensus method (Cao *et al.*, 2009) has a major flaw: it mixes all triads common to all the optimal trees together, including triads that are not combinable. In contrast, the RCC takes into account the fact that triads cannot be combined every time. Triads included in each tree of the RCC profile can always be combined together. Therefore, all trees of the RCC profile have a hierarchical structure.

The RCC possesses all the three properties listed above (table 1.2), namely sensitivity, non-ambiguity and strict retention of information. As such, the RCC profile is the flawless solution to the soft consensus aim. On the other side of the coin, the RCC method can produce more than one tree. Two non-combinable triads must be retained separately because they convey non-combinable positive phylogenetic information. The RCC is a strict *n*ts consensus equivalent to a strict triad consensus fulfilling the property of combinability of triads. Because all *n*ts common to a set of optimal trees are represented by the RCC profile, and because no ambiguity is created by mixing non-combinable triads in the RCC, there is no loss of phylogenetic information (i.e. sensitivity; table 1.2). Therefore, the RCC profile always has a maximal Consensus Efficiency index, and the CRI is always equal to 1. The CRI shows that a flawless consensus cannot always be represented by one single consensus tree. The RCC can be computed using RadCon 1.1.2 (Thorley and Page, 2000) and REDCON 3.0 (Wilkinson, 2001). We propose here an R script for computing RCC using Wilkinson's (1994) algorithm (https://github.com/JeremieBardin/Valentin-Rineau-PhD-scripts; author: Jérémie Bardin). The script displays the consensus profile graphically in addition to a tree file, with the discarded taxa for each tree of the consensus profile.

4.7. Conclusion

We formalized the differences between two incompatible interpretations of consensus trees: hard consensus aim (i.e., a consensus is a summary of the information conveyed by a set of optimal trees) and soft consensus aim (i.e., a consensus represents only that information shared by all the optimal trees). The hard consensus aim implies that all optimal trees are potentially true. The resulting consensus tree is false, since it does not represent any of the true trees. It is merely a practical tool, a compressed set of information that can be interpreted as a set of trees, but not as a phylogeny. The soft consensus aim implies that there is potentially one true tree hidden within the whole set of optimal trees: each node of the consensus refers to a clade present in all optimal trees. The resulting consensus tree is true, since it conveys only the non-ambiguous phylogenetic information common to the set of optimal trees. As such, it can be interpreted as a phylogeny and used in evolutionary studies. We show that the hard consensus aim should be discarded, because it considers disqualifiers as independent phylogenetic information. Negative phylogenetic information conveyed by disqualifiers can be retained in the consensus despite the fact that positive statements from which they are derived are not retained. The soft consensus aim is accurate, because orthogonal relationships between *n*ts and disqualifiers are always preserved. The origin of statements is taken into account.

A consensus tree with a perfect retention of information (CRI equals to one) must verify three properties: strict retention of information (unanimous agreement among the optimal trees is required), sensitivity (representation of all possible *n*ts) and non-ambiguity (hierarchical structure preserved by assembling combinable *n*ts only). Nestings are undesirable by-products generated by breaking the rule of combinability. Therefore, consensus methods that include nestings (i.e. Adams consensus and Triad consensus) should be discarded because they lose the hierarchical structure of the original information. The RCC is the only consensus method that can be regarded as 'perfect' because it fulfills the three properties listed above. An R script for computing RCC is proposed in supplementary online material 1.

5. Conclusion du chapitre

La théorie cladistique est la théorie de l'histoire des objets en évolution. Elle inclut une méthode, l'analyse cladistique, permettant de reconstruire cette histoire évolutive, par l'analyse et la synthèse de relations de parenté entre les objets et leurs parties. La 3ia est la méthode d'analyse cladistique qui reconstruit sur un critère congruence le cladogramme contenant le maximum de relations élémentaires de degré de parenté (représentés par des triplets pondérés).

La figure 1.46 montre les étapes de l'analyse de 3ia. Une analyse de 3ia nécessite de spécifier les objets à classer. La 3ia possède un fort pouvoir explicatif, puisqu'elle s'applique à n'importe quel type d'objets dont l'évolution est en accord avec la théorie cladistique. Ces objets se différencient et se diversifient. La 3ia peut s'appliquer entre autres aux taxons, aux aires biogéographiques ou encore aux langues humaines. Ces objets doivent pouvoir se décomposer en parties. Les taxons se décomposent en homologues, les aires en taxons. Des relations de parenté peuvent ensuite être proposées à partir d'arguments externes à l'analyse. Pour une analyse phylogénétique, ces arguments peuvent être de nature anatomique, morphologique, biochimique, comportementale, mais doivent toujours correspondre à des propriétés intrinsèques à ces taxons. Ces relations entre parties sont standardisées en relations entre porteurs de ces parties. Ces relations, les caractères, sont des hypothèses de clades. On choisit le cladogramme qui répond à la question initiale de la spécification des relations de parenté entre objets comme la hiérarchie de clades retenant le maximum d'hypothèses formulées en amont (homologies puis caractères). Si plusieurs cladogrammes sont également optimaux, le RCC permet de résoudre le problème du choix de l'arbre et de proposer un ou plusieurs phylogénies non-ambigües.

Le but de la 3ia est de reconstruire une hiérarchie de clades. Elle est une méthode historique dont le but est de reconstruire le passé d'entités en évolution. La 3ia est une méthode scientifique, puisqu'elle a pour but la construction d'énoncés réfutables. Un cladogramme est vrai ou faux s'il est possible de se rendre compte de la non-monophylie de certains groupements. Plus particulièrement, la 3ia permet de faire des prédictions. La capacité d'une théorie à faire des prédictions permet de mettre en évidence son pouvoir explicatif. Par exemple, la théorie de la relativité générale développée par Einstein a permis la production d'énoncés qui se sont révélés vrais après coup par l'expérience. Ces énoncés

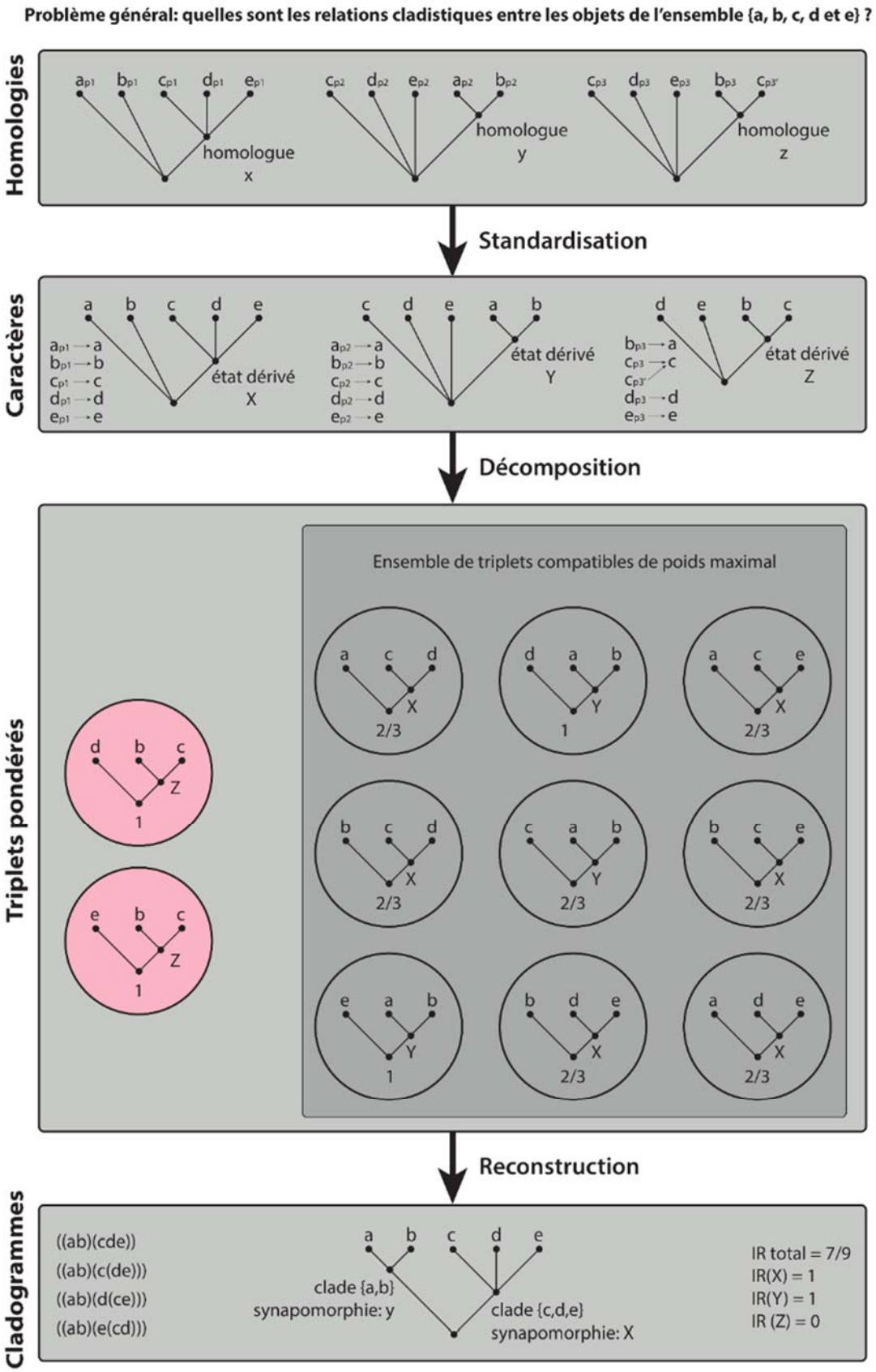


Figure 1.46. Schéma général d'une analyse en 3ia. Les homologies sont standardisées en caractères. Les caractères sont décomposés en triplets (à chaque triplet est associé une PF). Le cladogramme répond à la question des relations de parenté au sein de $\{a, b, c, d, e\}$.

sont des prédictions scientifiques. Une des plus célèbres est le mirage gravitationnel. Un mirage gravitationnel est produit par la présence d'un corps céleste massif générant un champ gravitationnel ayant pour effet de dévier la lumière. Un observateur verra une image déformée par un champ gravitationnel d'un objet situé derrière le corps massif. Le premier mirage gravitationnel est observé en 1979, soit 42 ans après la prédiction de Fritz Zwicky de l'effet gravitationnel de galaxies sur la lumière des sources qu'elles cachent (Zwicky 1937; Walsh *et al.* 1979). La théorie cladistique permet, notamment par le biais de la 3ia, de prédire l'existence d'homologues chez certains taxons. Une analyse génère un cladogramme. Un de ces clades est soutenu par une synapomorphie X, et on constate que l'une des UT de ce clade ne porte pas l'homologue X_h. On postule l'existence d'une réversion pour expliquer la présence de cette UT dans ce clade. La 3ia permet la prédiction suivante : l'UT porte l'homologue X_h. Après un travail de laboratoire d'évo-dévo, on se rend compte que l'homologue X_h était bien présent dans l'UT, mais que celle-ci n'était pas disponible à l'observation directe du fait d'une mutation inhibant l'expression de X_h. La 3ia, en tant que méthode d'analyse cladistique, permet de prédire l'existence d'homologues chez certaines UT par ce biais. Elle permet de la même manière de prédire quels étaient les homologues d'une UT possédant des données manquantes.

Chapitre 2. Analyse à trois éléments et parcimonie : éléments de comparaison

La partie 3 de ce chapitre a été écrite en collaboration avec A. Grand, R. Zaragüeta i Bagils et M. Laurin, la partie 4 avec R. Zaragüeta i Bagils et M. Laurin.

1. Introduction 137

2. Méthode de parcimonie et 3ia : comparaison théorique et méthodologique 138

3. Experimental systematics: sensitivity of cladistic methods to polarization and character ordering schemes 161

4. Impact of errors on cladistic inference: simulation-based comparison between parsimony and three-taxon analysis 191

5. Conclusion du chapitre 215

1. Introduction

Le premier chapitre de cette thèse présentait une introduction générale aux concepts de la théorie cladistique et au fonctionnement de la méthode de 3ia. Nous avons défini notre corpus théorique de manière cohérente à partir du principe selon lequel seuls les groupes complets en termes de filiation permettent, selon le principe d'hérédité, de reconstruire la phylogénie. Nous avons également porté une attention particulière à la structure mathématique des hypothèses phylogénétiques, celle-ci devant être nécessairement hiérarchique (Williams et Ebach 2006; Zaragüeta i Bagils et Bourdon 2007; Prin 2012). De ce premier principe, fondement de la vision de la théorie cladistique, découle le rejet des groupes paraphylétiques (et polyphylétiques) au profit des groupes monophylétiques (Hennig 1965, 1966). Le premier chapitre traitait de la 3ia comme d'une méthode autonome. Cependant, d'autres méthodes sont associées à la cladistique : la parcimonie standard et la compatibilité ont été définies et conceptualisées en regard de cette théorie. De ces deux familles de méthodes, seule la parcimonie standard est aujourd'hui utilisée. Le but de ce chapitre est de comparer cette la méthode de parcimonie à la 3ia afin, selon nous, de déterminer laquelle est la plus pertinente en phylogénétique cladistique. Ce chapitre se veut une argumentation de notre choix de traiter la reconstruction phylogénétique des rudistes en 3ia uniquement. Sous le terme de « *méthode de parcimonie* », il ne sera en fait question que de la parcimonie de Wagner, autorisant convergences et réversions, et caractères additifs (Kluge et Farris 1969; Farris 1970). (i) Dans un premier temps, nous présenterons une compilation de quelques-uns des arguments théoriques et méthodologiques donnés dans la littérature, en accord avec la théorie cladistique telle que nous l'avons définie dans le chapitre 1. Nous porterons l'emphase sur la structure mathématique des hypothèses initiales, notamment sur la notion de *pas de transformation* (ou *pas évolutif*) et ce qu'elle implique au niveau théorique, méthodologique et empirique. Nous montrerons que la différence initiale de structuration des caractères entre méthode de parcimonie – partition et pas de transformation – et 3ia – hiérarchie et classe de hiérarchie – permet de comprendre la plupart des points de divergences subséquents entre les méthodes : les données manquantes et non-applicables, la polarisation du cladogramme, le test des états dérivés (« optimisation » en parcimonie), le calcul de l'indice de rétention,

ainsi que l'analyse cladistique biogéographique. (ii) Dans un second temps, nous proposons de nouveaux arguments de comparaison empiriques, basés sur des simulations informatiques, suivant la recommandation de Siebert et Williams dans leur article « *Recycled* » : "*Ultimately, the relative merit of 3ta will be resolved empirically, by comparison of classifications produced from it with classifications using other methods"* (Siebert et Williams 1998: 339). Nous présentons ici le résultat de simulations basées sur des modèles d'évolution afin de tester l'efficacité respective de la méthode de parcimonie et de la 3ia. Nous proposerons également de comparer ces deux méthodes via des exemples dont on connait *a priori* l'histoire évolutive, et donc la phylogénie. Les différences empiriques seront à chaque fois reliées à des spécificités théoriques et méthodologiques des méthodes.

2. Méthode de parcimonie et 3ia : comparaison théorique et méthodologique

2.1. Cohérence théorique

2.1.1. Pas de transformation versus caractères hiérarchiques

La méthode de parcimonie adopte les termes plésiomorphie et apomorphie tels que définis par Hennig : « *We will call the characters or character conditions from which transformation started (a, b) in a monophyletic group plesiomorphous, and the derived conditions (a', a'', b', b'') apomorphous* » (Hennig 1966 : 89). Par la suite, la méthode de parcimonie, aux fondations hennigiennes, a toujours nécessité l'utilisation d'états mutuellement exclusifs (Farris 1970; Colless 1985; Pimentel et Riggins 1987; Pogue et Mickevich 1990). Le principe général de la méthode de parcimonie est exposé en figure 2.1 : soient n caractères-partitions, on choisit l'arbre – enraciné ou non – qui minimise le nombre de passages d'un état à un autre pour l'ensemble des caractères. Cet arbre est celui qui distord le moins l'ensemble des caractères, et est appelé arbre le plus parcimonieux (ou encore le plus court).

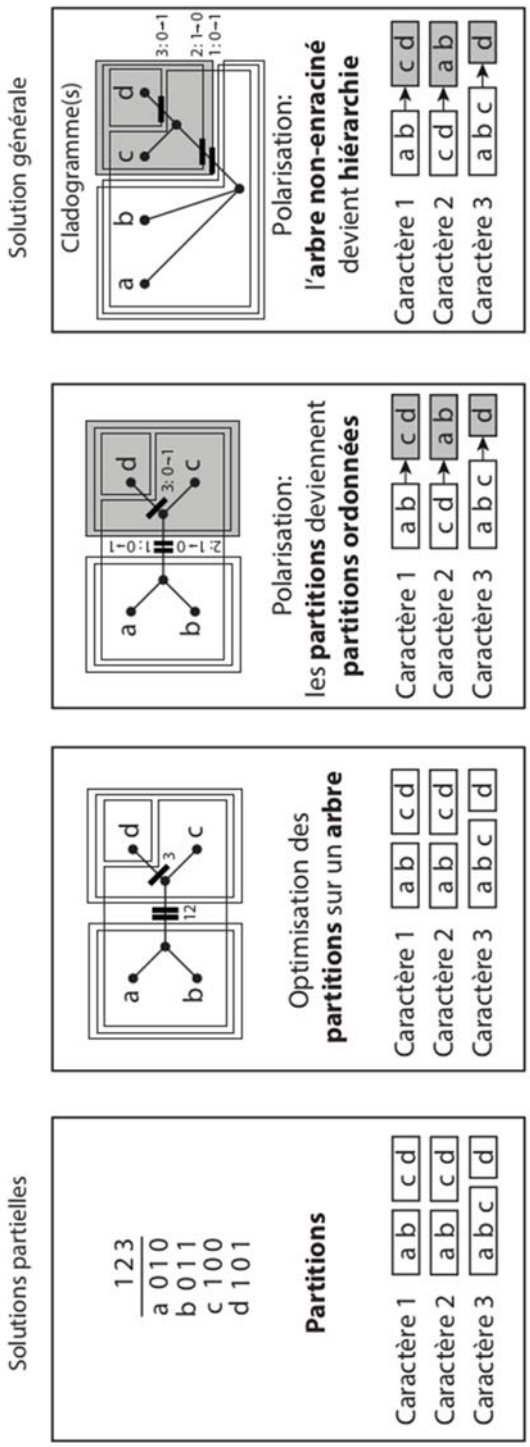


Figure 2.1. Schéma général de la méthode d'analyse cladistique en parcimonie. (i) Les caractères sont des partitions non ordonnées (langage mathématique : non-ordonnée ; langage des systématiciens : non polarisée), rassemblées en une matrice. (ii) Les partitions sont ensuite optimisées sur un arbre non-enraciné de manière à diminuer au maximum la distorsion des partitions – minimisation du nombre de classes créées. Ces structures ne sont pas mathématiquement équivalentes et n'ont pas le même contenu sémantique. (iii) Les partitions sont ordonnées (i. e. polarisées), le plus souvent grâce à l'un des taxons désigné comme extra-groupe. (iv) La polarisation des caractères entraîne la polarisation de l'arbre qui devient hiérarchique et sur lequel est possible de déduire si un groupe est monophylétique ou non. Le cladogramme optimal est celui qui implique le minimum de pas de transformation. Les caractères sont tous paraphylétiques par leur état plésiomorphe – l'état ancestral d'un caractère-partition ne contenant jamais tous ses descendants.

Dans le cadre de la théorie cladistique, le fait que les états soient mutuellement exclusifs est un problème qui entraîne des répercussions sur toute la méthode de parcimonie. Pour illustrer cela, supposons un caractère K, dit binaire, à deux états 0 et 1 et un ensemble $A = \{a, b, c, d\}$ de quatre UT distinctes et tel que chacune de ces UT peut être décrite par au moins l'un des états de K. Supposons ensuite que l'application de K à A nous conduise à assigner 0 à a et b et 1 à c et d. Nous obtenons alors bel et bien un caractère selon la définition standard donnée par la méthode de parcimonie. L'ensemble $P = \{\{a, b\}, \{c, d\}\}$ des classes associées à K est en effet une partition (car ces classes sont non vides, recouvrent A et sont mutuellement exclusives); le fait que les états soient mutuellement exclusifs interdit donc de fait que les caractères soient des hiérarchies inclusives. Plusieurs problèmes initiaux se posent. Supposons en effet que 0 soit plésiomorphe et 1 apomorphe (par le recourt à des taxons extra-groupes par exemple); on a alors une relation d'ordre de type ancêtre-descendant entre les états de K; or si cela est vrai, on obtient alors également une relation d'ordre de même type sur les classes de P; dit autrement, on est passé de la partition non ordonnée P à une partition ordonnée $O = \{a, b\} \rightarrow \{c, d\}$. Par conséquent, le taxon $\{a, b\}$ correspondant aux UT décrites par l'état plésiomorphe 0 est nécessairement paraphylétique au sens de Hennig (1965, 1966). Ainsi, même si l'étape de polarisation rend la partition ordonnée, cette dernière n'en reste pas moins une partition. Nous avons montré comment seule la différenciation permettait de rendre compte efficacement de l'histoire évolutionnaire de taxons grâce à l'hérédité d'homologues acquis puis transmis à leurs descendants. Dans un cadre cladistique de type différenciation, il apparaît évident que le taxon $\{a, b\}$ défini par l'état plésiomorphe 0 de K est paraphylétique. Or, un principe de la théorie cladistique est d'interdir les groupes paraphylétiques et, corollairement, qu'on ne groupe pas par un état plésiomorphe. Il reste que cet état est nécessaire à l'ensemble du caractère, et supprimer $\{a, b\}$ revient à supprimer la partition entière. Les caractères construits pour la méthode de parcimonie sont donc constitués à la fois de groupes monophylétiques et paraphylétiques. La seule manière d'obtenir des caractères cladistiques constitués de groupes monophylétiques uniquement est que ceux-ci s'assemblent en une hiérarchie; la 3ia impliquant nécessairement d'utiliser des caractères hiérarchiques, cette dernière n'est donc pas exposée à cette incohérence.

Les caractères-partitions ayant été construits, ils forment l'ensemble de solutions partielles qu'il convient de synthétiser en une solution générale, dont la théorie cladistique impose la forme hiérarchique. Si la méthode de parcimonie est une méthode cladistique, alors elle doit satisfaire au principe d'*analyse logique* qui veut que la somme des solutions partielles ne disent rien de plus que la solution générale (Hennig 1966; Nelson 1994; Cao *et al.* 2007; Zaragüeta i Bagils et Pécaud 2016). Dans le cas de la méthode de parcimonie, l'arbre final ne devrait rien dire de plus que l'ensemble des partitions. Or, comment est-il possible que la mise en congruence d'un ensemble des partitions donne une hiérarchie ? (i) Soit la solution générale n'est pas une hiérarchie mais une partition ordonnée. Ce point n'a jamais été abordé nulle part, tous les cladistes s'accordant par définition sur l'objectif de la reconstruction phylogénétique, celui-ci étant la construction de hiérarchies. (ii) Soit il y a une incohérence entre les solutions partielles et la solution générale qui n'est ni expliquée par la théorie cladistique ni justifiée par la méthode de parcimonie. Une connaissance *a priori* externe à l'analyse est nécessairement ajoutée lors de la synthèse des parties en tout. Cette connaissance *a priori* est celle de la hiérarchie des clades. On impose donc *a priori* la structure hiérarchique comme la solution générale. Mais dans ce cas, pourquoi ne pas l'avoir également imposé aux solutions partielles, qui ne doivent être rien de plus que des hypothèses sur la phylogénie, c'est à dire des hypothèses de clades ? Dans ce cas, il faut donc renoncer au couple de principes d'*analyse logique* et d'*analyse du tout, synthèse des parties*, puisque dans ce cas tout et parties doivent avoir logiquement la même forme mathématique, les parties étant des hypothèses sur le tout.

2.1.2. « *Optimisation* » : quelques problèmes découlant de l'incohérence partition/hiérarchie

2.1.2.1. Principe d'analyse logique et l'hypothèse de réversion en parcimonie

Nous avons vu précédemment que la méthode de parcimonie recherche le cladogramme optimal en comptant le nombre de pas de transformation sur un arbre non-enraciné, puis en choisissant celui qui minimise le nombre de transformation et en le polarisant. Minimiser les transformations peut encore être décrit comme une minimisation de la distorsion totale des partitions – caractères de la parcimonie. La distorsion arrive

lorsqu'un caractère n'est pas compatible avec un cladogramme. Dans ce cas, le caractère est reconstruit de manière à minimiser la création de classes supplémentaires. Prenant un caractère (*a*, *b*, *c, d*) → (*e*, *f, g*) (correspondant à un pas de transformation; fig. 2.2), son optimisation sur un cladogramme peut être parfaitement congruente sans distorsion, comme sur l'arbre (*a*, (*b*, (*c*, (*d*, (*e*, (*f, g*)))))). Puisque l'arbre est choisi comme celui qui minimise la distorsion des caractères, toute classe d'une partition définissant un état dérivé soutient un clade qui lui est associé. La classe (*e*, *f, g*) de la partition est synapomorphie du clade (*e*, *f, g*). Par exemple, sur l'arbre (*a*, (*b*, (*e*, (*f*, (*g*, (*c*, *d*))))), la partition doit être distordue de manière à minimiser la création de classes supplémentaires : ici, (*a*, *b*) → (*e*, *f*, *g*) → (*c, d*) (deux pas de transformation, soit trois classes au lieu de deux). Ce type de distorsion est généralement discuté comme étant une « réversion ». On réinterprète le caractère comme un état ancestral (*a*, *b*), transformé en état dérivé (*e*, *f*, *g*), lui-même transformé en un nouvel état dérivé (*c, d*). Plusieurs points sont à discuter à partir de cet exemple :

1. L'état dérivé (*e*, *f*, *g*), postulé comme étant monophylétique, se retrouve paraphylétique. Il est donc incomplet en termes de filiation, et ne devrait d'après la théorie cladistique pas servir au soutien d'un clade. Pourtant, il soutient ici le clade (*e*, *f*, *g*, *c*, *d*). N'importe quel état dérivé distordu de manière à produire une réversion devient de fait paraphylétique (voir Patterson 1982).
2. L'état dérivé (*c, d*) est une hypothèse qui n'a pas été construite par le systématicien. La signification du caractère était « *je postule l'existence d'un clade regroupant e, f et g sur la base d'un homologue partagé par ces taxons qui n'est pas partagé par les autres taxons a, b, c, d* ». Selon le principe d'analyse logique qui veut que la solution finale ne dise rien de plus que l'ensemble des solutions partielles, tout ce qu'on peut déduire des cladogrammes est qu'ils rejettent ou acceptent les hypothèses construites en première instance. Dans le cas d'une réversion, c'est l'optimisation seule qui va réévaluer le caractère et produire de nouvelles hypothèses d'homologie. L'algorithme répond alors au chercheur ayant produit ses caractères : « *l'état que tu as construit doit être distordu sur le cladogramme optimal. J'ai donc réévalué tes hypothèses phylogénétiques pour toi et je suis en mesure de te dire que* c *et* d *forment un clade basé sur l'existence d'un nouvel état dérivé que tu n'avais pas perçu*». La génération d'hypothèses *ex nihilo* n'apparaît

pas désirable pour une méthode d'analyse cladistique, puisque la solution générale n'est plus seulement la somme de toutes les solutions partielles.

La question des réversions a été un point de cristallisation des débats opposant les tenants de la méthode de parcimonie et de la 3ia (Kluge 1994; Farris *et al.* 1995; Farris 1997; De Laet et Smets 1998; Farris et Kluge 1998; Siebert et Williams 1998). Nous rediscutons des réversions dans la partie 3 de ce chapitre (p. 179) pour interpréter les résultats de nos simulations. La 3ia, au contraire, respecte l'ensemble des principes de la théorie cladistique : à cet effet, chaque triplet – compris comme hypothèse phylogénétique minimale – est accepté ou rejeté. Aucune hypothèse supplémentaire n'est générée *ex nihilo*.

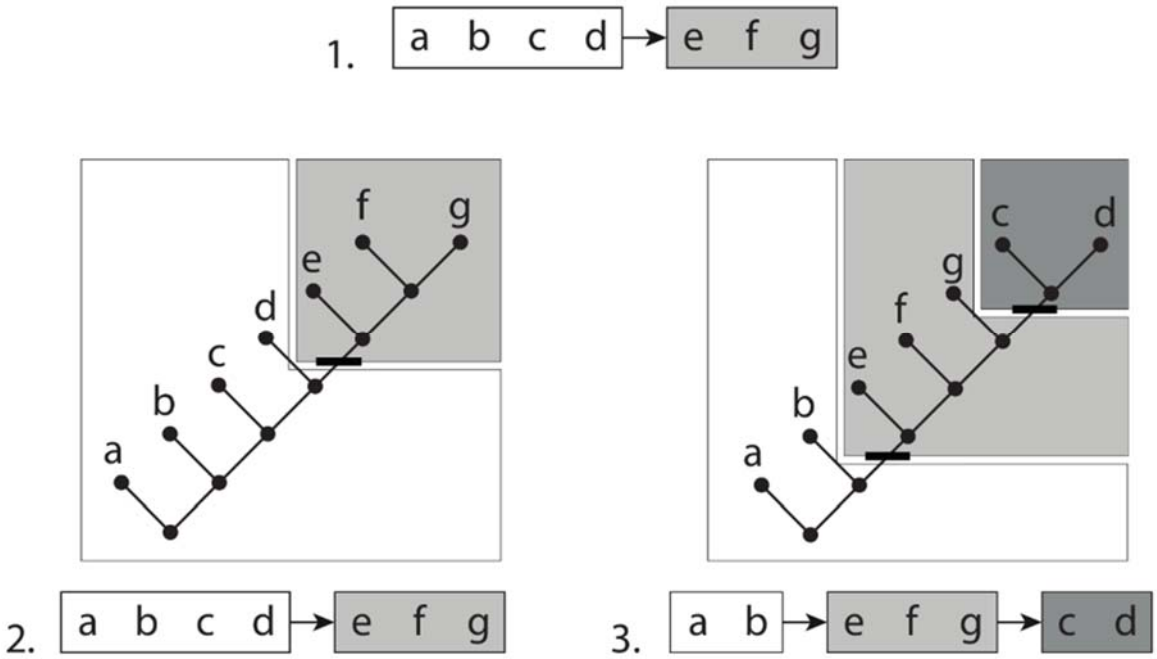


Figure 2.2. Exemple d'optimisation en parcimonie d'une partition sur un arbre. (1) Partition ordonnée (*a*, *b*, *c*, *d*) → (*e*, *f*, *g*) optimisée sur les arbres (2) et (3) (en haut, cladogramme ; en bas, partition résultant de l'optimisation. Sur l'arbre (2), la partition n'est pas distordue : l'état dérivé est synapomorphique du clade (*e*, *f*, *g*), et l'état ancestral (*a*, *b*, *c*, *d*) est paraphylétique – et donc artificiel. Sur l'arbre (3), la partition est distordue en trois classes au lieu de deux (distorsion de type « réversion »). Une nouvelle hypothèse de synapomorphie (*c*, *d*) est générée à partir de l'état ancestral paraphylétique. L'état dérivé (*e*, *f*, *g*) devient paraphylétique – et donc artificiel.

2.1.2.2. Données manquantes, données non-applicables et polymorphisme

Les réversions ne sont pas le seul problème de l'optimisation de structures mathématiques sur d'autres structures mathématiques non-équivalentes. Partitions et hiérarchies n'ont pas le même contenu épistémologique. Les données non-applicables en sont un premier exemple (Zaragüeta i Bagils et Bourdon 2007). En 3ia, les données non-applicables sont gérées de la manière suivante : lors de la construction d'un caractère, si un taxon ne possède par l'état dérivé, il est directement branché à la racine. Ainsi, l'hypothèse est que le taxon ne possède aucun des états dérivés. Il permet la création de triplets (chapitre 1, p. 37) supplémentaires, car c'est une connaissance sur ce taxon de savoir qu'il ne possède que l'état ancestral. Au contraire, une donnée manquante diminue le contenu informatif d'un caractère hiérarchique, puisqu'il nous manque des connaissances pour placer ledit taxon dans le caractère. Le taxon est donc supprimé, ce qui induit une diminution du nombre de triplets : le taxon peut se placer n'importe où dans le caractère. En 3ia, les hypothèses « *tous les états sont possibles* » (données manquantes) et « *aucun état n'est possible* » (données non applicables) pour un taxon donné sont donc gérées de façon complètement différentes (et pertinentes au regard de ce que ces hypothèses signifient). Au contraire, la méthode de parcimonie nécessite toujours une optimisation, quelle que soit la source du problème : donnée non applicable, ou donnée manquante. Dans le cas d'une donnée manquante, on ne renseigne pas d'état : l'algorithme propose un état au taxon de manière à minimiser la distorsion totale des caractères. Dans le cas d'une donnée non applicable, les algorithmes de parcimonie vont de la même manière optimiser l'état du taxon, alors qu'on avait comme hypothèse qu'aucun état ne s'appliquait à lui. La méthode de parcimonie ne peut donc pas gérer de manière différente ces deux problèmes (Zaragüeta i Bagils et Bourdon 2007). Cette critique peut s'étendre plus globalement à toutes les méthodes de reconstruction phylogénétique à matrice.

Nous avons également vu que le polymorphisme, c'est-à-dire l'assignation de deux états à un taxon, se gère pour des caractères hiérarchiques de manière non-ambigüe (chapitre 1, p. 48). Un nouveau problème se pose lors de l'utilisation de partitions. Une partition se définit par une collection de classes mutuellement exclusives. Le polymorphisme contrevient à cette règle en posant l'hypothèse d'un chevauchement entre

deux états (fig. 2.3). Cependant, s'il y a chevauchement, alors le caractère n'est plus une partition, et il faut choisir qu'elle partition donne la meilleure représentation de l'hypothèse faite par le chercheur. Là, les algorithmes de parcimonie contournent le problème. Plutôt que de construire un caractère ou le taxon possède effectivement les deux états comme en 3ia, la parcimonie va optimiser le caractère en considérant que le taxon polymorphique peut posséder n'importe quel état : pas tous les états en même temps, seulement *un*, celui qui minimise la distorsion totale des caractères par rapport à un cladogramme. Le polymorphisme se gère donc en parcimonie exactement de la même manière à la fois que les données manquantes et que les données non-applicables. C'est-à-dire qu'il ne se gère pas, mais qu'il est contourné : on ne peut pas vraiment coder du polymorphisme en parcimonie, celui-ci étant supprimé par l'utilisation d'un critère non pertinent. On note d'ailleurs qu'il se gère avant même la polarisation de l'arbre. Cette optimisation revient en fait à ne pas décider de ce que veut vraiment dire un caractère en termes de qui est plus proche de qui. L'hypothèse est déterminée en fonction du cladogramme-résultat ce qui est encore une défaillance dans le principe d'*analyse logique*. C'est aussi une violation du principe d'indépendance (la source des caractères est indépendante et externe à la mise en congruence). La 3ia offre des réponses différentes adaptées à chacun de ces problèmes : exclusion du taxon pour une donnée manquante, branchement à la racine pour une donnée non applicable, analyse en sous-arbres libres de paralogie pour le polymorphisme. Chacun

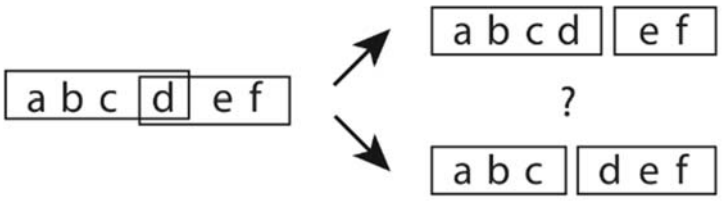


Figure 2.3. Exemple de polymorphisme en parcimonie. Une partition impose des classes d'états mutuellement exclusives. Or, un taxon polymorphique possède les deux états à la fois, et génère une pseudo-partition. La pseudo-partition ne peut pas être utilisée en l'état : elle doit être interprétée en partition. Deux interprétations sont possibles, conduisant à la partition 1 ou à la partition 2. Le choix de la bonne partition devrait se faire au moment même de la construction du caractère, selon des critères intrinsèques au caractère, par exemple sa structure. Aucun algorithme de parcimonie n'arrive à gérer le problème du polymorphisme, et le contourne en utilisant un critère extrinsèque totalement indépendant du problème : le choix de la partition se fait selon le critère de minimisation de la distorsion de l'ensemble des caractères.

de ces traitements, au contraire de la parcimonie, modifie le contenu informatif des caractères impliqués.

2.1.2.3. Ambiguïtés d'optimisation

Pour terminer la revue des problèmes de cohérence partition/hiérarchie, une dernière ambiguïté peut apparaître lors de l'optimisation de caractères contenant de la distorsion. Dans certains cas, un caractère homoplasique par rapport à un cladogramme donné peut avoir deux résolutions possibles (appelées optimisations) de la distorsion d'un caractère. Les optimisations sont : *transformation accélérée* (« *ACCTRAN* » pour « *accelerated transformation* »), *transformation retardée* (« *DELTRAN* » pour « *delayed transformation* ») et *minimisation de la valeur-f* (« *MINF* » pour « *minimization of f-value* »; Farris 1970, 1972; Swofford et Maddison 1987). La figure 2.4 montre un exemple simple d'optimisation ambigüe d'un caractère-partition. Soit un cladogramme (a, (b, (c, d))) et un caractère (a, d) → (b, c). L'optimisation ACCTRAN consiste à faire en sorte que la distorsion des caractères soutienne les nœuds les plus inclusifs possibles; pour notre exemple, cela implique donc d'attribuer l'état apomorphe décrivant b et c à la classe $\{b, c, d\}$ puis l'état plésiomorphe à la classe $\{d\}$. L'optimisation DELTRAN au contraire fait soutenir les nœuds les moins inclusifs possibles; pour notre exemple, cela implique donc d'attribuer l'état apomorphe de façon indépendantes aux deux classes $\{b\}$ et $\{c\}$. On remarque alors que quelle que soit l'optimisation choisie, ce caractère « coûte » deux pas et non un seul comme supposé initialement. L'ambiguïté réside ici dans le fait qu'aucun critère ne permet de choisir une optimisation plutôt que l'autre. En effet, la distorsion, si elle est qualitativement différente entre les deux possibilités, est quantitativement identique puisqu'elle induit la même distorsion d'une classe supplémentaire (deux pas de transformation). Nous avons montré en revanche qu'aucune ambiguïté de placement des états dérivés n'existe en 3ia. Dans cet exemple, on a posé l'hypothèse que b et c sont apparentés en vertu du partage d'un homologue (caractère (a, d, (b, c)). Dans le cas d'un arbre pectiné, b et c sont effectivement apparentés puisqu'il n'y a pas de nœud paralogue, et donc pas d'instance d'indépendance qui permettrait de rejeter l'état (Cao 2008).

2.2. Adéquation empirique

2.2.1. Nullius in verba

La méthode de 3ia, proposée en 1991 par Nelson et Platnick, est née de la volonté de palier aux faiblesses de la méthode de parcimonie par l'utilisation de triplets. Dans cet article, et plusieurs qui suivirent, des exemples ont été présentés comme permettant d'augmenter la précision de l'analyse de parcimonie et proposant de meilleurs résultats. Nous présentons ici l'un de ces exemples publié à compte d'auteur par Nelson (1996) et intitulé « *Nullius in verba* » pour illustrer ces débats à l'origine de la 3ia (Kitching *et al.* 1998). Cet exemple, très simple, nécessite quatre taxons *a*, *b*, *c*, *d* plus un extra-groupe *o* (défini comme un « *all-zero outgroup* »; ce taxon n'est actuellement nécessaire qu'en parcimonie), ainsi que trois caractères (fig. 2.5). Analysons en détail ce que ces caractères donnent comme information phylogénétique. Ici, 0 représente à chaque fois l'état ancestral,

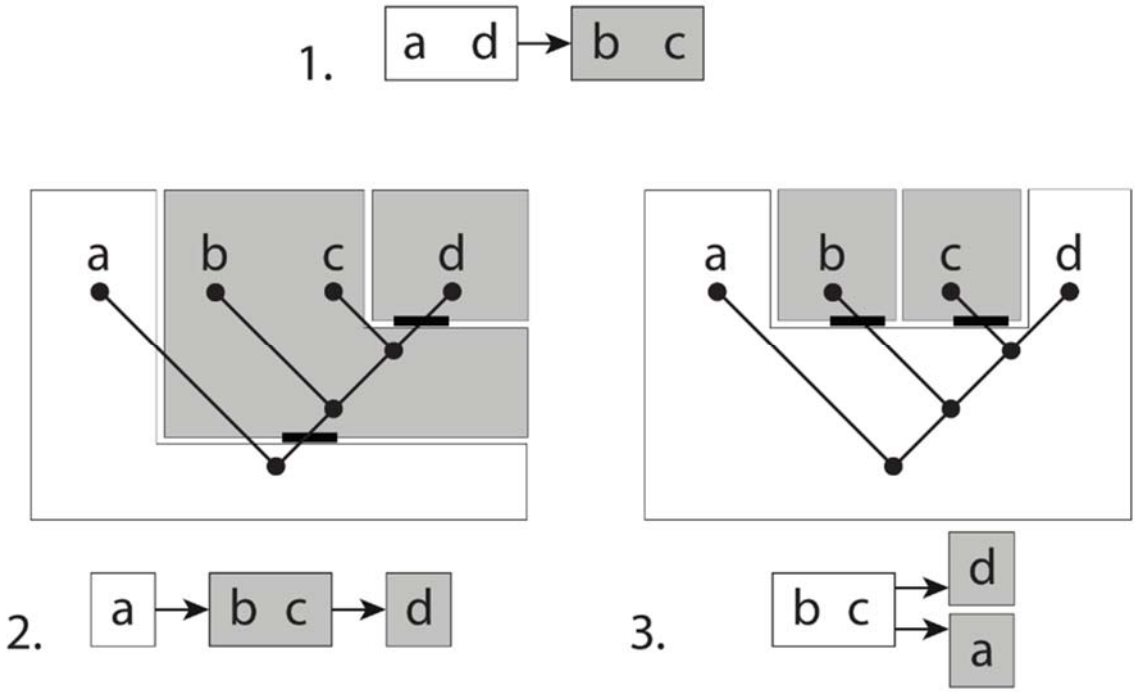


Figure 2.4. Exemple d'optimisation ambigüe en parcimonie d'une partition (*a*, *d*) → (*b*, *c*) sur un arbre : il y a incertitude sur quels clades sont soutenus par le caractère-partition. (2) Optimisation de type *transformation accélérée* (« *ACCTRAN* »). Le clade (*b*, *c*, *d*) est soutenu par l'état dérivé. (3) Optimisation de type *transformation retardée* (« *DELTRAN* »). L'état dérivé ne groupe pas de taxons au sein d'un clade et est non-informatif. Savoir si l'état dérivé (*b*, *c*) permet de former un groupe monophylétique est indécidable.

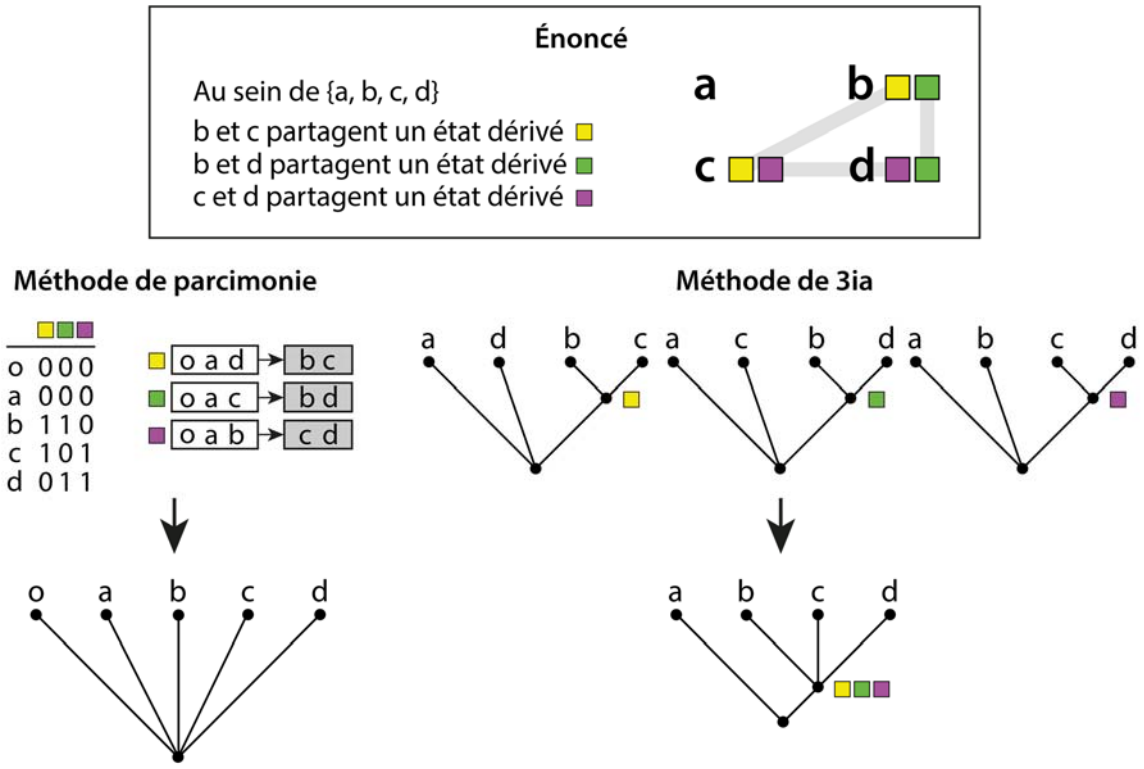


Figure 2.5. Problème énoncé par Nelson en 1996 et sa résolution en 3ia et en parcimonie. A partir de l'énoncé cladistique selon lequel au sein de l'ensemble {*a*, *b*, *c*, *d*}, on peut proposer des hypothèses de regroupement suivantes : {*b*, *c*} (carré jaune), {*b*, *d*} (carré vert), {*c*, *d*} (carré violet). En dessous, codage en matrice (parcimonie) et en hiérarchies (3ia), et consensus strict. Sous la méthode de parcimonie, aucune information non-ambigüe ne se dégage des hypothèses proposées. En 3ia, un clade (*b*, *c*, *d*) est retrouvé de manière non-ambigüe.

et 1 représente à chaque fois l'état dérivé. Dans le caractère 1, l'état dérivé est partagé entre *a* et *b*, à l'exclusion de tous les autres taxons. Ce caractère propose donc l'hypothèse d'une relation de parenté entre *b* et *c*. Le second caractère propose cette fois-ci le regroupement de *b* et *d*. Le dernier caractère propose le regroupement de *c* et de *d*. *a* ne possède jamais d'état dérivé; tous ses états sont ancestraux, à l'identique de l'extra groupe *o*. La question est alors de savoir si une information phylogénétique non-ambigüe est contenue dans cette matrice. En 3ia, la réponse est claire. La matrice est équivalente aux trois caractères suivants : (*a*, *d*, (*b*, *c*)), (*a*, *c*, (*b*, *d*), (*a*, *b*, (*c*, *d*). Un caractère propose de regrouper *b* et *c*, un autre de regrouper *b* et *d*, et un dernier de regrouper *c* et *d*. La solution est donc le regroupement en un clade de *b*, *c* et *d*. *a* ne portant que des états ancestraux, il n'y a jamais aucun argument pour l'inclure dans un clade, et celui-ci se branche à la racine. De l'ensemble des caractères, on peut extraire les triplets *a*, (*b*, *c*), *a*, (*b*, *d*) et *a*, (*c*, *d*) qui

forment la solution la plus congruente (IR = 0,77). Ce résultat apparaît intuitif : en effet, même si le clade {*b*, *c*, *d*} n'est pas soutenu par une synapomorphie partagée par les trois, le résultat est logique par raisonnement transitif. *b*, *c*, *d* partagent des états dérivés en commun que *a* ne possède jamais. Au contraire, la méthode de parcimonie, en optimisant les pas de transformation, donne six arbres : (*o*, *a*, *d*, (*b*, *c*)), (*o*, *a*, *c*, (*b*, *d*), (*o*, *a*, *b*, (*c*, *d*), (*o*, *a*, (*d*, (*b*, *c*))), (*o*, *a*, (*c*, (*b*, *d*))), (*o*, *a*, (*b*, (*c*, *d*))). Le consensus strict est complètement non-informatif, ce qui veut dire qu'aucune information non ambigüe n'émerge de l'analyse de parcimonie. La méthode de parcimonie ignore ici que *b*, *c*, et *d* partagent trois états dérivés deux à deux. Dans l'exemple de Nelson, la multiplication à l'infini de cette matrice donnera d'innombrables états dérivés entre *b*, *c* et *d* deux à deux (fig. 2.6). *a* ne possède jamais d'état dérivé, exactement comme l'extra groupe. La méthode de parcimonie ne trouve jamais de clade non ambigu, au contraire de la 3ia.

2.2.2. Indice de rétention et contenu informatif

L'indice de cohérence permet de mesurer l'augmentation du nombre de pas de transformation par rapport au caractère initial (Kluge et Farris 1969); c'est donc une mesure de la distorsion de la partition initiale sur le cladogramme obtenu. Cet indice présente plusieurs problèmes : (i) il est sensible au nombre de taxons, (ii) au nombre de caractères, (iii) ne peut jamais atteindre une valeur nulle (Archie 1989; Farris 1989; Goloboff 1991; Chang et Kim 1996), et (iv) ne reflète que la distorsion et pas si le caractère permet de soutenir des clades (Zaragüeta i Bagils 2002 : 181-182). Un caractère distordu peut quand même soutenir des clades, ou n'en soutenir aucun si par exemple les pas de transformation se retrouvent autapomorphiques. Par exemple, sur le cladogramme (*a*, (*b*, (*c*, (*d*, (*e*, *f*))))), le caractère (*a*, *b*, *d*, *e*) → (*c*, *f*) ne soutient aucun clade, puisqu'il est optimisé en (*a*, *b*, *d*, *e*) → (*c*) et (*a*, *b*, *d*, *e*) → (*f*), deux autapomorphies ne permettant pas le regroupement de taxons. Au contraire le caractère (*a*, *b*, *d*) → (*c*, *e*, *f*) est optimisé en (*a*, *b*, *d*) → (*c*) et (*a*, *b*, *d*) → (*e*, *f*) et soutient donc le clade (*e*, *f*). Or, l'indice de cohérence est de 0,5 à chaque fois. Il faut donc tenir compte de la distorsion du caractère-partition, mais

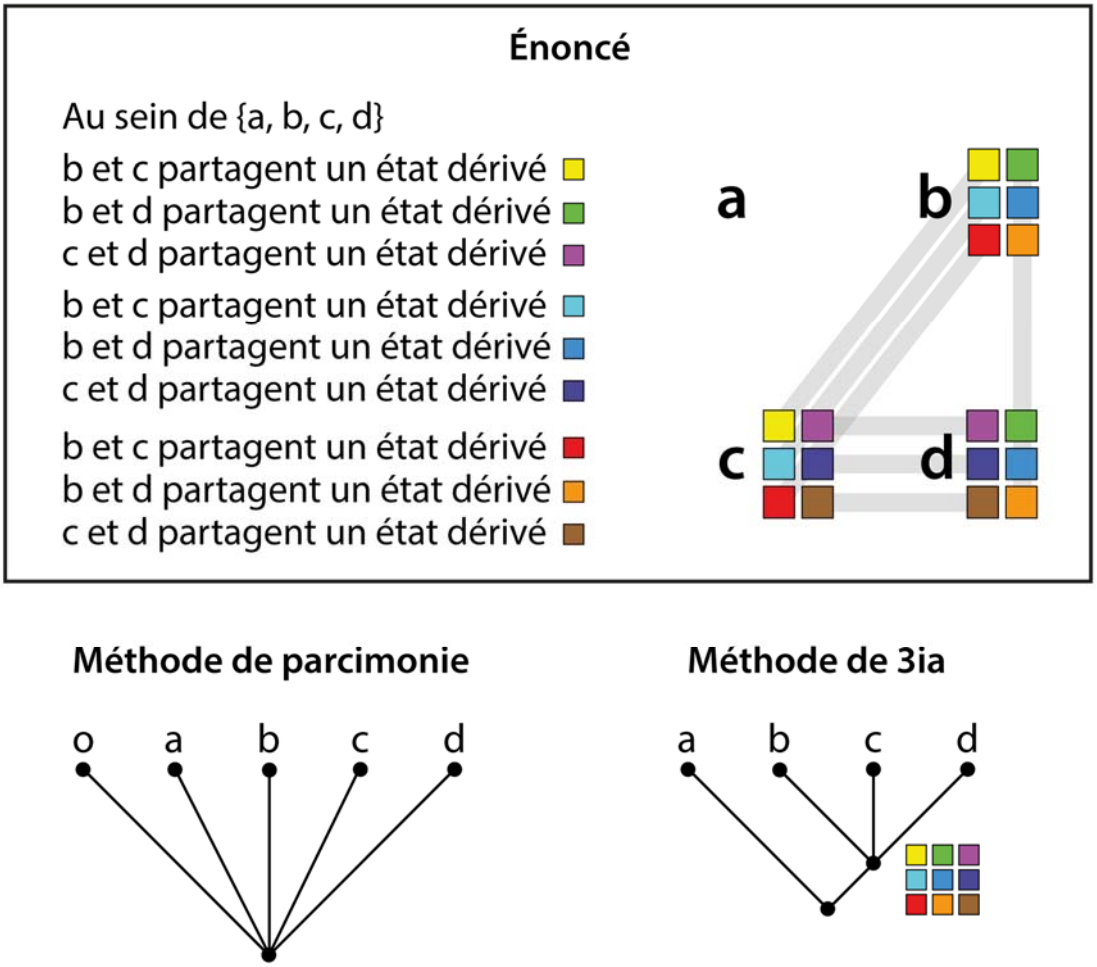


Figure 2.6. Suite du problème énoncé par Nelson en 1996 et sa résolution en 3ia et en parcimonie. Les hypothèses de l'énoncé peuvent être multipliées à l'infini : le taxon *a* ne porte jamais d'état ancestral, et les taxons *b*, *c* et *d* peuvent potentiellement partager deux à deux une infinité d'états dérivés. En dessous, consensus strict relatif à la parcimonie et à la 3ia. Sous la méthode de parcimonie, aucune information non-ambigüe ne se dégage des hypothèses proposées. En 3ia, un clade (*b*, *c*, *d*) est retrouvé de manière non-ambigüe.

également de la répartition de cette distorsion sur l'arbre (« *fit* » selon Farris 1989). L'ambigüité des partitions est telle qu'elle nécessite l'ajout d'un élément supplémentaire, l'ajustement. L'indice de rétention (Archie 1989; Farris 1989) selon Farris mesure la rétention de l'information portée par un caractère sur un cladogramme. En d'autres termes, l'IR mesure la fraction de synapomorphie d'un caractère, c'est-à-dire la part du caractère qui permet de soutenir un ou des clades (Farris 1983). Il se calcule pour un caractère optimisé sur un cladogramme selon la formule suivante :

$$\mathrm{IR}_{\mathrm{parcimonie}} = \frac{\mathrm{G} - \mathrm{L}_{\mathrm{obs}}}{\mathrm{G} - \mathrm{L}_{\mathrm{m}}}$$

Ou *G* est le nombre de pas de transformation minimal optimisé sur un arbre totalement irrésolu (distorsion maximale), L_{obs} est le nombre de pas effectif sur le cladogramme (distorsion réelle) et L_m le nombre de pas minimal que puisse générer le caractère (distorsion minimale). Si nous revenons à l'exemple précédent, le premier caractère possède un indice de rétention nul, puisqu'il ne permet le soutien d'aucun clade. Le second caractère, en revanche, permet de soutenir un clade, malgré de la distorsion : son indice de rétention est de 0,5. L'indice varie bien entre 0 (distorsion minimale du caractère) et 1 (distorsion maximale du caractère). Cependant, la compréhension de cette mesure est difficile, car on ne comprend pas exactement selon sa définition quels sont les objets théoriques manipulés. Quel est le sens de l'ajustement d'un caractère à un arbre plus sa distorsion en termes de synapomorphie et d'homoplasie ? Le but de l'indice de rétention selon Farris est de calculer la fraction de synapomorphie d'un caractère. Cependant, dire que l'ajustement d'un caractère à un arbre plus sa distorsion = la fraction de synapomorphie contenue dans un caractère par rapport à un arbre apparaît difficile à tenir. Nous rappelons que le calcul de l'indice de rétention en 3ia se fait selon la formule :

$$\mathrm{IR}_{3\mathrm{ia}} = \frac{\mathrm{T}_{\mathrm{retenus}}}{\mathrm{T}_{\mathrm{total}}}$$

Ou $T_{retenus}$ est le nombre de triplets indépendants du caractère retenus dans le cladogramme, et T_{total} est le nombre total de triplets indépendants générés par le caractère. L'indice de rétention calcule bien en 3ia la fraction de l'information contenue dans le caractère *retenue* dans le cladogramme (Kitching *et al.* 1998). Ici, le nombre de triplets indépendants du caractère retenus dans le cladogramme = la fraction de synapomorphie contenue dans un caractère. A 0, aucune information provenant du caractère ne participe au soutien du cladogramme (aucun triplet retenu), tandis qu'à 1, l'ensemble de l'information contenue dans le caractère est utilisé pour soutenir le cladogramme (tous les triplets sont retenus)

Certains exemples permettent de juger de la pertinence de l'IR en parcimonie et 3ia. En 3ia, l'IR augmente avec la quantité de triplets retenus ($T_{retenus}$), la parcimonie, elle, peut voir son IR gonflé de manière artéfactuelle. L'IR en parcimonie est sensible de manière

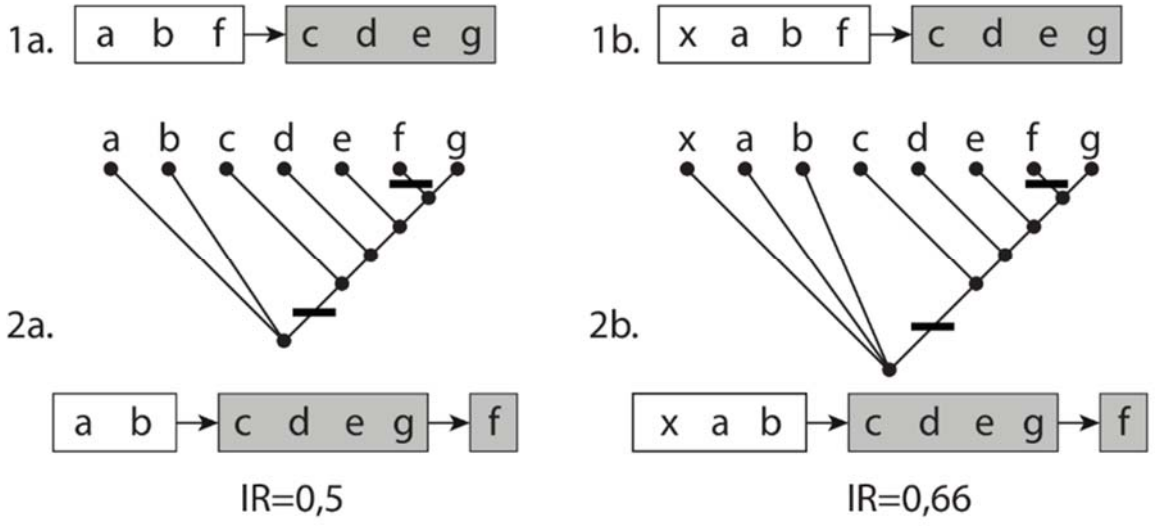


Figure 2.7. Illustration de la sensibilité de l'indice de rétention (IR) en parcimonie au nombre de taxons. L'indice de rétention calcule l'information cladistique du caractère-partition (1) retenu sur un cladogramme (2). (a) Exemple d'un caractère optimisé sur un cladogramme dont l'IR vaut 0,5. Un clade (*c*, *d*, *e*, *f*, *g*) est soutenu. De l'information est cependant perdue à cause de la distorsion du caractère. (b) En rajoutant un taxon *x* porteur de l'état ancestral à la racine, l'IR du caractère augmente et passe à 0,66. Cette augmentation est artificielle puisque l'information retenue ne change pas : le caractère est distordu de la même manière qu'en (a) avec deux pas de transformation, et soutient le même clade (*c*, *d*, *e*, *f*, *g*).

inexplicable au nombre de taxons, comme illustré en figure 2.7. En figure 2.7-1, un caractère se retrouve distordu sur une topologie, mais permet de soutenir le clade (*c*, *d*, *e*, *f*, *g*) : son IR vaut 0,5. En figure 2.7-2, on rajoute un taxon *x* qui vient se brancher à la racine, et qui possède l'état plésiomorphe. Le caractère est distordu de la même manière – deux pas de transformation au lieu d'un seul – et l'ajustement est identique – c'est exactement le même clade (*c*, *d*, *e*, *f*, *g*) qui est soutenu. Cependant, l'IR change, et passe à 0,66. La méthode de parcimonie ne propose aucune explication au comportement de l'indice de rétention dans ce cas particulier; il s'agit donc d'une défaillance de la méthode. Au contraire, l'IR s'explique toujours en 3ia de manière intuitive par une simple proportion de triplets indépendants.

La cohérence de la 3ia vient de son utilisation des triplets, quand les partitions sont systématiquement responsables des ambiguïtés, voire des défaillances techniques de la méthode de parcimonie. De nombreux exemples dans le chapitre précédent illustrent le pouvoir heuristique des triplets, pouvoir que n'ont pas les pas de transformation :

1. En 3ia, sachant deux caractères identiques en termes de nombre de feuilles et de nœuds internes, l'arbre qui présente plus de nœuds paralogues est toujours moins informatif, puisqu'un paralogue est une instance d'indépendance qui fait perdre de l'information sur l'inclusion des états les uns dans les autres – des états disjoints, et donc indépendants, apparaissent (chapitre 1, p. 66). La méthode de parcimonie, en utilisant des partitions, ne permet pas de représenter des hypothèses d'orthologie – il n'y a pas d'équivalence mathématique, l'orthologue est un terme associé aux hiérarchies. De la même manière, elle ne permet pas de différencier paralogues et orthologues au sein des caractères.
2. Au sein de caractères multi-états totalement pectinés, la gestion par PF supprime toute ambigüité aux triplets. Si le caractère comporte des polytomies, une ambiguïté réside (chapitre 1, p. 66). En parcimonie, aucun critère ne permet de choisir entre un caractère multi-états et plusieurs caractères binaires.
3. Nous avons précédemment discuté l'exemple de Pleijel (1995) en montrant (i) qu'il y a un critère objectif pour choisir une représentation des observations plutôt qu'une autre (le triplet) et (ii) que s'il y a plusieurs solutions optimales en termes de contenu informatif, toutes ces solutions sont strictement interchangeables car elles donnent, par définition, la même information (chapitre 1, p. 75; Faure-Brac *et al.* in prep). Au contraire, de nombreux codages en partition ont été proposés dans la littérature (Pleijel 1995; Forey et Kitching 2000). Suivant l'argument précédent, l'absence de critère de choix en parcimonie entraîne une ambiguïté lors de la construction de caractères. Aucun critère ne permet de trancher définitivement en faveur d'une manière de représenter les hypothèses phylogénétiques plutôt qu'une autre (Maddison 1993; Strong et Lipscomb 1999; Brazeau 2011) qui serait meilleure que les autres. Cependant, tous ces codages sont extrêmement différents les uns des autres, ainsi que leur contenu phylogénétique. Il est donc anormal de ne pas pouvoir décider de la meilleure représentation possible.

2.3. Étendue du pouvoir explicatif des méthodes cladistiques

La théorie cladistique possède un pouvoir explicatif qui ne se restreint pas à l'inférence phylogénétique. La reconstruction de l'histoire des aires est un domaine scientifique sur lequel peut s'appliquer la théorie cladistique (Nelson et Platnick 1981). Une aire biogéographique, ou aire d'endémisme, est définie par des UT. Si on considère que les taxons évoluent, et que les taxons sont les constituants des aires biogéographiques, alors les aires évoluent avec ces taxons. Cette évolution doit cependant pouvoir être représentée de manière hiérarchique, signifiant qu'il y a eu au cours du temps des aires biogéographiques qui se sont différenciées en d'autres aires biogéographiques. Les aires biogéographiques définies par des UT ne sont pas à confondre avec des aires géographiques définies par des critères spatiaux. Utiliser des taxons implique que ces taxons aient une histoire commune au niveau biogéographique *en tant qu'aires*. Cette histoire commune peut se retrouver par *maximisation de la congruence*. Si des taxons phylogénétiquement et écologiquement éloignés semblent avoir la même distribution biogéographique, alors le fait qu'ils aient subi un unique évènement est le raisonnement le plus parcimonieux. Un évènement externe est alors à chercher qui puisse s'appliquer à tous ces taxons, et qui produit une barrière divisant une aire en plusieurs – l'aire se différencie. La barrière détectée est extrinsèque aux taxons, comme par exemple une barrière géologique. Pour des taxons terrestres, l'ouverture de l'océan Atlantique au Mésozoïque peut jouer le rôle de barrière, ou encore la fermeture de l'Isthme du Panama au Néogène pour des taxons marins. Plusieurs taxons subissent le même évènement biogéographique (diversification d'aires), générant de la congruence dans leur répartition. Cette congruence est le reflet que « *la terre et la vie évoluent ensemble* » (Croizat 1958).

On appelle *biogéographie cladistique* la recherche de relations de degré de parenté entre aires biogéographiques (Nelson et Platnick 1981). Les *Unités Biogéographiques* (UB) sont les objets à classer de la biogéographie. En phylogénie, on pose des hypothèses sur des parties d'UT pour construire des homologies. Les homologies sont ensuite standardisées en caractères où les parties d'UT sont remplacées par leurs UT respectives. En biogéographie cladistique, les objets à classer sont les UB. Une UB comme une UT se

définit *a priori* d'une analyse. Une UB est une aire biogéographique, c'est-à-dire une aire d'endémisme. Une aire d'endémisme correspond aux limites de distribution congruentes et non-aléatoires de taxons différents (Cecca et Zaragüeta i Bagils 2015, d'après Platnick 1991). Si l'UT est définie par des homologues autapomorphiques, l'UB est définie par des UT. Une UT est donc une autapomorphie en biogéographie : elle permet de définir les objets à classer a priori de l'analyse. Une UB minimale est nécessairement définie par une ou plusieurs UTM. La monophylie d'une UB est également nécessaire. Elle est dépendante dans ce cas de la monophylie présumée des UT. Plus largement, dans le cadre de la théorie cladistique, ce que l'on sait à propos des aires biogéographiques provient uniquement de ce que l'on sait à propos des taxons de ces aires. Par exemple, on peut construire une UB à partir de deux UTM *a* et *b* dont la distribution spatiale est congruente. L'extension de cette UBM est l'ensemble d'UT $\{a, b\}$.

Les aires biogéographiques étant des entités monophylétiques, leur existence rentre dans le domaine de la théorie cladistique. Une méthode d'analyse cladistique cohérente doit pouvoir construire des cladogrammes à partir de n'importe quel type d'objets cladistiques. Dans les deux parties suivantes, nous montrerons de manière détaillée comment fonctionne les méthodes de 3ia et de parcimonie pour la construction de cladogrammes d'aires biogéographiques.

2.3.1. Analyse à trois aires

La biogéographie utilise les cladogrammes comme supports de l'analyse. Une UT est ici considérée comme une partie d'UB. Une analyse de biogéographie en 3ia, ou 3aa pour *three-area analysis* (à opposer à la 3ta pour *three-taxon analysis*), est donc une méta-analyse, c'est-à-dire une analyse cladistique fondée sur d'autres analyses cladistiques : on va se servir de plusieurs cladogrammes résultant d'analyses phylogénétiques pour construire une analyse de biogéographie. Les cladogrammes sont choisis en fonction des aires à étudier. Tous les cladogrammes ne sont pas adaptés à une analyse donnée. Par exemple, des taxons marins ne peuvent pas être utilisés pour inférer des relations entre aires biogéographiques continentales.

Chaque cladogramme est ensuite standardisé en un caractère biogéographique. Les UT sont remplacées par les UB qui y sont associées. La figure 2.8-1 montre un exemple de

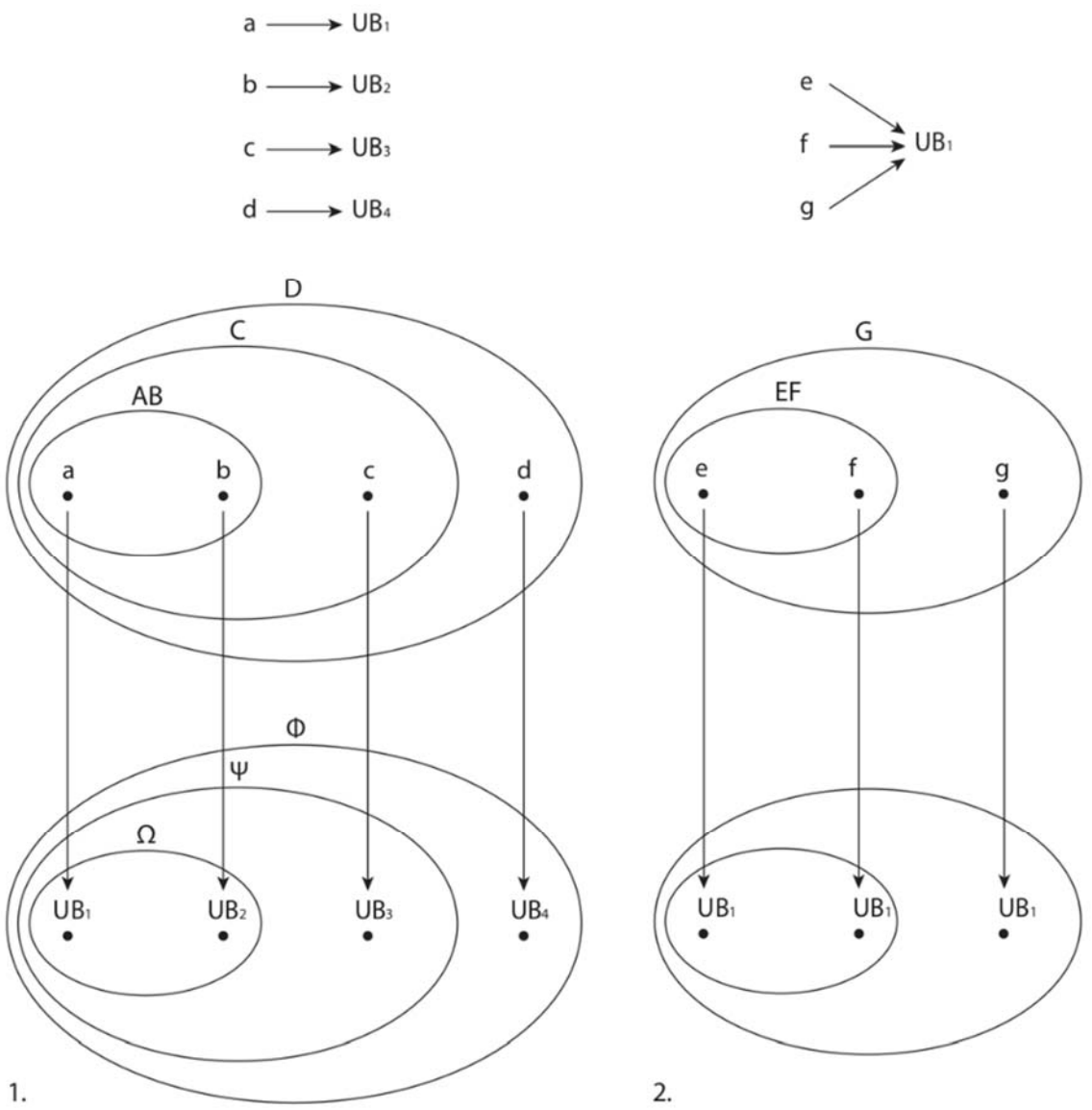


Figure 2.8. Standardisation en biogéographie. En haut, correspondance entre UT (*a*, *b*, *c*, *d*, *e*, *f*, *g*) et UB (UB_1, UB_2, UB_3, UB_4). On remplace sur un cladogramme de taxons les UT par les UB, ce qui forme un caractère biogéographique. Le caractère biogéographique pose des hypothèses de clades d'UB. (1) Standardisation complète. (2) Standardisation non informative due à une répétition de terminaux ne permettant pas la génération d'un caractère bioghéographique. On constate l'identité totale de la procédure entre phylogénie et biogéographie.

standardisation d'un cladogramme en caractère biogéographique. On a au départ un cladogramme d'UT (*d*, (*c*, (*b*, *a*))). Les UT *a*, *b*, *c* et *d* sont remplacées par les UB qui leur correspondent, respectivement UB_1, UB_2, UB_3 et UB_4. Par exemple, l'UT *a* est un des constituants de l'UB_1. Le caractère biogéographique (UB_4, (UB_3, (UB_2, UB_1))) exprime une hypothèse de relation de parenté entre UB. Le clade {UB_1, UB_2} (ici en tant qu'aire

inclusive monophylétique) signifie que UB_1 et UB_2 sont apparentés par rapport à UB_3 et UB_4. Biogéographiquement, on postule que ces UB formaient historiquement une AH Ω.

La standardisation peut parfois occasionner des pertes d'information due à l'homonomie. C'est le cas lorsque plusieurs UT proviennent d'une même UB, responsable du « *polymorphisme biogéographique* » (parfois appelé paralogie biogéographique). Le second exemple de la figure 2.8 montre un cas ou trois UT *e*, *f* et *g* proviennent toutes trois de la même UB. Dans ce cas, aucune relation entre UB ne peut être spécifiée. Cette répétition d'UB peut être due à une diversification de taxons sans diversification des aires qui les concernent. La différenciation voire la diversification de taxons au sein d'une aire unique ne permet d'inférer aucune parenté entre aires. La répétition des UB doit donc être supprimée de la même manière que pour la répétition des UT en phylogénétique. La standardisation peut donc être complète – à chaque UT correspond une UB – ou incomplète – plusieurs UT correspondant à la même UB. En biogéographie, il est également possible de se retrouver avec une UT assignable à plusieurs UB. Ce problème est celui des Taxons à Large Répartition ou TLR (en anglais, MAST pour « *Multiple Areas on a Single Terminal* ») et correspond à une UT à large répartition. Il n'y a cependant pas actuellement de consensus sur la manière de gérer les problèmes des TLR en 3ia (Nelson et Platnick 1981; Zandee et Ross 1987; Ebach et Humphries 2002; Ebach *et al.* 2005; Cecca et Zaragüeta i Bagils 2015).

L'aréagramme représente la solution au problème de l'inférence de relations de parenté entre UB. La figure 2.9 reprend les étapes de la 3ia et la correspondance entre phylogénie et biogéographie. Une fois les caractères biogéographiques construits à partir de phylogénies, les relations de parenté entre UB sont décomposées en triplets et recomposées de manière à obtenir l'aréagramme contenant le plus grand nombre de triplets congruents. Le corolaire est que l'incongruence va générer le rejet de triplets et/ou d'états de caractères (précisons que si rejeter des triplets n'implique pas de rejeter un état de caractère, l'inverse se vérifie toujours). Une des plus importantes sources d'erreur en biogéographie cladistique est liée au phénomène de dispersion. On rappelle que la congruence arrive en biogéographie lorsqu'un phénomène *extrinsèque* modifie les aires de plusieurs UT à la fois, comme la vicariance. La dispersion est le phénomène de modification de l'aire d'une unique UT. Les facteurs de la dispersion sont *intrinsèques* aux

taxons. Ils peuvent être par exemple fonctionnels, biologiques ou écologiques. Le mode de déplacement des organismes, la dispersion des larves, la durée de vie des juvéniles ou des adultes, la capacité à se propager dans un écosystème sont autant de faits générant de l'homoplasie biogéographique, et donc de la polyphylie. Ces faits et leur interprétation pour expliquer les phénomènes de dispersion se situent en dehors de la théorie cladistique. De même, le centre d'origine d'un taxon, c'est-à-dire une aire mais pas les aires qui en descendent, est une notion associée à la dispersion, et est par définition toujours paraphylétique. Cette notion ne fait donc pas partie de la biogéographie cladistique. Une aire biogéographique inclut toujours les aires qui se différencient à partir d'elle. Plus généralement, la biogéographie cladistique traite de l'histoire des aires biogéographiques et des taxons qui les caractérisent, mais pas de l'histoire des taxons individuellement.

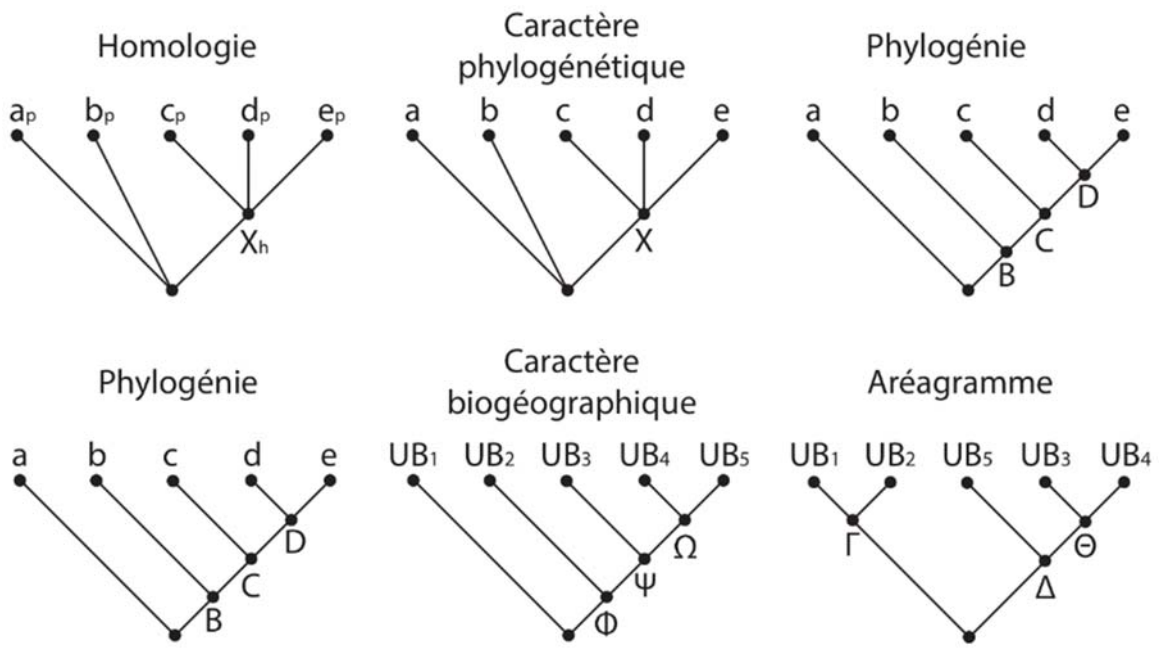


Figure 2.9. Correspondance entre étapes de 3ta et étapes de 3aa. L'homologie présente des hypothèses de parenté entre parties d'UT. L'homologie est standardisée en caractère ou les parties d'UT sont remplacées par les UT elles-mêmes. L'ensemble des caractères est décomposé en triplets. La recombinaison de l'ensemble de congruence maximale permet de générer la phylogénie, cladogramme de taxons. (3aa) Le cladogramme présente des hypothèses de parenté entre UT. Le cladogramme est standardisée en caractère biogéographique ou les UT sont remplacées par les UB. L'ensemble des caractères biogéographiques est décomposé en triplets. La recombinaison de l'ensemble de congruence maximale permet de générer l'aréagramme, cladogramme d'aires.

2.4. Analyse de parcimonie de Brooks

Notre argumentaire nécessite de comparer les méthodes de 3ia et de parcimonie sans modification de leur fonctionnement, afin de comprendre leur comportement lorsque le seul paramètre modifié est la nature des objets à classer. En l'occurrence, la méthode de parcimonie de Wagner appliquée à la biogéographie cladistique s'appelle l'*Analyse de Parcimonie de Brooks* (APB; Wiley 1986, 1988a, 1988b; Brooks 1990; Brooks *et al.* 2001). La méthode d'APB est expliquée sur la figure 2.10, et permet d'introduire le premier problème, qui a déjà été discuté dans la partie cohérence théorique et se retrouve donc logiquement ici, avec une ampleur encore plus importante. Encore une fois, la source des connaissances en biogéographie cladistique sont les cladogrammes. Les cladogrammes, quels que soient la méthode, doivent être convertis en hypothèses de parenté entre aires : la source de la connaissance de ces relations étant les cladogrammes, les caractères biogéographiques sont intrinsèquement hiérarchiques (fig. 2.9 et 10. Or, la méthode APB nécessite d'avoir des partitions. Pour pouvoir réaliser une analyse de parcimonie sur des aires, on va donc devoir transformer les caractères hiérarchiques en partitions afin de les coder dans des matrices qui seront lues par les algorithmes de parcimonie (fig. 2.10). Il

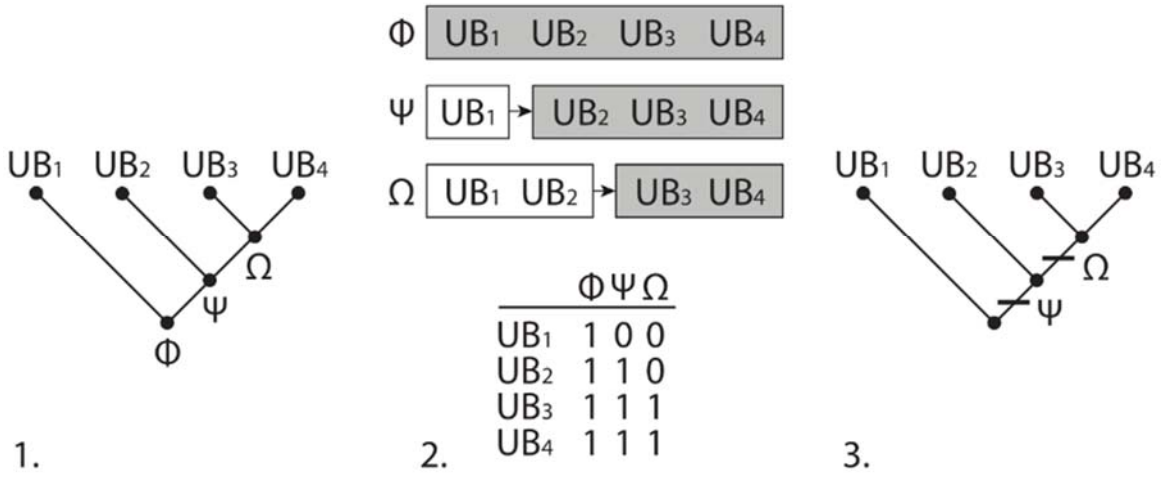


Figure 2.10. Illustration de la méthode de parcimonie en analyse biogéographique (Analyse de Parcimonie de Brooks). (1) La source des hypothèses sur les aires biogéographiques provient de la connaissance des relations de parenté entre taxons. (2) Le caractère biogéographique, intrinsèquement hiérarchique du fait de son origine, est transformé en caractère-partition. (3) Les partitions sont optimisées sur un arbre. Polarisé, l'arbre devient un aréagramme. L'étape (2) modifie le contenu sémantique des hypothèses biogéographiques.

apparaît évident ici que les hypothèses de départ, par exemple (UB_1, UB_2, (UB_3, UB_4)), verront leur contenu sémantique lors du passage à la partition (UB_1, UB_2) → (UB_3, UB_4). (UB_1, UB_2, ((UB_3, UB_4), (UB_5, UB_6))) constitue un exemple encore plus flagrant car la classe (UB_3, UB_4, UB_5, UB_6) étant sans instances immédiates, la seule classe qu'on peut lui associé dans la partition ordonnée correspondante est l'ensemble vide, ce qui est impossible selon la définition d'une partition (pas de classe vide). Les partitions sont ensuite optimisées sur un aréagramme. Là encore, à partir de partitions orientées, on obtient une structure hiérarchique. Cette structure hiérarchique n'est pas équivalente aux partitions, seulement à d'autres structures hiérarchiques, comme les caractères initiaux (fig. 2.10). Le codage de la matrice apporte donc une forte distorsion des hypothèses initiales en modifiant de manière infondée – car rien dans la théorie cladistique ne permet les partitions ordonnées qui contiennent intrinsèquement des classes ancestrales paraphylétiques – leur structure mathématique.

Les partitions, en plus d'être injustifiées, vont apporter toute l'ambigüité inhérente à leur structure dans le cadre de la méthode cladistique dont nous avons apporté les arguments au sujet des taxons. On reprendra pour illustrer ces défaillances le problème des réversions (chapitre 1, p. 94). Le résultat d'une APB peut être un aréagramme dont certains clades seront soutenus par des convergences, mais également par des réversions, de la même manière qu'en parcimonie de Wagner standard. Il s'ensuit qu'un clade d'aires peut être soutenu par la perte de la présence d'un taxon sur l'aire. Le fait que l'APB permette le soutien d'un clade par l'absence de taxons permet de mettre en évidence pourquoi certains auteurs ont critiqué les méthodes à matrices – sous-entendu les méthodes utilisant les partitions – comme étant intrinsèquement des procédures phénétiques (Williams et Ebach 2006), au sens où elles peuvent grouper par l'absence (et plus généralement par l'état plésiomorphe). Puisqu'il n'y a aucune différence intrinsèque faite par la théorie entre taxons et aires, la critique du regroupement par l'absence s'applique autant à la reconstruction phylogénétique qu'à la biogéographie cladistique.

3. Experimental systematics: sensitivity of cladistic methods to polarization and character ordering schemes

Cette partie correspond à un article publié dans la revue *Contributions to Zoology*. L'ensemble du texte, ainsi que les annexes notées de S1 à S8, sont disponibles à l'adresse ci-dessous.

Rineau, V., Grand, A., Zaragüeta, R. & Laurin, M. (2015). Experimental systematics: sensitivity of cladistic methods to polarization and character ordering schemes. *Contributions to Zoology*, **84**(2), 129-148. http://www.ctoz.nl/vol84/nr02/a03.

3.1. Introduction

The reconstruction of the Tree of Life is one of the most important quests in natural sciences, and it is essential for various biological and geological fields. Phylogenies are widely used in comparative and evolutionary biology (Hennig, 1966; Felsenstein, 1985) and even in conservation biology (Faith, 1992). Even though most contemporary phylogenetic studies of extant taxa are based on DNA sequence characters, phenotypic traits continue to play an important role because molecular sequences are unavailable for most long-extinct organisms, given that the oldest 'ancient' DNA extracted so far is less than 1 Ma old (Callaway, 2011). According to various estimates of extant and extinct biodiversity, and depending on how species are conceptualized, approximately 99% to 99.9% of life forms that have existed on Earth are now extinct (Miller and Foote, 1996; Newman and Sibani, 1999; Alroy, 2001). Thus, it is important to continue using and improving methods of phylogenetic inference from phenotypic data. The dominant phylogenetic theory for phenotypic characters is cladistics. This theory is based on the assumptions that the epistemic access we have to discover monophyletic taxa is through phylogenetic arguments called characters. Characters are based on hypotheses of homology, *i.e.* hypotheses that relate the observation of different features as being the same. This identity is explained by their evolutionary origin as parts evolved from the same part

in the exclusive (last) common ancestor. The method of argumentation is different in parsimony analysis and 3ta. In parsimony analysis, characters-states transform into one another, which is consistent with Hennig's original description of cladistic theory. In 3ta, character-states differentiate from one another, which is consistent with the relationships of taxa represented as hierarchies: if homologs are parts of taxa and taxa differentiate, then character-states should also differentiate. Theoretically, the criterion of choice among contradictory hypotheses is maximizing congruence. Congruence has been interpreted as minimizing a particular tree-distance (in parsimony method) or as maximizing compatibility among characters (as in the compatibility method; not tested in this paper). In 3ta, the search for optimal trees maximising parsimony or compatibility of triplets gives exactly the same results (Wilkinson, 1994b): this shows that the treatment of characters is contained in the method of treatment of homology and homoplasy of 3ta and parsimony, and that the representation of characters as matrices, graphs or hierarchies is only a consequence of this treatment.

According to Wiens (Chappill, 1989; Stevens, 1991; Thiele, 1993; Wiens, 2001), most morphological characters are fundamentally quantitative. Continuous characters are usually discretized because most cladistic software cannot handle them directly. Several methods have been proposed for discretizing characters, such as segment coding (Simon, 1983; Farris, 1990) and gap coding (Mickevich and Johnson, 1976; Almeida and Bisby, 1984; Archie, 1985) to limit the arbitrariness of the state delimitation, even though these do not usually result in capture of all the phylogenetic information (Laurin and Germain, 2011) unless ordering schemes are used to order the series of states or set transition costs, and a distinct state is recognized for each taxon (Wilkinson, 1992; Wiens, 2001). Thus, the coding, as a formalization of character hypothesis of the analysis (ordination, discretization, etc…) is of primary importance for phylogenetic inference.

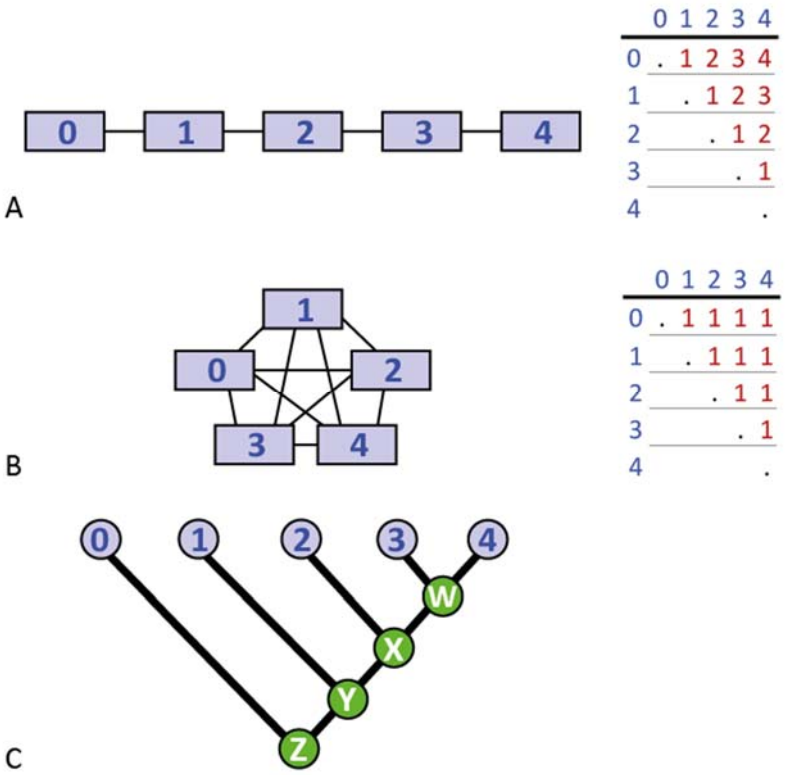


Figure 2.11. Three examples of character representation: unordered in parsimony (A; non-additive) and ordered in parsimony (B; additive) form the class of parsimony characters, associated with their stepmatrix (Sankoff and Rousseau, 1975) necessary in a matrix representation, and the other classes of character trees, hierarchies treated in 3ta (C). A symbolizes unordered character states, were all transformations between states have the same cost. B represents ordered states as an unrooted tree and the equivalent stepmatrix, and C represents a character tree, were nodes are homology hypotheses under the hierarchy. The differences in their representations (graph, unrooted tree, hierarchy) traduce both their content and the way in which corresponding methods treat them (how they deal with incongruence they can generate; for example, in parsimony graphs A and B, a step can correspond to transformation from 1 to 2 or from 2 to 1, which make no sense in the hierarchical representation C, symbolizing a 3ta treatment). Letters and numbers (0-5, Z-W) represent named nodes as homologs.

Characters can be separated into at least two classes (fig. 11). The first class comprises parsimony characters, treated as partitioned graphs, oriented or not, with cycles or not. Figure 2.11 shows to 'extreme' cases whose will be discussed subsequently, as unordered characters (fig. 2.11-A), which imply equal transition cost between all states (also called 'maximally connected'; Slowinski, 1993), and ordered characters (fig. 2.11-B), in which relationships between states are typically depicted by linear series (as 'minimally connected' characters in Slowinski's terminology) or unrooted trees,. Ordered characters were firstly labelled as 'additive characters' because the number of steps are added with

the number of transformations (Fitch, 1971). The second class of character is hierarchies (Cao *et al.*, 2007), where inclusion relationships between states can be represented by rooted trees, which are used in 3ta as implemented in LisBeth (Zaragüeta *et al.*, 2012). The example in figure 2.11-C is close to the ordered character shown in figure 2.11-B: it is a completely dichotomous hierarchy, with the maximum number of internal nodes for the tree. The results of phylogenetic analysis of quantitative characters, as well as of qualitative characters, is impacted upon by the coding of characters (Laurin and Germain, 2011). Knowing that the analysis of different characters can yield different results (*e.g.*, a different number of optimal trees, different node support by synapomorphies or bootstrapping, different topologies, different retention indices, etc…), differences in coding approaches may hamper some comparisons between trees obtained by various studies (Hauser and Presch, 1991). The advantages and drawbacks of the various ordering schemes have not received as much attention by systematists as they deserve (Wilkinson, 1992; Brazeau, 2011). The assumptions of ordered (*e.g.*, with the criteria of ontogeny or similarity) and unordered parsimony (Hauser and Presch, 1991), and of searching trees with the fewest transformations are widely known and need not be repeated here. Slowinski (1993) noted that in the 1970s and early 1980s, most multistate characters were ordered, with some authors viewing unordered states as 'agnostic' or 'nihilistic' (Mickevich, 1982). Since then, the situation has reversed and ordered characters have been used only in a minority of recent studies (Grand *et al.*, 2013).

The assumptions underlying the three-item analysis (hereafter, 3ta) are less well known and probably need to be presented briefly. Initially described as a 'more precise use of parsimony' (Nelson and Platnick, 1991), the three-taxon analysis can be understood as a completely different method (Nelson *et al.*, 2003; Cao *et al.*, 2007; Zaragüeta and Bourdon, 2007; Zaragüeta *et al.*, 2012). The method is available in LisBeth (Zaragüeta *et al.*, 2012), but can be performed using any parsimony software after the parsimony matrix has been converted into a triplet matrix, an operation that is automated in LisBeth. Within 3ta, characters are represented as hierarchies (Cao *et al.*, 2007). The phylogenetic trees are reconstructed after the hierarchical characters are treated as sets of three-item statements (3ts, triplets herein). Triplets (*e.g.*, (A, B) C)) are statements about the phylogenetic affinities between three taxa (*e.g.*, A is more closely related to B than to C). The

reconstructed cladograms are those that contain the largest subsets of compatible triplets; two triplets are incompatible if the three terminals considered are identical, but the relationship is different: *i.e.* (A(BC) and (B(AC)). Systematists generally store the character states in matrices; each parsimony-state in 3ta is placed in a terminal node of a hierarchy (Cao *et al.*, 2007; Zaragüeta and Bourdon, 2007) (fig. 2.11-C), which represents the character. The plesiomorphic state is also defined, as the root. This representation differs from the step matrices and their assumptions of 'proximity' (fig. 2.11-A, 2.11-B; their pattern of transformation, and the transformation costs). Some authors consider that hierarchical representation is a more relevant and intuitive way to express homologies (Williams and Ebach, 2006; Cao *et al.*, 2007) in a cladistic analysis, where hierarchies of homologs thus leads to a hierarchy of taxa. The relationship homolog/taxa is thus equivalent to the relationship character/cladogram: the first is the part and the second is the whole.

3ta involves a second analytical stage that has sometimes been misunderstood. Characters, *i.e.* hierarchies of character-states, may be converted into triplets, *i.e.* minimal hierarchies relating two terminals that are closer to each other than any is to a third. This decomposition may imply some redundancy, *i.e.* there are more triplets produced than necessary to recompose the character. In these cases, Nelson and Ladiges (1992) proposed what they called a fractional weighting (FW hereafter). FW quantifies the contribution of each triplet to the character. In the example shown in figure 2.12-A, only two triplets out of three in each group of three triplets are necessary because if taxon C is closer to D than to A, and if C is closer to E than to A, it follows that D is closer to E. Hence, the third triplet (shown in grey) is redundant and could be omitted. However, in the absence of external criteria, it is impossible to objectively determine which triplet is redundant (in the above example, any of the two other triplets could be omitted, instead of the one in grey). FW quantifies the contribution of each triplet to the character. Triplets are not characters, contrary to what has often been stated in the literature (Kluge and Wolf, 1993). Thus, even when using parsimony software, the constraint of the logical independence (Wilkinson, 1995) of the different columns of the matrix does not apply to the different triplets derived from the same character. On the contrary, FW reflects the

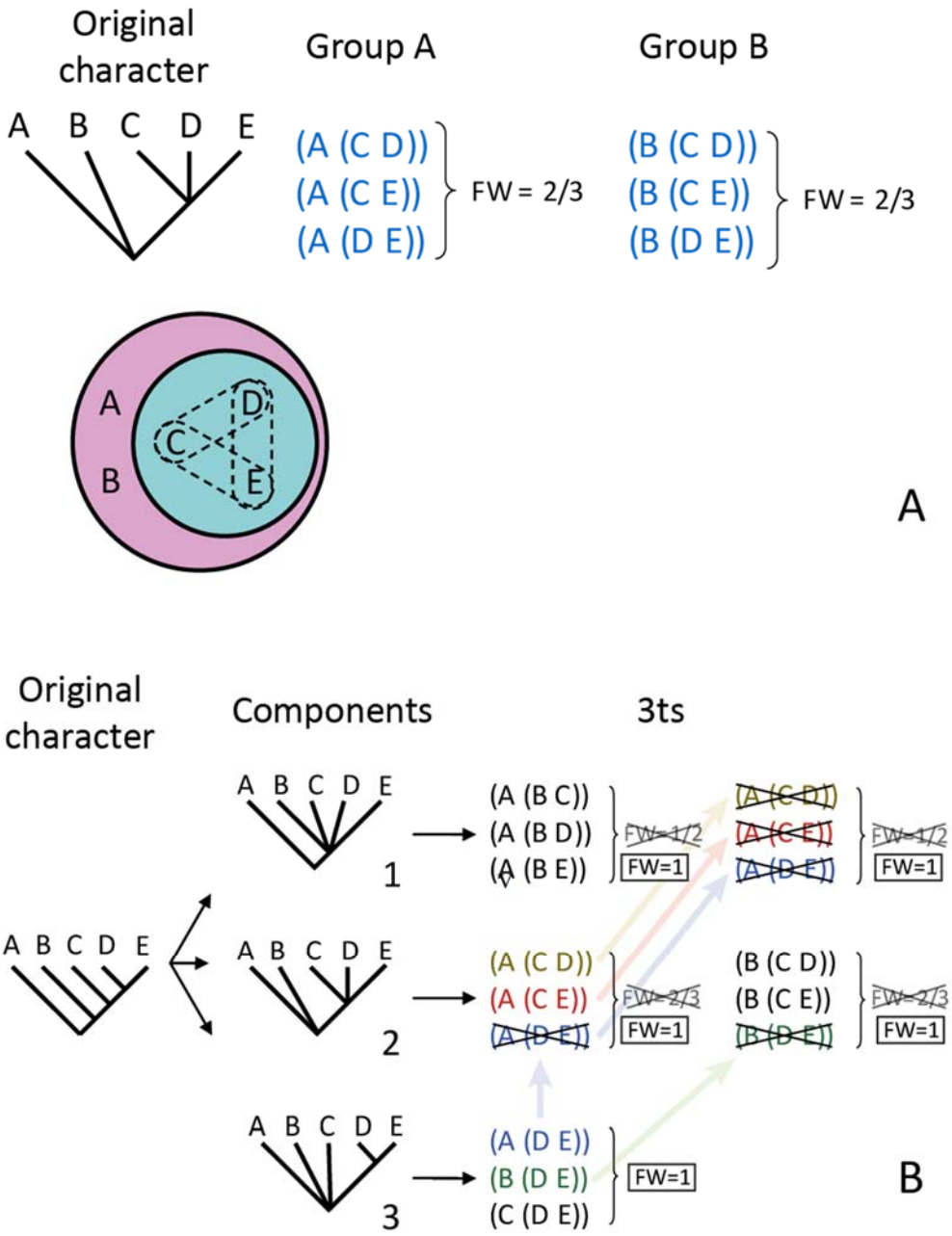


Figure 2.12. Fractional weighting (FW) and dependence among triplets derived from the same character. A. A character relating five terminal taxa is decomposed into its elementary relationships, expressed as a list of triplets. The decomposition results into a list of six triplets. However, only two of the three triplets [*e.g.* (A (C D)) and (A (C E)) that relate three terminals (C D, E) to other, more distant ones (*e.g.* A] are necessary to build the initial hierarchical character, as shown in the set diagram. FW assigns an information content/weight of 2/3 to each triplet of the two groups. B. A multi-state character is analyzed into components and then into triplets. 3ta takes into account the fact that each component is not independent. Thus, the triplets derived from different components of the same character may be redundant. Consider the triplets (A (D E)): it appears in the three components. However, it represents a single node in the original character. The triplets derived from the third component have a FW=1, *i.e.* they are non-redundant. The same triplets in the first and second components are thus redundant and can be removed. The FW of the remaining triplets of the component is then recalculated from FW=2/3 to FW=1 in the second component, and FW<1 in the first. Etc… Note that the matrix treatment of the triplets is immaterial with their status of dependent or independent, contrary to characters in parsimony analysis.

logical mutual dependency (Williams and Siebert, 2000), as shown on figure 2.12-B (but see Wilkinson *et al.*, 2004).

With perfectly congruent data, parsimony and 3ta give always the same tree. Operationally, the only difference between them is how they deal with incongruence in the data (*e.g.* in Nelson *et al.*, 2003). The history of wings in stick insects (phasms) highlighted by Whiting *et al.* (2003) can illustrate the differences and similarities between parsimony and 3ta (fig. 2.13). The cladogram of phasms shows evidence for many losses and independent reacquisitions of a complex structure (fig. 2.13A), the pterygote wing (fig. 2.13B). In figure 2.13C, this hierarchy (character tree drawn a posteriori from the analysis) is simplified so as to illustrate the different points of view. In parsimony analysis, the character 'wing' is traditionally coded as binary with the states absent and present. In 3ta, the character is coded as an inclusive hierarchy with the state present nested within the state absent (fig. 2.13A). The states 'absent' and 'present' are both considered as complementary homologues (de Pinna, 1991) by parsimony proponents whereas only the acquisition is considered to be a homologue by the 3ta proponents (Cao *et al.*, 2007): the class 'absence' includes 'everything else' and is not considered as an homologue (the plesiomorphic state in 3ta). 3ta proponents consider 'homologs' the nodes of a character state tree (fig. 2.11-C, 2.12-C). For instance, in figure 2.11-C, W is a homolog (*i.e.*, states 3 and 4 are considered homologous), X is another homologue (*i.e.*, 2, 3 and 4 are homologous), etc. The character tree shows 'polarized' (nested) homologies. With the practical example of our figure 2.13 (Whiting *et al.*, 2003), all wings can be considered homologous *at a specific level of inclusiveness* (fig. 2.13-C, first appearance of state 1, represented as slightly above the root). In parsimony, this binary character supports six clades (on figure 2.13-C, six classes are synapomorphic: 1, 0*, 1*, 0**, 1**, 1****, and one class is autapomorphic: 1***), whereas in 3ta, the character is one homology, accepted by the analysis, and support only one clade (in figure 2.13-A, the sister group of *Cancer pagurus,* the class 1 in figure 2.13-C, is synapomorphic). In 3ta, only the primary homology hypotheses proposed by the systematist are tested: the hypothesis was that the wings where homologous. Either the character is accepted as secondary homology or it is rejected as a hole. No new hypotheses generated by the algorithm are considered, especially homoplasies, *i.e.*, 'synapomorphic' reversals with multiple reappearances are not considered as evidence for support of clades.

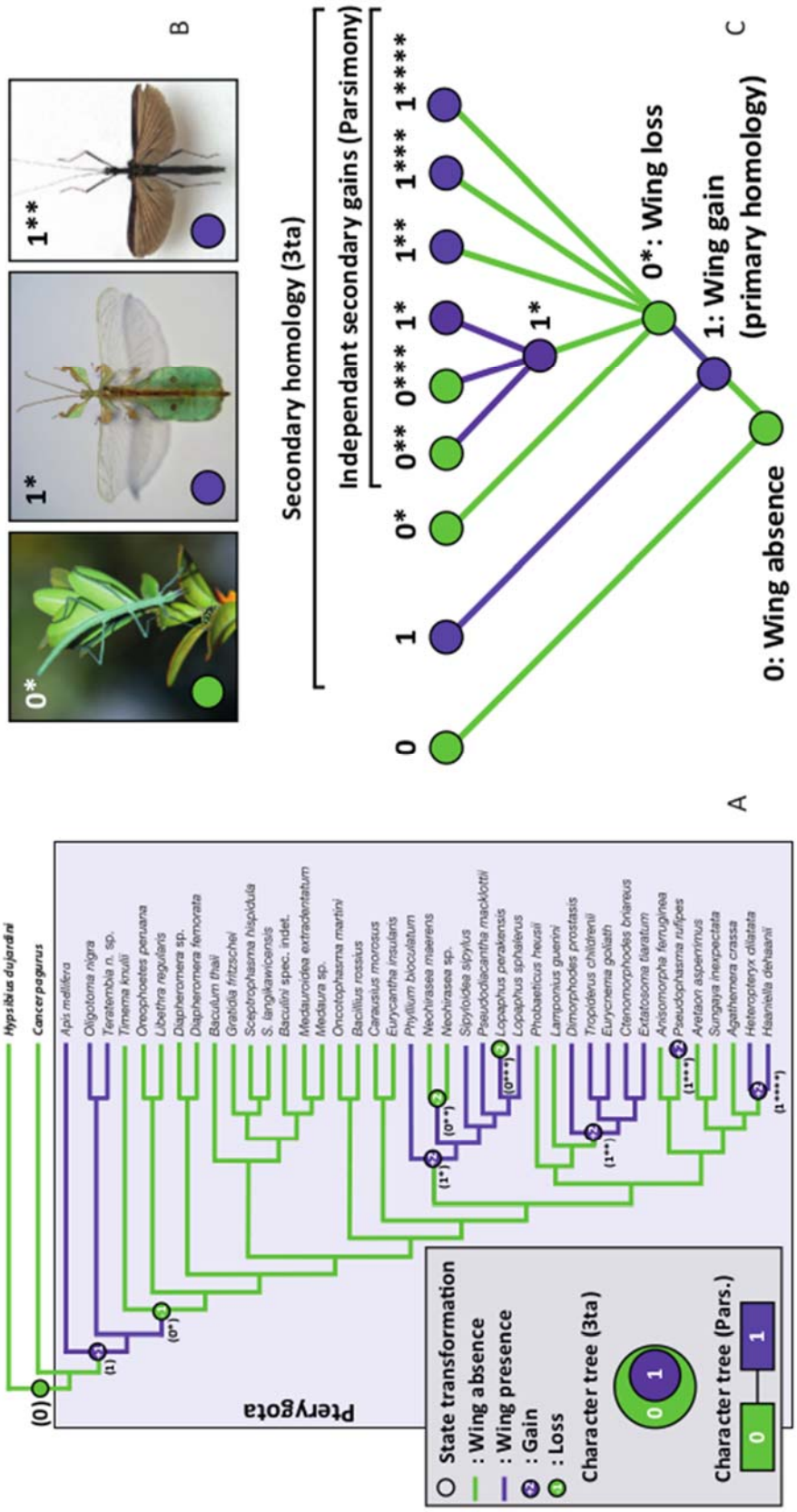


Figure 2.13. An example of character history according to cladistics. A is a modified phylogeny from Whiting *et al.* (2003) with transformations (obtained using parsimony under acctran optimization) including reversals (denoted by a '-' sign) and acquisitions or reacquisitions ('+') from a binary character absence/presence (the 3ta character tree is represented as a hierarchy). The degree of gains and losses are denoted by numbers (1 for primary, 2 for secondary, etc.) in the circles placed at the corresponding nodes. The degree of the resulting state (presence or absence, appearing in parentheses below and to the left of the symbols for gains or losses) is denoted by the asterisks; primary gains or losses have none; secondary gains or losses have one; ternary ones have two, etc. B shows the wingless Phasmatodea *Leptynia hispanica* (Bolivar, 1878), *Phyllium bioculatum* (Gray, 1832) with wings and *Pseudophasma acanthonotum* (Redtenbacher, 1906) with wings and *Pseudophasma acanthonotum* (Redtenbacher, 1906) with convergently re-acquired wings (copyrights, respectively: Fritz Geller-Grimm and Felix Grimm - CC BY-SA 3.0, Drägus - GFDL and Drägus – PD-self). C represents with a character tree the history reconstructed a posteriori from the analysis (with the mapping of the character on the optimal tree) of the character 'wing' of phasms. Each node represent a parsimony apomorphy, contrary to the 3ta for which the only apomorphy is the node 1 (as only secondary homology because the only primary homology postulated was the homology between all wings).

The algorithm can reject (or not) the proposed hypothesis, given the general framework. In this case, the hypothesis that all phasm wings are just pterygote wings cannot be rejected. The generation by the algorithm of new hypotheses about multiple transformations, appearances and disappearances, is considered spurious. Also, 3ta contains character are inherently polarized because this information is necessary to the analysis, while the polarization of characters or tree can be made before or after the analysis in parsimony. The assessment of plesiomorphy is considered in 3ta as necessary to establish phylogenetic relationships between states, and is part of the homology hypothesis. The treatment as triplets is not equivalent to unrooted quartets, which are partitions, and parsimony allows partitions to represent relationships, instead of 3ta. A long debate has led proponents on both sides to provide arguments on whether parsimony or 3ta was the best cladistic method. This debate has mostly relied on theoretical considerations and contrived examples based on very small datasets (Farris and Kluge, 1998; Farris, 2012; Harvey, 1992; Nelson and Ladiges, 1996; Platnick *et al.*, 1996; Zaragüeta and Bourdon, 2007), although a few studies have compared the methods with empirical datasets (Patterson and Johnson, 1995; Udovicic *et al.*, 1995; Williams, 1996; Bourdon, 2006; Corvez, 2012).

Recently, Grand *et al.* (2013) studied the relative merits of 3ta and parsimony with ordered and unordered states through simulations (100 matrices of six characters and eight taxa each) and two empirical examples. Their results suggested that 3ta has a greater efficiency to retrieve correct relationships (resolving power) than ordered parsimony, but that it also yields more incorrect relationships (artefactual resolution). They also clearly demonstrated that, under the conditions investigated in the simulations, unordered parsimony had the least resolving power and that it had greater artefactual resolution than ordered parsimony (*i.e.*, unordered parsimony can be considered the least effective of the three methods for treating multistate characters reflecting underlying continuous data). The results of the empirical studies of Grand *et al.* (2013) are more difficult to interpret because of the absence of a known reference phylogeny and the small taxonomic sample. Thus, it would be interesting to perform a more thorough study (based on more extensive simulations) with data generated especially to better discriminate between the three character coding schemes.

In this study, we test various hypotheses in order to refine characterization of the behavior of parsimony (ordered and unordered) and 3ta on discretized continuous characters. Thus, we test hypotheses the following hypotheses:

1. Given that our simulated data are intrinsically continuous, we expect that ordered methods (parsimony and 3ta) will outperform unordered parsimony. The relative performance between 3ta and ordered parsimony is more difficult to predict, but a greater accuracy of ordered parsimony could be expected for model-based simulations that allow genuine reversals.
2. We expect that errors in character polarity affect differently the methods, with the performance of 3ta deteriorating faster with increasing polarity errors than ordered parsimony because 3ta uses rooted character state trees, whereas ordered parsimony uses unrooted character state trees (fig. 2.11). The transformations are multidirectional in parsimony, contrary to 3ta, which should result in more synapomorphies being recognized in parsimony. 3ta might be more affected by polarization errors because 3ta disregards reversals as synapomorphies, and if a character is wrongly polarized, a genuine synapomorphy would be disregarded (fig. 2.14). In parsimony, if the polarity is wrong, all changes would be used, even though their interpretation might be erroneous (fig. 2.14). The performance of unordered parsimony with polarization errors is more difficult to predict. These variable levels of polarization errors are obtained by changing outgroup branch lengths.
3. Tree balance and branch lengths of the real ingroup phylogeny have an effect on the behavior of the methods: performance of phylogenetic methods can be influenced by terminal and internal branch length and tree symmetry. Based on the ratios of terminal to internal branch lengths (which can be considered a ratio of homoplasy to synapomorphy), then for ultrametric trees, performance should be best for the symmetrical tree, which has the lowest terminal/internal branch length ratio, and worst for a fully pectinate tree, which has the highest ratio. This should prevail in all three methods (throughout this paper, the 'three methods' will always refer to ordered parsimony, unordered parsimony, and 3ta). If we consider triplets as the minimal phylogenetic information (or if we use a metric, such as ITRI, which relies on triplets), we also expect that a long branch leading to a clade of several taxa is

more important than a long branch length leading to a clade of two taxa, because fewer triplets are involved in the latter. Thus, the structure of a phylogenetic tree may be expected to have an impact.

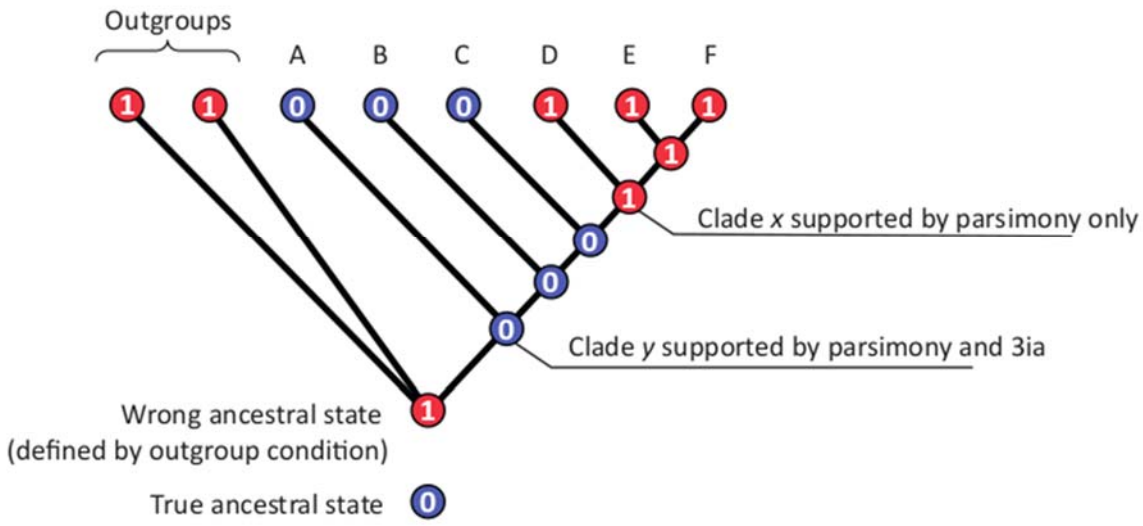


Figure 2.14. An example of optimal cladogram obtained both in parsimony and 3ta with effect of polarization error with an outgroup criterion. The true ancestral state is 0, and 1 is a true synapomorphy of D, E and F, but the outgroup criterion leads to an erroneous ancestral condition (1). Parsimony may be less sensitive to errors in polarization than 3ta, which cannot support a clade with what appears to be a reversal from 0 to 1.

3.2. Material and methods

3.2.1. Simulated character sets

3.2.1.1. Reference topologies

Data was simulated on three topologies of 20 ingroup OTUs and one outgroup OTU and a total of 18 branch length settings (fig. 2.15). One of the reference topologies is fully pectinate (*i.e.*, fully asymmetrical), a second is fully symmetrical for the ingroup (balanced) and a third is an equiprobable (randomly generated) tree, intermediate in asymmetry between the two others (fig. 2.15-A to F). In this paper, branch lengths represent evolutionary time, but given that we simulated characters using Brownian motion (see below), branch lengths also reflect expected character variance (Felsenstein, 1985).

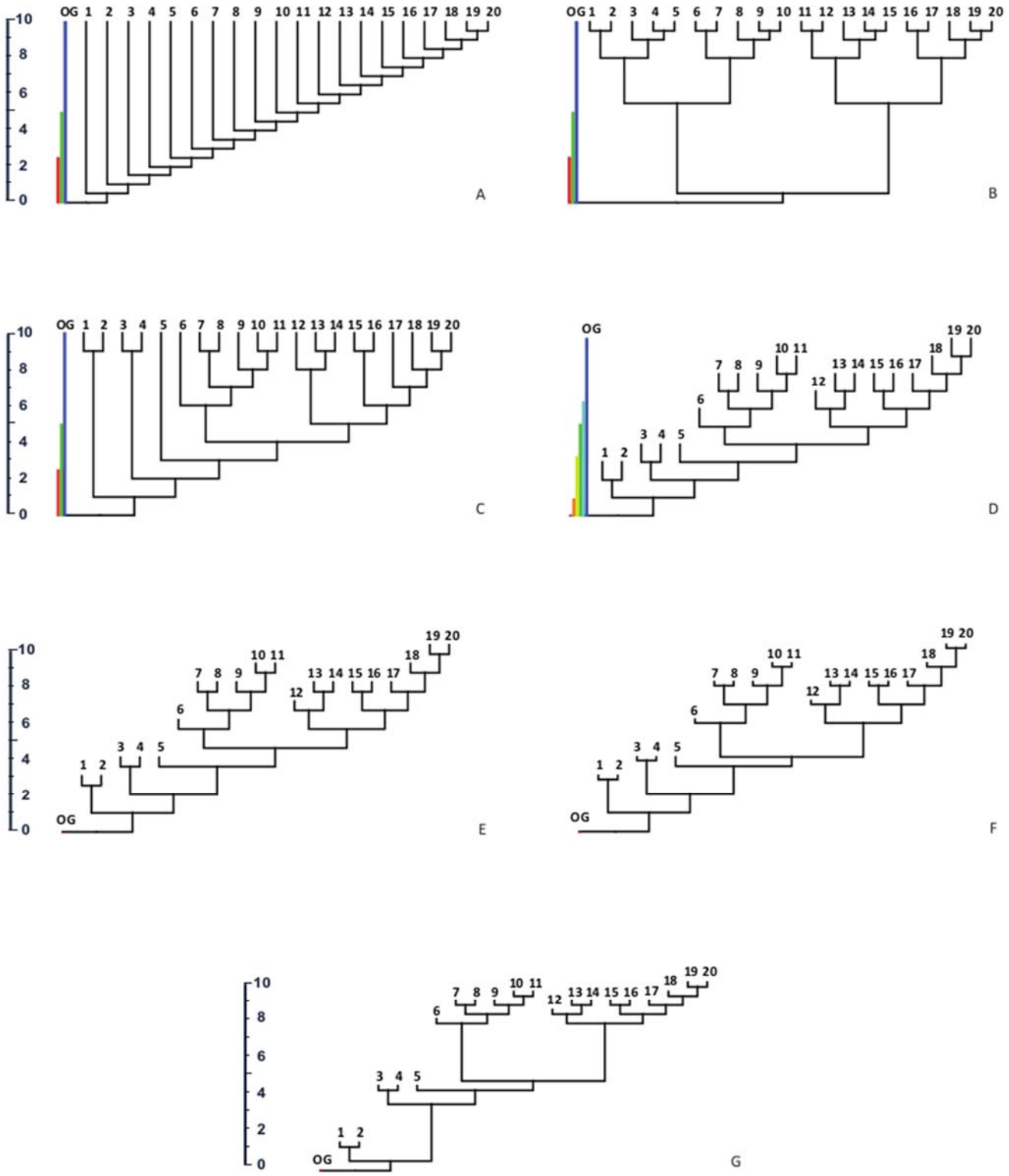


Figure 2.15. Trees used for our simulations: A, pectinate; B, symmetric; C, equiprobable; D, equiprobable with branch length set to 1; E-G, equiprobable with steady increase of internal/external branch length ratio. Each color represent a specific outgroup branch length expressed as a proportion of total tree depth (A-C: blue, 1; green, 1/2; red, 1/4; D: dark blue, 1; light blue, 2/3; green, 1/2; yellow, 1/3; orange, 1/10; red, 0). A–C represent trees with a neontological ingroup; trees D–G represent paleontological trees (with diachronous tips).

3.2.1.2. Reference trees

Three outgroup branch lengths were used on each of the reference topology, leading to nine ultrametric reference trees (fig. 2.15-A to C). Nine other branch lengths differing in both ingroup and outgroup were specified on the equiprobable tree with taxa of various geological ages, leading to 9 additional non-ultrametric reference trees (fig. 2.15-D to G). For six out of these nine trees, all terminal and internal branch lengths of the ingroup were set to one, and the outgroup branch was set at zero, one, three, five, six tenths, or the full tree depth (fig. 2.15-D). Three out of the nine trees (fig. 2.15-E to G) were generated by modifying the ingroup branch lengths on the tree with the outgroup branch set at zero length (actual ancestor).

3.2.1.3. Simulated matrices

From each of the nine reference trees illustrated in fig. 2.15-A to C, 100 matrices of 100 characters x 10 states were simulated (*i.e.*, a total of 900 matrices). Similarly, 100 matrices of 100 characters x 10 states were simulated for each of the nine paleontological trees illustrated in fig. 2.15-D to F (*i.e.*, a total of 900 paleontological matrices). We thus simulated a total of 1800 matrices, which were produced using Supplementary On-line Materials 1 (S1) on data that were discretized using Excel spreadsheets (S2 for parsimony; S3 for 3ta) and that were compiled into S4 (which are the matrices in the parsimony format).

3.2.1.4. Character coding

The characters were simulated with continuous Brownian motion in Mesquite (Maddison and Maddison, 2001) to represent data inherently ordered as morphoclines (such as size or shape characters). Simulations were made using this evolutionary model because it is one of the simplest and most widely used in evolutionary biology to study the evolution of continuous phenotypic characters. For example, phylogenetic independent contrasts (Felsenstein, 1985) and squared-change parsimony (Maddison, 1991) assume this model. Characters simulated through Brownian motion are continuous; they were then discretized into 10 equal intervals representing character states, in order to simulate morphoclines following the simple procedure described in (Laurin and Germain, 2011). Because Brownian motion has no tendency, the resulting distribution is Gaussian; thus, gap coding cannot be used, and the limits between states are arbitrary. The primitive condition is

determined by the outgroup criterion. The variable outgroup branch lengths allow us to assess the influence of polarization errors, whereas the variable ingroup branch lengths allow assessment of the impact of geological age of ingroup taxa on tree resolution (presumably by altering support of the clade subtended by the various branches), thus enabling a comparison of paleontological and neontological datasets. Each of the 100 characters (in the 1800 matrices) was coded in three different ways corresponding to unordered parsimony, ordered parsimony (with linear character states) and 3ta.

3.2.2. Tree searches

Characters were analyzed with the three methods. Thus, 5400 phylogenetic analyses were performed (three analyses for each of the 1800 matrices). Cladistic analyses were performed with PAUP* 4.0b10 (Swofford, 2003) for both ordered and unordered parsimony. Three-item analyses were also performed on PAUP* 4.0b10 but parsimony matrices were transformed into three-item statements matrices with fractional weighting (Mickevich and Platnick, 1989; Nelson and Ladiges, 1992) using LisBeth 1.3 (Zaragüeta *et al.*, 2012). All analyses were done with a heuristic search of 50 replicates (a number of searches that our preliminary analyses suggested was sufficient to recover all optimal trees for our matrices), using the TBR algorithm. A strict consensus (Sokal and Rohlf, 1981) of all optimal trees was then constructed using PAUP*.

3.2.3. Tree comparisons

The optimal tree yielded by the data analyses (or strict consensus tree when several optimal trees were found) were compared with the reference phylogenies, to assess accuracy of the results. We compared the behavior (resolving power and artefactual resolution) of ordered parsimony, unordered parsimony and three-item analysis (3ta) in phylogenetic inference, as recently done by Grand *et al.* (2013).

Several methods to compare trees (Robinson and Foulds, 1981; Estabrook *et al.*, 1985), but we prefer using rooted trees because they are the only classificatory structures which can convey unambiguous information of phylogenetic relationship (Rohlf, 1982). An unrooted tree conveys information, but it is ambiguous in the sense that it is compatible with several alternative relationships between taxa. Tree comparisons were

performed with the 'Inter-Tree Retention Index' (ITRI) proposed by Grand *et al.* (2013) to measure the degree of congruence between two trees. The ITRI is based on the proportion of relationships (*i.e.* triplet relationships between three OTUs, irrespective of their relationships with other OTUs) that are common to two trees. An advantage of this method is that the computed tree-to-tree similarities are not symmetrical (*i.e.*, the proportion of relationships is either the proportion for one tree or for the other). It is thus possible to discriminate between resolving power (power to find correct relationships) and artefactual resolution (incorrect relationships), by comparing trees obtained from simulated data with the tree used to generate these data. ITRI is equivalent to the retention index (RI) (Farris, 1989; Archie, 1989), calculated as a proportion of triplets (Kitching, 1998). The ITRI is defined as:

$$RI = \frac{(2n - (X + 2(n - X)))}{(2n - n)} = \frac{X}{n}$$

Where X stands for the sum of fractional weighting (FW) of triplets implied by a character compatible with a given tree; n is the sum of FW of all triplets obtained from the character. In each pairwise tree comparison, consensus trees were used to summarize congruence between results of an individual analysis (on a single matrix of 100 characters). See (Grand *et al.*, 2013) for a more detailed explanation. A strict consensus tree was built whenever analysis of a matrix produced more than one tree.

Our results are presented without mention of the number of optimal trees because the resolving power and artefactual resolution are the direct expression of data congruence and of the ratio homology/homoplasy, and that the ITRI expresses it well. Generally, a strict consensus tree of a thousand trees will show lower resolving power and artefactual resolution than a consensus of ten trees, because of data ambivalence. But there is always the possibility that an analysis will result in one false tree versus another analysis with one hundred slightly better trees. Kearney (2002) argued that 'resolution of relationships is obviously a goal of phylogenetic analysis', but her results show that phylogenetic performance cannot be measured by the number of trees. To avoid confusion with artefactual resolution, we complete her sentence by stating that it is *correct* resolution which is a goal of phylogenetic analysis. We think it is a necessary adjunction to remove

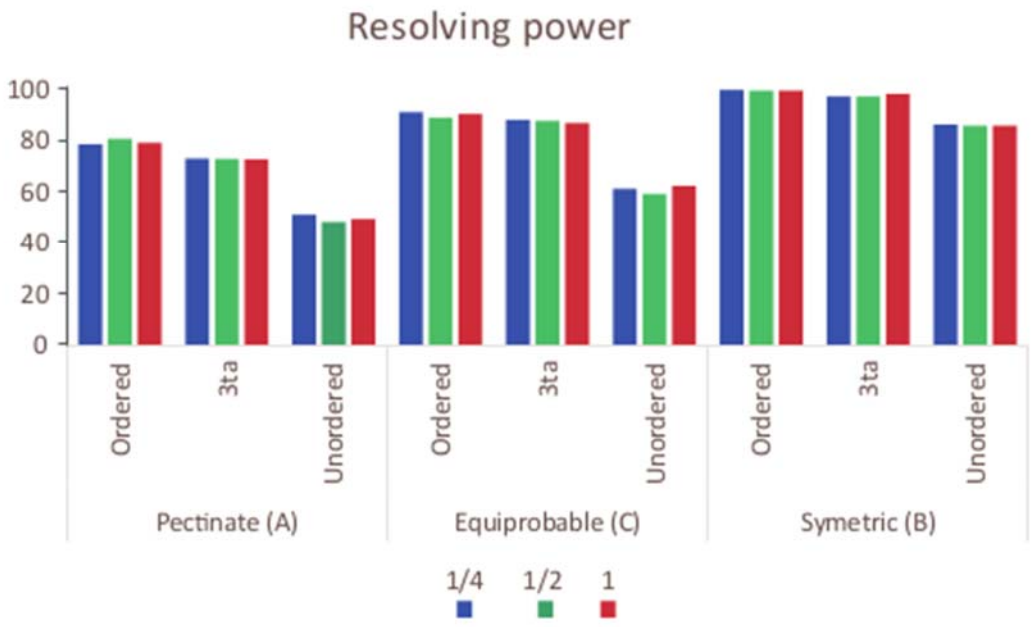
the ambiguity between resolution and number of trees. True resolution does not display a simple relationship with the number of optimal trees.

3.2.4. Testing the results

Each consensus tree summarizing a phylogenetic analysis is compared with the reference tree using the ITRI to assess resolving power and artefactual resolution yielded by each method, each set of outgroup branch length, and each topology. Resolving power is calculated as the proportion of FW for triplets of the reference tree that are also present in the consensus tree, and can be understood as the proportion of 'true information' the analysis has retrieved. Artefactual resolution is calculated as the proportion of FW for triplets of the consensus tree obtained from an analysis that is not present in the reference tree, and this represents artefactual resolution yielded by the analysis. We calculated means of ITRI (*i.e.*, we got a mean value for resolving power artefactual resolution on each reference topology). The statistical significance of differences in resolving power and artefactual resolution of the various methods of analysis was tested with the Wilcoxon signed-rank test (for paired samples) to compare means of ITRI, because there was no sample with normal distribution (Shapiro-Wilk test) and homoscedasticity of variances (Fisher test). Differences in topologies and branch lengths were tested using a Mann-Whitney test, and linear regressions were performed to test tendencies were the external/internal branch length ratio varies, as in trees D, E, F and G (fig. 2.15). Because many comparisons were made, we use the false discovery rate procedure (Benjamini and Hochberg, 1995) to assess the statistical significance of the differences in performance.

3.3. Results

All of the 2,700 analyses with equal branch lengths yielded strict consensus trees that are at least partly resolved; they all carry phylogenetic information. For ordered parsimony resolving power has a mean of 99.8% (sd: 1,1%) on the symmetric tree (fig. 2.15-B), 78.8% (sd: 8,8%) on the fully pectinate tree (fig. 2.15-A) and 91.1% (sd: 8%; fig. 2.16) on the equiprobable tree (fig. 2.15-C with blue outgroup branch length). Artefactual resolution is respectively about 0.2% (sd: 0,7%), 17.9% (sd: 7,2%) and 9.9% (sd: 6,3%. Resolving power is always a little lower and artefactual resolution greater in 3ta than in



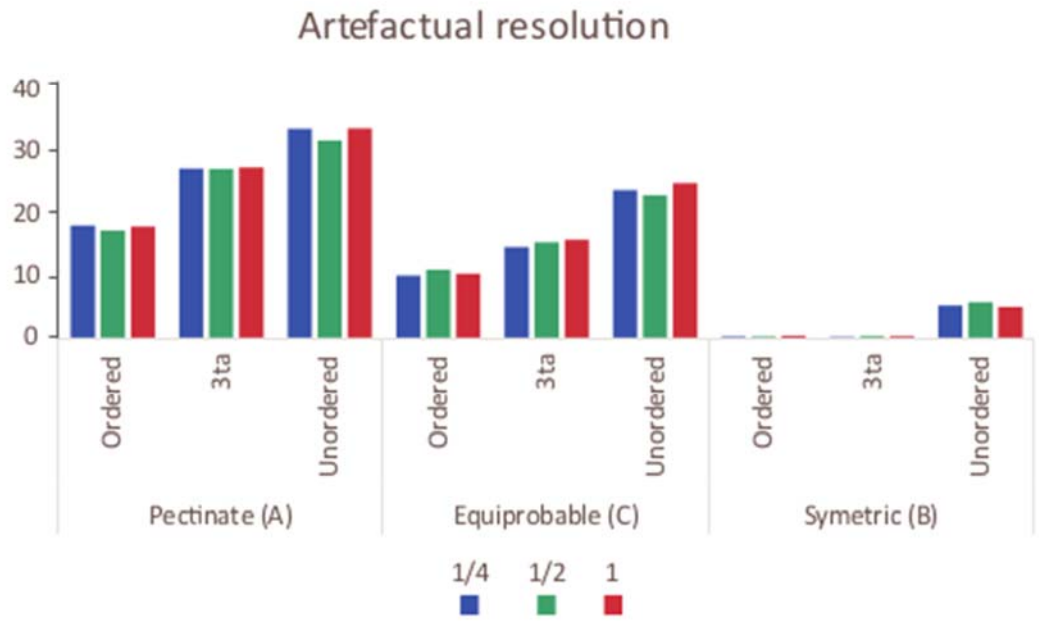


Figure 2.16. Histograms showing resolving power and artefactual resolution obtained by ITRI mean percentages sorted by outgroup branch length (blue, 1/4; green, 1/2; red, 1), by method (ordered parsimony, 3ta and unordered parsimony), and by topology (pectinate, equiprobable and symmetric; see S8 in http://www.ctoz.nl/vol84/nr02/a03). Each ITRI mean represents 100 matrices of 100 characters and 21 taxa (outgroup included).

ordered parsimony ($p < 0.0001$; table 2.1). Unordered parsimony has the lowest resolving power ($p < 0.0001$), and produces the greatest artefactual resolution ($p < 0.0001$). The outgroup branch length does not have a significant effect on resolving power or artefactual resolution for the first three topologies (fig. 2.15-A to C). Among the three reference topologies, the symmetrical topology seems to be the easiest to retrieve whatever cladistic method is used (greatest resolving power and lowest artefactual resolution), followed by the equiprobable topology, and by the fully pectinate topology (table 2.1). With the 1,800 analyses performed on the paleontological tree (fig. 2.15-D) and whatever method used, resolving power decreases and artefactual resolution increases with outgroup branch length ($p < 0.0001$; table 2.2; fig.). Ordered parsimony seems to be less sensitive than 3ta and unordered parsimony to polarization errors (reflecting outgroup branch lengths).

From the 2,400 analyses with different branch lengths (fig. 2.15-D to G), the differences between the resolving power and artefactual resolution values calculated on trees from a same reference topology cannot be explained by a linear model; results do not even vary consistently with the ratio between internal and terminal branch lengths (S5).

The impact of outgroup branch lengths was generally non-significant when we used trees with contemporary ingroup taxa (fig. 2.15-A to C). In 54 tests based on these trees, only 4 rejected the null hypothesis that outgroup branch length has no impact on performance (fig.5; table 2.1). All 4 significant results concern the trees built from the symmetrical reference topology. Differences in the other 50 comparisons on the same trees (fig. 2.15-A to C) do not show any tendency of increasing or decreasing resolving power or artefactual resolution when the outgroup branch length increases. Tests on trees built from the equiprobable paleontological reference tree (fig. 2.15-D) with a variable outgroup branch length show that the performance of all methods decreases when outgroup branch length increases (fig. 2.17; table 2.2). Results of ordered parsimony and 3ta are very similar, with unordered parsimony yielding much lower performance.

Finally, we have compared the performance for each method on two trees with the same topology but in which the taxa were extant (fig. 2.15-C) or extinct (fig. 2.15-D), to simulate neontological and paleontological trees, respectively. The results suggest that the paleontological tree is far easier to reconstruct than the neontological tree for ordered parsimony and 3ta in resolution power and artefactual resolution ($p = 0.0004$ for

resolution power in 3ta and $p < 0.0001$ for resolution in parsimony, as well as for artifactual resolution with both methods). Differences between trees in unordered parsimony were not significant.

3.4. Discussion

3.4.1. Brownian motion in cladistics

Under the conditions examined in our simulations, ordered parsimony performed best, thus confirming our first hypothesis. Simulating characters on a phylogeny has the advantage that the reference phylogeny is known without error. However, the Brownian motion, evolutionary model that is widely used in simulations of evolution, and which we used here, impacts upon the results, and reflects theoretical assumptions. Under Brownian motion, character evolution is stochastic and unpredictable, as are many historical events, but follows a general pattern that reflects the phylogeny, which can be inferred by analyzing character state data. Here, because of the generation of discretized continuous characters, the distribution of character states is unimodal (S6). They are intrinsically ordered and thus represent morphoclines. Brownian motion, like most other models of molecular evolution, such as GTR+I+Γ (Tavaré, 1986) or the speciational model, is not directional (*i.e.*, it implies no trends). Thus, the relationships between character states can be represented by unrooted trees (fig. 2.15-A). Brownian motion thus leads to an intrinsically ordered and unpolarized modeling (when the root condition is not specified in the simulation, as is the case here), contrary to models implying trends or irreversible evolution, which are intrinsically polarized. Under Brownian motion, the probability for a character state 0 to evolve into 1 is greater than the probability for 0 to evolve directly into 2; it was thus expected to favour ordered parsimony over unordered parsimony, and to a lesser extent, 3ta. However, this model, one of the simplest, seems applicable to various characters, such as ontogenetic sequence data (Poe and Wake, 2004; Poe, 2006).

3.4.2. Reversals in phylogenetics

A particularly controversial issue in cladistics concerns the treatment of reversals. Proponents of parsimony (Kluge, 1994; Farris *et al.*, 1995; Farris, 1997; Farris and Kluge, 1998) and 3ta (De Laet and Smets, 1998; Siebert and Williams, 1998) have been deeply

RESOLVING POWER

Significance of differences between methods

	Symetric			Pectinate
	Ordered/Unordered	Ordered/3ta	Unordered/3ta	Ordered/Unordered
1/4	< 0,0001	< 0,0001	< 0,0001	< 0,0001
1/2	< 0,0001	< 0,0001	< 0,0001	< 0,0001
1	< 0,0001	< 0,0001	< 0,0001	< 0,0001

Significance of differences between polarizations

	Symetric			Pectinate
	0,25/0,5	0,25/1	0,5/1	0,25/0,5
Ordered	0.0048	0.4915	0.3031	0.0973
Unordered	0.0628	0.8488	0.6356	0.9406
3ta	0.3477	< 0,0001	< 0,0001	0.1252

Significance of differences between topologies

	Ordered			Unordered
	Symetrical/Pectinate	Symetrical/Equiprobable	Pectinate/Equiprobable	Symetrical/Pectinate
1/4	< 0,0001	< 0,0001	< 0,0001	< 0,0001
1/2	< 0,0001	< 0,0001	< 0,0001	< 0,0001
1	< 0,0001	< 0,0001	< 0,0001	< 0,0001

ARTEFACTUAL RESOLUTION

Significance of differences between methods

	Symetric			Pectinate
	Ordered/Unordered	Ordered/3ta	Unordered/3ta	Ordered/Unordered
1/4	< 0,0001	0.0894	< 0,0001	< 0,0001
1/2	< 0,0001	0.0015	< 0,0001	< 0,0001
1	< 0,0001	0.0126	< 0,0001	< 0,0001

Significance of differences between polarizations

	Symetric			Pectinate
	0,25/0,5	0,25/1	0,5/1	0,25/0,5
Ordered	0.5541	0.8681	0.6676	0.3481
Unordered	0.8373	0.4260	0.3297	0.3999
3ta	0.0073	0.2094	0.2275	0.6242

Significance of differences between topologies

	Ordered			Unordered
	Symetrical/Pectinate	Symetrical/Equiprobable	Pectinate/Equiprobable	Symetrical/Pectinate
1/4	< 0,0001	< 0,0001	0.0088	< 0,0001
1/2	< 0,0001	< 0,0001	< 0,0001	0.0008
1	< 0,0001	< 0,0001	< 0,0001	< 0,0001

Table 2. 1. . Statistical significance of differences in resolving power and artefactual resolution (calculated through mean ITRI) between methods, outgroup branch lengths, and topology of the reference tree. The probabilities were calculated using Wilcoxon signed-rank tests for the comparisons between methods and with a Mann-Whitney test for the outgroup branch length and the shape of the reference tree. Non-significant results with α=0.05 are shown in blue, and non-significant results after correction for multiple tests using the false discovery rate (Benjamini and Hochberg, 1995) are shown in brown. Individual values can be found in S7.

RESOLVING POWER				
Significance of differences between methods				
Pectinate		Equiprobable		
Ordered/3ta	Unordered/3ta	Ordered/Unordered	Ordered/3ta	Unordered/3ta
< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
< 0,0001	< 0,0001	< 0,0001	**0.036**	< 0,0001
< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
Significance of differences between polarizations				
Pectinate		Equiprobable		
0,25/1	0,5/1	0,25/0,5	0,25/1	0,5/1
0.2791	**0.5833**	**0.0538**	**0.3112**	**0.2389**
0.5173	**0.3945**	**0.7058**	**0.9883**	**0.1008**
0.6512	**0.7563**	**0.1307**	**0.3170**	**0.3633**
Significance of differences between topologies				
Unordered		3ta		
Symetrical/Equiprobable	Pectinate/Equiprobable	Symetrical/Pectinate	Symetrical/Equiprobable	Pectinate/Equiprobable
< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
< 0,0001	0.0002	< 0,0001	< 0,0001	< 0,0001
< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001

ARTEFACTUAL RESOLUTION				
Significance of differences between methods				
Pectinate		Equiprobable		
Ordered/3ta	Unordered/3ta	Ordered/Unordered	Ordered/3ta	Unordered/3ta
< 0,0001	0.0015	< 0,0001	< 0,0001	< 0,0001
< 0,0001	0.0285	< 0,0001	< 0,0001	< 0,0001
< 0,0001	0.0009	< 0,0001	< 0,0001	< 0,0001
Significance of differences between polarizations				
Pectinate		Equiprobable		
0,25/1	0,5/1	0,25/0,5	0,25/1	0,5/1
0.9134	**0.4033**	**0.4026**	**0.8003**	**0.2962**
0.6320	**0.5934**	**0.6938**	**0.6759**	**0.4391**
0.5453	**0.9698**	**0.0949**	**0.1794**	**0.8145**
Significance of differences between topologies				
Unordered		3ta		
Symetrical/Equiprobable	Pectinate/Equiprobable	Symetrical/Pectinate	Symetrical/Equiprobable	Pectinate/Equiprobable
< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
< 0,0001	0.0006	< 0,0001	< 0,0001	< 0,0001

Continuation of table 2. 1.

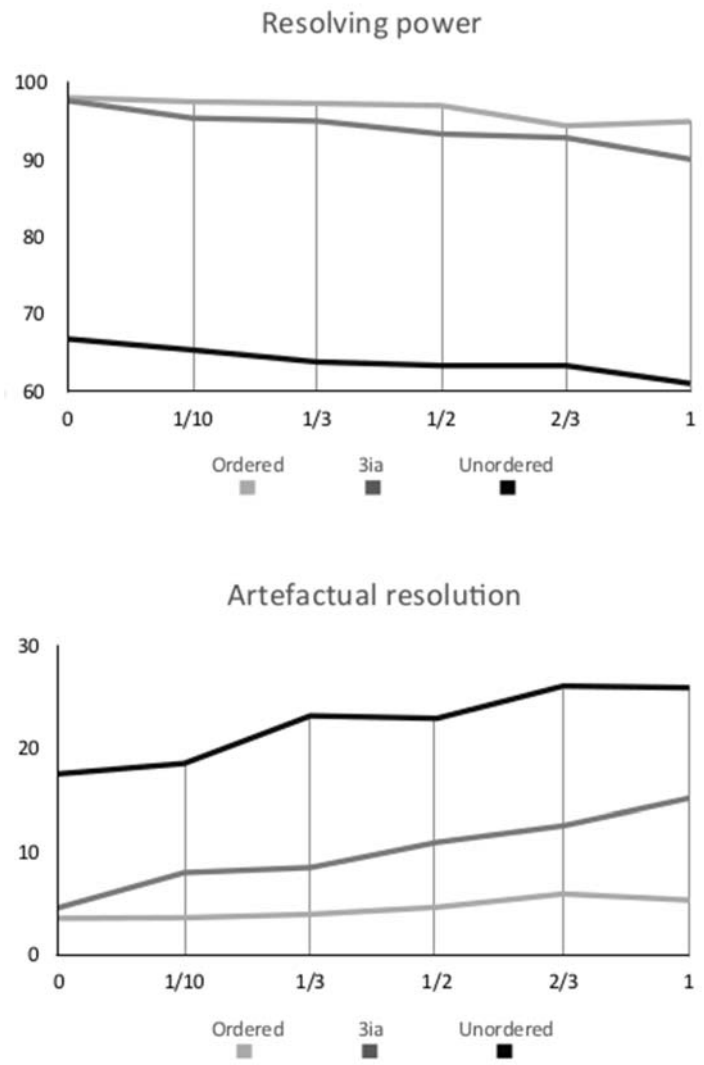


Figure 2.17. Impact of outgroup branch length expressed as a proportion of total tree depth (0, 1/10, 1/3, 1/2, 2/3, 1) on resolving power and artefactual resolution. This was obtained through ITRI mean percentages from the tree shown in fig. 2.15D (also see S8 in http://www.ctoz.nl/vol84/nr02/a03).

divided on this particular issue. In parsimony, a transformational approach to homology using the Wagner (Farris *et al.*, 1970) and Fitch (Fitch, 1971) parsimony algorithms treat characters from the perspective of unrooted character-transformation trees (Slowinski, 1993). Reversals provide information and can serve for clade support because they are evidence of secondary homology with the appropriate test of maximizing congruence. This maximization of congruence leads to search the pattern with the minimum of *ad hoc* hypotheses that are convergences and reversals. For Farris (2012), reversals can be inferred a priori in the inference of primary homologies but also from an analysis: *'More*

Linear regressions on polarization (tree D)

	Ordered	Unordered	3ta
Resolving power	0.02306	0.001758	0.0008212
Artefactual resolution	0.02505	0.008793	0.0009794

Table 2. 2. Impact of outgroup branch length (values of 0, 1/10, 1/3, 1/2, 2/3, 1, expressed as a proportion of total tree depth, were tested) on resolving power and artefactual resolution. Probabilities of the null hypothesis (no impact) were tested through linear regressions of ITRI means on outgroup branch length. These results were obtained from the tree shown in figure 2.15D. Resolving power, slope: -7.2 (ordered), -10.1 (unordered), -13.5 (3ta); intercept (1000 indicates complete power or error): 991.9 (ordered), 675.3 (unordered), 989 (3ta). Artefactual resolution, slope: 4.6 (ordered), 18.3 (unordered), 19.8 (3ta); intercept: 28.9 (ordered), 160 (unordered), 30.8 (3ta). Individual values can be found in S7.

fundamentally, even if (as I have seen other authors suggest) Hennig would have preferred to distinguish apomorphies from plesiomorphies before starting to construct the tree, he was obviously willing to revise assessments of plesiomorphy during tree construction, for Hennig did in fact recognize reversals and apply them as synapomorphies'. This is the most widespread point of view in cladistics, which prevails in systematic paleontology, and the only point of view represented in probabilistic methods, which prevail in molecular systematics and are starting to be used on phenotypic characters as well (Müller and Reisz, 2006). Assumptions of 3ta are much less familiar. In 3ta, hierarchical hypotheses of homology, *i.e.* a nested set of character states, are submitted to a test of congruence. The test either accepts or rejects the relevance of the hypothesis. Convergence is one of the multiple explanations of rejection. 'Parsimony-like reversals', *i.e.* the generation of hypotheses of homology not proposed by the systematist but generated by the method, violate hierarchical classifications. Thus, they cannot be justified in 3ta rationale. Evolutionary reversals, *i.e.*, losses of instances of character-states, are not used in 3ta to support nodes; they represent only homoplasies, *i.e.* mistaken hypotheses of homology.

For instance, the evolutionary hypothesis (generated after an initial analysis by inferring character history on a tree) involving three conditions deduced a posteriori from an analysis: 0 ('absent'), 1 ('present') and 0* ('secondary absence'; scored the same in a matrix but interpreted differently from primitive absence on a tree) is interpreted differently under parsimony and 3ta. The secondary absence can be explicitly represented as an apomorphy in the primary homology hypothesis (0(1(0*))) (under parsimony). Another interpretation (3ta) consists in disregarding secondary absence as synapomorphic

but to consider it as a particular case of absence: (0,2(0*)). Here, neither the absence nor the reversal is considered as a state (neither plesiomorphy, nor apomorphy) because the absence is not a state in 3ta (in fig. 2.13, 0*, 0** and 0*** are not considered in 3ta). Parsimony proponents favor the first option, which yields support for six clades in the phasmatodea phylogeny (fig. 2.13-A). To summarize, the first interpretation (parsimony) considers a loss as an homology and a synapomorphy (because it supports a clade), an homoplasy (because the primary hypothesis is falsified by the distribution of the other characters) and a plesiomorphy (as defined in the matrix), according to (Brower and de Pinna, 2014). The second interpretation considers a loss as uninformative: it is neither an homology nor a synapomorphy (because it supports no clade), it is not an homoplasy (because the primary hypothesis is in agreement with the distribution of the other characters) and it is not a plesiomorphy (because the absence is not a state in 3ta; it is the root, including all). 3ta proponents favor this interpretation: only one clade in the Phasmatodea phylogeny is supported, and the only synapomorphy is the homology reflecting the first appearance of wings. These two interpretations are thus in perfect opposition. Here we emphasize that Brownian motion is only coherent with the assumptions entailed by the first interpretation: reversals (*i.e.*, secondary absence) are treated as apomorphies in the primary homology hypotheses (as an order with parsimony, or as a hierarchy with 3ta). Our simulations produce informative reversals under Brownian motion, which can be exploited only under a parsimony viewpoint of these reversals: our results present a quantification of the loss in resolving power and artefactual resolution in 3ta if true and informative reversals are present (*i.e.* if true reversals are simulated and 'hidden' into the same state as the plesiomorphy but which support a clades of the known tree). Thus, our results must be interpreted accordingly. Irreversible characters might yield different results and will be tackled in another study.

We take this opportunity to propose a nomenclatural clarification about reversals (based on the example in figure 2.13-A) as secondary homology hypotheses; thus, this clarification is valid both for parsimony and for 3ta. First rounds of reversals are generally called 'secondary losses' (*e.g.* Carine and Scotland, 1999), when in fact, only the absence should be considered secondary and the loss in itself should be considered as an event that appeared for the first time (*i.e.*, primary). Thus, a character state is primitively

absent (primary absence; state 0 on figure 2.13). It can then appear; this is a primary appearance (of state 1), denoted +1 on figure 2.13-A. It can be subsequently lost (-1, reversal to state 0, but identified as 0* on figure 2.13-A, for greater clarity); this should be called a primary loss, which results in a secondary absence. After this, a secondary gain (+2) can lead to secondary presence (1* in figure 2.13-A), and a secondary loss (-2) can lead to ternary absence (0** in figure 2.13), etc.

3.4.3. Uncertain polarization

We failed to find significant results on the impact of outgroup branch lengths and uncertainty in polarization, except on some trees built from the symmetrical reference topology.

The effect of outgroup branch length appears to be much stronger on paleontological trees, perhaps because of the shorter branches in the ingroup. Tests on a non-ultrametric version of the equiprobable tree (fig. 2.15-D) with a variable outgroup branch lengths show that the performance of all methods decreases when outgroup branch length increases (fig. 2.17; table 2.2). Results of ordered parsimony and 3ta are very similar, with unordered parsimony performing much more poorly. 3ta is more sensitive to outgroup branch length, an effect that might be linked to reversal treatment, but that in any case confirms our second hypothesis. It is however surprising to see that an incorrect polarization has so little effect on resolving power and artefactual resolution.

3.4.4. Character states and ordering schemes

Our simulations clearly show that unordered parsimony performs far worse than the two other methods when reliable criteria for character state ordering are ignored (fig. 2.16). All states were ordered (except for unordered parsimony) using a similarity criterion, whereby transition costs (step-matrices in ordered parsimony) or state hierarchy (3ta) reflect similarity (and outgroup condition, for 3ta). Other ordering criteria exist (Hauser and Presch, 1991), but our results clearly indicate that ordering character states is preferable when characters can be shown to form morphoclines. These results suggest that the actual tendency not to order characters in phylogenetic analyses is suboptimal, and shows that important benefits could arise from considering ordering schemes when it appears

biologically justified. Note that such ordering requires no prior knowledge of the phylogeny; only knowledge of the character distribution or likely evolutionary model (which can be gained from genetic or developmental data, among others) is required. Ordered parsimony and 3ta share more similarities on this particular point than unordered parsimony.

3.4.5. Tree shape and branch length

Not all topologies are equally difficult to recover accurately. The pectinate topology (fig. 2.15-A), which shows the longest terminal branches (when its internal branches are all of about the same lengths), is most difficult, followed by the equiprobable (fig. 2.15-C) and the symmetrical topology (although this may be linked with the fact that internal branch lengths of that tree were roughly proportional with the number of descendant taxa, except for the branch below the ingroup; fig. 2.15-B; fig. 2.16; table 2.1). This confirms our second hypothesis. Modifying the distribution of branch lengths yielded results that are more difficult to interpret (S5). Our first assumption was that the results would improve (*i.e.*, the resolving power would be greater and the artefactual resolution lower) as the terminal/internal branch length ratio decreases. This assumption is corroborated for the trees D, E and F (fig. 2.15) for resolving power under ordered parsimony and 3ta, but only for trees D and E for unordered parsimony. Tree G yielded worse results, perhaps because some internal branches are shorter. Results on artefactual resolution are more complicated to interpret. Along with the branch lengths of a tree, the triplet content of clades is another important parameter to consider when performance is assessed using the ITRI. It is directly connected to the tree shape, *i.e.* the number of terminal taxa inside and outside each clade (Nelson and Ladiges, 1992). As a consequence, some clades and characters they support will affect more the ITRI than others. More specifically, clades with few taxa within them or with few taxa outside them have less triplet content than clades containing an intermediate number of taxa (in our trees, ten inside, eleven outside). These imbalanced clades will impact results only slightly compared to balanced clades. This explains why tree shape influences phylogenetic reconstruction, where a balanced tree has more balanced clades than a pectinate tree and is therefore more easily reconstructed.

Comparisons of results between an ultrametric tree (fig. 2.15-C) and a paleontological tree of the same topology (fig. 2.15-D) show that the latter features better resolving power and less artifactual resolution for ordered parsimony and 3ta. This result is congruent with the claim that adding extinct taxa breaks long branches, which results in important improvement on the resolution of the optimal trees. This claim is supported both on empirical data (Gauthier *et al.*, 1988) and simulations (Huelsenbeck, 1991).

3.4.6. Implications on simulation-based studies and evolutionary models

Our results highlight the advantages of models under which the relationships between character states can be represented by an unrooted tree (all molecular models) over 3ta hierarchical coding, if reversals can be simulated as in our study. Our simulations under Brownian motion can be represented as in figure 2.18. Firstly, characters are generated

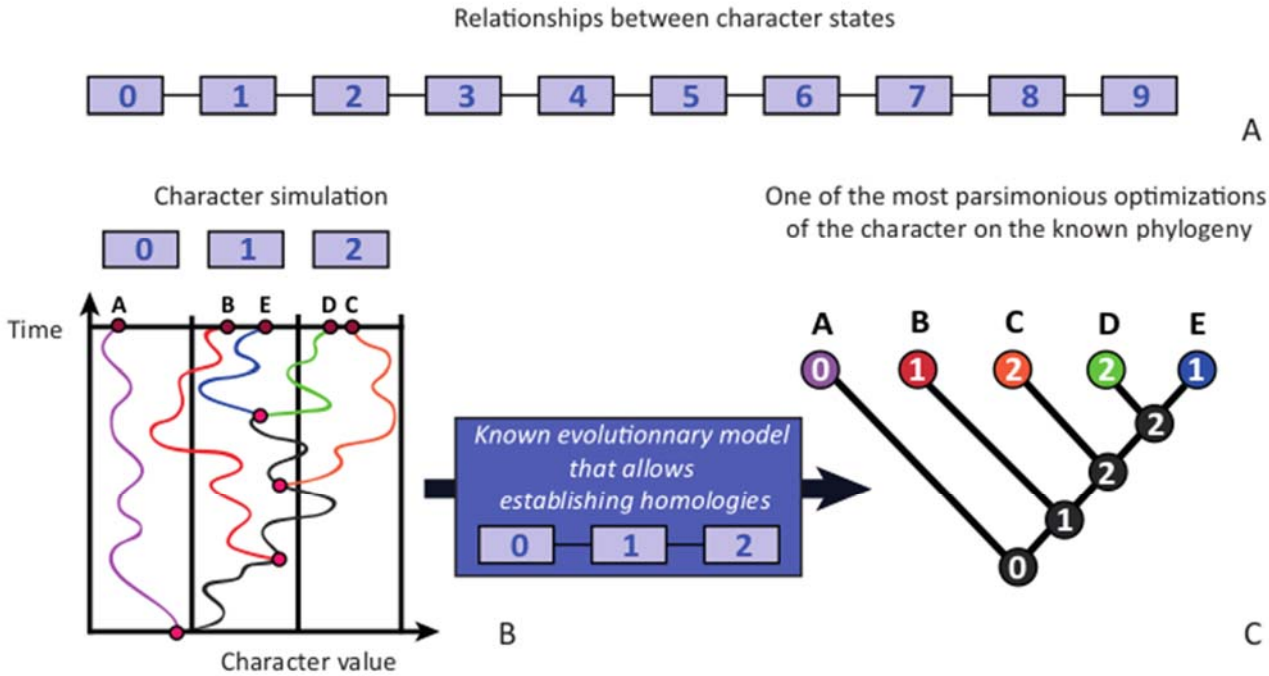


Figure 2.18. Character evolution under Brownian motion. In A, the unrooted tree represents the relationships between states of a character simulated under Brownian motion and discretized into 10 states. B shows the simulation of characters on a known phylogeny under Brownian motion. The left graphic represents fluctuations of a continuous character, with cladogenesis and terminal taxa shown as red nodes. After discretization through states 0, 1 and 2, the known evolutionary model and a topology (known or inferred) allow inferring character history. Here, the example treats characters under the parsimony method, with acctran optimization.

under a priori assumptions (here, the evolutionary model represented in figure 2.18-A and the reference phylogeny shown on figure 2.18-B). In a second step, these characters are interpreted as primary homology hypotheses by discretization and (for 3ta) conversion into hierarchical structures. Our procedure for simulation of characters, and by extension the use of evolutionary models, reflects quantitative characters that display informative reversals.

In parsimony, states that can be hypothesized to form morphoclines should be ordered if there is evidence. In this case, reversals can be represented by primary homology hypotheses. Empirical studies suggest that Brownian motion is a reasonable model for several types of characters, such as body size (Laurin, 2004), bone microanatomy (Canoville and Laurin, 2010), etc. Some characters do not seem to follow a strictly Brownian motion model, but instead may follow a speciational model (in which change occurs in both daughter-lineages at or near speciation, but no anagenesis takes place), such as morphological shape data in ratites (Laurin *et al.*, 2012), a punctuational model (similar to the speciational model, but change occurs only on one branch) or an Ornstein-Uhlenbeck model (Uhlenbeck and Ornstein, 1930; Felsenstein, 1988). Some of these alternative models can be obtained from Brownian motion by modifying the branch lengths of a tree (Garland *et al.*, 1993). Hence, data produced using such models might give results close to those that we report below, which should be applicable to much biological data.

In 3ta, the same representation of change is used for groups of parts (homologues) and for groups or whole organisms (taxa). In a transformation, the transformed part is considered a different object. In differentiation or modification, the differentiated or modified part is a particular form of the unmodified part. Systematists accept this difference for taxa (*e.g.* birds as a differentiated taxon nested in Dinosauria) but not for parts (*e.g.* feathers as a modified part, nested in scales). In this context, reversal as a loss is perfectly acceptable. However, reversal as synapomorphy is viewed as pointless.

All this suggests that irreversible characters (Nopcsa, 1923; Goldberg and Igić, 2008; Kohlsdorf *et al.*, 2010) are 'more compatible' with 3ta. Further work should be done to compare resolving power and artefactual resolution yielded by parsimony and 3ta when an irreversible-evolution model is used (which can be understand as hierarchic). Another interesting field of research consists in developing methods that better simulate

Darwinian evolution applied to digital life forms, as in the software Avida (Adami and Brown, 1994) to understand better the behavior of parsimony and 3ta.

3.5. Conclusion

Our simulated characters evolved gradually; they are continuous and have a unimodal distribution (S6). They were discretized because phenotypic data are often available in discrete form (in descriptions, for instance), and most available parsimony and all 3ta programs require this. Thus, these reflect the data routinely analyzed by systematists. In this case, we created 10 states per character to retain much of the original information while producing data that can easily be analyzed by almost any phylogenetic package (Laurin and Germain, 2011), and to allow detection of the spurious relationships yielded by homoplasy in the data (Bardin *et al.*, 2013). This procedure yielded many significant results (tables 2.1, 2.2). We show that unordered parsimony performs far worse than ordered methods on such data, with a loss in resolving power between 13.4% and 29.3% compared to ordered parsimony and with between 4.7% and 15.6% more artefactual resolution (results always statistically significant). Thus, a significant decrease in performance is expected when characters are not ordered as morphoclines. This result highlights the information content of character ordering schemes (Wilkinson, 1992), and we predict that coding continuous characters as fully unordered may significantly decrease resolving power and increase artefactual resolution. We also quantified the differences in resolving power between ordered parsimony and 3ta, which differ in the way they handle reversals as evidence for the support of a clade (among others). These differences vary between 1.3% (non-significant) and 6.3% (table 2.1) for the resolving power, and between 0.0% (non-significant) and 9.4% for the artefactual resolution. We also have quantified resolving power when polarization errors, produced by varying outgroup branch length, are introduced (*i.e.*, when polarization is based on an outgroup including a variable proportion of plesiomorphies). Further work could be done to quantify the effect of incorrect state ordering schemes as incorrect morphoclines assumptions.

Recently, Grand *et al.* (2013) found that 3ta yielded significantly greater resolving power and more artefactual resolution than ordered parsimony. In our study, based on an extended set of matrices, taxa and characters, ordered parsimony yielded the

greatest resolving power and the fewest artefactual resolutions. The use of exact hypothetical ancestors to root trees in Grand *et al.* (2013) may explain the contrasting results. Moreover, the present work yields a better understanding of the impact of evolutionary assumptions about character state order and reversals. Further simulations could be done, using different evolutionary models, such as irreversible characters, to see if they yield different results about the performance of parsimony (with ordered stares or not) and 3ta.

4. Impact of errors on cladistic inference: simulation-based comparison between parsimony and three-taxon analysis

Cette partie correspond à un article soumis à *Contributions to Zoology*.

Auteurs : Valentin Rineau, René Zaragüeta i Bagils, Michel Laurin

4.1. Introduction

Phylogenies serve as a basis for most macroevolutionary studies. The reliability of phylogenetic trees depends partly on the characters used to construct them. Characters are the core of systematics, and questions such as 'what is a good character?' or 'how to code characters to best capture their evolutionary information?' are of critical importance for phylogenetic inference. Even though most recently published phylogenies have relied mostly or exclusively on molecular data, the phenotypic data on which we concentrate here are regaining a vital importance in systematics because some of the most promising recent methods to date the Tree of Life require use of morphological data. These are tip dating, also known as total evidence dating (Pyron, 2011; Ronquist *et al.* 2012) and the more recent total evidence dating under the fossilized birth-death process (Zhang *et al.*, 2016). Phenotypic data are required in these cases because extinct taxa need to be placed in the phylogeny, and for most of these, molecular data are unavailable, given that the oldest uncontroversial DNA sequences are less than 1 Ma old (Prüfer *et al.* 2010; Willerslev & Cooper 2005), whereas metazoan body fossils that are sufficiently diagnostic to be easily placed in phylogenies date back to the Cambrian (e.g. Vinther *et al.* 2014), about 495–541 Ma (Gradstein *et al.* 2012). Moreover, it may not be sound to do without the knowledge that morphological research has accumulated over several centuries.

Among little-investigated factors that can influence the outcome of phylogenetic analyses are the various kinds of errors that can creep in at various stages of the process. Among these, scoring errors are a central problem of phylogenetic methods. Errors can produce incongruence between a character and the phylogeny, i.e. homoplasy, and hence,

change the most parsimonious cladogram. The sources of coding and scoring errors that can inflate the apparent extent of homoplasy are potentially unlimited; this includes, among others, typing errors in the matrix or erroneous interpretation of features (erroneous DNA alignment, data degradation, etc.). However, homoplasy can also result from difficulties of interpretation of the evolutionary process, as in lateral gene transfer, loss of morphological or molecular features, or convergence.

Errors can also creep into all other stages of the phylogenetic analysis, through problematic species delimitations, errors in OTU specifications, and errors in ordering schemes. Such errors are certainly more common than generally believed. Errors in coding phenotypic characters are due, in part, to the subjective judgement and unique knowledge of the researcher. In molecular analyses of nucleotide sequences, this problem can occur in subjective judgement during alignment and through different interpretations of insertions and deletions. Finally, suboptimal coding can lead to information loss, for instance when a character is coded with too few character states, which fails to capture part of the phylogenetic information.

Beyond plain errors, differences between phylogenies can reflect character treatment, which differs fundamentally in Maximum Parsimony (MP; Farris, 1970) and Three-taxon analysis (3ta; Nelson & Ladiges, 1992; Nelson & Platnick, 1991). The differences in how they deal with homoplasy originate from the particular way they manage hierarchical characters. The congruence between characters in MP is calculated by scoring the minimal number of transformations from a state to another, usually, i.e. in unordered MP, without specifying the putative synapomorphy. Transformation events from any state to any other state can arise, disregarding their status as plesiomorphic or apomorphic. For instance, a clade can be supported in MP by a reversal, whereas the hierarchical coding of 3ta fails to offer such a support. Under 3ta, a character coded as (0(1)), is decomposed and can only provide triplets (or three-taxon statements) composed of (A:0(B:1,C:1)), where A, B and C are terminal taxa. A triplet of (A:1(B:0,C:0)) from an initial hypothesis of (0(1)) is meaningless in 3ta and represents a rejection of the triplet (Nelson & Platnick, 1991). The differences in character representation between MP and 3ta cause significant differences in the way homoplasy and error are managed, and finally on the resulting cladogram. MP with a fully unordered coding is used in many phylogenetic studies on

morphological datasets, especially in paleontology (*e.g.*, Liu, 2016; Lu *et al.*, 2016). Model-based approaches such as Maximum Likelihood and Bayesian Inference are not treated in this paper because they are much more computationally demanding and would have necessitated reducing the number of taxa and characters, which would probably have decreased the statistical significance of our results. Such studies need to be adapted from other material constraints. A study comparing both MP, 3ta, Maximum Likelihood and Bayesian Inference with adapted datasets is in preparation.

Simulation-based systematics can help choose an appropriate phylogenetic inference method, or a specific coding scheme. This branch of systematics relies on comparisons of the results of analyses of datasets simulated on known phylogenies; typically, a few parameters vary to make the results more general. A first attempt to understand the differential behaviour of MP and 3ta was made by Grand *et al* (2013). Their study rested on characters simulated under Brownian motion, allowing them to compare three cladistic methods of character treatment. They showed for intrinsically ordered characters that unordered parsimony is the method with the worst efficiency. Between ordered MP and 3ta, 3ta provided the highest resolving power, but also the highest artefactual resolution. Rineau *et al.* (2015) performed more extensive simulations by comparing, on larger data sets in continuous Brownian motion, the impact of polarization, branch length and tree shape on phylogenetic performance. They showed that ordered MP always performed better than 3ta for characters with 10 states. The results were explained by the differences between MP and 3ta in their treatment of homoplasy, and more precisely of reversals.

We propose here an empirical study of the impact of homoplasy and scoring errors in phylogenetic studies with the same tools of experimental systematics. In order to complement the previous studies that our team performed on phylogenetic methods (Grand *et al* 2013; Rineau *et al* 2015), we deal here with the following inter-related issues:

1. The differential efficiency of MP and 3ta to coding variations. Our first analyses were made with an unrealistic number of states (10 per character) to exaggerate differences in performance. Even if morphological variation may be considered unlimited, morphological characters are often described using only two states.

Molecular characters (DNA sequences) have four or five states (five when coding insertions and deletions). We can hypothesize that when the number of states is reduced to better reflect most empirical studies, 3ta loses less phylogenetic information than MP because of its hierarchical coding. Indeed, in the simulations of Grand *et al* (2013), where the number of character states was lower (binary characters), triplets had in all cases a higher resolving power than MP. We thus test character sets with two, four, six, eight and 10 states.

2. Following the coding of character states and their number, we test the resilience of methods to the addition of errors directly in the character coding, more precisely in the ordering scheme. The aim of these experiments, by adding more and more errors to the ordering scheme, is to determine under which circumstances unordered parsimony becomes preferable to ordered parsimony and 3ta. Ordered MP and 3ta are supposed to become less efficient than unordered MP as the ordering scheme degrades.
3. Our first simulations in Rineau *et al* (2015) used Brownian motion, and showed greater efficiency for MP than for 3ta. We here attempt to compare these results on Brownian motion, with generation of convergences and reversals, to simulations under a model of irreversible evolution with only convergences. A heated debate focused on the theoretical justifications of how to deal with reversals arose among systematists (Farris & Kluge, 1998; Farris, 2012; Harvey, 1992; Nelson & Ladiges, 1996; Platnick *et al.*, 1996; Zaragüeta and Bourdon, 2007). In MP, reversals are not treated in a special way; they are transitions between states like any other. In 3ta, only derived character states are relevant for clade support by clade ancestry. Reversals may be interpreted on a tree in 3ta but are never used to positively support a clade. Our simulations address this problem by giving empirical results to compare with predictions based on the theory.
4. We also assess the impact of random scoring errors in the matrix. This has been done infrequently (if at all) before because previous studies in experimental systematics have focused on molecular data, and sequencing errors have not been of great concern. However, numerous problems can arise that are not correlated to the phylogeny at all, such as contaminations or alignment problems. In comparative anatomy, errors of interpretation by the systematist are the most crucial problem.

No datasets are exempt from such difficulties. Thus, we generated random errors in the data, which we termed 'uncorrelated homoplasy' because it is uncorrelated to the true evolutionary tree. To increase the uncorrelated homoplasy ratio, we generated errors randomly with respect to their value and position in the initial matrix. We tested the effect of uncorrelated homoplasy through comparisons with datasets lacking homoplasy (with perfectly fitting datasets without homoplasy) or with others showing correlated homoplasy (with reversals and convergences). We test our hypothesis that 3ta may be more efficient at extracting phylogenetic information when such random errors arise.

5. Model-based simulations are relevant to the extent that they mimic genuine biological data. To assess the performance of phylogenetic inference methods through another approach, we analyse two empirical datasets for which the true evolutionary history is known without error. We used these two empirical examples to compare the simple evolutionary models we used in our simulations. The two examples detailed below are the known HIV transmission history, and an experimentally generated phylogeny of manuscript texts copied and recopied with typos.

A synthesis of all these results allows us to compare the efficiency of MP (ordered or not) and 3ta in various situations that have not been examined before.

4.2. Material and methods

4.2.1. Simulated character sets

4.2.1.1. Reference tree

The data sets were simulated on a randomly generated topology by Mesquite (Maddison & Maddison 2011). The tree comprised 20 OTU (fig. 2.19), plus an outgroup to root the trees (to polarize character trees in 3ta). Our tree is ultrametric, and the information about topology is supplemented by 39 branch lengths for the ingroup and a branch length leading to the outgroup. The branch lengths are directly proportional to the expected variance of characters for the Brownian motion evolutionary model (Felsenstein 1985). Branches allow a visual quantification of transformations, and can be equated to

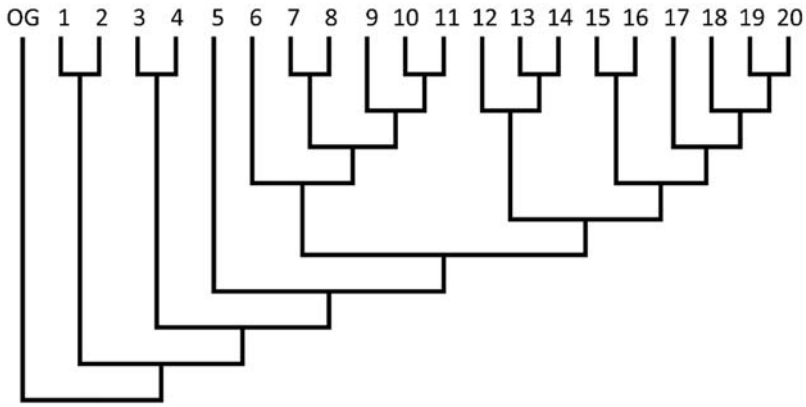


Figure 2.19. Ultrametric tree used in our simulations. The randomly generated phylogeny comprises an ingroup of 21 OTU and a single outgroup OTU.

evolutionary time (OTU are of the same geological age for our simulations given that our tree is ultrametric) under Brownian Motion.

4.2.1.2. Simulations and matrix coding

Continuous characters were simulated with Brownian motion on Mesquite. This simple model is very often used in evolutionary biology because it allows representation of characters with stochastic evolution, without selection pressure. This model is assumed in square-changed parsimony (Maddison 1991), among other comparative methods. Simulated characters are inherently ordered because their continuous value is discretized into several states (in this case, of equal amplitude). The stochastic nature of Brownian motion also constrains the character states to display a unimodal distribution. Discretization is thus arbitrary, and methods such as gap coding (Almeida & Bisby 1984; Archie 1985; Mickevich & Johnson 1976) are not especially helpful in this situation. First, 100 matrices of 100 characters were modelled. The 100*100 continuous characters were then discretized (into states with equal intervals). The discretization of each character into eight, six, four and two states allows us to test the efficiency of ordered MP, unordered MP and 3ta with various numbers of states, for a total of 500 matrices; this is our first test.

The second test with Brownian motion addressed coding errors, and more precisely errors in the ordering scheme. The simulation consists of 100 new matrices of 100 characters of 10 states (conditions used to test topology and polarization in Rineau *et al.*, 2015), each with the coding of the true ordering reflecting the ideal situation for an analysis.

Then, the true scheme was modified according to two alternative scenarios to introduce errors, as shown in figure 2.20. The first scenario implied few errors: we produced 100 matrices with, for each character, a single permutation between adjacent states (fig. 2.20, scenario 1), and then added permutations iteratively to a maximum of five permutations per character. The second scenario differed in the severity of errors; in this case, the permutations were between states at opposite extremes of the transformation series, which should be far worse for ordered parsimony and 3ta (fig. 2.20, scenario 2). Each scenario of five coding iterations plus the perfect (initial) ordered scheme led to a total of 1100 matrices, analyzed in ordered parsimony, unordered parsimony and 3ta.

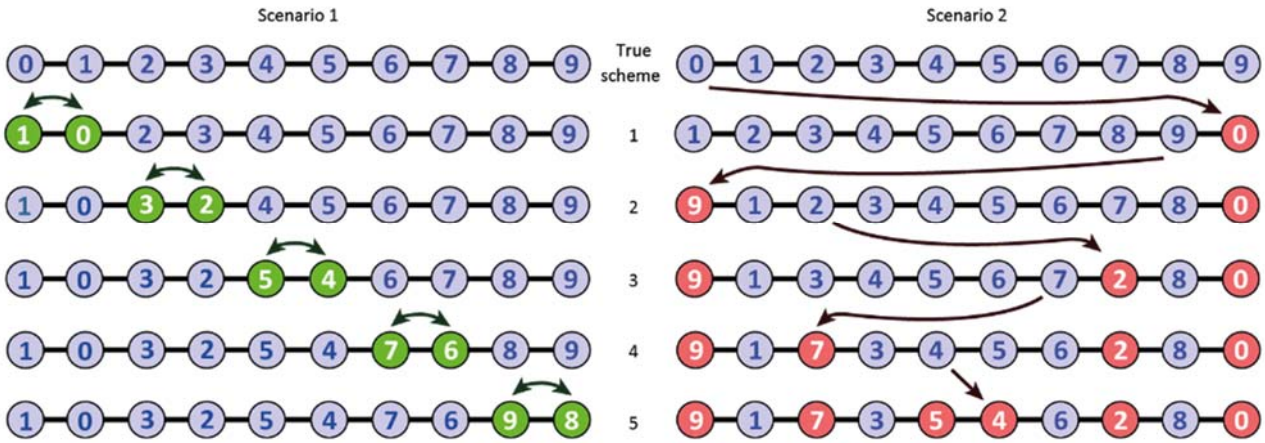


Figure 2.20. Two scenarios of errors in character ordering (in a character with 10 states) with five stages of degradation. Scenario 1 represents the case with slight permutation errors involving neighbors. Scenario 2 represents the worst case possible, with states displaced as far as possible from its neighbors. Coloured states represent permuted states and arrows represent permutations.

Another set of 100 matrices of 100 binary characters was generated with another evolutionary model, the asymmetrical Markov-k model. Parameters of the model were set to simulate irreversible characters by fixing the probability of going from states "0" to "1" at 0.1 and the probability of going from states ‘1’ to ‘0’ at 0.0 (forbidden).

These simulations tested only the effect of correlated, phylogeny-dependent errors and homoplasy. Finally, we took the 100 original matrices generated by Brownian motion with 100 characters and four states per character. We generated for each cell of each matrix a probability p=0.20 of change from the correct value generated under Brownian motion to

a random value. This was done using the RAND function of Microsoft Excel 2013©. This allowed the generation of completely uncorrelated errors.

4.2.1.3. Tree searches

All 1700 matrices of 100 characters were analyzed with both MP (ordered and/or unordered when characters had more than two states) and 3ta. The result of each analysis was represented by a strict consensus tree whenever more than one tree was found. All analyses were undertaken using PAUP* 4.0b10 (Swofford, 2003). In order to perform 3ta analyses, the matrices of triplets with fractional weighting were generated using LisBeth 1.0 (Zaragüeta *et al.*, 2012). All analyses performed on PAUP used the heuristic TBR algorithm with taxon addition and 50 random replicates after verification of their convergence.

4.2.1.4. Tree comparisons

The optimal trees resulting from the analyses were then compared to the initial reference topology (fig. 2.19). To obtain percentages of accuracy for each analysis, the 'Inter-Tree Retention Index' (ITRI thereafter) introduced by Grand *et al.* (2013) was computed. Several methods have been developed to compare phylogenetic trees. However, most use unrooted trees (Estabrook *et al* 1985; Robinson & Foulds 1981) to calculate distances between trees. Phylogenetic information, i.e. relationships between taxa, is nevertheless represented by a rooted tree. The ITRI measures a degree of congruence by using triplets as elementary phylogenetic assertions. The method differs from that advocated by Critchlow (1996) by using fractional weighting, a correction to eliminate triplet redundancy (Mickevich & Platnick 1989; Nelson & Ladiges 1992). The ITRI has also the advantage of differentiating true resolution (TR - a percentage of the true relationships, the relationships present in the reference tree that also occur in the obtained trees) from false resolution (FR - a percentage of the false relationships, the relationships that are not present in the reference tree that occur in the trees obtained by our searches). When necessary, we synthesized the general efficiency of an analysis using a standardized efficiency percentage, computed as (TR-FR+100)/2 (TR and FR are percentages; the efficiency TR-FR is comprised between -100 and 100; added to 100 and divided by 2, the efficiency is standardized in percentage. In such a measurement, an efficiency of 50% is

uninformative (the number of correct clades equals the number of wrong clades). A lower efficiency indicates more artefactual than correct resolution. However, in these analyses, we are interested only in comparing the relative efficiency of ordered MP, unordered MP, and 3ta.

4.2.1.5. Test of the significance of the results

For each set of 100 matrices, two means, one for TR and one for FR, are computed. The Wilcoxon test with signed ranks allows us to compare the ITRI computed from the same matrices but analyzed by different methods, and with different kinds of discretization. We selected this non-parametric test because none of the samples follows a normal law (as shown by a Shapiro-Wilk test), and the variances were uneven (shown by a Fisher test). Linear regressions were computed to test the effect of the number of states and the number of errors on ordering schemes. The false discovery rate procedure (Benjamini & Hochberg 1995) was used because of the high number of tests, to discard false positives. All tests were computed on R 3.0.3 and XLSTAT Pro 2014.2.

4.2.2. HIV character set

The first empirical comparison between parsimony and 3ta was first made using the HIV molecular sequences (population sequences from the *env* V3 and $p17^{gag}$ regions, known to be sufficiently variable for phylogenetic analyses) of Leitner *et al.* (1996), available on GenBank (access numbers: from U68509.1 to U68521.1), for the ingroup. Two outgroup sequences were used, one African HIV dated from 1957 (GQ431830.1) to ensure that the virion was outside of the ingroup, whose last common ancestor is dated from 1981, and a SIV (Simian Immunodeficiency Virus) sequence from a macaque (AF041984.1), far more temporally distant from the ingroup. The sequences of 439 pb were manually aligned. Insertion and deletion events were coded as binary characters. The two matrices include 14 taxa and 23 parsimony-informative characters (all unordered) with the HIV outgroup (29 with the SIV outgroup), and were analyzed with parsimony on PAUP* 4.0b10 and in three-taxon analysis on LisBeth 1.0, both in branch and bound. Some sequences included polymorphism. In 3ta, polymorphism produces a repetition of terminal taxa in the character-tree because a polymorphic terminal taxon is assigned to more than one terminal state. This repetition is caused by paralogy (Fitch, 2000, Nelson, 1994). Paralogy-free

subtree analysis (Nelson & Ladiges 1996; Zaragüeta *et al* 2012), initially proposed for biogeographic paralogous taxa (created by the presence of taxa in more than one area) was used to deal with repeated terminal polymorphic taxa in the 3ta analyses.

4.2.3. Copied manuscript character set

The evolution of copied texts through time can be studied through theories and techniques very similar to cladistics (de Pinna *et al.*, 2016). It is our second empirical comparison between parsimony and 3ta. The protocol of their generation is described in Spencer *et al* (2004). A text used as an ancestral sequence was copied by volunteer scribes. Then, the previously handwritten copies were copied, and so on, thus generating a known phylogeny of the manuscripts. We aligned the words of the various text sequences by hand. The ancestral sequence contains 834 words (49 sentences). We coded a new character set revealing four types of possible events: mutations (copying errors implying one or several letters of a single word; 'ordinance' becoming 'audience' is, for example, coded as a single event); permutations ('what, single-handed' becoming 'single-handed, what'); fragmentation of a word into several ('gadflies' in 'gad flies'); and insertions/deletions of words. The coding resulted in a parsimony and a 3ta matrix. To limit search time (and keeping the matrix small enough to use a branch and bound algorithm for both methods), several OTU's were randomly suppressed to produce the matrices for parsimony (PAUP) and 3ta (Lisbeth). The matrix contains 13 taxa and is rooted on the known ancestral sequence. It contains 88 informative characters.

4.3. Results

4.3.1. Ordering and number of states

Efficiency increases linearly with the number of states (from two to four, six, eight and 10 states) for ordered MP and in 3ta. In contrast, efficiency in unordered MP decreases linearly with the number of states (fig. 2.21; table 2.3). The efficiency of 3ta does not change exactly as MP. It is the most efficient method with binary matrices (77.3%, compared with 73.2% for MP). With four states, performance of both methods is similar (3ta: 78.3%, ordered MP: 78.0%) because MP efficiency increases quicker with additional states. Finally, MP becomes more efficient than 3ta when more than four character states

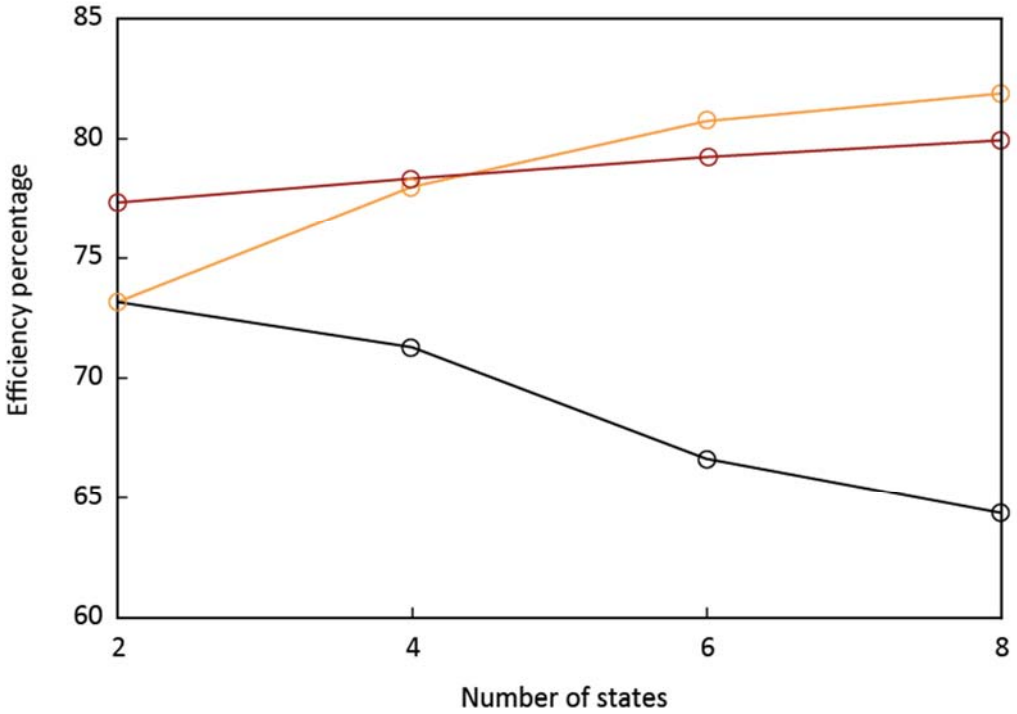


Figure 2.21. Impact of number of states in ordered, unordered MP and 3ta on the efficiency of the methods. The difference in efficiency between methods was assessed by iterative recoding of the same matrices with two, four, six and eight states. Orange: ordered MP; red: 3ta; black: unordered MP.

number of states	Ordered			Unordered			3ta		
	TR	FR	Efficiency	TR	FR	Efficiency	TR	FR	Efficiency
2	69.89	23.55	73.17	69.90	23.55	73.17	81.90	27.20	77.35
4	76.34	20.33	78.00	68.22	25.63	71.29	83.14	26.43	78.35
6	81.91	20.42	80.74	64.47	31.20	66.63	84.02	25.52	79.25
8	84.28	20.51	81.88	58.60	29.90	64.35	84.27	24.39	79.94
p-value	.	.	0.0387	.	.	0.0139	.	.	0.0032
intercept	.	.	71.23	.	.	76.64	.	.	76.56
slope	.	.	1.44	.	.	-1.56	.	.	0.43
R^2	.	.	0.92	.	.	0.99	.	.	0.97

Table 2. 3. Impact of the number of states on MP and 3ta efficiency. Probabilities of the null hypothesis (no impact of number of states on efficiency) were tested through linear regressions.

are coded, though the difference between the methods is small (for six states, 3ta: 79.2% and ordered MP: 80.7%). The differences between 3ta and ordered MP can be analyzed more precisely: the TR is higher in 3ta than in ordered MP (81.9% vs 69.8% for binary coding) in most cases. Only with at least eight states do we see an identical TR of 84.3%. However, the FR is always slightly lower (between 3 and 6%) in ordered MP than in 3ta.

4.3.2. Ordering and errors

To test the impact of errors in the ordering scheme, the reference was unordered MP (efficiency: 69.0%). For the first set of errors (scenario 1, minor errors), ordered MP is always superior to unordered MP (about 18.3 to 21.2% more efficient), with only a very slight decrease of 2.6% from the smallest to the greatest number of errors (fig. 2.22). For the second set of errors (scenario 2, major errors), ordered MP becomes drastically less efficient than unordered MP between the second and the third error. The third to the fifth errors are dramatic in terms of efficiency, where ordered MP can lose a maximum of 42.3% efficiency. Linear regressions of these data show how efficiency decreases as the number of ordering errors increases (table 2.4).

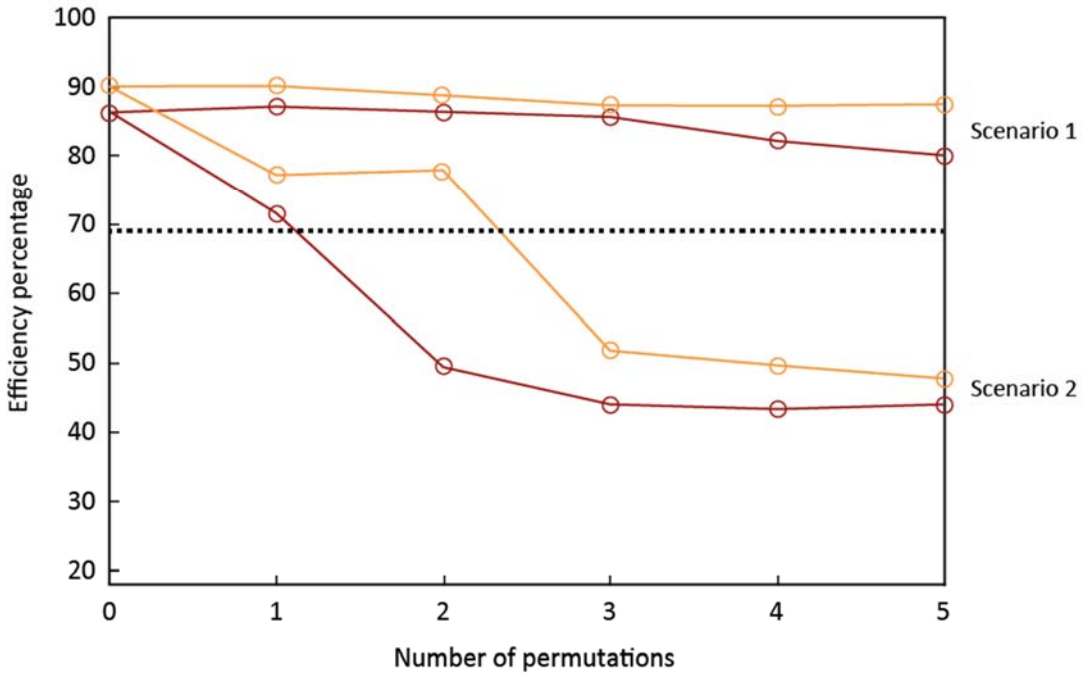


Figure 2.22. Impact of ordering errors shown in figure 2 on MP efficiency. The difference in efficiency between methods was assessed by iterative recoding of the same matrices with the various ordering schemes with more and more permutations. Scenario 1 shows MP's behavior with slight errors; scenario 2 shows the worst possible errors in the ordering scheme. Orange: ordered MP; red: 3ta. Unordered MP is constant because the ordering scheme is not coded, and is represented as the black dotted line.

	scenario 1		scenario 2	
number of errors	MP	3ta	MP	3ta
0	90.14	86.47	90.14	86.47
1	90.17	87.36	77.26	71.73
2	88.83	86.63	77.94	49.41
3	87.39	85.86	51.83	44.02
4	87.34	82.35	49.7	43.37
5	87.56	80.36	47.81	44.02
p-value	**0.0121**	**0.0203**	**0.0048**	**0.0185**
intercept	90.2	88.15	88.67	78.12
slope	-0.65	-1.32	-9.15	-8.65

Table 2. 4. Impact of errors in the ordering scheme on MP and 3ta efficiency, based on characters with 10 states. Probabilities of the null hypothesis (no impact) were tested through linear regressions. Two kinds of errors were tested: minor (scenario 1) and major (scenario 2) errors. See figure 2 for explanations of both scenarios. Unordered parsimony for these data yields an efficiency of 69.0%.

4.3.3. Irreversible characters

As shown in figure 2.23-1 and table 2.5, MP efficiency is significantly higher under a model generating only convergent characters than under a Brownian motion model (BM hereafter) generating convergences and reversals (78.6% and 73.2%, respectively; p-value: < 0.0001). 3ta efficiency appears to be the same in both models (76 to 77.3%, p-value: 0.2572). The loss of efficiency of MP results from a 20.2% drop in TR (in BM compared to irreversible characters), partly compensated by a loss of 9.3% of false resolution (FR). Reversals in 3ta create the opposite effect: an increase of true resolution of 9.2% and an increase of 6.5% of FR. Unordered MP was not tested (separately from ordered MP) here because the data used for this test are binary.

4.3.4. Uncorrelated errors

The control is here on BM with four states (78.0% efficiency for ordered MP, 71.3% for unordered MP and 78.3% for 3ta; table 2.6). Uncorrelated homoplasy (error randomly added to the data matrix) added to simulations under BM decreases efficiency of all methods (fig. 2.23-2). Unordered MP scores are the worst (44.4%) in this context. Efficiency of 3ta also decreases to 49.2%, but its performance remains significantly ($p < 0.0001$) superior to MP (46%). We highlight another interesting result: 3ta always yields a single most parsimonious tree in these specific simulations, though not always fully

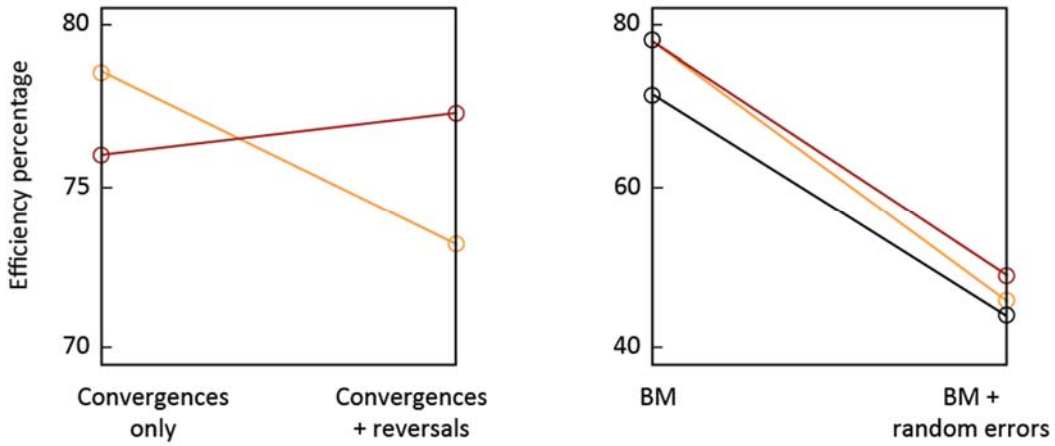


Figure 2.23. 1 – Impact of reversals on phylogenetic reconstruction methods. Left: the evolutionary model used is an irreversible model, generating only convergences, showing better performance of MP than 3ta (p-value: 0.02153). Right: a standard Brownian Motion model is used, generating both convergences and reversals; 3ta performs better than MP (p-value: 0.0004). 2 – Impact of randomly generated errors on phylogenetic reconstruction methods. Left: standard Brownian motion model (no significant difference between methods; p-value: 0.4402). Right: same model with random errors added; 3ta performs better than MP (p-value: <0.0001). Orange : ordered MP; red: 3ta; black: unordered MP.

resolved (in all of the 100 analyses), whereas MP yields an average of three most parsimonious trees per analysis.

4.3.5. Empirical tests

All characters are unordered in the analyses of empirical datasets. In the HIV example, 3ta is more efficient than unordered MP with both outgroups (table 2.7). 3ta efficiency is 96.3% for the HIV outgroup and at 57.6% for the SIV outgroup. It is between 17.6% and 5.0% more efficient than MP, which shows an efficiency of 79.1% with the HIV dataset, and 52.6% with SIV. With the published dataset from Spencer *et al* (2004) as we recoded it, the results are more similar between methods (table 2.8) and the dataset seems equally well handled by both methods. MP is slightly more efficient than 3ta with a higher TR and a lower FR, and an efficiency of 92.7% versus 91.7% for 3ta.

	Random error test	
	BM	BM with random error
Ordered MP	78.0	46.0
Unordered MP	71.3	44.4
3ta	78.3	49.2
p-value Ordered MP/3ta	0.7002	> 0.0001

Table 2. 6. Impact of homoplasy by random errors on phylogenetic reconstruction on ordered MP, unordered MP and 3ta. Random errors are generated on matrices generated with Brownian motion on four states and compared to analyses of the same matrices without random error. A Wilcoxon test is used to test the significance between ordered MP and 3ta.

	Reversal empirical test	
	BM	irreversible model
MP	73.2	78.6
3ta	77.3	76.0
p-value Ordered MP/3ta	0.0004	0.0215

Table 2. 6. Impact of reversals in the evolutionary model on MP and 3ta. An irreversible model (matrices modeled without reversals and binary characters) is analysed with both methods, and compared to Brownian motion with binary characters. A Wilcoxon test is used to test the significance of the difference of the results obtained with ordered MP and 3ta. Given that the characters used for this simulation are binary, there is no freedom in ordering scheme and hence, no difference between ordered and unordered MP.

4.4. Discussion

4.4.1. Effect of the number of states

When comparing ordered and unordered MP using data generated through BM (fig. 2.21), our results show a linear correlation between the number of states and the efficiency of all methods, with contrasting tendencies (table 2.3). The efficiency of ordered MP increases with the number of states (when the ordering scheme is correct). This is expected because the phylogenetic information content of character states increases linearly with the number of states if the latter are ordered (*i.e.* a binary character can support only one clade, but a ternary ordered character will support two nested clades). However, this is not true for unordered states because in the worst-case scenario in which the number of states equals the number of OTUs, an unordered character has no phylogenetic information content, whereas an ordered character has a maximal information content (it would specify a fully-resolved tree, if not contradicted by other characters). Thus, splitting characters into many

		VIH phylogeny				
		Number of trees	RI	TR	FR	Efficiency
MP	VIH	49 (95 steps)	0.69	58.2	0	79.1
	VIS	156 (217 steps)	0.64	38.6	33.5	52.5
3ta	VIH	1	0.74	97.1	4.6	96.2
	VIS	1	0.85	54.9	39.9	57.5

Table 2. 8. Empirical test of cladistic analysis by unordered MP and 3ta of HIV virion. The analysis is made with two alternative outgroups, a close HIV and a much more remote SIV. Results present the number of trees for each analysis, the number of steps for MP (the concept being meaningless in 3ta), the retention index (RI) and the efficiency composed of TR and FR.

	Manuscripts phylogeny				
	Number of trees	RI	TR	FR	Efficiency
MP	2 (185 steps)	0.68	100	14.6	92.7
3ta	1	0.76	99.8	16.5	91.6

Table 2. 8. Empirical test of cladistic analysis by MP and 3ta of copied manuscripts. Results present the number of trees for each analysis, the number of steps for MP, the retention index (RI) of the characters on the obtained tree, and the efficiency composed of TR and FR.

states should not be undertaken when relationships between character states are poorly constrained. On the contrary, when relationships between states appear clear, as for linear morphoclines, maximizing the number of ordered character states will maximize the general efficiency of MP to retrieve the correct tree.

Previous results on the comparative efficiency of the two ordered methods, MP and 3ta, have been contradictory (Grand *et al* 2013; Rineau *et al* 2015). Our simulations show a relationship between the number of states and the efficiency of both methods, which explains the contradictory results in previous studies. MP is more efficient than 3ta when the number of states per character is higher than four (fig. 2.21); this was the case for the study of Rineau *et al.* (2015), which simulated characters with 10 states each. With four character states, as in most molecular studies, ordered MP and 3ta have the same ability to find the true phylogeny (p-value: 0.4402). However, 3ta performs better than MP with binary characters, which is the most common situation in paleontological studies, with 4.2% more efficiency (p-value: 0.0004). This explains why Grand *et al.* (2013) found more favourable results for 3ta, because most of their characters contained three or fewer states. In simulations with binary characters, the artefactual resolution is always significantly

higher in 3ta than in MP (3.6 to 6.1% higher; table 2.3). This difference in resolution reflects the fact that the number of most parsimonious trees is always lower in 3ta, leading to a more resolved strict consensus. Nevertheless, 3ta retains more correct phylogenetic information than MP with decreasing number of states. The true resolution at eight states is equivalent; 2.1% better in 3ta with six states, 6.8% better with four states and 12% better with binary characters. The true resolution decreases in MP with decreasing number of states mostly because resolution decreases. To sum up, the 3ta seems able to deal with the information of binary or ternary characters, while MP might better capture the information contained in the ordering scheme of characters with more than four states.

4.4.2. Errors in ordering schemes

The objective of the second set of simulations was to compare the performance of ordered and unordered methods when the ordering scheme contains errors (fig. 2.22; table 2.4). Adding errors into the coding scheme of simulations under Brownian motion decreases the efficiency of both ordered MP and 3ta, as expected, but not to the same extent. The comparison with unordered MP shows that ordering characters strongly constrains the hypothesis. To investigate precisely the change in efficiency in the relationship with the number of errors, the analyses were performed on characters with 10 states. 3ta under BM with more than four states is always less efficient than MP, so ordered and unordered MP are emphasized in this discussion. Errors of ordering can vary in many ways. In the first scenario, when errors are minor (fig. 2.20), ordered MP has always a strong advantage compared to unordered MP (approximately 20% more efficient). In the second and worst case scenario, ordered MP efficiency decreases dramatically when three or more inversions are present (six or more states are wrongly placed in the scheme) to become less powerful than unordered MP. Thus, even with one or two dramatic errors (inversions) in the coding scheme, ordered MP remains more efficient than unordered MP. More than half of the character-state positions (one of the worst-case scenarios when ordering states) need to be mistaken to make ordering disadvantageous with MP. With 3ta, results are always slightly worse than for ordered MP, with the advantage of the ordering scheme being lost, in the case of severe errors, with two (rather than three) or more inversions between states (table 2.4). This poor efficiency of 3ta in this case might be partly due to the high number of character states used here (10), a behaviour already highlighted by Rineau *et al.* (2015).

This result was expected because in 3ta, hypotheses of homology are phylogenetic arguments addressed to support a phylogenetic hypothesis. If the arguments are massively inconsistent, one of the main theoretical premises of the method is violated.

Thus, several subtle errors in the ordering scheme are not enough to suppress the advantage of ordered analyses. For this advantage to be lost, the ordering scheme must be severely flawed. An example would be using a similarity criterion where in reality, genetic determinants cause a discontinuous evolution that does not yield clines, such as merosity in flowers (the fact for floral pieces to be organized in multiples), in which case going from three axes to six is probably easier in certain clades than going from three to four axes (Decraene & Smets 1994).

4.4.3. Effect of evolutionary reversals and convergences

The aim of this set of simulations was to test the efficiency of cladistic methods on phylogenies constructed with a model forbidding reversals of attribute (only convergences). These results were then compared to those obtained by analysing datasets constructed with a model that allows reversals. We aimed to test their differential performance because proponents of MP (Farris *et al* 1995; Farris & Kluge, 1998; Farris, 2012; Harvey, 1992; Kluge 1994) and 3ta (Nelson, 1996; Nelson & Ladiges, 1996; Nelson *et al.*, 2003; Platnick *et al.*, 1996; Siebert & Williams 1998; Williams & Ebach, 2005; Zaragüeta and Bourdon, 2007) have crystallized heated debates on how methods must manage reversals on a theoretical basis. The handling of reversals in 3ta and MP reflects deep differences in how characters are conceptualized in both approaches. Taxic homology used in 3ta diverges in the treatment of plesiomorphies from the transformational homology used in MP (Brower 2014; Carine & Scotland 1999; Farris 2014; Scotland 2000). For 3ta proponents, systematists propose putative hypotheses of synapomorphy. The root of a character may, or may not, be a state; in the former case, it is a plesiomorphy (Nelson, 1994). At a defined taxonomic level of an analysis, 3ta may consider a loss as uninformative because it is neither a homology nor a synapomorphy (Rineau *et al* 2015). However, the most common way to deal with homology is that used in MP, in the context of transformational homology. If 0 is a plesiomorphy and 1 an apomorphy, the transformation from 0 to 1 supports a clade. However, MP also supports a clade if another transformation appears in the tree from 1 to

0. Moreover, MP may support both transformations at the same time, for different clades. Thus, evolution of the characters as in matrices (MP) or their nested status (3ta) need to be carefully assessed under both approaches.

One of the main arguments used against 3ta was its inability to recognize 'reversals' (De Laet and Smets, 1998; Farris, 2010; Farris 2012; Farris and Kluge, 1998; Farris *et al.*, 2001; Farris *et al.*, 1995). Critics have called reversals plesiomorphic character-states that are considered synapomorphies at less inclusive nodes through the minimization of the Manhattan distance implemented in MP. Simulations with irreversible evolution were made to provide new empirical data relevant to the theoretical discussions on reversals and 3ta. Their purpose was to verify how 3ta, which does not modify proposed hypotheses of homology so that they fit the tree, compares in performance with MP. If 3ta was unable to manage reversal in phylogenetic reconstructions, the method should have a drastically lower efficiency in the presence of reversals than in irreversible evolution. Our results (fig. 2.23-1; table 2.5) falsify this assumption because MP efficiency collapses when adding reversals in the evolutionary model, whereas 3ta efficiency remains stable (p-value: 0.2572 between 3ta with and without reversals in the evolutionary model). This result might be linked with the fact that in 3ta, reversals increase the number of incompatible triplets, whereas under MP, reversals increase the number of evolutionary steps. Supernumerary steps may distort the entire character, whereas triplets divide the character into minimal parts that can be independently retained or rejected. Our results show for the first time that the main criticism against 3ta, may be empirically unfounded. Thus, in the absence of reversals, the performance of 3ta should have been best.

Conversely, with an irreversible evolutionary model, MP performs statistically better than 3ta (2.6% better, p-value: 0.0215). The explanation for this phenomenon is probably due to the divergence of character representation between steps (instance of transformations between states) and triplets (statement of kinship relationship based on a derived state). Take the example of a homoplastic character state that support two clades in a tree understood as a convergence. These two clades in MP are supported by independent steps. It does not matter where these convergent transitions supporting both clades are situated on the topology. In 3ta, the closer the clades are to each other, the higher the number of triplets will support the tree. The example on figure 2.24 illustrates the difference

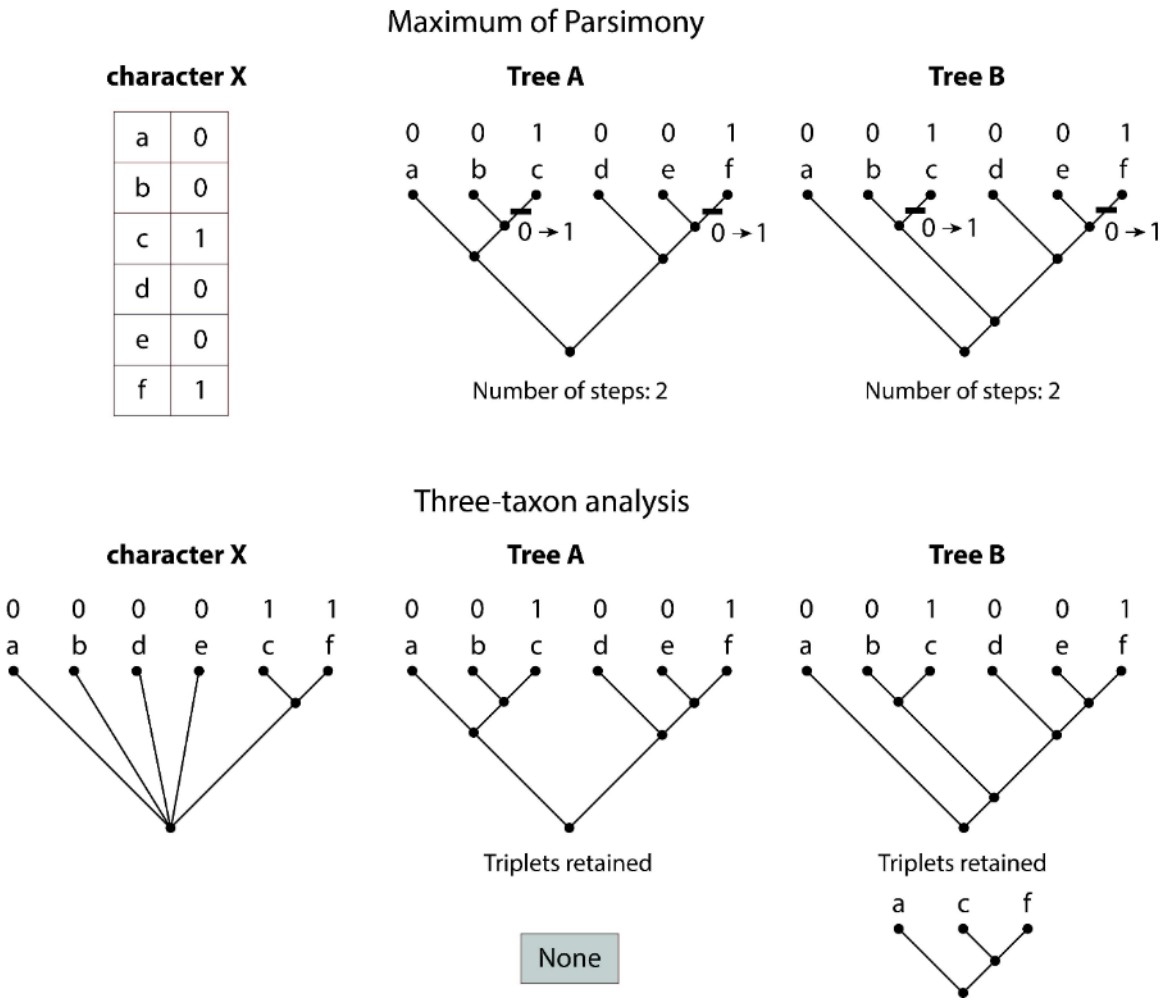


Figure 2.24. Example of differential behaviors between MP and 3ta. The example shows two phylogenies, A and B, and a character *X*. Character *X* is represented in a matrix in MP, and through a hierarchical character (*a*, *b*, *d*, *e*, (*c*, *f*), with (*c*, *f*) as derived character state. In MP, the optimization of both trees always leads to two steps. Regarding only character *X*, they are both suboptimal. In 3ta, the optimal tree is chosen as the structure retaining the maximal number of hypotheses (i. e. triplets resulting from the character decomposition). Tree A is completely incompatible with character *X*, as no triplet is retained. Regarding only character *X*, tree B is a better phylogeny as it retains one triplet from the initial character. In 3ta, triplets tend to 'attract' nodes that independently acquire a derived state, whereas this phenomenon does not occur in MP.

between MP and 3ta. Regarding only the character X, both trees A and B are equivalent under MP, as the character distribution implies two steps on both topologies. The hierarchical 3ta character (*a*, *b*, *d*, *e*, (*c*, *f*)) is completely incompatible with the phylogeny A ((*a*, (*b*, *c*)))), (*d*, (*e*, *f*))). On this character, no triplet can be retained. Regarding only that character, this phylogeny is less optimal that the second phylogeny B (*a*, ((*b*, *c*), (*d*, (*e*, *f*)))), which is compatible with one triplet *a*, (*c*, *f*), because of the proximity between the clades

that bear the derived state. The criterion for choosing a phylogeny in MP is to minimize the number of steps. In 3ta, it is to choose the most congruent phylogeny, the one which agrees with the maximum number of phylogenetic hypotheses represented by triplets. While the supernumerary transformation steps are the same in both topologies under MP, there is a 'node attraction' between clades sharing homoplasic convergent states in 3ta. This may reflect the difficulty of 3ta to deal with convergent features, especially when they are far from each other on the tree.

4.4.4. Random errors, homoplasy and phylogenetic inference

The purpose of our final set of simulations was to propose an alternative to the standard experimental simulations usually performed to compare phylogenetic methods (Wright & Hillis 2014). In standard evolutionary models, homoplasy is always of evolutionary significance. In evolutionary models, errors are always considered only as convergences and/or reversals. In these final simulations, we added random errors to matrices produced by BM. These random errors simulate errors unrelated to evolution, such as human errors. These include errors of interpretation of the DNA alignment, errors of interpretation of structure due to taphonomical problems (Sansom *et al* 2011), erroneous character coding or simply mistakes when scoring cells in a matrix (such as ambiguity of the sequencing at a particular position, or a measurement error on a structure, etc.). Problems of homoplasy generating events that do not follow the dichotomy reversal/convergence, such as lateral gene transfert (generating homoplasy when forced on a hierarchical topology), can be assimilated to these "random errors" (a phylogenetic network becomes partially random from a hierarchical point of view). The differences between errors in the matrix and genuine homoplasy – phylogenetically structured homoplasy such as convergences and reversals – have been little investigated, but this topic is relevant to compare methods in different situations. The interpretation of characters in DNA sequences, for example, must be a little less sensitive than comparative anatomy to interpretative human errors, although other sources of error, such as alignment ambiguity or sequencing errors, may be common. The impact of such errors on phylogenetic analyses needs to be minimized by using methods that does not overinterpret the dataset. Differences in efficiency were detected following addition of interpretative errors to the dataset (both with BM and with perfectly congruent datasets without convergences or reversals). Even

though MP and 3ta lose efficiency when 20% of random errors are added, 3ta is 3.2% more efficient than ordered MP in the presence of such errors to reconstruct the true phylogeny (fig. 2.23-2; table 2.6).

The fact that reversals are never used to support clades in 3ta may limit overinterpretation of their evolutionary significance. 3ta integrates into its definition the possibility of error, defining homology as only "*the relationship among parts of organisms that provides evidence for common ancestry*" (Brower & de Pinna 2012, 2014; De Pinna 1991). MP, in contrast, will maximize phylogenetic information through minimization of supernumerary steps. Using this method means assuming that if a hypothesized synapomorphy is rejected, the method might generate several other synapomorphies (implying supernumerary evolutionary steps), to fit inconsistent character-states, which may lead to an overinterpretation of the original characters and an increase of false resolution.

Homoplasy generated by interpretative errors and all types of human errors may be crucial in phylogenetic reconstructions given that such errors may not be uncommon in empirical datasets (e.g. Marjanovic & Laurin 2008, 2009). The comparison of phylogenetic methods should address their robustness to such errors in the data. We highlight for the first time through simulations the capacity of 3ta to handle this issue significantly better than MP. In systematics based on phenotypic data, human errors are added to other sources of homoplasy such as convergence and reversals, and should be better taken into account in future studies.

4.4.5. Empirical studies

The first set of analyses on HIV virion supports our conclusions drawn from the simulations. The analyses using the nearest HIV outgroup show 3ta to be much more efficient than MP (table 2.7). The main reason for the low performance of MP is certainly due to the high number of optimal trees (49 MPT), as shown in all our previous simulations. However, we note than neither of the 49 trees is the true tree in MP, nor the unique 3ta tree. With a farther outgroup (SIV), both methods decrease in efficiency, but 3ta remains more efficient than MP. However, 3ta efficiency seems to decrease more quickly with a more distant outgroup (38.7% decrease against 26.5% for MP). This sharp decrease may reflect

difficulties of 3ta in dealing with distant outgroups with long branches, as shown by Rineau *et al.* (2015). These results may also reflect the different ways in which MP and 3ta use outgroups. In 3ta rationale, outgroup rooting is better used to root trees of states. This strategy is more sensitive to outgroup choice than the MP handling because it assumes the outgroup to bear only plesiomorphic states, an unrealistic supposition.

Finally we show with 3ta analyses that a small number of informative characters (23) is sufficient to infer correctly a phylogeny when the polarization is reliable (outgroup criterion here).

The dataset on manuscript characters favors MP, which reaches its highest efficiency in this study (92.7%). 3ta shows slightly less efficiency (91.6%), perhaps because the true tree contains polytomies (Spencer *et al* 2004). Given that our results are synthetized by a strict consensus, the fact that MP retains more optimal trees than 3ta is advantageous here. Given that 3ta yields a less optimal, fully resolved tree, polytomies (uncertainties) are not represented, pushing up the score of false resolutions.

4.5. Conclusion

We used simulations here to propose new empirical arguments to fuel methodological discussions on character coding and phylogenetic methods and to develop new and more realistic evolutionary models. Debates between proponents of MP and of 3ta, and also on character coding and ordering have been mainly based on theoretical considerations. Model-based studies allow comparison of phylogenetic analyses using simulations of characters following models on reference trees. They provide a way to compare analyses by varying a single parameter. Empirical examples can also be found in the literature of known genealogies (Leitner *et al* 1996; Spencer *et al* 2004), and our study has used both approaches.

The effect of the number of states per character is highlighted by our simulations (fig. 2.21). Logically, more character states in unordered MP lead to a greater number of compatible topologies and to a quick decrease of efficiency, when the underlying evolutionary model logically leads to ordered states. The ordering of character states appears to be essential to retrieve the true tree. Both 3ta and MP with ordered characters

see their efficiency increase with the number of character states. However, the two methods do not perform optimally under the same conditions. Their ability to retrieve the phylogeny is roughly equal with four states per character, MP is more efficient with more than four states, whereas 3ta is more efficient with fewer than four states. Because multistate characters with five or more states are rare in matrices, 3ta seems more efficient for most phenotypic datasets.

The simulations referred to in the previous paragraph were made with ordering schemes that were known and error-free. We performed additional simulations with controlled errors to determine under which conditions (with how much ordering error) ordering characters is useful. The results show that analyses using ordered datasets are unwise only when several severe ordering errors are introduced. Unordered MP performs correctly only when we can suspect several drastically different but equally plausible ways to order characters. Otherwise, ordering following morphoclines (for example) seems preferable.

Irreversible simulations were used to offer empirical arguments to fuel the debate on reversal management in phylogenetics, which has relied mostly on theoretical considerations so far. 3ta using taxic homology seems to perform better than MP and transformational homology in the presence of reversals. MP may over-interpret data by supporting clades with reversals (Brower & de Pinna 2012), leading to artificially supported false clades and a drastic increase of the number of most parsimonious trees, compared to 3ta. Alternatively 3ta may succeed better than MP at sorting the primitive absence from losses.

Our final simulations assessed the impact of random errors (not correlated with the phylogeny) that might be explained by errors that have nothing to do with evolution, such as mistakes, interpretation errors, alignment problems, taphonomical losses (Sansom *et al* 2011), etc. In our simulations, the first to include this new parameter (as far as we know), 3ta is also more efficient than MP, possibly because of the over-interpretation of inconsistent data by MP. This new source of error should be better integrated in future model-based simulations for experimental systematics because it can be a parameter of critical importance (*e.g.*, Marjanovic & Laurin 2008, 2009). Future simulation-based studies may include probabilistic methods such as Bayesian inference and maximum

likelihood. All probabilistic methods implement transformational homology. Consequently, they might react in a way similar to MP to human error, but this will need to be assessed.

Lastly, empirical studies put into perspective the fact that the ability of 3ta to retain fewer optimal trees may be advantageous or not, depending on the real tree. The greater resolution provided by 3ta proved useful to deal with the HIV dataset, but deleterious with the manuscript dataset, in which MP performed a bit better.

5. Conclusion du chapitre

Dans ce chapitre, nous avons proposé une comparaison entre la 3ia utilisant des caractères hiérarchiques et la parcimonie standard utilisant des partitions. La 3ia a été créée au départ comme une manière d'augmenter la précision phylogénétique des résultats d'analyse cladistique; de nombreux exemples ont été proposés afin de mettre en évidence les différences de comportements d'avec la parcimonie (Nelson and Platnick 1991; Platnick 1993; Nelson 1996; Platnick *et al.* 1996; Nelson *et al.* 2003; Williams and Ebach 2005). Ces arguments prennent place au niveau de la méthode. Au regard du développement de l'homologie taxique par rapport à l'homologie transformationnelle (Carine and Scotland 1999; Scotland 2000; Brower 2014, 2014), de nouvelles critiques de la parcimonie sont apparues, au niveau théorique cette fois. La partition et le pas de transformation impliquent nécessairement de la paraphylie, contredisant ainsi un des principes les plus fondamentaux de la théorie cladistique, celui que seuls les groupes monophylétiques sont naturels, les groupes monophylétiques étant artificiels. La 3ia utilise la hiérarchie (Cao *et al.* 2007, 2007; Zaragüeta i Bagils *et al.* 2012) pour la construction des caractères, excluant de fait la génération d'entités paraphylétiques. Plus généralement, les fondements théoriques et méthodologiques de la 3ia permettent d'adresser de nombreuses critiques à la parcimonie; la réciproque n'est pas vraie.

De futures études en 3ia, notamment empiriques comme nous en avons produits ici, devront permettre de proposer de nouvelles voies de recherches afin de continuer le développement de la méthode de 3ia. Les simulations informatiques permettent de mettre en évidence des différences de comportement qui peuvent être reliées à des considérations

méthodologiques. Nos simulations nous ont par exemple permis de mettre en évidence le fait que deux clades caractérisés par le même état de caractère (convergence) ont tendance à essayer de se rapprocher le plus possible l'un de l'autre sur une topologie. Ce phénomène est intrinsèque à l'utilisation des triplets.

Pour mettre en perspective ces résultats empiriques, il faut prendre en compte le fait que les simulations que nous avons utilisées portent un biais potentiel, celui du modèle d'évolution. Un modèle d'évolution est une représentation simpliste d'un processus complexe. Les analyses sur le HIV ou les manuscrits recopiés sont un essai pour se dégager de cette influence. Cependant, des analyses de grand nombre permettant de tester statistiquement la significativité des résultats obtenus est impossible, faute d'études suffisantes de ce type. De futures simulations informatiques basées sur la génération de populations soumises à la sélection naturelle et à la dérive génétique (Ofria *et al.*; Adami and Brown 1994; Hagstrom *et al.* 2004) devraient permettre de simuler l'évolution de populations dans un cadre historique, et de manière plus réaliste. Des simulations de ce type confrontant 3ia et méthodes à modèles (Maximum de Vraisemblance et Inférence Bayésienne) permettront également d'argumenter du statut de ces méthodes, à la logique « transformationnelle » similaire à la parcimonie, et dont les fondations théoriques sont encore trop peu discutées.

PARTIE 2. LES RUDISTES: ANATOMIE COMPARÉE, PHYLOGÉNIE, SYSTÉMATIQUE, PALÉOÉCOLOGIE

There had been aeons when other Things ruled on the earth, and They had had great cities. Remains of Them, he said the deathless Chinamen had told him, were still be found as Cyclopean stones on islands in the Pacific. They all died vast epochs of time before men came, but there were arts which could revive Them when the stars had come round again to the right positions in the cycle of eternity. They had, indeed, come themselves from the stars, and brought Their images with Them.

These Great Old Ones, Castro continued, were not composed altogether of flesh and blood. They had shape — for did not this star-fashioned image prove it? — but that shape was not made of matter. When the stars were right, They could plunge from world to world through the sky; but when the stars were wrong, They could not live. But although They no longer lived, They would never really die...

The Call of Cthulhu. Howard Philips Lovecraft, 1928

Chapitre 3. Taxonomic revision of the genus *Ichthyosarcolites* Desmarest, 1812 (Bivalvia, Hippuritida): an example of high intraspecific variability

Ce chapitre a été écrit en collaboration avec L. Villier et correspond à un article soumis dans la revue *Cretaceous research*.

1. Introduction 223
2. Material and methods 225
3. Morphology 235
4. Systematic paleontology 239
5. Chapter conclusion 260

1. Introduction

The Cenomanian fossils *Ichthyosarcolites* are one of the elder defined genus of rudist bivalves. They were described by Desmarest (1812) from internal molds found in Charentes (Western France), although the first occurrence of these fossils in the literature can be retrieved in the name Cucroïdes by Guettard (1786) who identified them as "*corps marins fossiles*" with the morphology of "*cercles pierreux*" (fig. 3.1-A). Desmarest (1812, 1817) considered the fossils as a link between "*hippurites*" and "*orthoceratites*" (cephalopods) because of a structure interpreted as a siphon (fig. 3.1-B). Later, d'Orbigny (1850) showed that the structure assumed as a siphon by Desmarest was a small cavity filled by tabulations, without communication between the chambers. D'Orbigny (1850) erected the genus *Caprinella* (= *Ichthyosarcolites*) within the rudists, close to *Caprina* because of similar shell coiling and the presence of canals in the shell. With the discovery of novel forms of rudists during the nineteenth century and the better understanding of rudist anatomy, the *Ichthyosarcolites* affinities became more difficult to assess. The hinge of *Ichthyosarcolites* shares similarities with Radiolitids (Skelton and Smith, 2000) and the presence of canals occurs also in Caprinuloideids and Antillocaprinids (Skelton, 2013a). Douvillé (1887) erected the family Ichthyosarcolitidae in which *Ichthyosarcolites* was the only genus until the description of *Curtocaprina clabaughinkisorum* (Mitchell, 2013a). *Ichthyosarcolites* was continuously enriched with new species description, to a total of nineteen nowadays, the latest being *Ichthyosarcolites alatus* from Central America described by Filkorn (2002).

The fossil record of *Ichthyosarcolites* is rich, with shells found in abundance in shallow marine deposits of Cenomanian age. The preservation state of specimens is usually bad, either found as internal molds or stuck into hard limestones. The specimens are traditionally studied from shell sections. The genus *Ichthyosarcolites* can be recognized firstly because of the huge development of its narrow right valve that can approach one meter in length, but the overall shell shape development expresses a high variability leading to problems of species delimitation.

We reassess the morphology of the genus *Ichthyosarcolites* using specimens from institutional collections and newly collected shells from the presumed type locality in

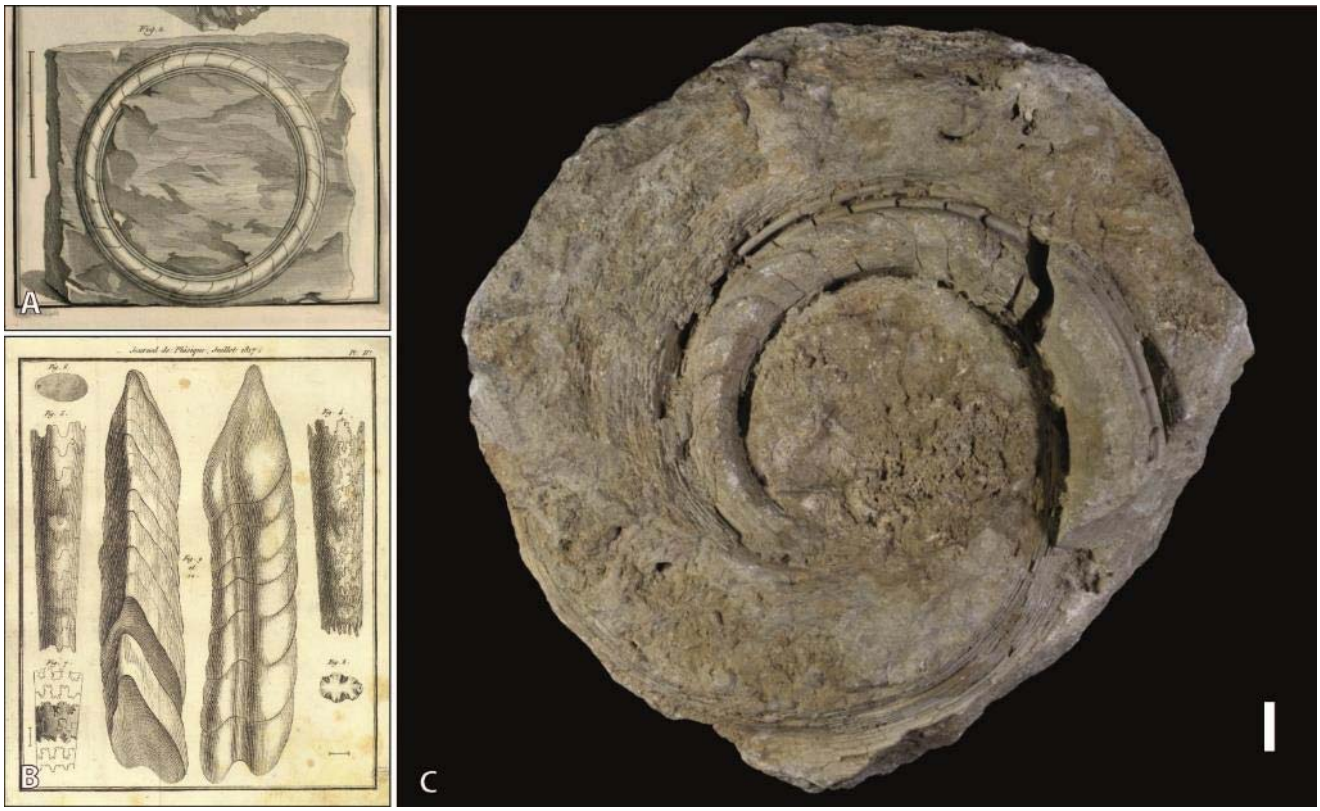


Figure 3. 1. A – Earliest illustration of *I. triangularis* as 'Cucroïdes' by Guettard (1786), who thought the fossil to be circular ('*pierres circulaires*'). B – Original drawings of the internal molds of *I. triangularis* (Desmarest 1817). C – Specimen of *I. triangularis* MNHN.F.R08003 from Île d'Oléron illustrated in d'Orbigny's Paléontologie Française (1847); this specimen is designated here as neotype of *Ichthyosarcolites triangularis* Desmarest 1812.

Charentes (France). These specimens were compared to the previously published descriptions. We also discuss the accuracy of a taxonomic framework grounded on flange counts and on the overall shell shape, that were previously used to define species within the genus. An analysis of flange's positions is used to compare the specimens and to test homologies among shells. We also made a discriminant analysis to see if any correlation or dependence exists between their internal anatomy and external flange's locations. Based on morphological character combinations that minimize information redundancy, a new taxonomic framework is proposed providing are a better phylogenetical significance.

2. Material and methods

2.1. Studied material

Investigations are based on literature survey, access to collection specimens at the AMU (Marseille, France) MNHN (Paris, France), UPMC (Paris, France), NMNH (Washington DC, USA), complimented with pictures of collection or field specimens for European and Mexican countries. Fieldwork investigations in Charentes focused on middle and upper Cenomanian levels in which *Ichthyosarcolites* is common as internal molds. *Ichthyosarcolites* are found on the north of Île Madame (45°96'N 1°12'W) in two distinct stratigraphic levels, the units C1 and C2 of the Cenomanian series (Moreau, 1993; Videt, 2003). *I. triangularis* occurs together with *Sphaerulites foliaceus* in grey wakestones and packstones of middle Cenomanian age that represent subtidal, quiet environments. The aragonite shells are dissolved but were often encrusted with oysters and other epibionts that preserved the shell outlines. At Pointe de Chaucre on the west coast of Île d'Oléron (45°98'N 1°39'W), the rocky shore exposes largely the unit B3 to C3 in which rudists (*Caprina adversa*, *I. triangularis* and *S. foliaceus*) are found in high density of individuals (Chéreau 1997). The rudists occur in carbonate deposits together with oysters (*Rhynchostreon suborbiculatum* and *Rastellum carinatum*), Nerineid gastropods, Chaetetid sponges and diverse benthic foraminifera depending on initial environmental conditions. At Pointe de Chaucre, a biostrome of large *C. adversa* and *I. triangularis* develops in beds representing high energy environment and shallow marine sandy bottoms, whereas quiet environments associate more often *I. triangularis* with *S. foliaceus*. Among the rudists, *Ichthyosarcolites* is the most widespread form, occurring in varied types of sedimentary environments, from shallow marine carbonate sands to upper offshore muds. The species is interpreted environmentally tolerant, being the first rudist to settle on shallow marine mobile and soft sediments (Chéreau 1997).

D'Orbigny (1850) provides good descriptions and illustrations of *Ichthyosarcolites* among which two figured specimens of *I. triangularis* are from Charras (Charente-Maritime, France; MNHN.F.R08003; fig. 3.1-C). An additional specimen of *I. triangularis* from Charras is available in UPMC collections. The d'Orbigny's collection includes

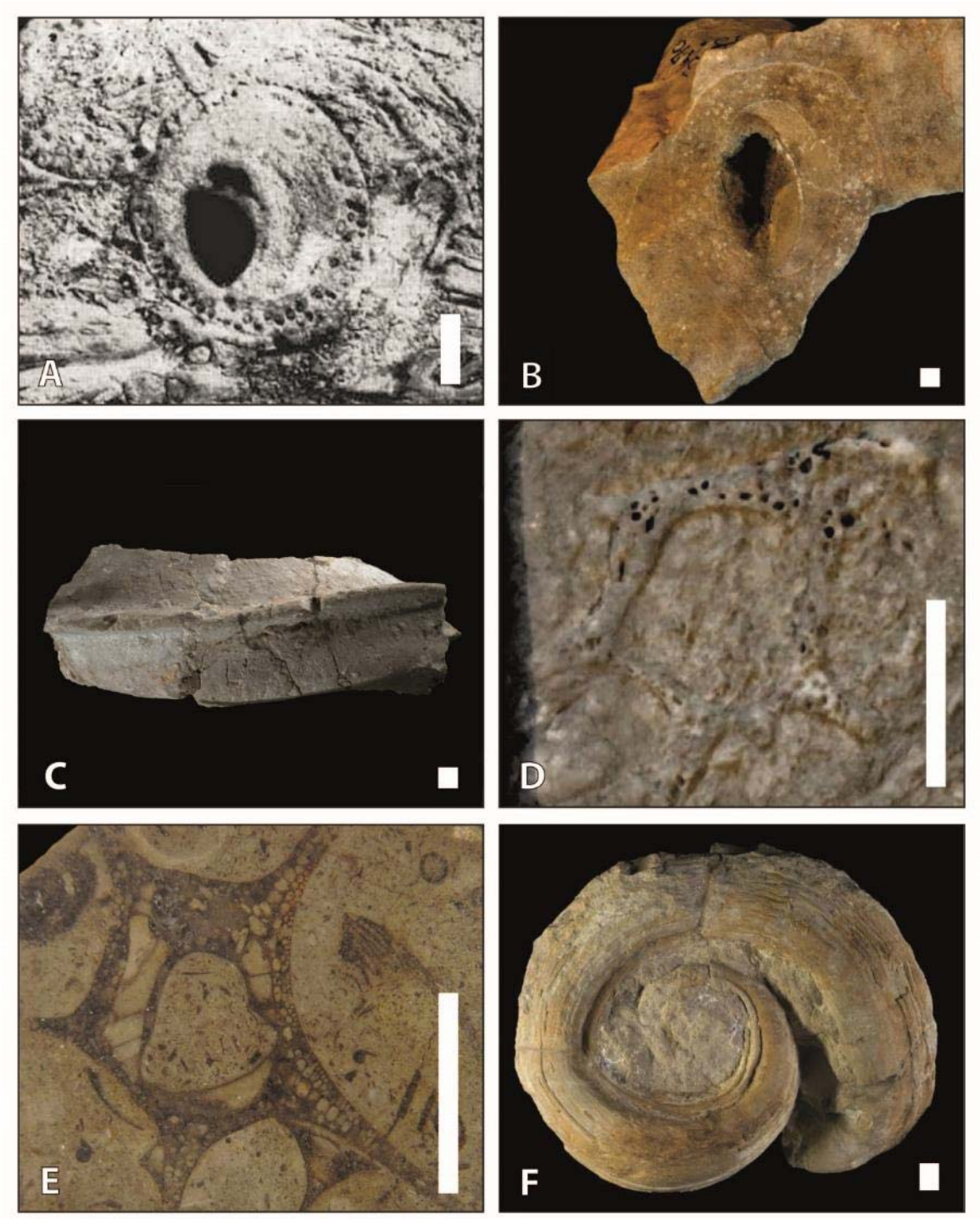


Figure 3. 2. Types of *Ichthyosarcolites* species. All scale bars = 1cm. A – Holotype of *Ichthyosarcolites rotundus* Polšak, 1967 (LGPZ.778) showing a cylindrical shell. Picture from original publication. Cenomanian, Fontanela bay (Istria, Croatia). B – Holotype of *Ichthyosarcolites monocarinatus* Sliskovic, 1966 (MB.3470) showing a small flange and round canals. Cenomanian, Dinaric Alps. Picture provided by Alceo Tarlao. C – Syntype of *Ichthyosarcolites bicarinatus* (Gemmellaro collection, MGUP.021.2-110) showing two flanges and a twisted body along the umbo-aperture axis. Cenomanian, Piemonte (Italy). Picture provided by Tony Costagliola. D – Holotype of *Ichthyosarcolites polycarinatus* Sliskovic, 1966 (MB.3482) showing seven elongated flanges and small capillary canals. Upper Cenomanian, Dinaric Alps. Picture provided by Alceo Tarlao. E – Holotype of *Ichthyosarcolites alatus* Filkhorn, 2002 (UNAM.6979) showing five flanges and a presumed juvenile attached to its ventral side. Left valve. The myocardinal apparatus and its associated cavities are visible. Lower Cenomanian, Guerrero (Mexico). Picture provided by Javier Aguilar-Pérez. F – Lectotype of *Ichthyosarcolites doublieri* d'Orbigny, 1847 (MNHN.F.R07971), lacking tabulae. Internal mold showing the highly developed right valve and the short, left valve. Upper Cenomanian, Martigues (Bouches-du-Rhône, France).

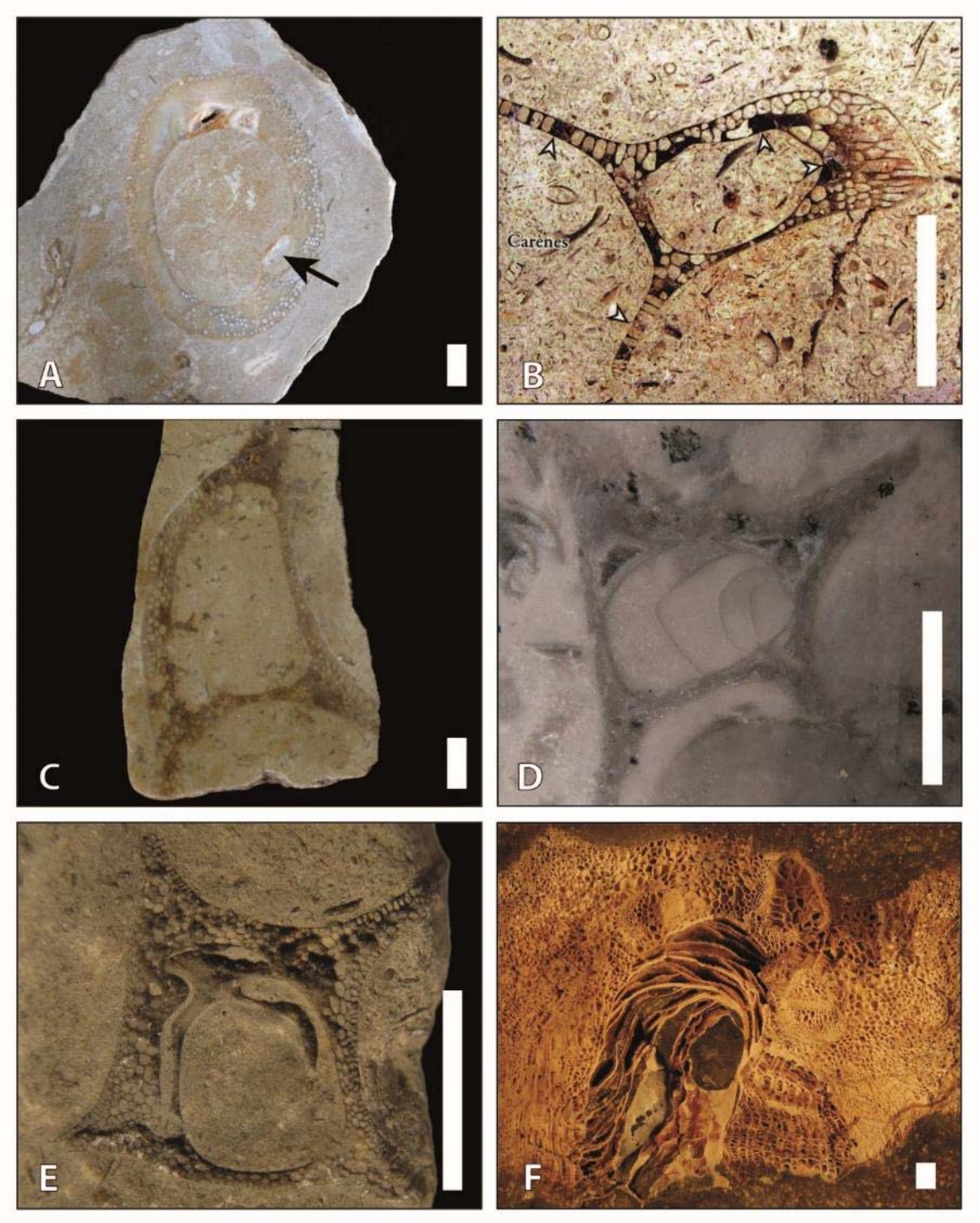


Figure 3. 3. Shells internal features of *Ichthyosarcolites*. All scale bars = 1cm. A – *I. rotundus* without any flange (AMU.ICH1) from La Bédoule (Provence, France). Arrow: posterior accessory cavity. B – Picture of *I. bicarinatus* from Chiki-Aouimeur (2010) showing two elongated flanges and big round canals (Aqaba, Jordania). C – Sliced right valve of *I. triangularis* with three flanges, the two sockets and the posterior myophoral cavity (AMU.HBES 85-a). D – Specimen of *I. bicarinatus* from Cañón de la Servilleta (Tamaulipas, Mexico) showing four flanges and thin tabulae. Picture provided by Javier Aguilar-Pérez. E – Mexican right valve with four flanges *I. bicarinatus* (from Filkhorn, 2002). The myocardinal apparatus from the left valve is visible in connection (UNAM.6970). Picture provided by Javier Aguilar-Pérez. F – Sliced specimen of *I. coraloidea* (GMUS.Mp1) showing general cavity filled of canals, a thick shell and three cavities invaded by capillary canals. Picture provided by Brian Pratt.

additional specimens from the Île Madame (MNHN.F.6517-B). Two three-dimensionally preserved specimens are available: one fragment of *I. triangularis* from an unpublished plate of the Guéranger's collection (MV.2003.1.10900; Le Mans, France), and the syntype of *Ichthyosarcolites bicarinatus* (MGUP.021.2-110; fig. 3.2-C). The lectotype of *Ichthyosarcolites doublieri* from Martigues (Bouches-du-Rhône, France) was also accessed (MNHN.F.R07971; fig. 3.2-F). One specimen of *Ichthyosarcolites rotundus* from Roquefort-la-Bédoule is available at Aix-Marseille University (AMU.ICH1; fig. 3.3-A). M. Floquet provided us field photographs of other *Ichthyosarcolites* specimens from Roquefort-la-Bédoule. Sliced specimens from the locality Jebel Selloum (Tunisia) can be found in the Bessaïs' collection (Aix-Marseille University): six slices of *Ichthyosarcolites bicarinatus* (AMU-HBES 84-a, 84b) and one slice of *Ichthyosarcolites tricarinatus* (AMU-HBES 85-a; fig. 3.3-C). The collection of J. Philip houses two slices (AMU-PCC 41-a) of *Ichthyosarcolites poljaki* from Koulovate (Greece). Several additional species are analyzed from photography taken from Aquismon Ciudad Valles, Canon de la Servilleta (Tamaulipas), San Joaquin Queretaro and Paso del Rio (Mexico) by Javier Aguilar-Perez (fig. 3.3-D, E), including the holotype of *Ichthyosarcolites alatus* (fig. 3.2-E), from Istria (Croatia) and from Bosnia Herzegovina, including the types of *Ichthyosarcolites monocarinatus* and *Ichthyosarcolites polycarinatus* by Alceo Tarlao (MB.3470, MB.3482, MB.4197 (3); fig. 3.2-B, D) and from Hrušica (Slovenia) by Bogdan Jurkovšek (BJ1901, BJ1914, BJ1942). Two small pieces of probable myophoral parts of a specimen labelled "*Ichthyosarcolites cornutus*" and several unidentifiable specimens from USA labelled *Ichthyosarcolites* sp. (YPM 24697, 24694 and 24696) were observed and compared to Old World specimens, completed by photographs of a Canadian specimen of *Ichthyosarcolites coraloidea* (GMUS.Mp 1; fig. 3.3-F; Caldwell and Evans, 1963).

Abbreviations. AMNH – American Museum of Natural History; AMU – Centre Européen de Recherche et d'Enseignement des Géosciences de l'Environnement, Aix-Marseille University; BJ – Paleontological collection of Dr. Bogdan Jurkovšek and Dr. Tea Kolar-Jurkovšek, Ministry of Culture of Republic of Slovenia and the Natural History Museum of Slovenia in Ljubljana; GMUS – Geological Museum of the University of Saskatchewan, Saskatoon; LGPZ – Laboratory of Geology and Paleontology of the Faculty of Science of Zagreb; MB – Zemaljski muzej BiH, Sarajevo; MGUP – Museo Geologico

G.G. Gemmellaro della R. Università Degli Studi Di Palermo; MNHN – Muséum National d'Histoire Naturelle, Paris; MV – Musée Vert, Muséum d'Histoire Naturelle du Mans; UNAM – Universidad Nacional Autónoma de México; UPMC – Université Pierre et Marie Curie; YPM – Yale Peabody Museum.

2.2. Morphological analysis

There are very few three dimensionally preserved shells of *Ichthyosarcolites* available. Fossils are usually internal molds of the general cavity or recrystallized shells observed in cross sections. The distinction of *Ichthyosarcolites* species in the literature is based on external features: occurrence of external flanges along the shell (fig. 3.2), shape of the internal cavity (figs. 3.2 and 3.3), and development of internal *tabulae* (fig. 3.4-A to D). All of these features were reappraised and analyzed using morphometry and statistical tools to improve the species definitions (table 3.1) and taken into account as Least Inclusive Taxonomic Units or LITU (Pleijel and Rouse, 2000). A LITU is a taxa, defined by a property, in which it is impossible to raise any other taxa.

Several species are defined by their number of flanges. However, the relative positions of flanges around the shell received very limited attention, whereas the location and growth patterns of flanges or other shell ornaments often support homologies in mollusks (Merle, 2005). The anatomical position of a flange is not recognizable unless the shell can be properly oriented. We quantify the position of a flange by its angle, measured from the central tooth of the right valve, or the central tooth socket, between the anterior and posterior teeth of the left valve (fig. 3.5-A). The method also requires to place a landmark at the center of the general cavity that we approximate as the center of the smaller circle including the general cavity. This procedure allows comparisons among shells of different sizes with flanges of different sizes, morphologies and directions. The posterior and anterior sides on the left valve can be identified thanks to the posterior myophore which is always shorter than the anterior myophore. On the right valve, a small cavity lies at the posterior side of the shell. The angle is always measured from the dorsal to the posterior part of the shell. All measurements were performed on pictures of sliced specimens using ImageJ 1.48v (Rasband, 1997). Measures were possible only on 20 shells for 58 angles measures. The Cosines and Sinuses of all angles are plotted on Figure 3.5-B using R

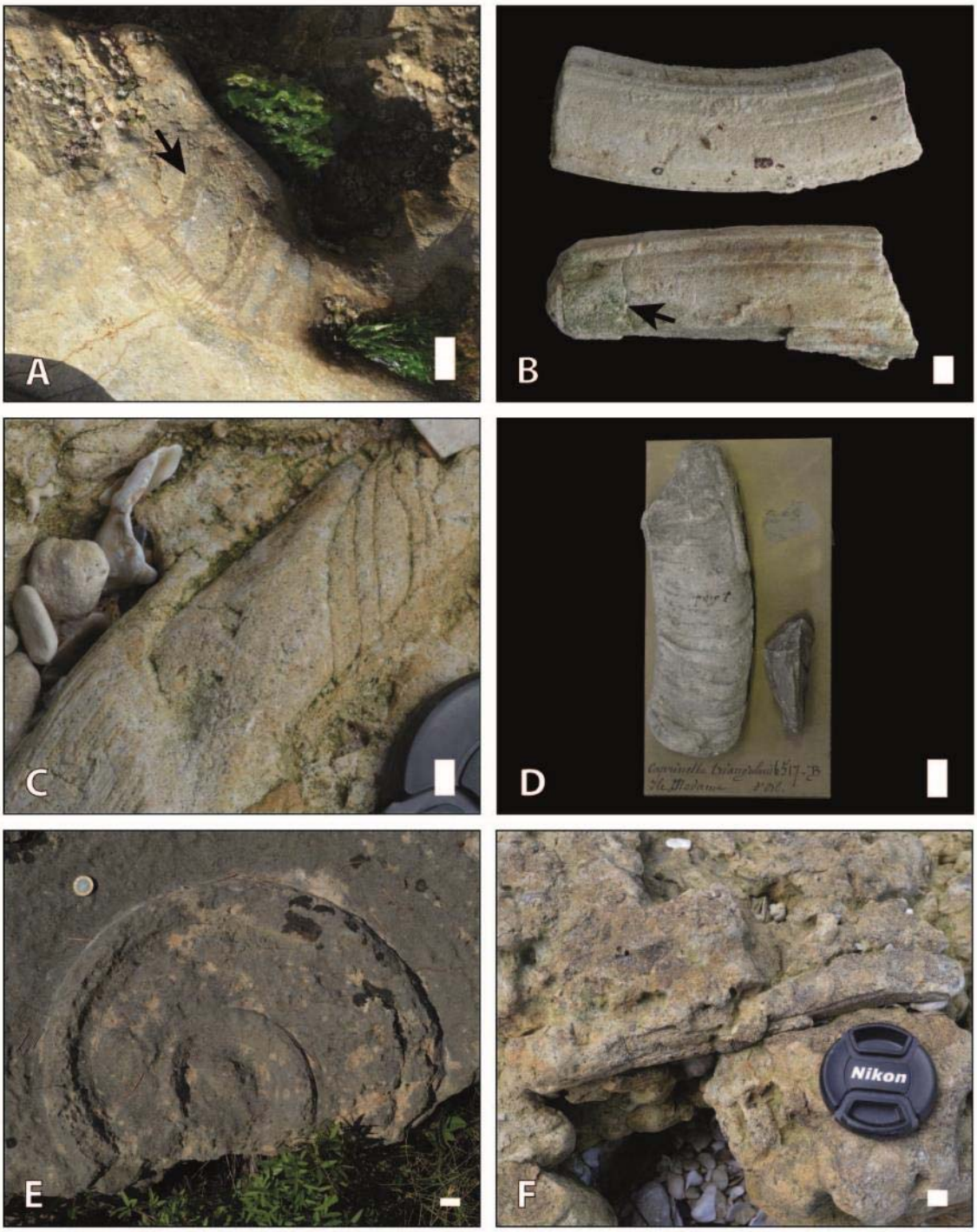


Figure 3. 4. Photographs of *Ichthyosarcolites* shells illustrating tabulae and coiling patterns. All scale bars = 1cm. A – Longitudinal slice of *I. triangularis* from Île d'Oléron showing thick tabulae in the general cavity and the sockets. Arrow: tabulae. B – Internal molds of *I. triangularis* from Île d'Oléron (Charentes, France) showing the partial or complete dissolution of the thin tabulae in the general cavity and the sockets. Arrow: tabulae. C – Specimen of *I. triangularis* from Île d'Oléron with intersecting tabulae. D – Crushed internal mold of *I. triangularis* from Île Madame, showing tabulae and one socket (d'Orbigny collection, MNHN.6517–B). E – Almost complete coiled right valve internal mold from Provence (France). F – Specimen of *I. triangularis* from Île d'Oléron with uncoiled internal mold.

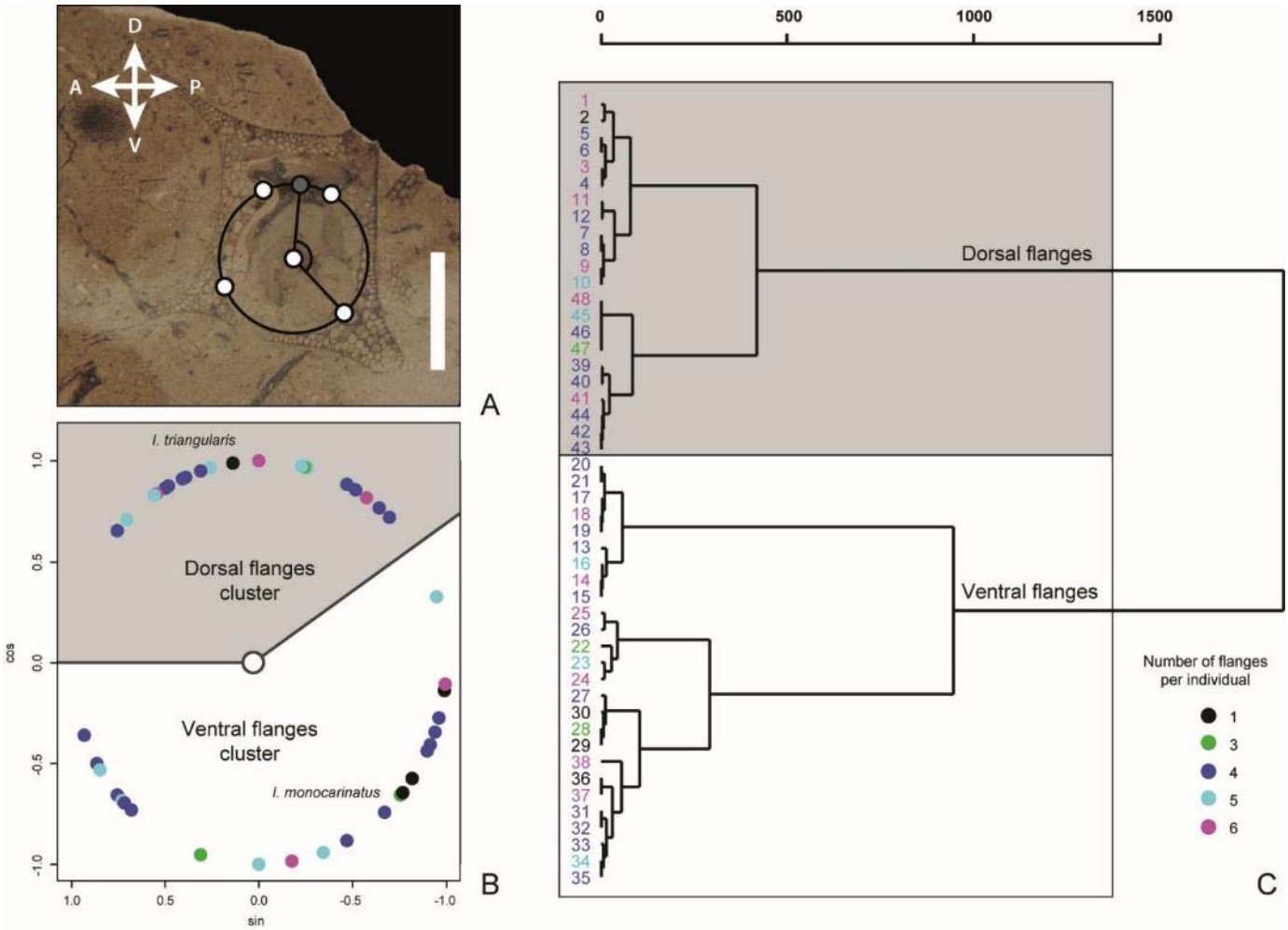


Figure 3. 5. A- Method for measurement of the angular positions of flanges on rudist shells. A circle approximates the general cavity. White circles correspond to the projection of the flange stem on the circle. Angles are measured clockwise with 0° corresponds to the presumed position of the central tooth or central tooth socket. Scale bar = 1cm. Specimen UNAM.6978, picture provided by Javier Aguilar-Pérez. B- Plot of flanges angle measurement by cos and sin coordinates. C- Hierarchical classification of flanges positions. The two main clusters represent the dichotomy between *I. monocarinatus* ventral flanges and *I. triangularis* dorsal flanges. Each point corresponds to a flange position, and the color corresponds to the number of flanges the shell from which the flange is from bear.

(R Core Team, 2013) to summarize the dataset. A hierarchical clustering using Ward's method is made on the angular positions of flanges (fig. 3.5-C) to test for homologous flange positions.

We carried out an analysis using geometric morphometry on the shape of the internal cavity. Only the specimens with clear orientation and no trace of taphonomic compaction were chosen. Twenty-three shells were considered: the shells used for analysis

Nominal species	Tabulae	Shell shape	Ligament	Localisation
Ichthyosarcolites triangularis Desmarest, 1812	Sub-plane to oblique	Coiled to uncoiled	No	Afghanistan, Bosnia and Herzegovina, Egypt, France, Greece, Libya, Spain, Tunisia, Turkey
Ichthyosarcolites obliqua Deshayes, 1825	Very oblique (<45° from commissure)	Unknown	Unknown	France
Ichthyosarcolites doublieri Orbigny, 1847a	No	Coiled	Unknown	France, Portugal
Ichthyosarcolites quadrangularis Tuomey, 1854	"regular, abruptly bent downwards"	Loosely coiled	Unknown	USA
Ichthyosarcolites loricatis Tuomey, 1854	Unknown	Unknown	Unknown	USA
Ichthyosarcolites cornutis Tuomey, 1854	"abruptly bent downwards, and towards the concave side"	Coiled	Unknown	USA
Ichthyosarcolites coraloidea Hall and Meek, 1855	Sub-plane to oblique	Unknown	Unknown	USA, Canada
Ichthyosarcolites bicarinatus Gemmellaro, 1865	Unknown	Uncoiled	No	Algeria, Bosnia and Herzegovina, Croatia, France, Greece, Italy, Libya, Slovenia, Turkey
Ichthyosarcolites marginatus Počta, 1887	Unknown	Loosely coiled	No	Czech republic
Ichthyosarcolites ensis Počta, 1887	Unknown	Unknown	No	Czech republic, Italy
Ichthyosarcolites tricarinatus Parona, 1921	Unknown	Loosely coiled	No	Afghanistan, Lybia, Bosnia and Herzegovina, Croatia, Greece, italy, Slovenia, Spain, Turkey
Ichthyosarcolites poljaki Polšak, 1964	Unknown	Unknown	No	Bosnia and Herzegovina, Croatia, Greece, italy, Slovenia, Turkey
Ichthyosarcolites rogi Pleničar, 1965	Unknown	Unknown	No	Slovenia
Ichthyosarcolites monocarinatus Slišković, 1966	Unknown	Unknown	No	Bosnia and Herzegovina, Croatia, Italy, Slovenia, Spain, Turkey
Ichthyosarcolites polycarinatus Slišković, 1966	Unknown	Unknown	No	Bosnia and Herzegovina
Ichthyosarcolites rotundus Polšak, 1967	Unknown	Unknown	No	Croatia
Ichthyosarcolites iokungensis Bobkova, 1968	Sub-plane	Unknown	No	Russia
Ichthyosarcolites quadratus Pleničar and Dozet, 1994	Unknown	Unknown	Unknown	?
Ichthyosarcolites alatus Filkorn, 2002	Unknown	Unknown	No	Mexico
Oryxia sulcata n. sp.	Unknown	Unknown	Yes	Slovenia

Table 3. 1. Comparison of the known morphological features and geographic distribution of all 19 nominal *Ichthyosarcolites* species, plus the herein recognized taxon *Oryxia sulcata* nov. gen. nov. sp. "Unknown" means that the information was unavailable from the literature or the surveyed specimens.

Nominal species	Flanges	Flange shape in cross section	General cavity shape	Pallial canals
Ichthyosarcolites triangularis Desmarest, 1812	1 (ventral)	Short and triangular	Triangular to ovoid	Capillary and palissading
Ichthyosarcolites obliqua Deshayes, 1825	Unknown	Unknown	Unknown	Unknown
Ichthyosarcolites doublieri Orbigny, 1847a	Unknown	Unknown	Circular	Unknown
Ichthyosarcolites quadrangularis Tuomey, 1854	Unknown	Unknown	Quadrangular	Capillary and palissading
Ichthyosarcolites loricatis Tuomey, 1854	Unknown	Unknown	Ovoid	Capillary and palissading
Ichthyosarcolites cornutis Tuomey, 1854	Unknown	Unknown	Unknown	Capillary and palissading
Ichthyosarcolites coraloidea Hall and Meek, 1855	Unknown	Unknown	Ovoid	Capillary and palissading
Ichthyosarcolites bicarinatus Gemmellaro, 1865	2	Long and narrow	Circular to sub-quadrangular	Capillary and palissadic or large ovoid
Ichthyosarcolites marginatus Počta, 1887	Unknown	Unknown	Unknown	Capillary and palissading
Ichthyosarcolites ensis Počta, 1887	1	Short and triangular	Unknown	Capillary and palissading
Ichthyosarcolites tricarinatus Parona, 1921	3	Long and narrow to short and triangular	Circular to polygonal	Capillary and palissading
Ichthyosarcolites poljaki Polšak, 1964	3	Long and narrow to short and triangular	Circular to ovoid	Capillary and palissading
Ichthyosarcolites rogi Pleničar, 1965	3	Long and narrow	Triangular	Capillary and palissading
Ichthyosarcolites monocarinatus Slišković, 1966	1 (dorsal)	Short and triangular	Ovoid	Capillary and palissading
Ichthyosarcolites polycarinatus Slišković, 1966	7	Long, narrow and elongated	Circular	Capillary and palissading
Ichthyosarcolites rotundus Polšak, 1967	0	Non applicable	Circular	Capillary and palissading
Ichthyosarcolites iokungensis Bobkova, 1968	1 (ventral)	Short and triangular	Ovoid	Unknown
Ichthyosarcolites quadratus Pleničar and Dozet, 1994	Unknown	Unknown	Unknown	Unknown
Ichthyosarcolites alatus Filkorn, 2002	3 to 5	Long and narrow to short and triangular	Circular to polygonal, irregular	Capillary and palissading
Oryxia sulcata n. sp.	3	Short and triangular in section	Circular to ovoid	Capillary and palissading

Continuation of table 3.1.

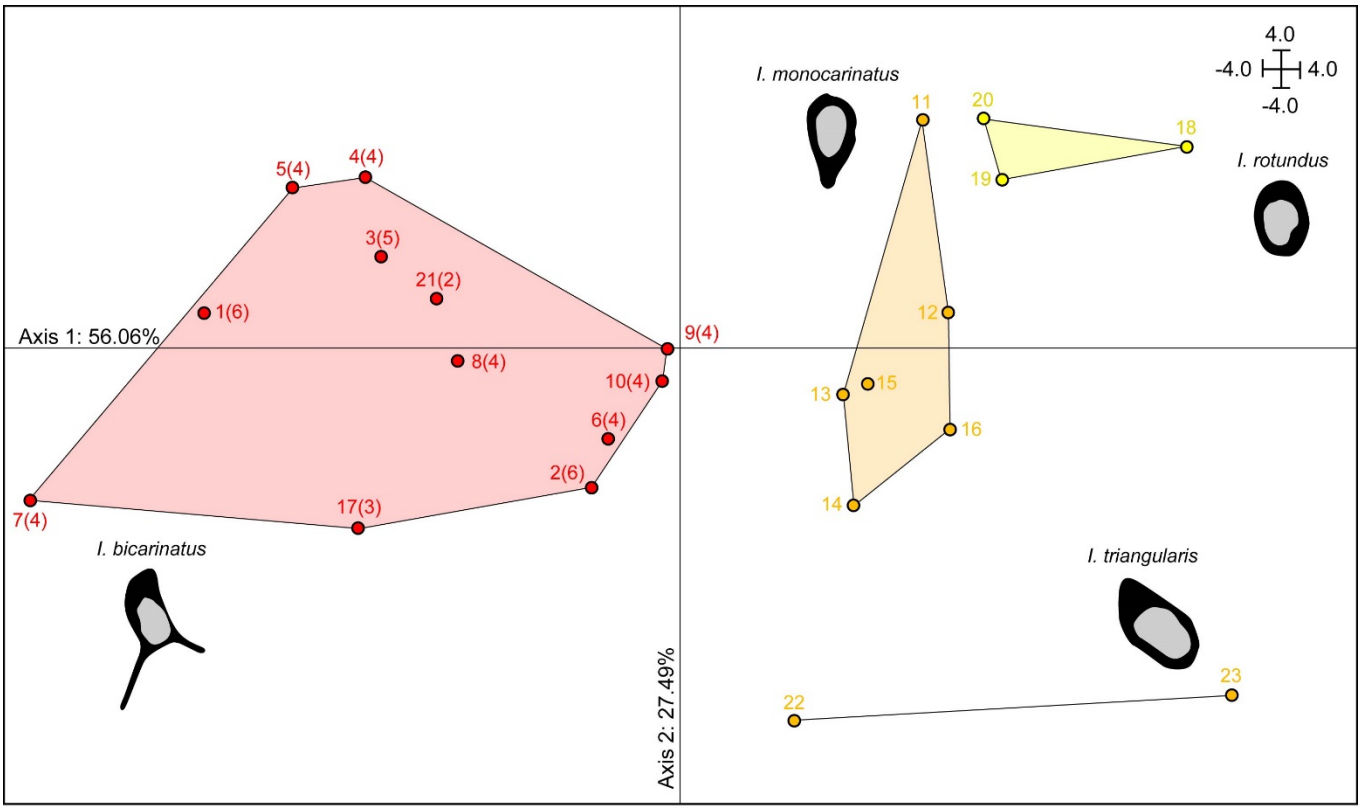


Figure 3. 6. Discriminant analysis plot of 23 general cavities of *Ichthyosarcolites* shells landmarks. Four clusters are designated from the number of flanges: *I. rotundus* (no flange), *I. monocarinatus* (one ventral flange), *I. triangularis* (one dorsal flange) and *I. bicarinatus*. Analyzed data are presented as supplemental Online Material.

of flanges orientation, plus three shells lacking flanges. The general cavities are described by 10 landmarks equally distributed on the outline. Landmark coordinates are digitalized from photographs using the website PhyloNimbus (appendix 1). The procruste-fitted landmark dataset were converted to procruste residuals by subtracting the mean shape using Past 3.15 (Hammer *et al.*, 2001). A principal component analysis (PCA) was implemented from the Procruste residuals (Hammer & Harper, 2008). The ten principal components from the PCA were used as uncorrelated variables for a linear discriminant analysis (LDA; fig. 3.6). An LDA is especially suited for maximizing inter-group separation and minimizing intra-group separation for predefined groups (McLachlan, 2004). The LDA is used here to check if the clusters constructed with the data of flanges positions are congruent with the analysis of internal anatomy.

3. Morphology

3.1. Pallial canals

The canals are often missing due to the usually poor preservation of fossils (figs. 3.2-C, 3.4-B, 3.4-D, 3.4-F). Wherever preserved, the patterns of pallial canals distribution are similar on the two valves. There is always a single outer row of small palisading canals, the rest of the aragonitic shell being filled by capillary canals of variable diameter, rounded to polygonal in cross section. Some Mexican specimens exhibit irregular rows of enlarged round canals inside the flanges (fig. 3.2-E). The juvenile specimens only have small rounded capillary canals (fig. 3.2-E). Canals seem to increase with shell size and flange developments. The specimen from Jordania illustrated as *I. bicarinatus* by Chikhi-Aouimeur (2010; fig. 3.3-B) exhibits exceptionally large, round to polygonal canals. This larger canal size leads to an accentuated polygonal shape of the canals and the relative reduction of shell material between the canals, and canal layers in the flanges. This canal size and distribution patterns are linked to ontogeny and the development of flanges. However, the high interindividual variability does not support the recognition of any distinct pattern that could be used to support any of the nominal species.

3.2. Tabulae

Tabulae are thin skeletal walls produced during shell growth and representing successive floors of the internal cavities of the shell. The species of *Ichthyosarcolites* are described with tabulae grossly parallel to the commissural plane, except for *I. obliquus* defined with obliquely oriented tabulation, and for *I. doublieri* lacking tabulations. The tabulae are very thin, fragile aragonitic structures that may disappear easily through taphonomic and diagenetic processes. The tabulations are usually recrystallized or appear as dissolution voids in internal molds. In best preservation conditions, tabulae are to be found in the general cavity, the sockets and the posterior myophoral cavity (fig. 3.3-D, 4-A). In practice, the traces of the tabulae are often erased (fig. 3.2F, 4E, 4F). On Figure 3.4-B, the tabulae of the lower shell are well visible on the left side, but disappear nearly completely on the right side. The partial preservation of tabulae suggests incomplete

dissolution prior to lithification. The lack of tabulae in the holotype of *I. doublieri* (fig. 3.2-F) is probably due to a complete dissolution of the tabulae during early diagenesis.

On the fossils, the tabulae may appear perpendicular or strongly oblique to the shell growth axis (fig. 3.4). The oblique tabulae have the shape of a three-quarter cornet with a pointing downward dorsal attachment, and an open rounded ventral attachment to the front wall of the general cavity. The obliqueness of the tabulae can be seen only in a dorsoventral view (fig. 3.4-C, F). The obliqueness of tabulae appears highly variable among specimens within a local assemblage. In the *I. triangularis* from Charentes, the expression of tabulae varies gradually among individual shells from parallel to the commissure to strongly oblique depending (fig. 3.4-A, D). The high variations in tabulae expressed in *I. triangularis* include the typical morphology of *I. obliquus*, which challenges the reliability of *I. obliquus* as a distinct species.

3.3. Shell coiling

There is a high diversity of coiling patterns in *Ichthyosarcolites* shells. Some species are described as completely uncoiled, other as simply curved or coiled. The coiling is always in a spiral plane, with loose to tight whorls (fig. 3.1-C, 4-E). The torsion of the flanges on figure 3.2-C suggests twisting in the coiling plane, probably in response to substrate irregularities. Shells are never straight along its entire development. In Figure 3.4F, the uncoiled shell begins by a coil at its right apex. The uncoiling seems to appear during the animal life in relationship with environmental constraints. A wide range of coiling patterns can be encountered at a given location, but the shell is most often uncoiled in coarser, sandy deposits.

3.4. Ligament

The ligament is totally missing in typical *Ichthyosarcolites* (figs. 3.2, 3.3, 3.4). Among the specimens described as *I. poljaki, I. bicarinatus* and *I. tricarinatus* by Pleničar & Jurkovšek (1999) from Slovenia, few shells exhibit a small dorsal protrusion that was considered a flange by the authors. This dorsal protrusion is bordered alongside by a ligamentary ridge. This protrusion should be understood as a bulge of shell stranded

posteriorly to the ligamentary ridge, not homologous to a flange. The ridge of the shells sinks to the anterior tooth socket and connect to a small ligamentary cavity. Similar dorsal ligamentary bulges are also recognized in some forms from Istria (Croatia) by Polšak (1967), even if the inside of the shell does let see a ligamentary cavity due to a poor preservation. The ligament is of critical importance in the phylogeny of rudists (Skelton and Smith, 2000; Yonge, 1967). Four evolutionary stages are recognized: (i) the ligament follows the commissure dorsally (*e.g. Diceras*); (ii) the verticalized ligament is perpendicular to the commissure plane (*e.g. Valletia*, Monopleurids); (iii) the ligament is internalized in a cavity that loses contact with the exterior (*e.g.* Trechmanellids); (iv) the ligament disappears completely with the ligamentary cavity and ridge (e.g. Titanosarcolitinae, Parasarcolitinae). *Ichthyosarcolites* represent the fourth stage. However, the ligamental ridge found in the Slovenian shells would place them in the second stage, out of the genus *Ichthyosarcolites*. Theses hypotheses are tested in chapter 4.

3.5. Flanges

Flanges, ribs, carinae and ridges are local thickening of the shell, appeared by differential growth patterns. Even if the expression of shell thickenings in *Ichthyosarcolites* varies from low rounded bulges or ribs to high and thin blade-like ridges (figs. 3.2, 3.3), all have a similar structure and likely appear following the same developmental processes. If the number and positions of flanges is consistent among individuals, then we can assume homology and use them to define species. Few attempts were made to homologize flanges in which each flange receive a number or a letter (Carbone *et al.*, 1971; Polšak, 1967). However, only the number of flanges are listed in the diagnoses of *Ichthyosarcolites* species, not their positions. *I. rotundus* is defined by the absence of flanges. A single flange occurs both in *I. monocarinatus* and *I. triangularis*, two in *I. bicarinatus*, three in *I. tricarinatus*, three to six in *I. alatus* (Aguilar-Pérez, 2008; Filkorn, 2002), and seven in *I. polycarinatus*. Flanges morphology can vary in multiple ways: location, orientation and relative development, which can be quantified using the length to width ratio of the flange in cross section. The recognition of flanges may become difficult when weakly developed. Filkorn (2002) illustrated adults *I. alatus* bearing either four, five or six flanges, but small

juveniles with only three flanges are also assigned to *I. alatus* (fig. 3.2-E), suggesting variation of flanges number during ontogenesis.

A consistent test of homology for the flanges would be to detect constancy in their positions. Flanges of same positions are more likely to be homologous. Among the species described with a single flange, *I. monocarinatus* can be defined by the presence of a ventral one, and *I. triangularis* on the presence of a dorsal one. This variation in flange position is corroborated by the main two clusters (fig. 3.5-C) of the hierarchical classification. The position of the flange could perfectly support the definition of these two species considering that the uniqueness of the flange is not homology.

On shells with more than one flange, groups of the hierarchical clustering may evidence pattern of flange distribution homologies. In case of two flanges, homology is assumed if the two flanges are consistently, or never, fall within the same cluster. The cluster analyses fail at recognizing consistent patterns of flange distribution among two flanged or three flanged shells. It is not possible to distinguish consistently homologous flanges among shells. The plasticity of flange distribution pattern and the number of flanges is not reliable in an evolutionary point of view. The definition of a species from its number of flanges must be avoided hereafter.

Ichthyosarcolites is a recumbent form (Skelton, 1976) lying on sandy to muddy substrates. This ecotype constrains the development of flanges around the shell. Our primary hypothesis was that no flange is likely to develop on the side of the shell lying on the seafloor, and that we can expect the flattest side of any *Ichthyosarcolites* shell to be in contact with the substrate. The hierarchical classification of flanges separates well ventral flanges from dorsal flanges, with rare occurrence of flanges on the anteroposterior side (fig. 3.5-B). Most *Ichthyosarcolites* were likely lying on their anterior side, with usually a stabilizing flange in ventral or dorsal position.

3.6. General cavity

Some *Ichthyosarcolites* species are defined from shell cross sections. The characters descriptions focus on external features including bulges, flanges, the relative shell thickness, and the general cavity outline. For example, *I. triangularis* is defined with a

triangular shape in cross section. As observed in many mollusks, the external features of the shell reflect the shape of internal structures such as the general cavity. Because flange number and general cavity shape are the two main features used to erect *Ichthyosarcolites* species, we test whether the groups defined from flange distribution can also be recognized in general cavity outline.

Following the results of the hierarchical classification on flanges, a discriminant analysis is performed on cavity outline descriptors, assuming two kinds of one-flanged *Ichthyosarcolites*: *I. triangularis* and *I. monocarinatus*, the flange lacking *I. rotundus* and a last cluster with all shells with two or more flanges. The results show a clear partition of *I. triangularis*, *I. monocarinatus*, *I. rotundus* and the multi-flanged *Ichthyosarcolites* in the morphospace (fig. 3.6). The congruence between the number and position of flanges and the internal cavity outline suggests geometric constraint on the shell development. Addition of flanges leads to a more or less polygonal cavity.

The Mexican populations of *Ichthyosarcolites* demonstrate a good variability in the number of flanges among adults (4-7), and also an ontogenetic variation, with juvenile of *I. alatus* carrying a lower number of flanges (3 to 4) than adults (4 to 5; Filkhorn, 2002; Aguilar-Perez, 2008). A similar plasticity in expression of flanges number is understood as intraspecific variability in Antillocaprinids (e.g. *Parasarcolites atkinsoni*; Mitchell and Gunter, 2006), a group close to Ichthyosarcolitidae (Skelton and Smith, 2000). For these reasons, we retain a single species of *Ichthyosarcolites* bearing two flanges or more. The shells bearing two flanges or more of our cluster analysis always have both a ventral and a dorsal flange. For specimens possessing more than one flange, flanges are located more randomly.

4. Systematic paleontology

Our analyses lead to keep four species characterized by a unique combination of homologous parts (Wheeler & Platnick, 2000): one with occurrence of a single ventral flange, one with a single dorsal flange, one with at least on dorsal and one ventral flanges, one lacking flange. The systematic arrangement of the Bivalvia used in this study follows Carter *et al.* (2011) and Skelton (2013a, 2013b).

4.1. Genus *Ichthyosarcolites* Desmarest, 1812

Class: Bivalvia Linnaeus, 1758

Infraclass: Heteroconchia Gray, 1854

Order: Hippuritida Newell, 1965

Suborder: Hippuritidina Newell, 1965

Superfamily: Caprinoidea d'Orbigny 1850

Family: Ichthyosarcolitidae Douvillé, 1887

Genus *Ichthyosarcolites* Desmarest, 1812

Type species. *Ichthyosarcolites triangularis* Desmarest, 1812

4.1.1. Diagnosis

Rudist with small left valve and extremely developed right valve. Both valves loosely coiled with non-contiguous whorls that may uncoil during shell growth. Presence of zero to seven flanges along the shell. The relative size and shape of the flanges vary from rounded costae to prominent ridges. Aragonitic layers filled with capillary canals. Tabulations fill all cavities of the two valves and the pallial canals. No ligamentary ridge or cavity. The two anterior and posterior sockets of the right valve are of equal size, small and rounded. Presence of a small anterior accessory cavity. The two myophores are similar to those of radiolitids, forming blades attached to the ventral side of the teeth on the left valve. The myophores of the left valve insert directly in the wall of the general cavity of the right valve.

4.1.2. Remarks

The monogeneric family Ichthyosarcolitidae was created by Douvillé (1887) because of the high differentiation of *Ichthyosarcolites* from all other rudists. *Ichthyosarcolites* possess so specific characters that confusion with other genera is excluded. *Stellacaprina gunteri* shows a similar morphology but the species has two additional toothlets, diagnostic feature of the Antillocaprinidae (Mitchell, 2013b). *Curtocaprina clabaughikinsorum* is the only other member of Ichthyosarcolitidae

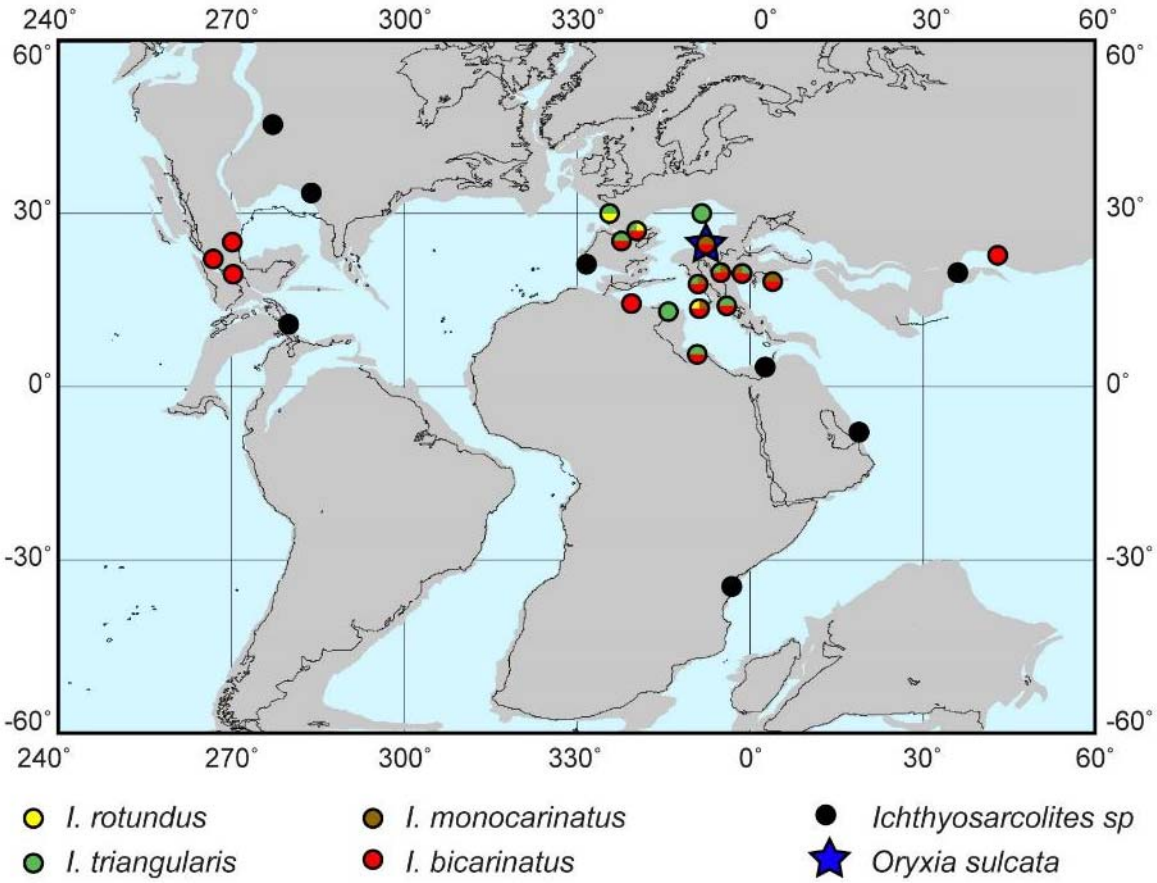


Figure 3. 7. Geographical map showing the fossil localities of Cenomanian age bearing *Ichthyosarcolites*. Yellow dots represent *I. rotundus*, green dots *I. triangularis*, brown dots *I. monocarinatus*, orange dots *I. bicarinatus*, red dots *I. polycarinatus* and blue stars *Oryxia sulcata* nov. gen. nov. sp. Black dots represent *Ichthyosarcolites* occurrences from which the identification to species level cannot be verified. The paleogeographic map was generated using the ODSN Plate Tectonic Reconstruction Service (http://www.odsn.de/odsn/services/paleomap/paleomap.html).

following Mitchell (2013a). The species completely differs from *Ichthyosarcolites* in the structure of its dentition, the size and shells shape (right and left valves of same size, approximately 2cm, a single ventral flange). If all identified *Ichthyosarcolites* species belong exclusively to the Cenomanian, *Ichthyosarcolites* sp. is cited from Albian in Sinai, Egypt (Zakhera, 2010). The map on Figure 3.7 shows the geographic location of all discovered specimens. *Ichthyosarcolites* is a recumbent form reported from peri-reefal and fore-reef environments with a high water-energy (Filkorn, 2002) to calm mudflats, where it plays a role of pioneer (Videt, 2003).

4.1.3. *Ichthyosarcolites triangularis* Desmarest, 1812

(figs. 3.1-C, 3.4-E)

Abridged citation list

1786 Cucroïdes; Guettard: 552-553: pl. XXVIII, fig. 3.2

1812 *ichtyosarcolites triangularis* Desmarest: 321-324

1817 *ichthyosarcolites triangularis*; Desmarest: 50-51, pl. II, figs 9-10

1821 Glossopètres; Defrance: 72

1821 Ichthyosarcolite; Defrance: 549-550

1825 *Ichthyosarcolites triangularis*, Desmarest; Deshayes: 501

1825 *Rhabdites triangularis* N.; De Haan: 41, 42, 52, 53, 57, 160

1826 *Ichthyosarcolites*, Desmarest; Orbigny: 167-168

1827 Ichthyosarcolite; Ducrotay de Blanville: pl. XI, fig. 3.2

1846 *Ichthyosarcolithus* A. G. Desmarest 1817; Hermannsen: 554-555

1849 *Caprinella triangularis*, d'Orb., 1847; d'Orbigny: pl. 542

1850 *Caprinella triangularis*, d'Orb., 1847; d'Orbigny: 192-194

1854 *Ichthiosarcolites quadrangularis* Tuomey: 172

1855 *Caprinella triangularis* Desm.; Woodward: 52-53, figs 25-27

1887 *Ichthyosarcolites marginatus* Počta: 207

1887 *Ichthyosarcolithus triangularis* Desmarets; Douvillé: 791-793, figs 15-17

1887 *Ichthyosarcolites triangularis* Desmarest; Fischer: 1057-1058, fig. 3.804

1888 *Ichthyosarcolithes*; Douvillé: 706, 728-729

1889 *Ichthyosarcolites marginatus* Poč.; Počta: 87, pl. VI, figs 6, 7

? 1905 *Ichthyosarcolites triangularis* Desmarets 1817; Paquier: 94-95, pl. IX, figs 7-9

? 1916 *Ichthyosarcolites triangularis*; Hennig: 461

1921 *Ichthyosarcolites triangularis* Desm.; Parona: 10-11, pl. I, figs 1a, 1b, fig. 3.9

1930 *Caprinella triangularis*; Repelin: 61, pl. VII, fig 5

1937 *Ichthyosarcolites triangularis* Desmarest; MacGillavry: 47-59

1968 *Ichthyosarcolites triangularis*; Damestoy: 2-6, figs 1-3

1969 *Ichthyosarcolites triangularis*; Dechaseaux: N795, figs 4a, 4b

1970 *Ichthyosarcolites triangularis* Desmarest; Quarto di Palo: 107, pl. XV, fig. 3.1

2000 *Ichthyosarcolites triangularis* Desmarest; Skelton & Smith: 110, 111, 117, 123, figs 4c, 4d

2010 *Ichthyosarcolites triangularis* Desmarest; Chikhi-Aouimeur: 96, fig. 3.87.1

2015 *Ichthyosarcolites triangularis* Desmarest; Troya García: 248-254, 267, 271, figs 139-141

4.1.3.1. Type material

The type specimen illustrated by Desmarest, 1817 has been lost (fig. 3.1-B). We assign the specimen MNHN.F.R08003 (fig. 3.1-C) as neotype of *Ichthyosarcolites triangularis* Desmarest 1812. Cenomanian, Charras (Charentes, France).

4.1.3.2. Diagnosis

Ichthyosarcolites species with small left valve and extremely developed right valve. Both valves loosely coiled with non-contiguous whorls. Presence of a unique dorsal flange somewhat triangular in cross section.

4.1.3.3. Description

Shell highly inequivalve; left valve very small (approximately 1/10 of the right valve length) and slightly coiled. Right valve narrow tubular, extremely developed and ventrally coiled (dorsal external side) with non-contiguous whorls (fig. 3.1-C). Planispiral growth of the shell (figs. 3.1-C, 3.4-E). Left valve loosely ventrally coiled and very small compared to the right valve. Right valve shape highly variable, ranging from 0.5 to 2 coils, the space between two coils ranging from a distance of 3 cm to 50 cm for completely coiled specimens (figs. 3.1-C, 3.4-E). Both valves of equal diameter. General cavity of both valves of equal diameter, between 1 cm and 8 cm for adults (fig. 3.4-E), flange not included. A unique flange is present, always in dorsal position (fig. 3.1-C), of triangular shape, as wide as high. Flange straight with a rounded external margin. External aspect of the shell completely smooth or with occurrence of small tightened costae of approximately 0,1mm height and 1mm width. Shell composed of an outer calcitic layer of approximately 0,5mm and an inner aragonitic layer with a thickness of approximately 15% of the diameter of the body cavity in transversal section, flange not included. The aragonitic layer is densely perforated with close small capillary canals (approximately 1mm in diameter) round to slightly ovoid (sometimes honeycombed) except for the myophores and

the anterior and posterior teeth. The external side of the aragonitic shell is bordered with a single row of sub-rectangular/palisading canals of same width as capillary canals. Tabulae present on both valve with highly variable thickness and surface orientation in the general cavity, pallial canals, sockets and the posterior accessory cavity (PAC). Tabulae flat to highly curved dorsally. When highly oblique, presence of a horizontal part in the middle of the tabulae in dorsoventral axis (fig. 3.1-C). Tabulae within the pallial canals; tabulae of a canal are regularly spaced and independent from them of other canals. Single other ovoid PAC present posteroventrally on the right valve, slightly bigger than sockets and filled with tabulae (fig. 3.4-A). No ligamental ridge and cavity.

Left valve myocardinal apparatus – Inverse dentition, composed on the left valve of an anterior and a posterior conical tooth of similar shape and size, separated by a central tooth socket never observable as a distinct cavity. Thin vertical wall-like myophore is ventrally adjoined to each tooth. The two myophores constitute a U-shape myophoral arcade, attached to the teeth, protruding directly in the general cavity of the right valve. Development of vertical myophores generating a space between it and the inner shell directly contiguous to it, looking like a cavity that may be filled with few thin blades directed with an anteroposterior axis. Whole myocardinal apparatus is U-shaped and extends itself to make a connection with the ventral edge of the shell.

Right valve myocardinal apparatus – Central tooth shape unknown, probably reduced, filled with pallial canals. Presence of two ovoid anterior and posterior tooth sockets directly connected to the general cavity and elongated in an anteroposterior axis. Sockets always very close one to another (approximately a 2mm separation space). Myophores undifferentiated from the inner shell, no known muscle scars.

4.1.3.4. Remarks

The type-species of the genus was primarily described by Desmarest (1812) as '*Coquille droite et épaisse, presque triangulaire, munie intérieurement de cloisons obliques en forme de mi-cônes ou cornets, et d'un sinus ou siphon longitudinal et latéral*'. *I. triangularis* is figured for the first time in Desmarest (1817); fig. 3.1-B). The species was primarily described from internal molds of the general cavity and sockets segmented by tabulae. The main distinctive feature was the triangular shape of the shells. The type locality

of Desmarest's species remains unknown to himself, but is certainly from Charentes (Western France). The dorsal thickening of the aragonitic shell was correctly represented in d'Orbigny's figured specimen (fig. 3.1-C).

With a single dorsal flange *I. marginatus* (Počta, 1887) completely fall within the variability of *I. triangularis*.

4.1.3.5. Occurrences

Exclusively Cenomanian. Living in tropical shallow waters on carbonate platforms of the northern and southern Tethyan margins. More precisely, *I. triangularis* was found in France (Charentes, western France and Provence southeastern France), north Spain, Croatia, Greece, Czech Republic, Italia, Tunisia, northwestern Turkey, Poland, and Libya (fig. 3.7; table 3.1).

4.1.4. *Ichthyosarcolites rotundus* Polšak, 1967

(figs. 3.2-A; 3.3-A)

Abridged citation list

1967 *Ichthyosarcolites rotundus* n. sp.; Polšak: 80-81, 187, pl. VI, fig 2, pl.9, fig. 3.2-3

1971 *Ichthyosarcolites rotundus* Polšak; Carbone *et al.*: 147, fig. 3.16

1981 *Ichthyosarcolites rotundus* Polšak; Carbone *et al.*: 147, fig. 3.16

1998 *Ichthyosarcolites monocarinatus* Sliškovic; Cestari *et al.*: 70, figs 1-2

1998 *Ichthyosarcolites rotundus* Polšak; Özer: 240, figs 6-7

1999 *Ichthyosarcolites rotundus* Polšak; Özer: 69, pl. I, fig. 3.4

2001 *Ichthyosarcolites rotundus* Polšak; Özer *et al.*: 860, figs 10-11

2015 *Ichthyosarcolites monocarinatus* Desmarest; Troya García: 260-264, figs 145-147

4.1.4.1. Type material

The holotype of *I. rotundus* (fig. 3.2-A) is hosted at the Laboratory of Geology and Paleontology of the Faculty of Science of Zagreb (collection number 778). Cenomanian, Fontanela bay (Istria, Croatia).

4.1.4.2. Diagnosis

Ichthyosarcolites species with very loosely coiled right valve. Complete absence of flange or thickening of the shell. Cross section of the shell and of the general cavity ovoid with a longer dorsoventral diameter. Presence of a distinct PAC on the right valve, slightly bigger than sockets. Aragonitic layers filled with capillary canals and an external row of palisading canals.

4.1.4.3. Description

Left valve unknown. Right valve narrow tubular and nearly straight. Right valve length unknown. Section circular (3.5 to 15cm in diameter) to elliptic in the dorsoventral axis (length/width: 6/4.5cm). Shell thickness constant on all the shell for approximately 1/3 to 1/4 of the diameter. Aragonitic layer of the right valve densely perforated with closely arranged, very small, round to slightly ovoid capillary canals, sometimes honeycombed (diameter: 0.5 to 3mm). The external side of the aragonitic shell is bordered with a single row of radial palisading canals of same width as capillary canals, but sub-rectangular in cross section. Outer calcitic shell layer unknown. Ovoid cavity (length/width: 0.7/1cm) present posteroventrally, of equal size than sockets (PAC; fig. 3.3-A) on the right valve. Very thin tabulae occurring in general cavity, canals, sockets and PAC. The blade between this cavity and the general cavity is so thin that it contains no canal, as the blades separating sockets from the general cavity. No ligamentary ridge and cavity.

Right valve myocardinal apparatus – Rudist with inverse dentition. Central tooth shape unknown, filled with canals of variable size, ranging between less than 1 to 10 mm in diameter. Anterior and posterior tooth sockets directly connected to the general cavity. Anterior tooth socket size between 5/5mm and 15/5mm, posterior tooth socket between 12/5mm and 15/5mm (length/width). Sockets always very close one to another (distance of 1 to 7 mm). Posterior socket sometimes sub-pentagonal: the socket presents a right angle on its posterodorsal side, two acute angles on the posteroventral side and the anterior side, and to open angles on the anterodorsal and anteroventral sides, with two small faces on the anterior side (fig. 3.3-A) making it stranded on the anteroposterior axis. Anterior of the posterior tooth socket of similar size. Myophores undifferentiated.

4.1.4.4. Remarks

The species *I. rotundus* was named because of a circular right valve, without any flange. Cestari *et al.* (1998) suggested synonymy of *I. rotundus* with *I. monocarinatus* assuming that the flange and shell outline was eroded in the specimens assigned to *I. rotundus*. The examination of samples from La Bédoule, southeastern France (fig. 3.3-A), invalidate this hypothesis: a continuous row of palisading canals is visible in cross section alongside the shell outline, while it should be discontinuous in case of abraded flange.

4.1.4.5. Occurrences

Exclusively Cenomanian. Living in shallow tropical waters of western Tethys, in western France (Charentes), southeastern France (Provence), and Italy (Sicily) (fig. 3.7; table 3.1).

4.1.5. *Ichthyosarcolites monocarinatus* Slišković, 1966

(fig. 3.2-B)

Abridged citation list

1966 *Ichthyosarcolites monocarinatus* Slišković: 177-178, fig. 3.1

1967 *Ichthyosarcolites monocarinatus* Slišković; Polšak: 80, 186, pl. VI, fig. 3.1, pl. VIII, figs. 3.1-5, pl. IX, fig. 3.1

1968 *Ichthyosarcolites iokungensis* Bobkova: 289-290, pl. 68

1976 *Ichthyosarcolites monocarinatus* Slišković; Praturlon & Sirna: 90-100, tab.1, fig.18

1983 *Ichthyosarcolites monocarinatus major* Slišković; Slišković: 19-21, fig. 3.1

1992 *Ichthyosarcolites monocarinatus* Slišković; Turnsek *et al.*: 220, pl. 14, fig.2

1998 *Ichthyosarcolites monocarinatus* Slišković; Cestari *et al.*: 70-72, figs 1, 2/1-6

? 1999 *Ichthyosarcolites monocarinatus* Slišković; Pleničar & Jurkovšek: 77, pl. 11, fig. 3.3a-b

2005 *Ichthyosarcolites* cf. *monocarinatus* Slišković; Pleničar: 34, pl. V, fig. 3.1

2015 *Ichthyosarcolites monocarinatus* Desmarest; Troya García: 260-264, 269, figs 145-147

4.1.5.1. Type material

The holotype MB370 is hosted at the Geological-Paleontological Collection of the National Museum in Sarajevo plus one paratype MB 4197(3). Cenomanian, Dinaric Alps.

4.1.5.2. Diagnosis

Ichthyosarcolites species with loosely coiled shell. A unique ventral flange of variable size, from short, triangular in cross section (1/1 cm) to thin and of the size of the general cavity. General cavity ovoid in the dorsoventral axis.

4.1.5.3. Description

Left valve unknown. Right valve planispiral. Right valve only known from a 40cm long butt-shaped fragment, rotating of 90 degrees in 20cm followed by 20cm sub-straight. Elliptic section in the dorsoventral axis: 3/4(+0.3) cm, 11/13(+2) cm, 4/4.4(+1) cm, the three values representing individual specimen length/width and flange length. Shell thickness approximately constant on all the shell for 1/2 to 1/5 of the diameter. Presence of a unique flange, always in anteroventral position. Flange shaped as a little bulge rounded and protruding of only 5mm to a more developed one forming an isosceles triangle of 2cm length with a largely rounded end (fig. 3.2-B). Because of this flange, the cylindrical shell presents a flat ventral side. The aragonitic layer of the right valve is densely perforated with close very small capillary canals round to slightly ovoid (diameter approximately of 1 mm). Outer calcitic shell layer unknown. A single other ovoid cavity 0.4/1 cm is present posteroventrally of the right valve, of equal size to slightly more important than sockets and PAC. Very thin tabulae present on the general cavity, canals, sockets and the PAC. Blade between PAC and the general cavity so thin that it contains no canal, as the blades separating sockets from the general cavity. No ligamentary ridge and cavity.

Right valve myocardinal apparatus – Rudist with inverse dentition. Central tooth of unknown shape, filled with canals. Diameter between 2 to 6 mm. Presence of two ovoid (approximately 4/8 mm) to polygonal very close anterior and posterior tooth sockets directly connected to the general cavity. One is generally a little more elongate than the other in the anteroposterior axis. Myophores undifferentiated from the inner shell.

4.1.5.4. Remarks

Ichthyosarcolites monocarinatus (Slišković, 1966) is a species described with one flange, and an ovoid general cavity. A subspecies *I. monocarinatus major* was raised to differentiate forms with an '*highly developed longitudinal edge*' (Slišković, 1983). *I. iokungensis* is assumed to differ from *I. triangularis* by an ovoid section of the general cavity, the absence of costae and a poorly developed ventral flange (Bobkova, 1968). Both the descriptions of *I. monocarinatus* and *I. iokungensis* match the definition of *I. monocarinatus*, assuming variability in expression of flange development.

4.1.5.5. Occurrences

Exclusively Cenomanian. Living in the tropical, shallow waters carbonate platforms of the European Tethyan margins, in Slovenia, Italy (Abruzzo, Latio, Sicilia), Croatia, Bosnia Herzegovina and Turkey (Denizli).

4.1.6. *Ichthyosarcolites bicarinatus* Gemmellaro, 1865

(figs. 3.2-C to D, 3.3-B to E, 3.5-A)

Abridged citation list

1865 *Caprinella bicarinata* Gemmellaro: 236-237, pl. IV, figs 5, 6

1898 *Caprinella bicarinata*; Douvillé: 150

1914 *Ichthyosarcolites bicarinatus* Gemmellaro; Parona: 21

1921 *Ichthyosarcolites bicarinatus* (Gemm); Parona: 12-13, pl. II, fig. 3.1, figs 7, 10

1921 *Ichthyosarcolites tricarinatus* n. f.; Parona: 13, pl. II, fig. 3.2, figs 8, 11

1937 *Ichthyosarcolites bicarinatus* (Gemmellaro); MacGillavry: 49, 51, 54-56

1937 *Ichthyosarcolites tricarinatus* Parona; MacGillavry: 49, 55, 54, 56

1964 *Ichthyosarcolites poljaki* n. sp.; Polšak: 66

1965 *Ichthyosarcolites rogi* n. sp.; Pleničar: 97-98, 100-101, figs. 3.8-10

1966 *Ichthyosarcolites polycarinatus* Slišković: 178, fig. 3.2

1967 *Ichthyosarcolites bicarinatus* Gemmellaro; Polšak: 186, pl. 5, fig. 3.1B-C

? 1967 *Ichthyosarcolites bicarinatus* Gemmellaro; Polšak: 186, fig. 3.21, pl. 5, fig. 3.2A

1967 *Ichthyosarcolites tricarinatus* Parona; Polšak: 76, 185, fig. 3.19, pl. 4, figs 5, 7, pl. 5, fig. 3.2

1967 *Ichthyosarcolites poljaki* Polšak; Polšak: 77, 185, pl. 4, fig. 3.6, pl. 5 fig. 3.1

1971 *Ichthyosarcolites bicarinatus* (Gemm.); Carbone *et al.*: 146-147, fig. 3.14

1971 *Ichthyosarcolites poljaki* (Gemm.); Carbone *et al.*: 147, fig. 3.15

1981 *Ichthyosarcolites bicarinatus* (Gemm.); Carbone *et al.*: 146, fig. 3.14

1981 *Ichthyosarcolites poljaki* Polšak; Carbone *et al.*: 147, fig. 3.16

1981 *Ichthyosarcolites tricarinatus* Parona; Carbone *et al.*: 147-148, fig. 3.17

1982 *Ichthyosarcolites bicarinatus* (Gemmellaro); Sirna: 84, fig. 3.7, pl. I fig. 3.f

1988 *Ichthyosarcolites bicarinatus* Gemmellaro; Accordi *et al.*: p. 139, fig. 3.5, pl. 1, fig. 3.7

1989 *Ichthyosarcolites bicarinatus* Gemmellaro; Accordi *et al.*: p. 165, pl. 2, fig. 3.7

1993 *Ichthyosarcolites bicarinatus* (Gemmellaro); Cherchi *et al.*: p. 93, fig. 3.2, pl. 4, fig. 3.5

1998 *Ichthyosarcolites bicarinatus* Gemmellaro; Özer: p. 240, fig.6-5

1998 *Ichthyosarcolites tricarinatus* Polšak; Özer: p. 240, fig. 3.6-6

1998 *Ichthyosarcolites poljaki* Polšak; Özer: 240, fig 6/4

? 1999 *Ichthyosarcolites poljaki* Polšak, 1967; Pleničar & Jurkovšek: 75-76, pl. 7, fig 2b

not 1999 *Ichthyosarcolites bicarinatus* (Gemmellaro); Pleničar & Jurkovšek: 75-76, pl. 8, figs 2a, 2b, pl. 11, fig. 3.2

not 1999 *Ichthyosarcolites tricarinatus* (Gemmellaro); Pleničar & Jurkovšek: 76, pl. 5, figs 1b, 3, pl. 8, fig. 3.3, pl. 9, figs. 3.1-6, pl. 10, figs. 3.1a, 1b, 2, tab. 11, figs. 3.1a, 1b

2001 *Ichthyosarcolites tricarinatus* Parona; Özer *et al.*: 860, fig. 3.10, 11-4

2002 *Mexicaprina alata* new species; Filkorn: 681-686, figs 4-7

2008 *Ichthyosarcolites alatus* (Filkorn 2002); Aguilar-Pérez: 113-116, fig. 3.64/1-4

2009 *Ichthyosarcolites bicarinatus* (Gemmellaro); Sarı & Ozer: 367, fig. 3.9a

2010 *Ichthyosarcolites bicarinatus* (Gemmellaro, 1865); Chikhi-Aouimeur: 96, fig. 3.87/2-4

2010 *Ichthyosarcolites tricarinatus* (Gemmellaro, 1865); Chikhi-Aouimeur: 97, fig. 3.88/2-3

2013 *Ichthyosarcolites alatus* (Filkorn); Mitchell: 85-87, fig. 3.1/f-g

2015 *Ichthyosarcolites tricarinatus* Parona; Troya García: 254-260, 269-271, figs 142-143, 148, 150

4.1.6.1. Type material

Because the type was not explicitly defined in the original publication by Gemmellaro, we assign the specimen MGUP-021.2-110 from the origin publication as lectotype of *I. bicarinatus* and MGUP-021.2-111 as paralectotype. They are conserved in the Museo Geologico G.G. Gemmellaro della R. Università Degli Studi Di Palermo. Cenomanian, Piemonte (Italy).

4.1.6.2. Diagnosis

Ichthyosarcolites with two to seven flanges, rarely more. The flanges number probably increases during ontogeny. Flanges highly plastic in their position and shape, generally distributed all around the shell, varying in shape from small bulge to very thin and narrow blades that can exceed in size the shell diameter. Orientation of flanges variable, generally perpendicular to shell margin. General cavity ovoid, quadrangular or polygonal. Aragonitic layers filled with capillary canals and an external row of palisading or pyriform canals.

4.1.6.3. Description

Left and right valves are of similar diameters (6/5 to 25/45mm, flange not included). Relative length between valves unknown. Coiling pattern unknown. Always a minimum of two flanges, and up to seven. Shell cross section varying in shape from grossly rounded to polygonal (triangular or quadrate) according to the number of flanges. When two flanges occur they usually appear at two contiguous vertices of a square (fig. 3.3-B). Shape section appearing grossly triangular in case of three flanges (fig. 3.3-C), more quadrate with four flanges (fig. 3.3-D, E). When flange number increases, shell sides change from convex to flat or even concave (fig. 3.2-E); Repartition of flanges globally equidistant when four or

more flanges are present (fig. 3.2-E). Flange development highly variable among individuals and among flanges of single individuals. Flanges regularly extend to a size larger than the diameter of the shell. Flanges usually narrow, parallel sided, and reaching a large size, occasionally short, triangular with rounded extremity. End of a flange always rounded, keel-like, sometime with enlarged margin. Number of flanges increases during the shell growth. Flanges generally orientated perpendicularly to the shell at their base, but can make an angle of 70° to 120° with the shell surface. Outer calcitic layer unknown. Inner aragonitic layer thickness varying between 6 and 25 percent of the diameter of the body cavity. Aragonitic layer densely perforated with, tiny round to honeycomb capillary canals (diameter of approximately 1mm), except in myophores, posterior and anterior teeth. External side of the aragonitic shell bordered by a single row of radial canals of same diameter as capillary canals, but subrectangular to pyriform in cross section. Single row of enlarged rounded canals in the middle of large flanges. Ovoid PAC on the posteroventral side on the right valve, slightly bigger than sockets and filled with tabulae. Tabulae present on both valves in the general cavity, the pallial canals, the sockets and the PAC.

Left valve myocardinal apparatus – Anterior and a posterior conical tooth of similar shape and size (fig. 3.2-E). Wall-like myophore ventrally adjoined to each tooth. Thin vertical myophores attached to the teeth, making a U-shape myophoral apparatus protruding in the general cavity of the right valve. Development of vertical myophores generating a space between it and the inner shell margin that may be occupied with few thin curved walls. Myophoral cavity of the anterior myophore can develop itself in a diverticulum dorsally to the anterior tooth. Myophoral cavity of the posterior tooth always smaller. Anterior myophore extending itself to make a connection with the ventral edge of the shell. Posterior myophore twice shorter, connected to the posterior shell wall of the general cavity by an open angle.

Right valve myocardinal apparatus – Central tooth shape unknown, probably very reduced and filled with pallial canals. Presence of two ovoid anterior and posterior tooth sockets directly connected to the general cavity, elongated in an anteroposterior axis. Myophores undifferentiated from the inner shell, no known muscle scars (fig. 3.3-E).

4.1.6.4. Remarks

Several species names were justified by a number of flanges superior or equal to two and a particular shape of the internal cavity, two elements that we have shown to be correlated according to clustering and discriminant analysis.

Ichthyosarcolites bicarinatus Gemmellaro, 1865; Cenomanian, Piemonte (Italy) – *I. bicarinatus* was the first species to be described with multiple flanges and thus has taxonomic anteriority. Several fossils from Hrušica (Slovenia) assigned in the literature to *I. bicarinatus* (Pleničar and Jurkovšek, 1999) exhibit a ridge connected to a very thin cavity marking a ligament groove that ends near to the posterior tooth socket. The specimens bearing a ligamentary ridge are excluded here from *Ichthyosarcolites*.

Ichthyosarcolites tricarinatus Parona, 1921 – *I. tricarinatus* was described by the combination of three flanges and a quadrangular general cavity. Some specimens described from Hrušica (Slovenia) in Pleničar and Jurkovšek (1999) must be excluded from *Ichthyosarcolites* because of the presence of a ligamentary ridge.

Ichthyosarcolites poljaki Polšak, 1964; Cenomanian, Southern Istria (Croatia) – *I. poljaki* was not described nor illustrated in Polšak (1964). It must be considered a *Nomen nudum*. Plenicar (1965) considered the specimens of Polšak's collection to be conspecific with his new species *I. rogi.*

Ichthyosarcolites rogi Pleničar, 1965; Upper Cenomanian, Kočevski Rog (Slovenia) – *I. rogi* was described as a variation of *I. triangularis*, as a form with three "*narrow and well-marked ribs*", but without the typical triangular cross section. *I. rogi* differs only from *I. bicarinatus* by additional of a flange, which is here considered part of intraspecific variation.

Ichthyosarcolites polycarinatus Slišković, 1966; Upper Cenomanian, Dinaric Alps – The species is based on a single specimen that exhibits seven flanges and a rounded general cavity in cross section. Following our analyses, all the available specimens of *I. polycarinatus* match the new definition of *I. bicarinatus*.

Ichthyosarcolites alatus Filkorn, 2002; Lower Cenomanian, Guerrero (Mexico) – The species *I. alatus* was primarily assigned to *Mexicaprina* before being revised as

Ichthyosarcolites by Aguilar-Pérez (2008). The individual shells are characterized by the presence of three, four or five mostly highly developed flanges. *I. alatus* cannot be differentiated from *I. polycarinatus,* both being characterized by "several" flanges. *I. alatus* from Mexico are known from both Cenomanian and Albian ages.

4.1.6.5. Occurrences

From Albian to Cenomanian (Aguilar-Pérez, 2008). Widespread in tropical shallow marine waters along the Tethyan margins and central Atlantic. *I. bicarinatus* was found in Mexico (Tamaulipas, Guerrero, San Luis Potosí, Querétaro y Colima), France (Provence), Northern Spain, Croatia, Greece, Czech Republic, Italia, Tunisia, Algeria, northwestern Turkey, Libya, and China (Xinjiang; fig. 3.7; table 3.1).

4.1.7. *Incertae sedis* species

The following names by their definitions cannot refer to distinct species. Their description is insufficient to assign them to a species inside the genus *Ichthyosarcolites*. Moreover, a specimen identified as *I. coraloidea* by Caldwell and Evans (1963) is an Antillocaprinid, and the taxa *I. ensis* is not an *Ichthyosarcolites* species as it bear features of *Caprinuloideidae*, and lacks characters required for a definitive taxonomic assignment.

1825 *Ichthyosarcolites obliqua* Deshayes: 501.

1937 *Ichthyosarcolites obliquus* Deshayes 1825; MacGillavry: 52.

Ichthyosarcolites obliquus Deshayes, 1825; Cenomanian, Charentes (France) – Deshayes was the first to understand that filaments accreted to the molds of *Ichthyosarcolites* were an integrant part of the shell, and must be understood as pallial canals. He also found that another species must be separated from *I. triangularis* on the basis of its tabulae morphology which way is more oblique. He erected the species *I. obliqua* but unfortunately, without any drawing. This species is a synonym of *I. triangularis* that already shows oblique tabulae (fig. 3.1-C) with highly variable shapes in local assemblages as at Île d'Oléron (fig. 3.4-A to D).

1850 *Caprinella doublieri*, d'Orb., 1847; d'Orbigny: 191-192, pl. 541

1967 *Ichthyosarcolithes doublieri* (d'Orb); Philip: 498

1973 *Caprinula doublieri* d'Orbigny; Berthou: 97, pl. 57, 58

Ichthyosarcolites doublieri d'Orbigny, 1847; Upper Cenomanian, Martigues (Bouches-du-Rhône, France) – *I. doublieri* was described as an *Ichthyosarcolites* without any tabulae. It was found in Charentes by d'Orbigny (fig. 3.2-F), in Provence (Philip, 1967) and in Portugal (Berthou, 1973), always as internal molds. The number of flanges is unknown in *I. doublieri*. The tabulae are fragile and very thin structures (fig 3-D). On the internal mold in Figure 3.4-B, we can see at the bottom of the mold the presence of tabulae that progressively disappear in the same individual. The thickness and development of tabulae varies among specimens (fig. 3.3-D, 3.4-A), some specimens showing tabulae at varied preservation state, some disappearing partially or completely, while they are entirely lost in other shells. It is likely that *I. doublieri* correspond to a preservation state, with dissolved tabulae of another *Ichthyosarcolites* species.

1854 *Ichthiosarcolites loricatis* Tuomey: 172.

1854 *Ichthiosarcolites cornutis* Tuomey: 172.

1854 *Ichthiosarcolites quadrangularis* Tuomey: 172.

Ichthiosarcolites loricatis Tuomey, *Ichthiosarcolites cornutis* Tuomey, *Ichthiosarcolites quadrangularis* Tuomey; Noxubie county (Mississipi, USA) – The three species were never formally described. Their age is unknown. Tuomey based his new species on the aperture shape (and by extension on the general shape of the shell), while all other characters similar to *I. triangularis*. In *I. quadrangularis*, the aperture is "[…] *terminating obliquely and rather abruptely*"; in *I. loricatis*, it is "[…] *somewhat oval, irregular, obtusely rounded at the base, giving the cast an imbricated appearance*". Finally, *I. cornutis* is "*distinguished from the preceding* [*I. loricatis*] *by the more distant and regular septa, and in the greater regularity of the curvature of the shell*" (Tuomey, 1854). The coiling and the tabulae can be highly variable in a single population of *Ichthyosarcolites*. *I. quadrangularis, loricatis* and *cornutis* do not present sufficiently different features to support the definition new taxa. Key features are unknown, but they could be identified to the family Antillocaprinidae (Stephenson, 1938).

1855 *Caprinella coraloidea* Hall & Meek: 380, pl. II, fig. 3.3, a-f

Ichthyosarcolites coraloidea Hall & Meek, 1855; Upper Campanian, Sage Creek (South Dakota, USA) – This species differs from *I. triangularis* by "*being curved not*

exactly in the same plane, in increasing much more rapidly in size from the apex, and in having the fibrous portion of the shell so thick upon the inner side of the volution as to bring the sides in contact if continued a single turn" (Hall and Meek, 1855). Coiling of the shell is variable enough not to be considered as a valuable criterion for species definition. Desmarest described *I. triangularis* from straight internal molds, but 30 years later, d'Orbigny (1850) illustrated a highly coiled right valve as *I. triangularis* in 'Paléontologie Française' (fig. 3.1-C). The large fossil shells of Charentes show that coiling pattern can change during ontogeny (fig. 3.4-F). In southeastern France, other specimens can be seen completely coiled up to a large size (fig. 3.4-E). In *I. coraloidea* "*The septa are also much more irregular than in the European species, those which are distinct upon the back of the shell often converging so that two unite in a single one on the inner side of the volution*" (Hall and Meek, 1855). The diversity of forms, expressed only in populations of Charentes show the diversity of tabulae obliqueness, some of them occasionally converging (fig. 3.4-C). The number of flanges is unknown in *I. coraloidea* and pending description of new complimentary material, the species name should be considered *incertae sedis*. The unique specimen described as *Ichthyosarcolites coraloidea* by Caldwell and Evans (1963; fig. 3.3-F) is related to Antillocaprinids and not to *Ichthyosarcolites*. This specimen has its sockets and myophoral cavities filled with pallial canals, a characteristic feature of Titanosarcolitinae and Parasarcolitinae.

1887 *Ichthyosarcolites ensis* Počta : 207

Ichthyosarcolites ensis Počta, 1887; Upper Cenomanian, Radovesnice (Central Bohemia, Czech Republic) – The species possesses capillary canals and a triangular flange similar to *Ichthyosarcolites*. However, *I. ensis* was erected for a shell described by Počta with three internal cavities of similar size. The development of the three, large and equally sized internal cavities (two sockets and the general cavity) is likely a feature of Caprinuloideidae. Initial illustrations and descriptions of *I. ensis* remain insufficient for a taxonomic reappraisal without access to the original material.

1994 *Ichthyosarcolites quadratus* Pleničar; Pleničar & Dozet: 188.

Ichthyosarcolites quadratus Pleničar, 1994 – *I. quadratus* was never formally defined and is considered a *nomen nudum* following the article 13 of the ICZN.

4.2. Genus *Oryxia* gen. nov.

Class: Bivalvia Linnaeus, 1758

Infraclass: Heteroconchia Gray, 1854

Order: Hippuritida Newell, 1965

Suborder: Hippuritidina Newell, 1965

Superfamily: Caprinoidea d'Orbigny 1850

Genus *Oryxia* gen. nov.

Type species. *Oryxia sulcata* gen. nov. sp. nov.

4.2.1. Derivation of the name

From the Mammal genus *Oryx* for similarity of its long lightly curved horns and the right valve of the new taxon.

4.2.2. Diagnosis

Rudist bivalve with right valve narrow, gently coiled anteriorly, with a large posteroventral flange and a very slight anteroventral flange with triangular shape in cross section. A ligamentary ridge in dorsal position is connected to a thin ligamentary cavity ending dorsally and parallel to the anterior tooth socket. A bulge is accreted posteriorly to the ligamentary ridge. Aragonitic layers filled with round capillary canals. Presence of an anterior accessory cavity. Small ovoid anterior and posterior sockets of equal size (5 to 10 mm). Central tooth very small (1 to 2 mm), filled with canals.

4.2.3. Remarks

Oryxia differs from *Ichthyosarcolites* in the possession of a small internal ligament cavity, near to the posterior tooth socket and communicating with the outside, where it leads to a longitudinal ligamentary ridge. It differs also by the presence of a small anterior accessory cavity in *Oryxia* (right valve) while it is posterior in *Ichthyosarcolites*. *Oryxia* is a recumbent taxa, as *Ichthyosarcolites*. However, whereas *Ichthyosarcolites* shells were

lying on their anterior side, *Oryxia* lies always on their ventral side due to a flat part extended by their posteroventral flange.

4.2.4. *Oryxia sulcata* sp. nov.

(fig. 3.8)

? 1967 *Ichthyosarcolites bicarinatus* Gemmellaro; Polšak: 186, fig. 3.21, pl. 5, fig. 3.2A.

? 1999 *Ichthyosarcolites poljaki* Polšak, 1967; Pleničar & Jurkovšek: 75-76, pl. 7, fig 2b.

1999 *Ichthyosarcolites bicarinatus* (Gemmellaro); Pleničar & Jurkovšek: 75-76, pl. 8, figs 2a, 2b, pl. 11, fig. 3.2.

1999 *Ichthyosarcolites tricarinatus* (Gemmellaro); Pleničar & Jurkovšek: 76, pl. 5, figs 1b, 3, pl. 8, fig. 3.3, pl. 9, figs. 1-6, pl. 10, figs. 1a, 1b, 2, tab. 11, figs. 1a, 1b.

4.2.4.1. Derivation of the name

From the latin "*sulcus*" because of the presence a longitudinal ligamentary ridge on its shell.

4.2.4.2. Type material

Holotype – Three slices cut in the right valve BJ.1901-HR2 (Pleničar and Jurkovšek, 1999: fig. 3.8-A, C). Cenomanian of Hrušica, Slovenia.

Paratypes – Three slices cut in the right valve BJ.1914-HR2 (Pleničar and Jurkovšek, 1999: fig. 3.8-D), two slices cut in the right valve BJ.1942-HR2. Cenomanian of Hrušica, Slovenia.

4.2.4.3. Diagnosis

Same as genus by monotypy.

4.2.4.4. Description

Left valve unknown. Right valve tubular, narrow and slightly anteriorly coiled (12mm of anterior coiling each 100mm length). Planispiral development. Longest

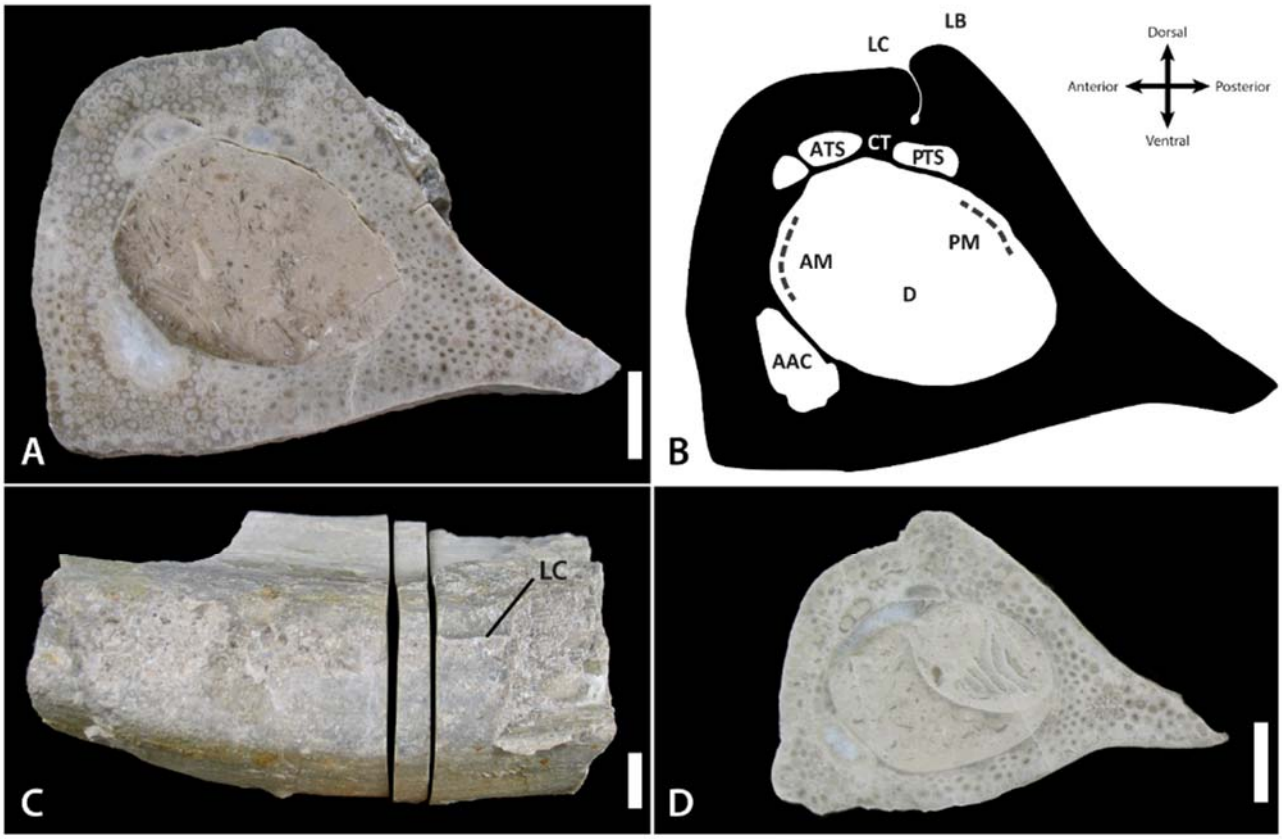


Figure 3. 8. Holotype of *Oryxia sulcata* nov. gen. nov. sp. (right valve; BJ.1901-HR2) sliced in three samples. A – Transversal view of a slice. B – Interpretative scheme of A with main morphoanatomical features and homologs. AM: anterior myophore; PM: posterior myophore; ATS: anterior tooth socket; PTS: posterior tooth socket; CT: central tooth; D: general cavity of the right valve; AAC: anterior accessory cavity; LC: ligamentary cavity; LB: ligamentary bulge. C – Dorsal view of the three samples of the right valve. D – transversal view of the paratype BJ.1914-HR2. All pictures provided by Bogdan Jurkovšek.

known fragment measuring 22cm in length (holotype). Diameter of the shell varying between 4.5 and 7.5cm. Diameter of the general cavity varying between 2.8 and 3.5cm. Shell shape subrectangular because of flanges and shell thickenings. Slight bead of 1.3cm width located dorsally to the posterior tooth socket and posterior to the longitudinal ligamentary ridge. Bead lopsided anteriorly, slightly covering the ligamentary ridge. Subtriangular anteroventral thickening, of 1.3cm width at its base. Presence of a posteroventral flange. Flange margin broken, but has a minimum of 3cm in length for a thickness of 0.5 to 5 mm. Ventral plane surface created by the simultaneous presence of the posteroventral flange and the anteroventral thickening. Very thin ligamentary ridge (8/0.2 mm) connected to the dorsal part of the shell, slightly anteriorly curved in cross section.

Ridge ending with a little anteroposteriorly elongated cavity (2/0.5 mm) located two millimeters dorsally to the posterior tooth. Shell aragonitic; aragonitic inner layer thickness comprised between 1cm and 3cm. Aragonitic layer completely filled with round to ovoid capillary canals with a mean diameter of 1mm. Calcitic outer layer missing, probably thin and smooth. No visible tabulae.

Right valve myocardinal apparatus – Central tooth filled with capillary canals (length/width: 2/1 mm to 3/2 mm). Presence of two flat anterior (4/1 mm to 10/3 cm) and posterior (4/1 mm to 8/2 mm) tooth sockets on the dorsal side of the shell and directly connected to the general cavity. Vertical myophores undifferentiated from the inner shell. Ovoid anterior accessory cavity present anteroventrally (under the anteroventral thickening), larger than sockets (12/5 mm).

4.2.4.5. Remarks

Anterior accessory cavity possibly homologous to the anterior myophoral cavity in Titanosarcolitids. Homology with the accessory posterior cavity in Ichthyosarcolitids (fig. 3.3-A) is impossible because of its location, presumed by the presence of a posterior ligament.

4.2.4.6. Occurrences

Exclusively Cenomanian. *Oryxia sulcata* is only known from the bioherm of Hrušica (Slovenia), on the northern part of the Dinaric carbonate platform. Polšak's (1967) specimens from Istria (Croatia) identified as *I. bicarinatus* may be also *O. sulcata* shells.

5. Chapter conclusion

The Albian-Cenomanian genus *Ichthyosarcolites*, firstly described by Desmarest in 1812, is known from all around the Tethys, from Mexico to China. Following our study and morphoanatomical analyses, we retain only four species of *Ichthyosarcolites* out of the 19 published nominal species. *I. rotundus* possesses no flange, *I. triangularis* a single dorsal flange, *I. monocarinatus* a ventral flange, and *I. bicarinatus* expresses a high polymorphism with the occurrence of two to seven flanges. Homology of the flanges are redefined from their position on the shell. The number of flanges can no

longer serves to delimitate species beyond two flanges. The internal cavity shape is correlated to the external flanges and corroborate the existence of two species bearing a single flange: *I. triangularis*, possessing a dorsal flange and *I. monocarinatus*, possessing a ventral flange. Specimens previously attributed to *Ichthyosarcolites* from Hrušica, Slovenia exhibit a longitudinal ligament ridge and a small and flat cavity dorsal to the anterior tooth outwardly opening by the ridge. The presence and the morphology of a ligament is critical in the evolution of the rudists, and is supporting the raise of *Oryxia sulcata* gen. nov. *Oryxia* is assigned to the superfamily Caprinoidea, and may be very close to families losing their ligamental ridge, such as Trechmanellidae, Antillocaprinidae and Ichthyosarcolitidae. *Oryxia* is important in the phylogeny of the rudists and the evolution of the ligament among rudists is certainly a key in the further phylogenetic explorations of Caprinoidea.

Chapitre 4. Homologies and the phylogeny of rudists (Bivalvia, Hippuritida): new insights from comparative anatomy

Ce chapitre a été écrit en collaboration avec J-P. Masse et L. Villier.

1. Introduction 265
2. Historical background 267
3. Material and methods 273
4. Results 317
5. Discussion 322
6. Chapter conclusion 345
7. Character list 347

1. Introduction

The rudists are a group of heterodont bivalves that appeared during the upper Jurassic, and spread all around the Tethys in warm shallow waters during the Cretaceous. They were eradicated during the end cretaceous mass extinction. Some rudist taxa reached one of the largest shell sizes known in bivalves and developed an outstanding range of morphologies. Enlargement, uncoiling, and development of internal cavities are one of the very specific features appearing during the group evolution. Most prominent changes are the strong development of the myocardinal apparatus (hinge and myophores), the loss of a ligament and the presence of pallial canals in the shell that can adopt a large diversity of shapes. The wide morphological diversity allows rudists to adopt varied life habits, as they are represented by three main ecotypes adapted to specific environmental conditions (hard or soft substrate, high to low sedimentation rate, turbidity, water energy; Skelton 1978). The development of rudists as dense gregarious congregations in shallow, warm waters during the Cretaceous contributed to the edification of particular carbonate platforms (Gili *et al.* 1995; Carannante *et al.* 1999; Skelton and Gili 2012), their shell debris becoming a major component of the carbonate factory.

Rudists were reported first probably by Palissy (1563), and quoted by Sauvages (1746) in southern France. Lamarck (1819, 1835) gathered several fossils under the name ‘les rudistes‘, and defined the group as inequivalve fossil shells that do not seems to have hinge, nor ligament and muscle scars. The initial diagnostic characters were inaccurately described. Rudists nowadays include all inequivalve bivalves with a pachyodont myocardinal apparatus and an outer calcitic layer with a fibrillar prismatic microstructure (Skelton 2013a). Rudists are gathered into the order Hippuritida (Newell 1965), recently redefined by Bouchet and Rocroi (2010). The first works on rudist evolution (Douvillé 1886, 1887, 1935) recognized two main lineages based only on the morphology of the posterior myophore of the right valve (fig. 4.1). However, the classification of rudists considers a high diversity of morphological characters. As an example, pallial canals originating from the accessory cavities up to invade the entire skeletal structures are considered as diagnostic characters for some taxa (Mac Gillavry 1937; Dechaseaux 1969; Damestoy 1970; Skelton and Smith 2000; Sano *et al.* 2012). Numerous informative

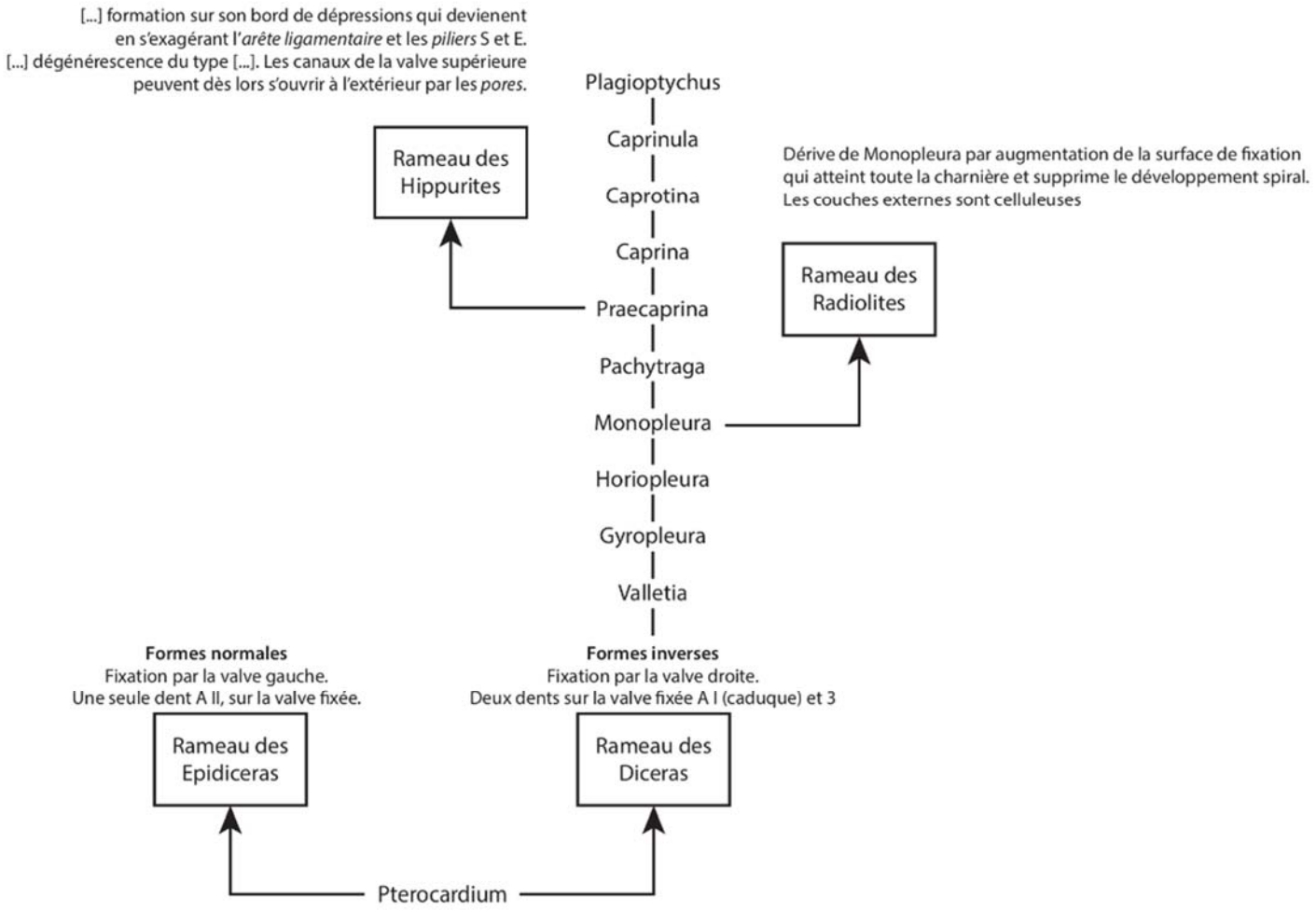


Figure 4. 1. Evolutionary scheme transcribing the view of Douvillé (1935). Four 'rameaux' (branches) evolve from a *Pterocardium* origin. Each branch is supported by morphological features. The 'Rameau des *Diceras'* is much detailed because it's the starting point to two other lineages, the *Hippurites* branch and the *Radiolites* branch.

characters can be derived for systematics and evolution, but the delineation of meaningful homologies is not straightforward on bivalve shells. The definition and inclusiveness of most taxonomic groups remains unsettled (Skelton 2013a).

The uncertainties in modern phylogenetic scenarios likely arise from incorrect assumptions of equivalence between the recognized 'morphological types' and the establishment of homologies, crucial in the goal of building a comprehensive evolutionary framework. The main objectives of this paper are the formalization of new homology hypotheses and the resolution of the main uncertainties of previous cladistic analyses on

rudists (Chartrousse 1994; Skelton and Smith 2000; Mitchell 2013b; Sano *et al.* 2014). There is a consensus on the main dichotomy between the suborders Requieniidina and Hippuritidina, our study will focus only on Hippuritidina, which is the most diverse group. The current taxonomy of the Hippuritidina counts 13 families that are traditionally assumed originated from the paraphyletic one, Monopleuridae. Two species of each of the 13 families are considered in our analysis. Refinement in the understanding of rudist characters should bring the keys to understand the evolutionary history of rudists and to refine the actual taxonomic scheme.

2. Historical background

The first comprehensive work concerning rudist evolution dated from 1886 with the 'Essai sur la morphologie des rudistes' by Douvillé, 105 years after the first description of rudists by Sauvages (1746). The recognition of rudists as a group with common ancestry and its taxonomic position have been of discussion for decades. The term 'les rudistes' was born with Lamarck (1819) to designate an assemblage of fossils with 'Ligament, charnière et animal inconnus. Coquille très inéquivalve. Point de crochets distincts' and '[…] ni charnière, ni ligament des valves, ni muscle d'attache, et qu'on n'aperçoit aucune trace qui indique la place où ces objets pourraient se trouver'. Rudist definition was thus grounded on missing morphological features and mixed taxa of varied affinities: 'Sphérulite, Radiolite, Calcéole, Birostrite, Discine et Cranie'. They also included brachiopods (*Crania* and *Discina*), corals (*Calceola*) but excluded the rudists *Diceras* and *Hippurites* that were described in other groups because of distinctive characters. Des Moulins (1826) understood Lamarck's assemblage as unnatural because there was no positive character linking them in an exclusive way. Deshayes (1825) and Des Moulins (1826) distinguished the characters shared by *Crania* and *Discina* with the other brachiopods, and kept the name 'Rudistes' for *Sphaerulites*, *Radiolites* and *Hippurites*. For the next decades, the rudists integrated some new groups with *Caprina* (Orbigny 1840), *Caprotina* and *Ichthyosarcolites* (Orbigny 1842a, 1842b) up to the modern configuration of the order Hippuritida (Newell 1965).

The highly modified shell morphologies were alternatively understood as deriving from Orthoceratids, Corals, Annelids, Brachiopods (see the review of Woodward 1855 for

early hypotheses), etc. The recognition of two valves nested rudists within bivalves (Deshayes 1832; Douvillé 1935), and their hinge and muscle scars are seen as typical of Heterodonta (Carter *et al.* 2011). The inclusion of the Chamidae in rudists is a hypothesis present in several papers such as 'Sur quelques formes peu connues de la famille des Chamidés' (Douvillé 1887) but is no longer retained in the modern works (Kennedy *et al.* 1970).

The first comprehensive evolutionary hypothesis grounded on precise characters was proposed by Douvillé (1886). He defined the 'normal' forms, with two teeth and one socket in the right valve and two sockets and one tooth in the left valve, by their attachment to the substrate by the left valve. They included *Diceras* and *Heterodiceras*. The other 'inverse' forms had two teeth and one socket in the left valve and two sockets and one tooth in the right valve. They were attached by the right valve. Douvillé (1886) described the 'normal' and 'inverse' as a lineage grading from *Diceras* to *Heterodiceras*, *Monopleura*, *Plagioptychus*, *Caprotina*, *Sphaerulites*, *Hippurites*, and to *Radiolites*. The 'normal' forms are presumed to be ancestors of the '*inverse*' forms. The posterior muscle on the left valve is supported by a myophoral blade projected in the right valve and inserted in an associated socket, which links *Plagioptychus* to *Caprotina*. Two cavities named O and O' link *Sphaerulites* with *Caprotina*, and the whole cardinal apparatus is seen as analogous in *Hippurites* and *Sphaerulites*. Douvillé's (1887, 1935) thoughts about rudist evolution radically changed later on (fig. 4.1). The hypothesis of a direct lineage was abandoned to the benefit of a dichotomy between the 'normal' and the 'inverse' forms. This dichotomy is still used in current classifications. The first division in the rudist taxonomy is made by the two suborders defined by their attachment valve: the left valve for the suborder Requieniidina and the right valve for the order Hippuritidina (Skelton 2013a). Within the group of inverse forms the '*gyropleurinés*' are separated from the '*monopleurinés*' (Douvillé 1889). In *Monopleura*, the posterior myophore of the right valve is depressed, whereas there is a prominent and blade-like posterior myophore in *Gyropleura*. This dichotomy guided assignment of other genera: *Polyconites* was phylogenetically assigned as close to *Monopleura* and *Horiopleura* to *Gyropleura*, because of their posterior myophores assuming convergence for other characters. The monopleurids finally included *Monopleura*, *Polyconites* and *Radiolitids*, and the gyropleurids *Gyropleura*, *Horiopleura*,

Caprina, *Caprinula*, *Caprotina*, hippuritids and *Plagioptychus*. In this context, the canals of *Plagioptychus*, *Caprina* and *Caprinula* appeared independently, and the 'accessory cavities' of *Caprotina* may be at the origin of the canals found in *Hippurites* left valve. The canals form pores in hippuritids because of the disappearance of an external shell layer. The idea of convergent canal appearance survived until today (Skelton 2013a). For Paquier (1905) and Cox (1933), the caprinid canals have an origin in the accessory cavities of caprotinids, whereas it cames from the canals of *Pachytraga* for Douvillé (1935). Paquier affiliated *Ichthyosarcolites* to the monopleurids because of the absence of the blade-like posterior myophore. Woodward (1855) suggested that an internal blade-like structure distinctive of the highly modified *Hippurites* and very similar, and close to the posterior tooth of the left valve might be homologous to a posterior myophore, the observed striations being interpreted as muscle scars. In this hypothesis, the anterior myophores would be divided in two equal parts. Bayle (1855) did not recognize the muscle scar and alleged posterior myophore to be a third tooth, the real posterior myophores migrating to the anterior side, close to the anterior myophores. Bayle's ideas are not retained by modern classifications.

Most works subsequent to Douvillé considered the main taxa (Radiolitidae, Caprinidae, etc.) as independent lineages originating from a huge paraphyletic stem group made of Monopleurids, Caprotinids and Polyconitids. The idea was formalized by Mac Gillavry (1937) and still endured in Skelton's scheme of 2003 (fig. 4.2). Thus, several common characters such as pallial canals, are assumed to be convergent features. MacGillavry recognized that 'the posterior muscle is attached in the left valve to a raised myophorous apophysis, which fits tooth-like into an alveole of the right valve' observed in *Caprotina*, *Dictyoptychus* and *Hippurites*, and that 'the dentition is almost perfectly symmetrical' in *Hippurites* and *Dictyoptychus* could represent phylogenetically informative features. The years 1969 and 1970 mark a rupture in rudist systematics for which Dechaseaux (1969) and Damestoy (1970) reset the value of pallial canals to support a single huge family: the Caprinidae. Damestoy goes further in the use of canal complexity as a marker of caprinid evolution, the most derived stage being expressed in *Dictyoptychus*

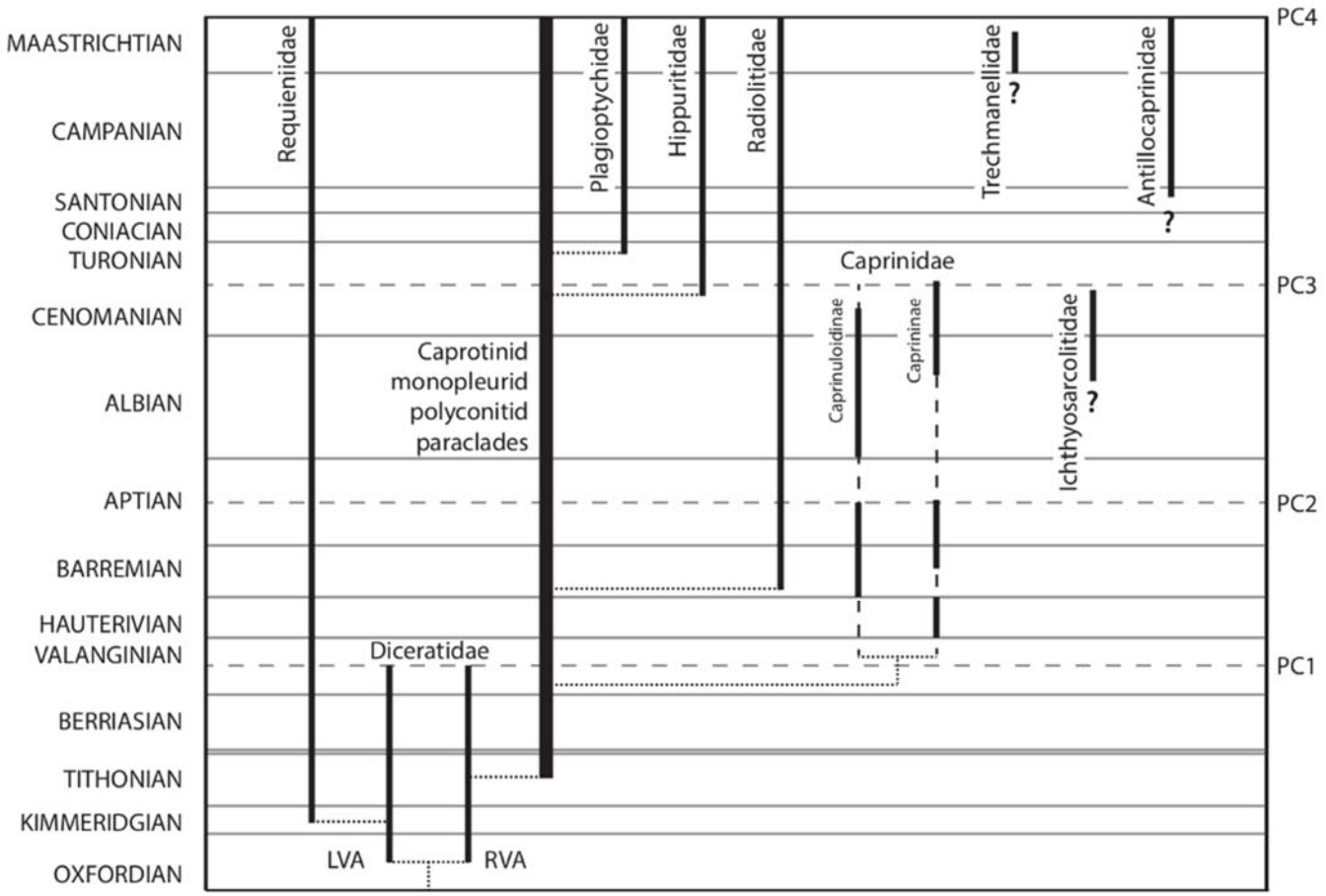


Figure 4. 2. Evolutionary scheme according to Skelton's 2003 tree, replacing all of the rudist families in an evolutionary context. Uncertainties remain in the three Trechmanellidae, Antillocaprinidae and Ichthyosarcolitidae families. The Diceratidae family is partitioned in two branches by their attachment valve. LVA: Left Valve Attachment; RVA: Right Valve Attachment; PC: platform crisis.

(cited as *Threchmanella*), *Paracaprinula*, *Antillocaprina*, *Titanosarcolites* and *Rousselia* with capillary canals. Subsequently, Skelton (1978) was very critical on Dechaseaux's work, and to the presumed monophyly of Caprotinidae. The Caprotinidae were merged with Monopleuridae. Skelton splited Dechaseaux's Caprinidae, assuming polyphyly, as MacGillavry and Douvillé did. Most of the families are seen as independent branches nested in Monopleuridae. The idea of independent lineages and the partition in irreducible families persisted for a long time (Skelton 2003, 2013a) and is still widespread today.

The first cladistic analyses were made by converting Skelton's evolutionary diagram into a cladogram, leading to the first formalization of characters (Smith 1994; Stone & Telford 1999). A 'pachytragiform myophore arrangement' and the presence of

accessory cavities (Smith 1994) became synapomorphic for the clade gathering Caprotinidae and Caprinidae, with Hippuritidae added by Stone and Telford (1999). Caprinidae and Caprinuloideidae formed a clade with pallial canals. These preliminary analyses however suffered from a limited number of characters, and were only a formalization of the diagnostic features of groups used in former evolutionary scenarios and taxonomies, as evidenced by an almost complete lack of homoplasy (Smith 1994; Steuber 1999; Stone and Telford 1999). The cladistic analysis from the thesis of Chartrousse (1994) focused on the relationships between Caprinidae and Caprinuloideidae (with Monopleuridae and Polyconitidae to test the monophyly of the clade Caprinidae plus Caprinuloideidae). 18 genera were used as terminal taxa but there was no attempt to address phylogeny of the entire rudist group. The cladistic analysis of rudist of Skelton and Smith (2000; fig. 4.3) was a leap forward the homology formalization, and remains today the most comprehensive in terms of characters and taxonomic sampling. It embraced all the diversity of rudists in terms of families. It confirmed Douvillé's dichotomy between normal and inverse rudists. Three main clades were recognized in the inverse rudists:

- Caprinidae and Caprinuloideidae.
- Polyconitidae, Caprotinidae, Plagioptychidae, and Hippuritidae.
- Radiolitidae, Ichthyosarcolitidae, Antillocaprinidae, and Trechmanellidae.

More recent analyses focused on specific groups (Mitchell 2013b; Sano *et al.* 2014) and do not integrate all Hippuritida families. In Mitchell's analysis, only Monopleuridae, Caprinidae, Caprinuloideidae, Ichthyosarcolitidae and Antillocaprinidae are sampled (5/13 families) for a total of 24 genus used as terminal taxa. In Sano *et al.* analysis, only Diceratidae, Monopleuridae, Polyconitidae, Caprinulidae, Plagioptychidae, Caprinidae and Caprinuloideidae are sampled (5/13 families) for a total of 13 species used as terminal taxa. We note other structures interpreted as evolutionary diagrams in the literature, generally on less inclusive taxa, and using other theoretical backgrounds than cladistics (Masse and Fenerci-Masse 2010; Masse *et al.* 2015). However, the last taxonomic synthesis of Skelton (2013a) does not strictly follow the results of the previous cladistic analyses, and erected the superfamily Caprinoidea, gathering the families Caprinidae, Caprinuloideidae, Ichthyosarcolitidae and Antillocaprinidae. Another superfamily, Radiolitoidea, designed

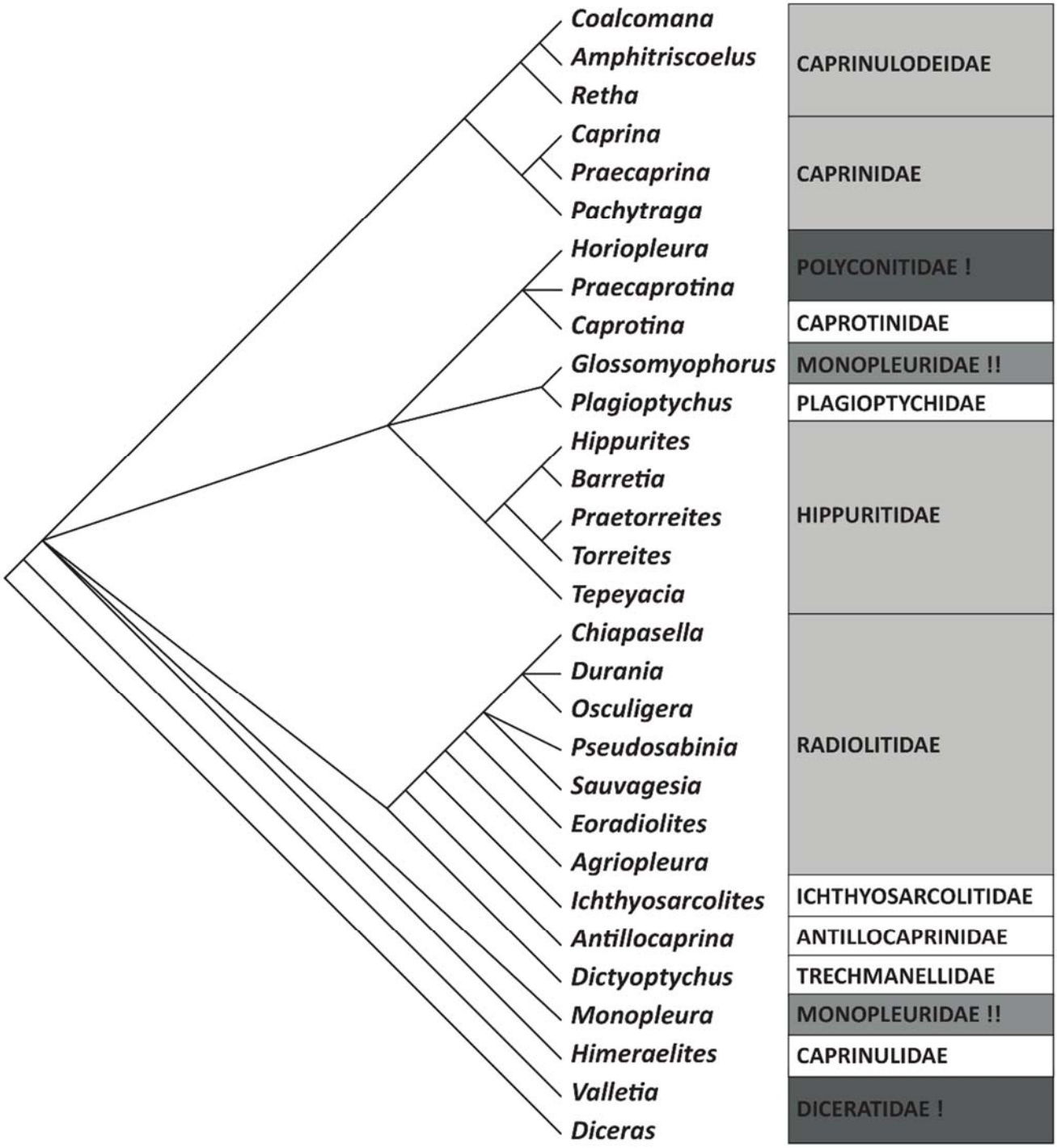


Figure 4. 3. Strict consensus tree redrawn from Skelton and Smith's (2000). The corresponding families are replaced for each genus. One exclamation (!) mark means paraphyly and two means polyphyly (!!).

paraphyletic, was erected to give a name to all non-Caprinoidea families. Skelton's conclusions were that 'Evolutionary relationships […] remain unresolved probably because of relatively frequent convergent and iterative evolution of characters within different lineages' (Skelton 2013a, p 27). To conclude, several of the actual families were never tested by cladistic analysis as it is the case with Trechmanellidae, Caprotinidae, and Plagioptychidae.

3. Material and methods

3.1. Taxonomic sampling

We assume Skelton's assertions about convergence among rudists reveals a problem that can only be solved by cladistics and its testing procedure. The goal of our analysis is to evaluate the phylogenetic relationships of rudists at a high taxonomic scale with an extended reappraisal of morphological characters. Two species were selected in each family to sweep the spectrum of morphological diversity expressed in the Hippuritidina, in which Skelton's review (2013a) enumerates 147 named genera. Four families (e.g. Diceratidae, Monopleuridae, Polyconitidae and Caprinuloideidae) are known or presumed to be paraphyletic (Skelton & Smith 2000; Skelton 2013a), and our sampling allow us to test the monophyly of the whole 13 families of Hippuritidina. For each family, we choose two species based on following criteria: the species must be well known to ensure the most precise coding, the specimens should preferably be preserved in 3D. Species are chosen as minimal taxa preferentially to genus to avoid the possibility of coding polyphyletic taxa. The species are chosen to recover the broadest possible morphologically diversity in terms of character assemblage and covering the maximal temporal range. We also choose species considered to be of critical importance to understand the origins of major taxa. This is the case for the genus *Tepeyacia*, hypothesized as a sister group of Hippuritidae (Skelton and Smith 2000) and *Pachytraga* as a sister group of Hippuritidae (Skelton and Masse 1998; Skelton and Smith 2000). *Sellaea* is retained as a key taxon for testing the family Caprinulidae and the link between *Himeraelites*, *Neocaprina* and *Caprinula* (Skelton and Masse 1998; Skelton and Smith 2000; Steuber and Bachmann 2002; Skelton 2013a). *Oryxia* is also retained for our analysis. This species has a unique

association between its morphology and its ligamentary cavity. Finally, a total of 27 species were considered for study. Thereafter in the text, the species are called by their genus name. All analyzed taxa are illustrated in figures 4.4 to 4.9 (photographies) and figures 4.10 to 4.13 (drawings). Characters analysis is either made from observations of 3D preserved specimens, drawings, figures, 3D reconstructions or descriptions from the literature (table 4.1). Additional specimens and taxa, sampling complementary character combinations, were however considered in homology definition and evolutionary interpretation. For example, *Kimbleia capacis* is not coded, but exhibits a distinct pattern of canals in the left valve (Aguilar-Pérez 2008) which is taken into account for character formalization. For rooting characters, we mainly use taxa from the family Diceratidae as outgroup, considered as stem group for the Hippuritidina. In the discussion section, the coded species are assumed representative of their genera (unless specific mention indicates the contrary). Thus, we assimilate the species to their genera to lighten the discussion about suprageneric relationships. The dating of the clades take into account the earliest reported occurrence of the genera, following Steuber *et al.* (2016).

Abbreviations. AMU – Centre Européen de Recherche et d'Enseignement des Géosciences de l'Environnement, Aix-Marseille University (JPMA: Jean-Pierre Masse collection; CC: Alexandre Chartrousse collection); BJ – Paleontological collection of Dr. Bogdan Jurkovšek and Dr. Tea Kolar-Jurkovšek, Ministry of Culture of Republic of Slovenia and the Natural History Museum of Slovenia in Ljubljana; BM (and NHM and BMNH) – Natural History Museum, London; MGUP – Museo Geologico G.G. Gemmellaro della R. Università Degli Studi Di Palermo; MNHN – Muséum National d'Histoire Naturelle, Paris; MV – Musée Vert, Muséum d'Histoire Naturelle du Mans; PUAB – Universidad Autònoma de Barcelona; UCBL – Université Claude Bernard Lyon 1; UNAM – Universidad Nacional Autónoma de México; UPMC – Université Pierre et Marie Curie; USNM – Smithsonian National Museum of Natural History; UT (and WSA) – University of Texas; UWIGM – University of West Indies Geological Museum.

Family	Species	References of used specimens	Status	Type of material
Diceratidae	*Valletia germani*	MNHN.F.J05798	Lectotype of Valletia tombecki, junior synonym (Munier-Chalmas, 1882)	Cleared valves
		MNHN.F.J15659	Paralectotype of Valletia tombecki, junior synonym (Munier-Chalmas, 1882)	Cleared valves
		MNHN.F.J05799	Paralectotype of Valletia tombecki, junior synonym (Munier-Chalmas, 1882)	Cleared valves
	Diceras arietinum	MNHN.F.A52439	Syntype (Lamarck, 1805)	Cleared valves
Monopleuridae	*Monopleura michaillensis*	UPMC.1272	NA	Cleared valves
	Glossomyophorus costatus	MNHN.F.R06779	Holotype (Masse et al., 1984)	Cleared valves
		MNHN.F.A27280	Paratype (Masse et al., 1984)	Cleared valves
Radiolitidae	*Radiolites radiosus*	MNHN.F.J09137	Figured (Toucas, 1908)	Cleared valves
		MNHN.F.J09021	Figured (Toucas, 1908)	Cleared valves
		MNHN.F.B17070	NA	Cleared valves
	Agriopleura marticensis	JPMA.15208	Figured (Masse and Fenerci-Masse, 2014)	Photographies (sliced specimen)
Hippuritidae	*Hippurites radiosus*	MNHN.F.A58757	NA	Cleared valves
	Barrettia monilifera	UWIGM.RUD.2006.222	Figured (Mitchell, 2010)	Photographies (sliced specimen)
Caprinuloideidae	*Mexicaprina minuta*	JPMA.14097-1, JPMA.14097-5	Figured (Chartrousse, 1998)	Cleared valves
	Amphitriscoelus waringi	NA	NA	Photographies (sliced specimen)
Caprinidae	*Pachytraga paradoxa*	JPMA.17738-1, JPMA.17738-2	NA	Cleared valves
	Pachytraga paradoxa	NA	Figured (Paquier, 1905)	Photographies
	Offneria rhodanica	UCBL.R781	Holotype (Paquier, 1905)	Cleared valves
		UCBL.R275	Figure (Paquier, 1905)	Sliced specimen
Trechmanellidae	*Dictyoptichus morgani*	NA	Figured (*Polyptychus morgani*; Douvillé, 1904)	Drawings
		NHM.L.58420, L.58422, L.58424, L.58427	Figured (Cox, 1934; Skelton, 2013)	Photographies (sliced specimen)
	Semailia smithi	BM.LL41931	Holotype Morris and Skelton, 1995)	Photographies (sliced specimen)
Antillocaprinidae	*Titanosarcolites giganteus*	UT.47533, UT.46895	Figured (Mitchell, 2013)	Sliced specimen
		BMNH.L44205	Figured (Chubb, 1971)	Photographies (sliced specimen)
		UWIGM.2004.1	Figured (Mitchell & Gunter, 2006)	Drawings
	Clinocaprina lowentsami	USNM.245241	Holotype (Mitchell, 2013)	Cleared valves
		USNM.245242	Paratype (Mitchell, 2013)	Sliced specimen
Ichthyosarcolitidae	*Ichthyosarcolites bicarinatus*	IGM.6970	Holotype (Filkhorn, 2002)	Photographies (sliced specimen)
		IGM.6972, IGM.6973	Paratype (Filkhorn, 2002)	Photographies (sliced specimen)
	Curtocaprina clabaughikinsorum	UT.10937	Holotype (Mitchell, 2013)	Cleared valves
Plagioptychidae	*Plagioptychus arnaudi*	UPMC.1273	NA	Cleared valves
		NA	Figured (Douvillé, 1888)	Drawings
	Coralliochama orcutti	USNM.186662	Paralectotype (Marincovitch, 1973)	Cleared valves
		USNM.186665	Paralectotype (Marincovitch, 1973)	Sliced specimen
Polyconitidae	*Polyconites verneuili*	Ch01, Ch02	Figured (Pascual-Cebrian et al., 2013)	3D-scan
	Tepeyacia corrugata	CAS.66657.01	Holotype (Palmer, 1928)	Photographies (sliced specimen)
		PWS collection	Figured (Skelton and Smith, 2000)	Photographies (sliced specimen)
Caprotinidae	*Caprotina semistriata*	MVM.2003.2.5147	NA	Cleared valves
	Chaperia costata	MNHN.F.R07942	Lectotype (d'Orbigny 1839)	Cleared valves
		MNHN.F.R07950	Paralectotype (d'Orbigny 1839)	
Caprinulidae	*Caprinula boissyi*	MNHN.6924, collection d'Orbigny	Figured (Douvillé, 1888)	Sliced specimen
		CC.Mar.001	Figured (Chartrousse, 1998)	Photographies (sliced specimen)
	Sellaea sp.	WSA3058 '*Sellaea americana* Van de Flied'	NA	Sliced specimen
NA	*Oryxia sulcata*	BJ.1901-HR2	waiting publication (holotype)	Sliced specimen
		BJ.1914-HR2	waiting publication (paratype)	Sliced specimen

Table 4. 1 List of specimens studied for the phylogenetic analysis. For each specimen is specified the family, species, reference, status, publication and type of material when available.

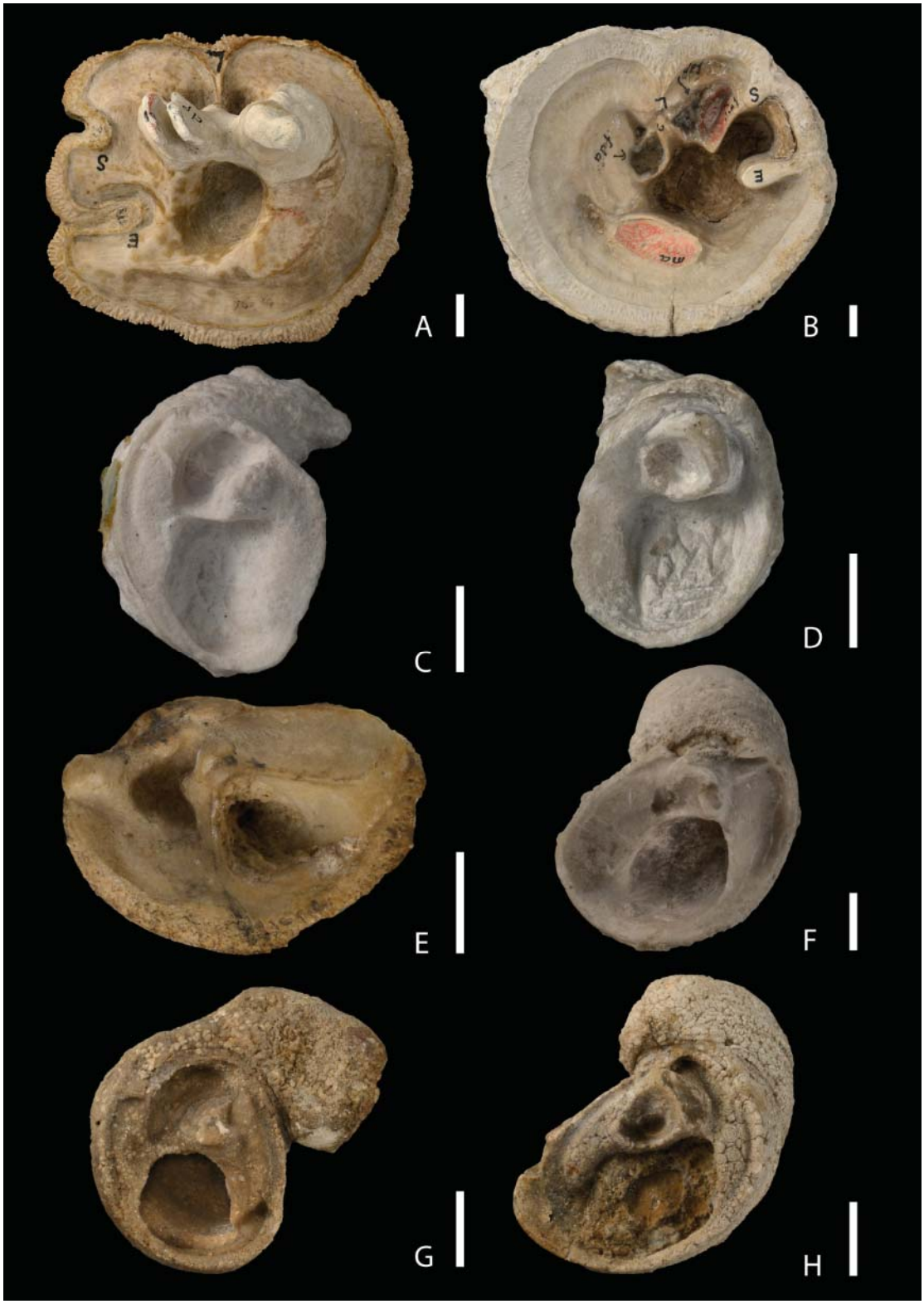


Figure 4. 4. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Hippurites radiosus*, interior of left valve; Senonian, Charentes, France (MNHN.F.A58757). B – *Hippurites radiosus*, interior of right valve; Senonian, Charentes, France (MNHN.F.A58757). C – *Monopleura michaillensis*, interior of left valve; Barremian, Brouzet, Gard, France (UPMC.1272). D – *Monopleura michaillensis*, interior of right valve; Barremian, Brouzet, Gard, France (UPMC.1272). E – *Plagioptychus arnaudi*, interior of left valve; France (UPMC.1273). F – *Plagioptychus arnaudi*, interior of right valve; France (UPMC.1273). G – *Valletia germani*, interior of left valve; lower Cretaceous, Le Fourney, Savoie, France (MNHN.F.J05798). H – *Valletia germani*, interior of right valve, lower Cretaceous, Le Fourney, Savoie, France (MNHN.F.J05798). Scale bars, 10 mm.

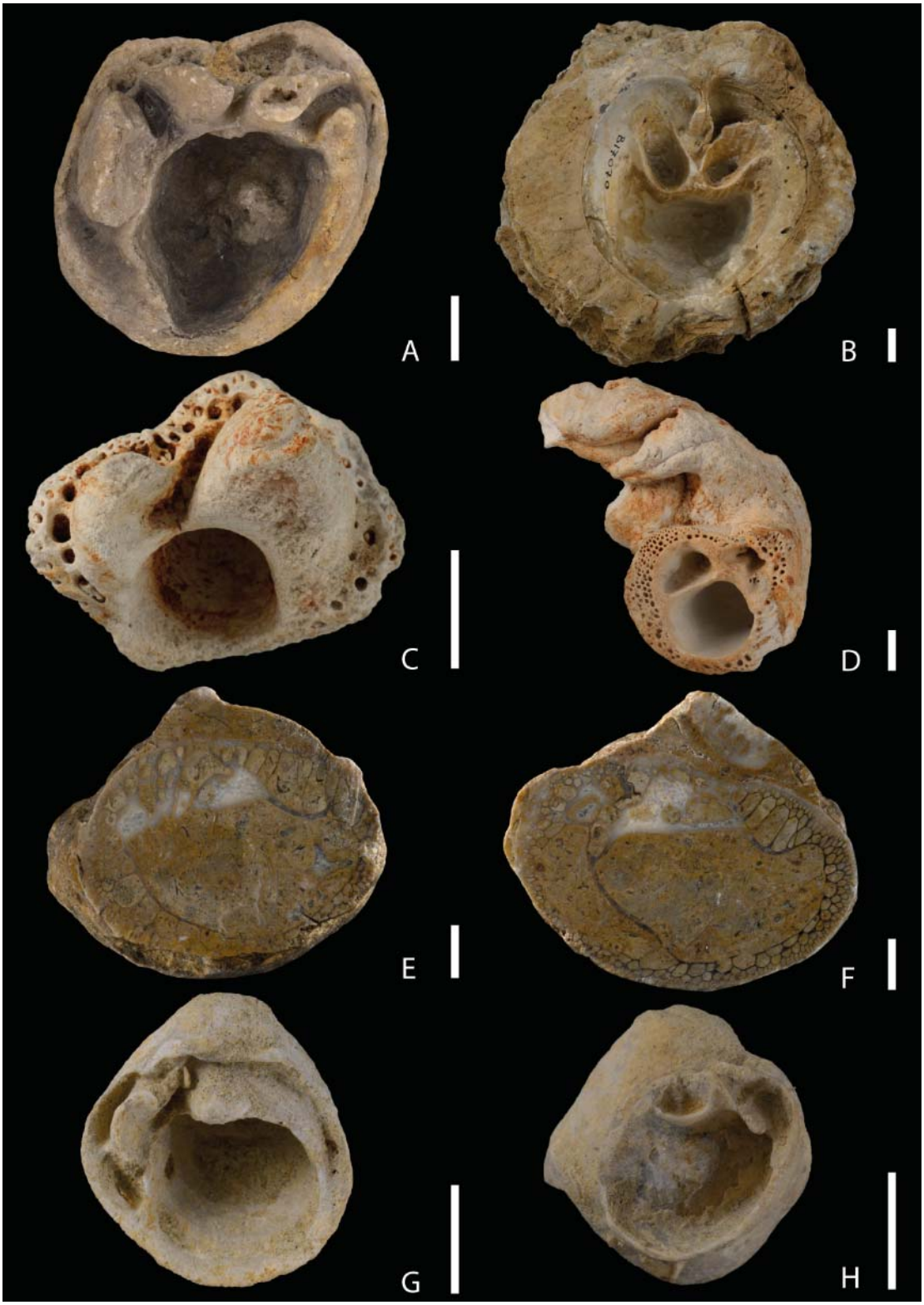


Figure 4. 5. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Radiolites radiosus*, interior of left valve; Turonian, Charentes, France (MNHN.F.J09021). B – *Radiolites radiosus*, interior of right valve; Turonian, Charentes, France (MNHN.F.B17070). C – *Mexicaprina minuta*, interior of left valve; Albian-Cenomanian, Querétaro, Mexico (JPMA.14097-5). D – *Mexicaprina minuta*, interior of right valve Albian-Cenomanian, Querétaro, Mexico (JPMA.14097-1). E – *Caprinula boissyi*, interior of left valve; Turonian, Corbières, Aude, France (Coll. D'Orbigny, MNHN.6924). F – *Caprinula boissyi*, interior of right valve; Turonian, Corbières, Aude, France (Coll. D'Orbigny, MNHN.6924). G – *Caprotina striata*, interior of left valve; Cenomanian, Le Mans, Sarthe, France (Collection Guéranger, MV 2003.2.5147). H – *Caprotina striata*, interior of right valve; Cenomanian, Le Mans, Sarthe, France (Collection Guéranger, MV 2003.2.5147). Scale bars, 10 mm.

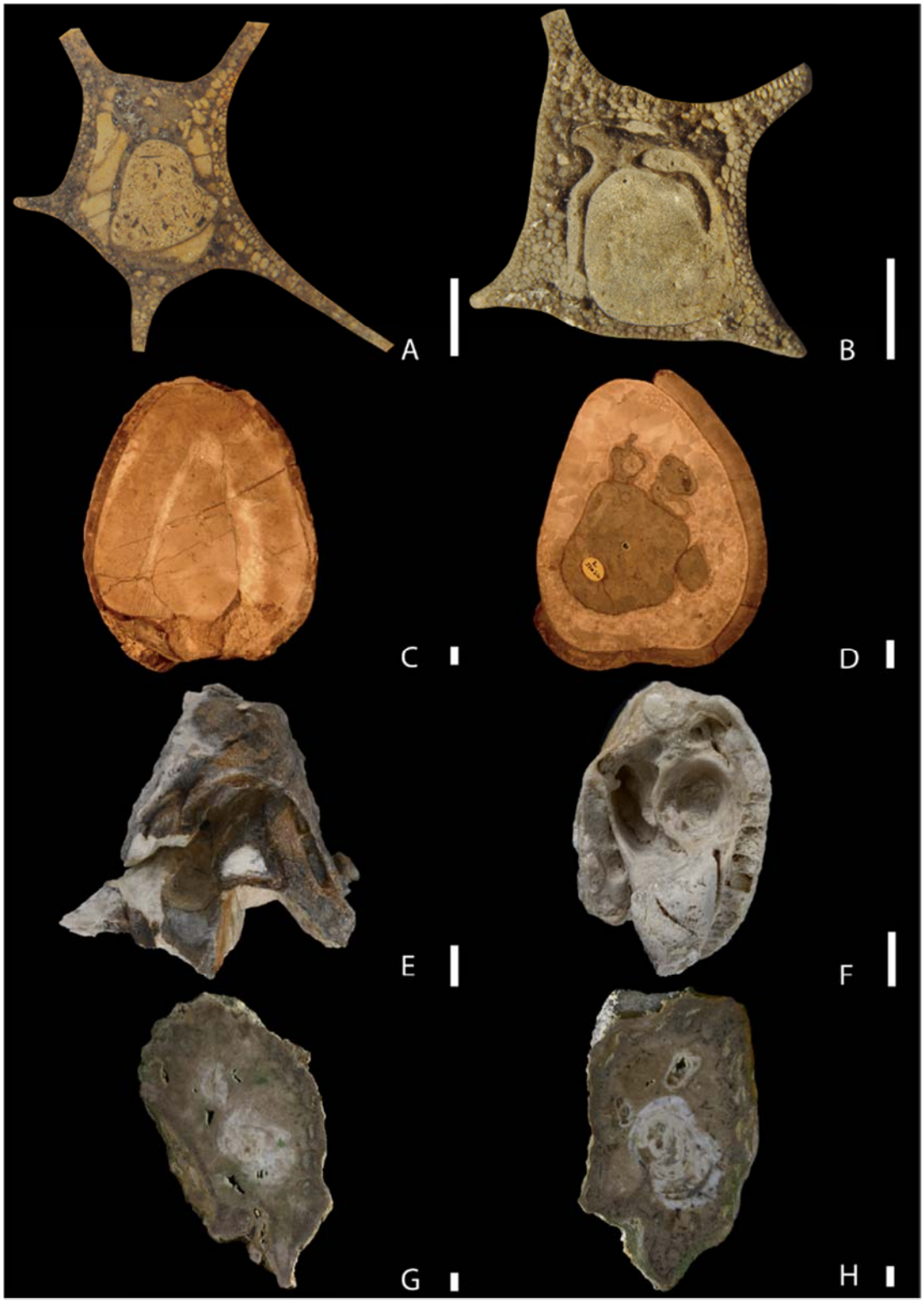


Figure 4. 6. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Ichthyosarcolites bicarinatus*, interior of left valve; Cenomanian, Guerrero, Mexico (IGM.6970). B – *Ichthyosarcolites bicarinatus*, interior of right valve; Cenomanian, Guerrero, Mexico (IGM.6972); pictures provided by Javier Aguilar-Pérez (A and B). C – *Dictyoptychus morgani*, interior of left valve; Maastrichtian, Bakhtiari country, Iran (NHM.L-58427). D – *Dictyoptychus morgani*, interior of right valve; Maastrichtian, Bakhtiari country, Iran (NHM.L-58424); pictures provided by Peter W. Skelton (C and D). E – *Pachytraga paradoxa*, interior of left valve; Aptian, Rustrel, Vaucluse, France (JPMA. 17738-1). F – *Pachytraga paradoxa*, interior of right valve; Aptian, Rustrel, Vaucluse, France (JPMA.17738-2). G – *Titanosarcolites giganteus*, interior of left valve; Maastrichtian, Ocozocuautla Formation, Chiapas, Mexico (UT.47533). H – *Titanosarcolites giganteus*, interior of right valve; Maastrichtian, Ocozocuautla Formation, Chiapas, Mexico (UT.47533). Scale bars, 10 mm.

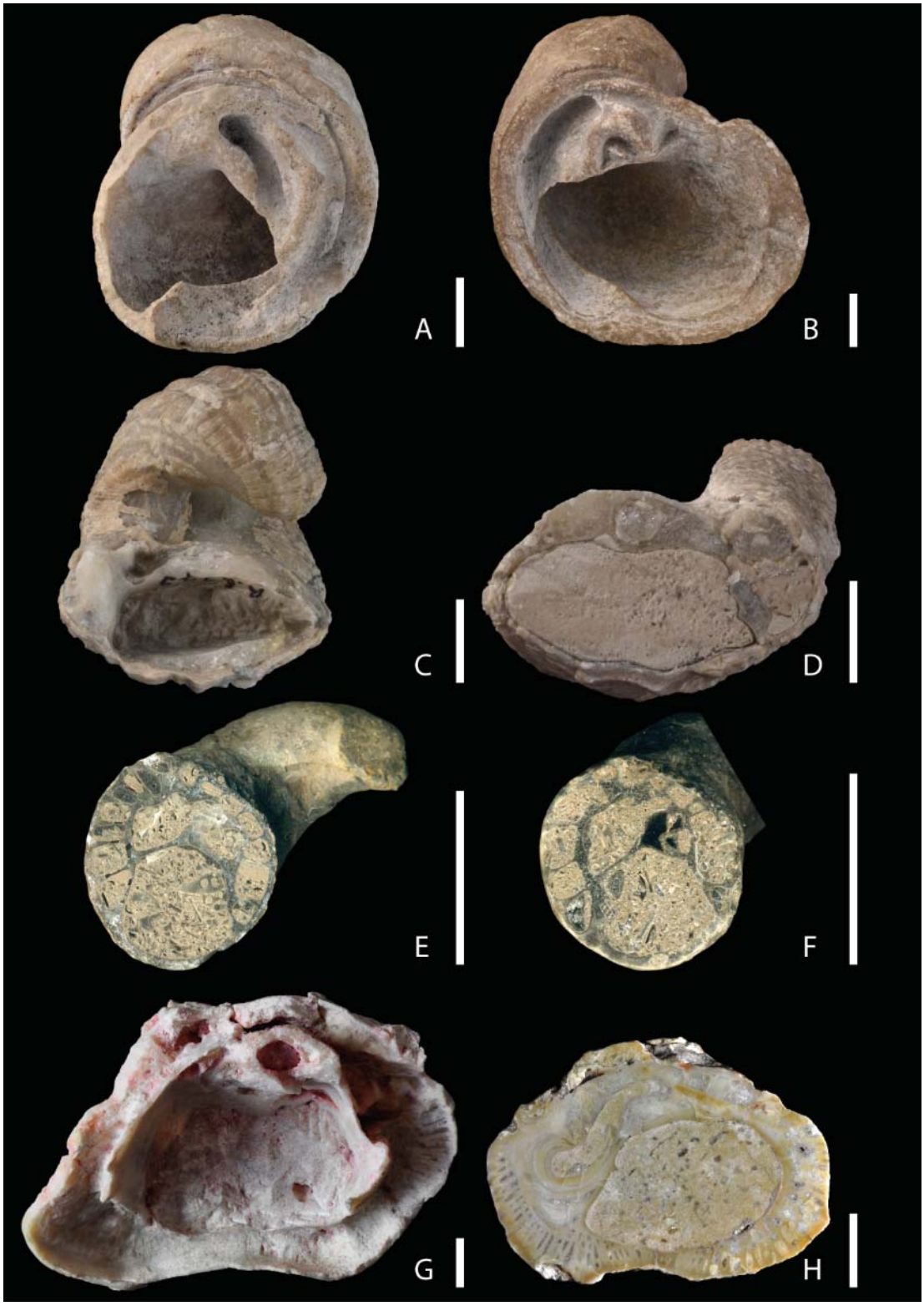


Figure 4. 7. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Diceras arietinum*, interior of left valve, Coulonges, Yonne, France (MNHN.without collection number). B – *Diceras arietinum*, interior of right valve; Oxfordian, Saint-Mihiel, Meuse, France (MNHN.F.A52439). C – *Glossomyophorus costatus*, interior of left valve; Aptian, Roub al Khali, Saudi Arabia (MNHN.F.R06779). D – *Glossomyophorus costatus*, interior of right valve, Aptian, Roub al Khali, Saudi Arabia (MNHN.F.A27280). E – *Amphitriscoelus waringi*, interior of left valve. F – *Amphitriscoelus waringi*, interior of right valve; pictures provided by Jean-Pierre Masse (E and F). G – *Offneria rhodanica*, interior of left valve; Urgonian, Saint-Montand, Ardèche, France (UCBL.R275). H – *Offneria rhodanica*, interior of right valve, Urgonian, Saint-Montand, Ardèche, France (UCBL.R781); pictures provided by Emmanuel Robert (G and H). Scale bars, 10 mm.

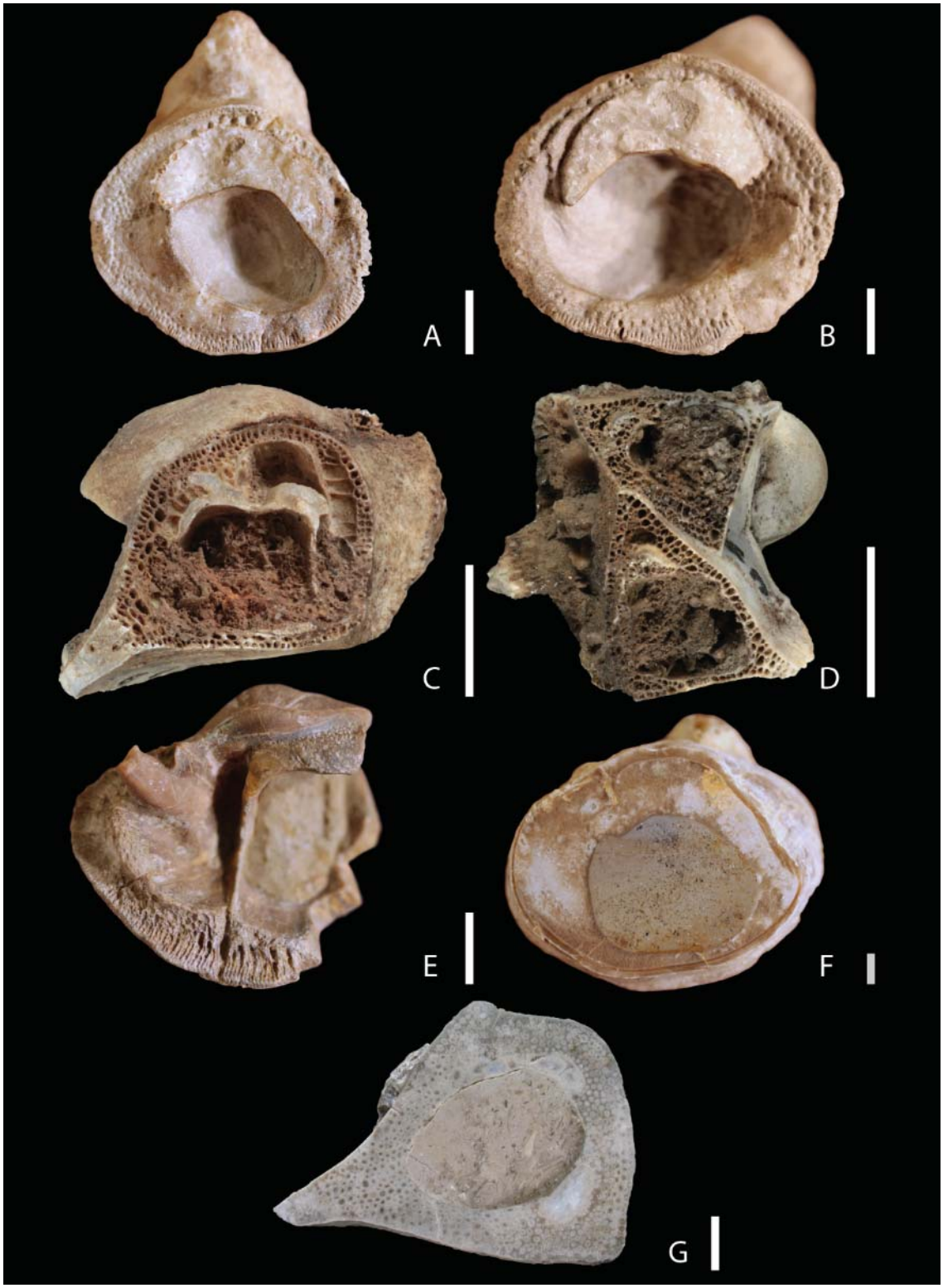


Figure 4. 8. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Clinocaprina lowentsami*, interior of left valve; Campanian, Stapleton Saint James, Jamaica (USNM.245241). B – *Clinocaprina lowentsami*, interior of right valve; Campanian, Stapleton Saint James, Jamaica (USNM.245241). C – *Curtocaprina clabaughikinsorum*, interior of left valve; Albian, Kerr County, Texas, USA (UT.10937). D – *Curtocaprina clabaughikinsorum*, interior of right valve; Albian, Kerr County, Texas, USA (UT.10922). E – *Coralliochama orcutti*, interior of left valve; late Campanian/early Maastrichtian, Punta Banda, Baja California, Mexico (USNM.186662); picture provided by Jean-Pierre Masse. F – *Coralliochama orcutti*, interior of right valve; late Campanian/early Maastrichtian, Punta Banda, Baja California, Mexico (USNM.186665). G – *Oryxia sulcata*, interior of right valve; Cenomanian of Hrušica, Slovenia (BJ.1901-HR2); picture provided by Bogdan Jurkovšek. Scale bars, 10 mm.

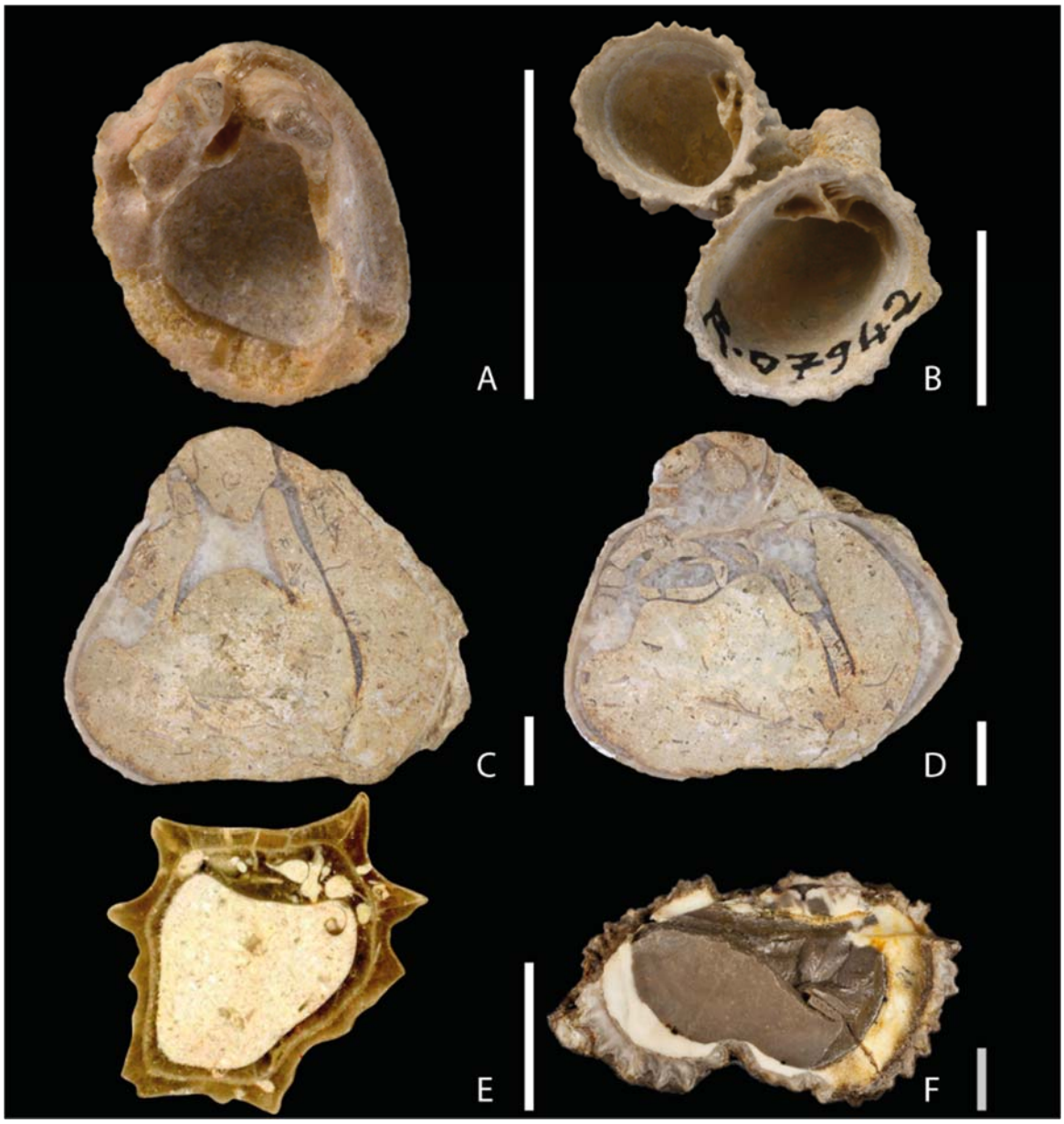


Figure 4. 9. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Chaperia costata*, interior of left valve; Cenomanian, île d'Aix, Charente-Maritime, France (MNHN.F.R07950). B – *Chaperia costata*, interior of right valve; Cenomanian, île d'Aix, Charente-Maritime, France (MNHN.F.R07942). C – '*Sellaea' sp.*, interior of left valve; Mc Lennan County, Texas, USA (WSA3058). D – '*Sellaea' sp.*, interior of right valve; Mc Lennan County, Texas, USA (WSA3058). E – *Agriopleura marticensis*, interior of right valve; Martigues, Bouches-du-Rhône, France (fig. 15-G in Masse and Fenerci-Masse 2014; JPMA.15172-2). F – *Tepeyacia corrugata*, interior of right valve; Turonian, Puebla, Mexico (CAS.66657.01). Scale bars, 10 mm.

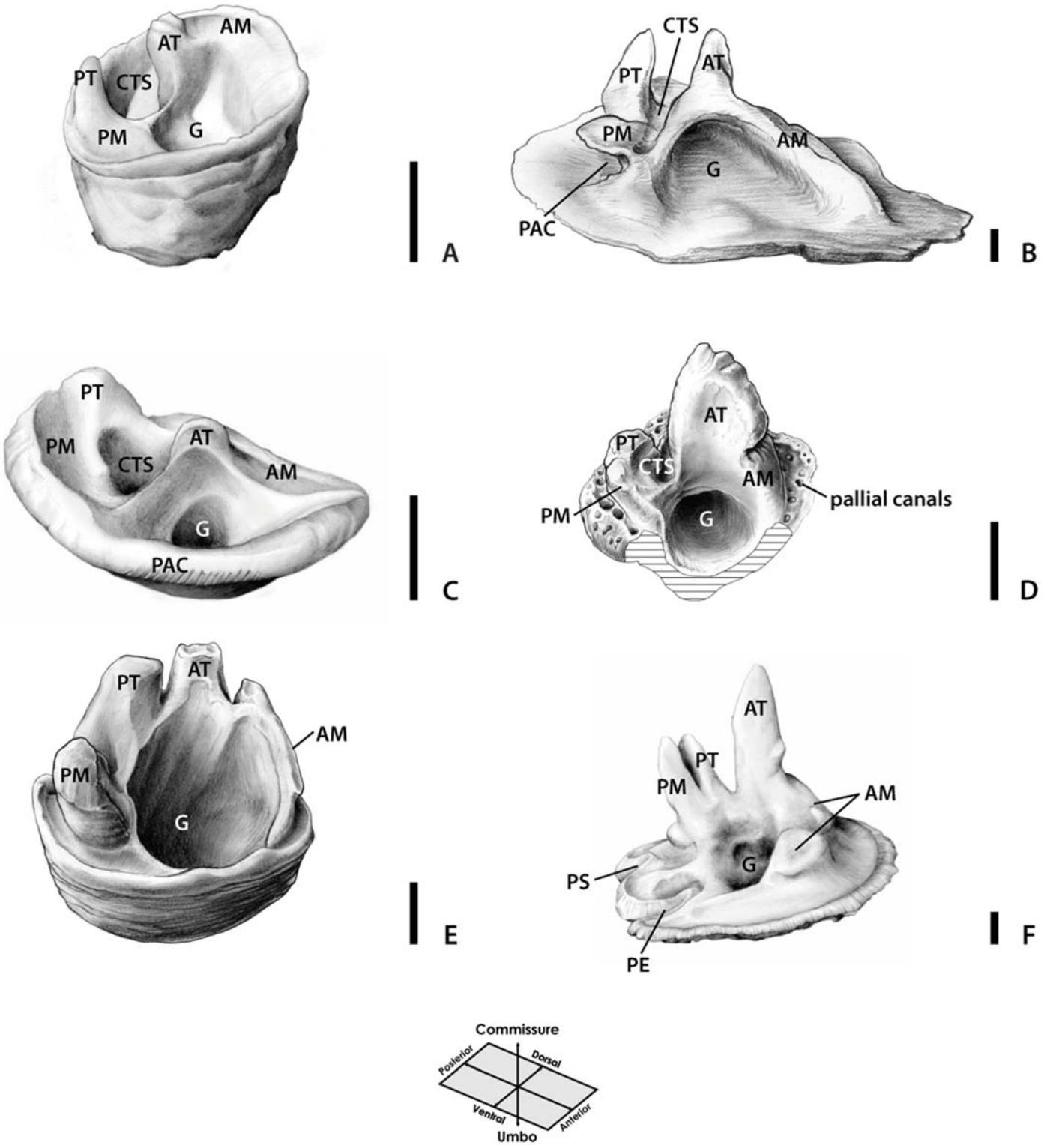


Figure 4. 10. Drawings of specimens from our taxonomic sampling showing orientation of various homologous features on left valve. A – *Monopleura michaillensis* (UPMC.1272). B – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013; Aptian, València, Spain). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Mexicaprina minuta* (JPMA.14097-5). E – *Radiolites radiosus* (MNHN.F.J09137). F – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm.

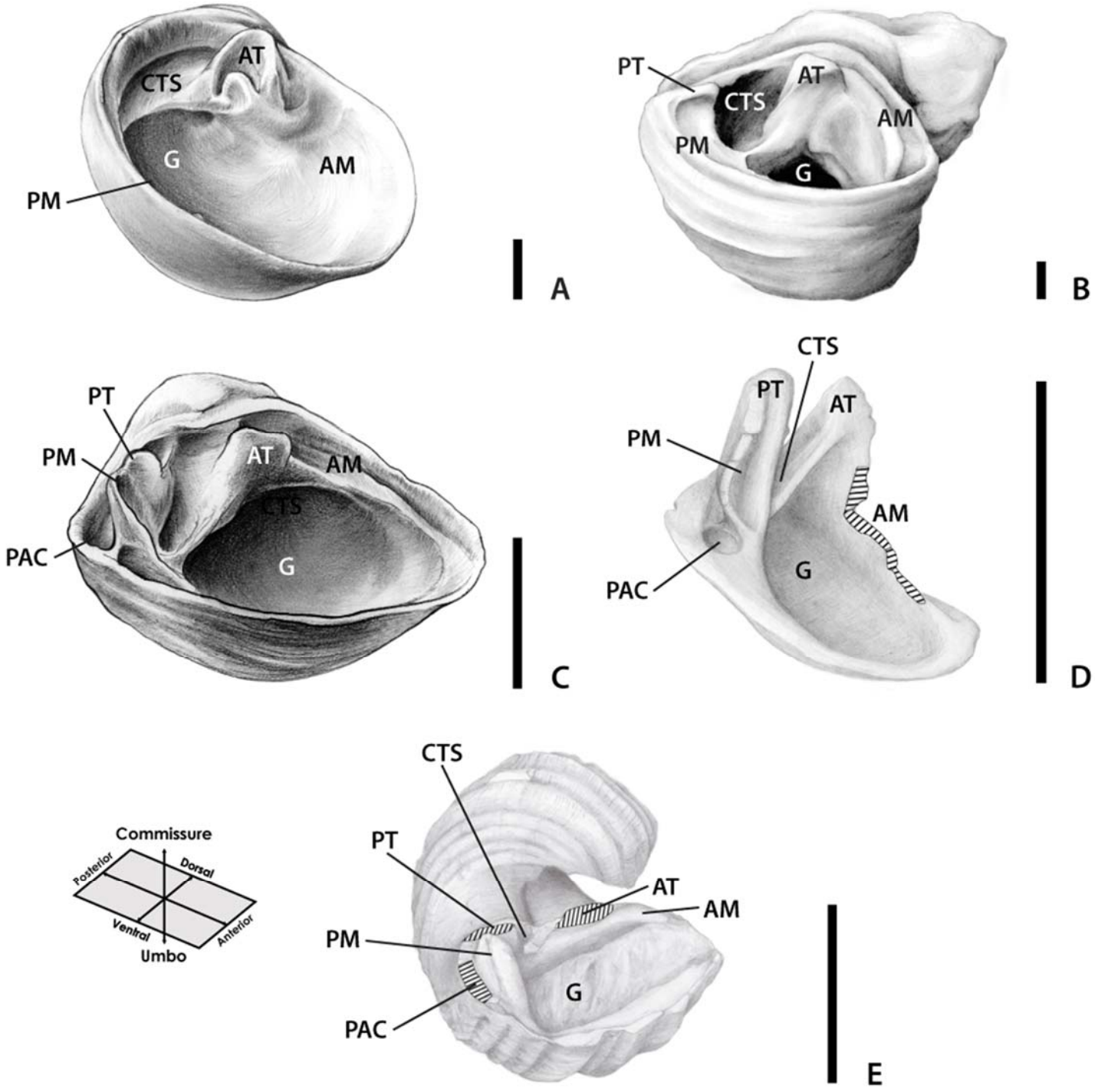


Figure 4. 11. Drawings of specimens from our taxonomic sampling showing orientation of various homologous features on left valve. A – *Diceras arietinum*. B – *Valletia germani* (MNHN.F.J05798). C – *Caprotina striata* (MV 2003.2.5147). D – *Chaperia costata* (MNHN.F.R07950). E – *Glossomyophorus costatus* (MNHN.F.R06779). Scale bars, 10 mm.

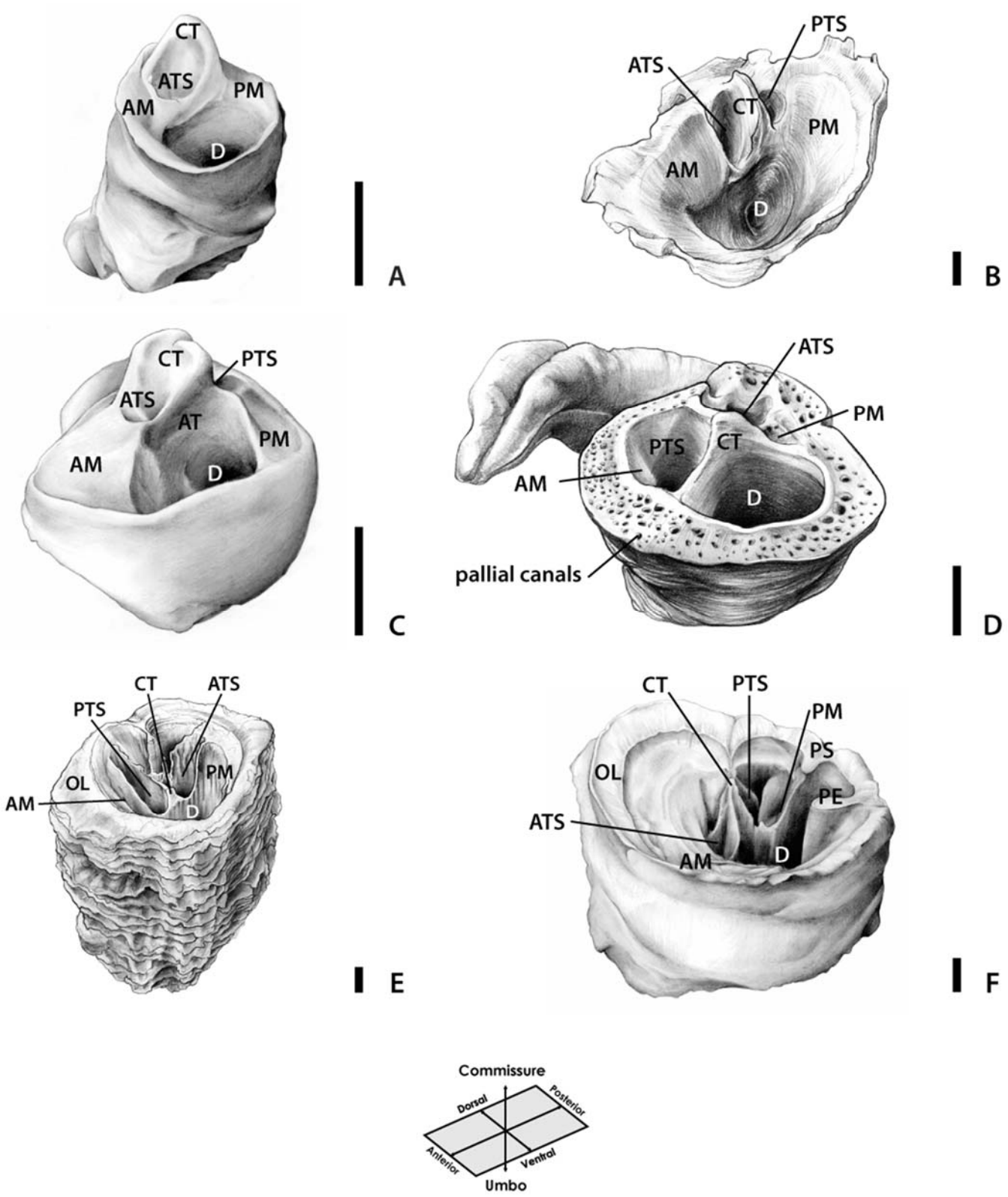


Figure 4. 12. Drawings of specimens from our taxonomic sampling showing orientation of various homologous features on right valve. A – *Monopleura michaillensis* (UPMC.1272). B – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013; Aptian, València, Spain). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Mexicaprina minuta* (JPMA.14097-1). E – *Radiolites radiosus* (MNHN.F.J09021). F – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm.

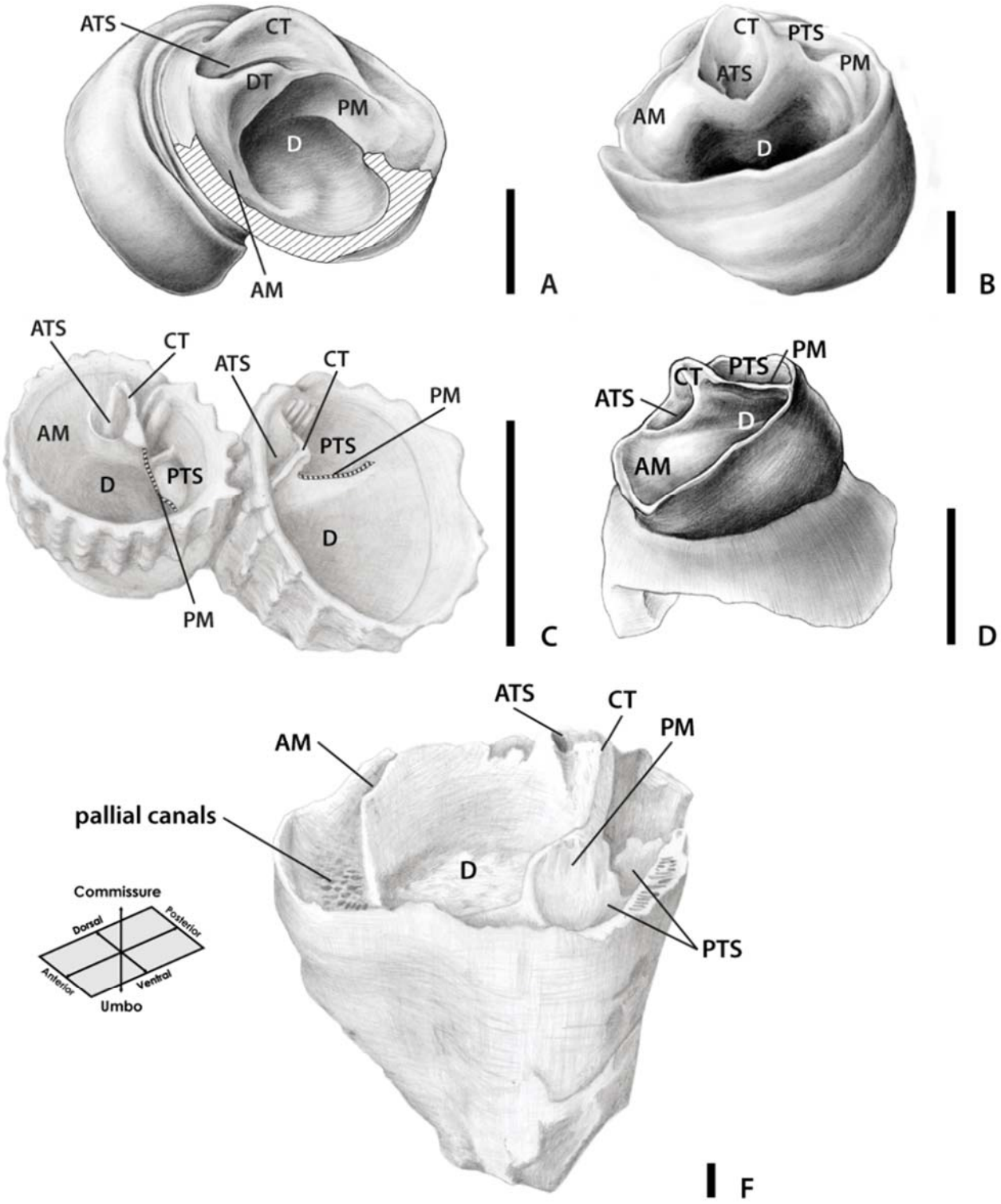


Figure 4. 13. Drawings of specimens from our taxonomic sampling showing orientation of various homologous features on right valve. A – *Diceras arietinum* (MNHN.F.A52439). B – *Valletia germani* (MNHN.F.J05798). C – *Chaperia costata* (MNHN.F.R07942). D – *Caprotina striata* (MV 2003.2.5147). E – *Offneria rhodanica* (UCBL.R781). Scale bars, 10 mm.

3.2. Homologies and characters

3.2.1. Approach

The formalization of independent characters is made difficult by the morpho-anatomy of bivalve shells. Contrary to taxa showing discrete parts, bivalves need to be conceptually divided into homologous regions varying in development and position from a shell to another. This delineation is one of the most problematic aspects of rudist phylogeny. The two main pitfalls we were confronted to are the definition of false (polyphyletic or paraphyletic) homologs, and the temptation to discard morphological features previously seen as irrelevant. The 'accessory cavities' are an example of such a problem.these problems. In Skelton and Smith (2000: 113), these features are hypothesized polyphyletic and induced by physical constraints caused by myophoral reorientation. They rejected all of the cavities close to the myophores ('ectomyophoral' and 'endomyophoral' cavities) assuming the feature to be fully dependent of myophores. However, their redefinition may refine homologs and could be used for phylogenetic analysis. We considered both the cavities and myophores as potentially evolutionary informative. A physical modification of the shell morphology can lead to several responses from the animal. The 'accessory cavities' might provide precise homologs after a careful study. One of the main goals of our study is to show the differential responses of rudist taxa to similar morphoconstructionnal constraints.

Some of the most important characters were previously used independently to the valve it occurs on, assuming a strong structural link between the structures of the two valves. Skelton's 'myocardinal configuration' characters (characters 20 to 28 in Skelton and Smith 2000) were considered irrelatively to specific variations on a particular valve. We show herein that a myocardinal arrangement is not an irreducible feature in which all elements are strongly dependent. Dependence is a mandatory consequence of the changes in the bioconstruction of a shell, but complex structures like myophoral and hinge structures of the rudist bivalves could be subject to differential evolutionary responses. Our characters are designed to authorize independent changes of each individual feature. The "myocardinal organization' kinds of characters can be divided into elementary features in

attempt to make more precise and complete assumptions about evolution. The character 'LV myophore a pedunculated plate, rooted on myocardinal platform, with flaring posterior rim widely separated from posterior valve margin by deep ectomyophoral recess and with adductor insertion on its inner surface opposing myophoral shelf or ridge in RV. 0, no; 1, RV myophore a flat to gently sloping shelf; 2, RV myophore a postero-dorsally facing ridge' (character 24 in Skelton and Smith, 2000) refers both to an 'ectomyophoral cavity' and to posterior myophores. Its redefinition will highlight Skelton and Smith's character 24 to be a consequence of the existence of several independent homologs evolving at different evolutionary rates. Their character 32 'Pallial canals regularly developed in aragonitic inner shell layers. 0, no; 1, yes in LV only. 2, in both valves' suffers of the same problem, as pallial canals of left and right valves must be separated as they have different evolutionary histories. Chartrousse (1994) published the only analyses in which every myophore is treated independently, as in the character three of his third analysis 'Situation de l'insertion myophorale postérieure de la valve inférieure: - état 0: horizontale (pallèle/subparallèle au plan de commissure); état 1: vertical (perpendiculaire au plan de la commissure sans le dépasser); état 2: vertical (perpendiculaire et dépasse le plan de la commissure)'. Our analysis follows a similar effort in character definition.

A first step is to decompose the classic morphological sets that are canals, teeth, myophores, in order to found the maximum of relevant and logically independent characters. The origin of each individual feature is traced by comparison of their morphologies, positions and ontogeny. A global synthesis of assumed homologies for both left and right valves is presented on figures 9 and 10. The origin of the canals is an important issue in resolution of rudist phylogeny that benefits from a decomposition of their parts and a comparative study of these minimal units. The pallial canals originate from partitioning of cavities. Descriptive names are avoided. For example, the ectomyophoral and endomyophoral cavities are used by Skelton and Smith (2000: p. 113). They are designed as unambiguous which is purely descriptive and will cause confusion into the problem of homologs delineation. The same problem arises from Chartrousse's dichotomy between myophoral cavity and accessory cavity. In our anatomical nomenclature we chose to keep names highlighting the origin of features, disregarding their function. Traditionally, the canals are supposed to emerge independently in parallel lineages but such putative

homologies are never tested. Therefore, we investigate the relationships between cavities and canals and give a name reflecting homology, even if the function changes or the morphology become highly modified. We also explore how the structures traditionally described under the same name are homologous or not among the taxa. For example, all cavities do not correspond to homologous structures. Numerous cavities labelled accessory cavity correspond to sockets. Homologous features are redefined in table 4.2 with their correspondence in Douvillé, Skelton and Chartrousse's works. Our primary homologies are thereafter tested by cladistic analysis.

The morphological, developmental and positional similarities are the key criteria for the establishment of homology hypotheses. The left and right valves are thus considered separately, as in the analyses done in the thesis of Chartrousse (1994). When recognized, a homolog on the right valve stay on the right valve in all taxa, by opposition to descriptive or functional terms that consider the relationship between right and left valve, as the 'endomyophoral' and 'ectomyophoral' cavities of Skelton and Smith (2000), or in Chartrousse's 'posterior accessory cavity' and 'posterior myophoral cavity' (1998; table 4.1). Thus, the model of canal evolution must be understood in each valve. We found several origins to each set of canals that are detailed in the subsequent character part and gain precision for defining the origin of features. It leads to a description of the two valves independently, and the previous characters based on both valves, as 'Myophore(s) always attached to hinge plates. 0, no; 1, yes' (character 23 in Skelton and Smith, 2000) are redefined. If a structure is the result of several movements on both left and right valves, as 'LV myophore projecting into RV, with adductor insertion rotated out to face anterior valve wall of RV. 0, no; 1, yes' (character 21 in Skelton and Smith, 2000), it needs to be decomposed and defined valve by valve. A more precise formulation will be to code the position of structures on each valve, and the alignment structure/cavity becomes only a consequence. The 'myocardinal configuration' defined by Skelton and Smith is here divided into several characters. A similar problem arises with characters about 'Pachyodont dentition' (characters 16 to 19 in Skelton and Smith 2000). Their character 16 'Two teeth in RV and one (plus one incipient) in LV. 0, no; 1, yes.' has an incorrect formulation, rather descriptive than traducing precisely homology, each bivalve tooth necessitating to be coded independently (#1, #2, #23; see character list at the end of the chapter). Only few features

Position	Abbreviation	Name	Correspondences with previous terminologies		
			Douvillé	Skelton (2000 and after)	Chartrousse 1998
Left valve	LV	Left valve		LV	
	G	Left valve general cavity	G	bc	
	AM	Anterior myophore	ma	am	
	PM	Posterior myophore	mp	pm	
	AT	Anterior tooth	B' (AII after 1902)	at	
	PT	Posterior tooth	B (PII after 1902, PIV in 1935)	pt	
	CTS	Central tooth socket	n (3b' after 1902, 3' in 1935), n' (n appendix) O, O', O'', Oma, Oa (O, O', O'': "*cavité accessoire* ")	cts, endo (endomyophoral cavity)	cs (central socket), pmc (posterior myophoral cavity), pac (posterior accessory cavity)
	PAC	Posterior accessory cavity	O, O'	ecto (ectomyophoral cavity)/canals	
	LC	Ligamentary cavity	L, O	L	
	pallial canals	pallial canals of the LV	Canaux palléaux	Pallial canals	
	calcitic canals	calcitic canals	canaux	hippuritid canals	
	pores	pores	pores	pores	
Right valve	RV	Right valve		RV	
	D	Right valve general cavity	D, O	bc	
	AM	Anterior myophore	ma'	am	
	PM	Posterior myophore	mp'	pm	
	CTS	Central tooth	N (3b after 1902, 3 in 1935)	ct	
	ATS	Anterior tooth socket	b' (AII' after 1902)	ats	
	PTS	Posterior tooth socket	b (PII' after 1902, PIV' in 1935), omp, op	pts, endo, ecto	ps, pac, pmc
	PMC	Posterior myophoral cavity		pm cavity	
	LC	Ligamentary cavity	L	accessory cavity	
Both valves	OL	Outer layer		ol	
	IL	Inner layer			
	GC	General cavity	cavité générale	GC	
	Wall(s)	Wall(s)			
	Tabulae	Tabulae			
	Flange(s)	Flange(s)	Lames externes		
	Costulations	Costulations			
	Tubes	Tubes	Lames externes anastomosées		
	Radial bands	Radial bands	bandes siphonales	radial bands	
	P1	posterodorsal pillar	S	P1	
	P2	posteroventral pillar	E	P2	

Table 4. 2. Glossary of terms used for description of rudist morphology in the present study, their description and their link to other terms that can be found in the literature (Douvillé, 1935; Skelton & Smith 2000; Skelton 2013; Chartrousse, 1998).

can be coded irrelatively to their valve when a common origin, for example flanges on both valves, can be linked to the mantle deformation of the animal. The deformation of the edge of the mantle affects both the right and the left valve, because the mantle constructs the feature by skeletal accretion. Finally, even presumed homoplasic features are kept, such as tabulae (Skelton and Smith 2000: 114). These characters are simple and likely arise in several taxa. However, only cladistic analysis is a relevant test for homoplasy.

In the next part are described the general sets of homologs: teeth, myophores, vertical blades, canals, horizontal blades (tabulae, etc.), calcitic layer, and ligament (figs. 4.14 to 4.17). For each set, we discuss homology of position, morphology, origin, function and development, wherever possible. Each set is then decomposed into minimal evolutionary entities which are used thereafter to construct characters.

3.2.2. Rudists morphology

The preservation of rudists does not allow generally to perfectly see internal characters. The rudist morphology is generally studied using sliced specimens. Slice terminology is given in Figure 4.18.

Rudists are bivalves with a pachyodont dentition, comprising few, large teeth. The teeth are part of the hinge, and their function is to maintain the two valves in connection. Most species of Hippuritidina possess an anterior tooth (AT) and a posterior tooth (PT; character #1) on the left valve (LV). The AT is always in connection with the left valve body cavity (G) and the anterior side of the posterior tooth socket (PTS). The PT is posterior to the AT and can be connected to G and the posterior side of the PTS or completely enclosed in the PTS without any connection with G. In Hippuritidina, only *Diceras* has a drastically different dentition which seems inversed when compared to another Hippuritidina species (Munier-Chalmas 1882; Malchus 1995). While other species possess two teeth in LV and one tooth in RV, *Diceras*, and all members of the suborder Requienidina, possess a single tooth on the LV and two teeth on the RV. *Diceras* exhibits an anterior tooth (AT) on the LV, a Central Tooth (CT) and a tooth we call here a Diceratid Tooth (DT; #23) on the RV. Teeth are cone shaped and have a great diversity of sizes. The PT is usually smaller, sometimes of equal size than AT (#1), but rarely taller. The PM shelf

produces bulges on the LV along the hinge margin, out of which evolved a tooth in *Valletia*, dorsally to the PM. A single tooth is found on the RV of Hippuritidina, the CT. Its position is constant, but its morphology accepts a high variation. The CT is always connected to D, stuck between the two sockets of the RV. Very small in most rudists, the CT can be tilted posterodorsally (#25). The CT can be so small that it has been coded as lost in some species in the phylogeny provided by Skelton and Smith (2000; character 19). Ancestrally, the CT partially surrounds the PTS and has a dorsoventral axis (#24).

Valves are connected by two adductor muscles, an anterior and a posterior. The *myophores* are the muscles attachment structures (#4, 6, 26, 27). In most bivalves, the muscle function is generally to close the two valves of the shell, and muscles work as a force opposed to the ligament which tends to open the shell. The function of the myophores is to carry the adductor muscles (Kennedy and Taylor 1968; Skelton 1978, 2014). They can be localized by muscle scars on the shell. The anterior muscle is attached to the anterior myophores (AM) on each valve, and the same happens with the posterior myophores (PM) for a total of four myophores. The AM is always located near the ventral side of the AT on the left valve or near the ATS on the right valve. The position is the same for the PM relatively to the PT and the PTS. The myophores can be horizontal (on a shelf), oblique or vertical (in a cavity or on a shell wall). In rudists, the muscle insertions can be directed as in almost all bivalves in a left/right valve axis, but also on the dorsoventral axis in *Antillocaprina*, or on the posteroanterior axis in Caprinids. The myophores can be developed in robust shelves in *Valletia* or *Monopleura*, strongly protruding tooth-like myophores in *Titanosarcolites* or in Radiolitids, or blade-like PM-LV in Hippuritids. All rudists possess two myophores on each valve, with a singular exception: the anterior muscle (and the corresponding AM) divided in two parts in all hippuritids (#6.5).

Vertical blades can develop in rudist to divide cavities in the aragonitic shell. Cavities can occur at various positions in the shell (#11, 12, 28, 29, 30). Cavities can form canals when they are filled with radial blades (blades oriented from the internal to the external part of the shell; #12, 31). The canals can invade the entire shell including the teeth (as in *Curtocaprina clabaughinkinsorum* or *Titanosarcolites giganteus*; #32) and the myophores (as in *Coralliochama orcutti* or *Caenosarcolites oddsensis*; #16). The

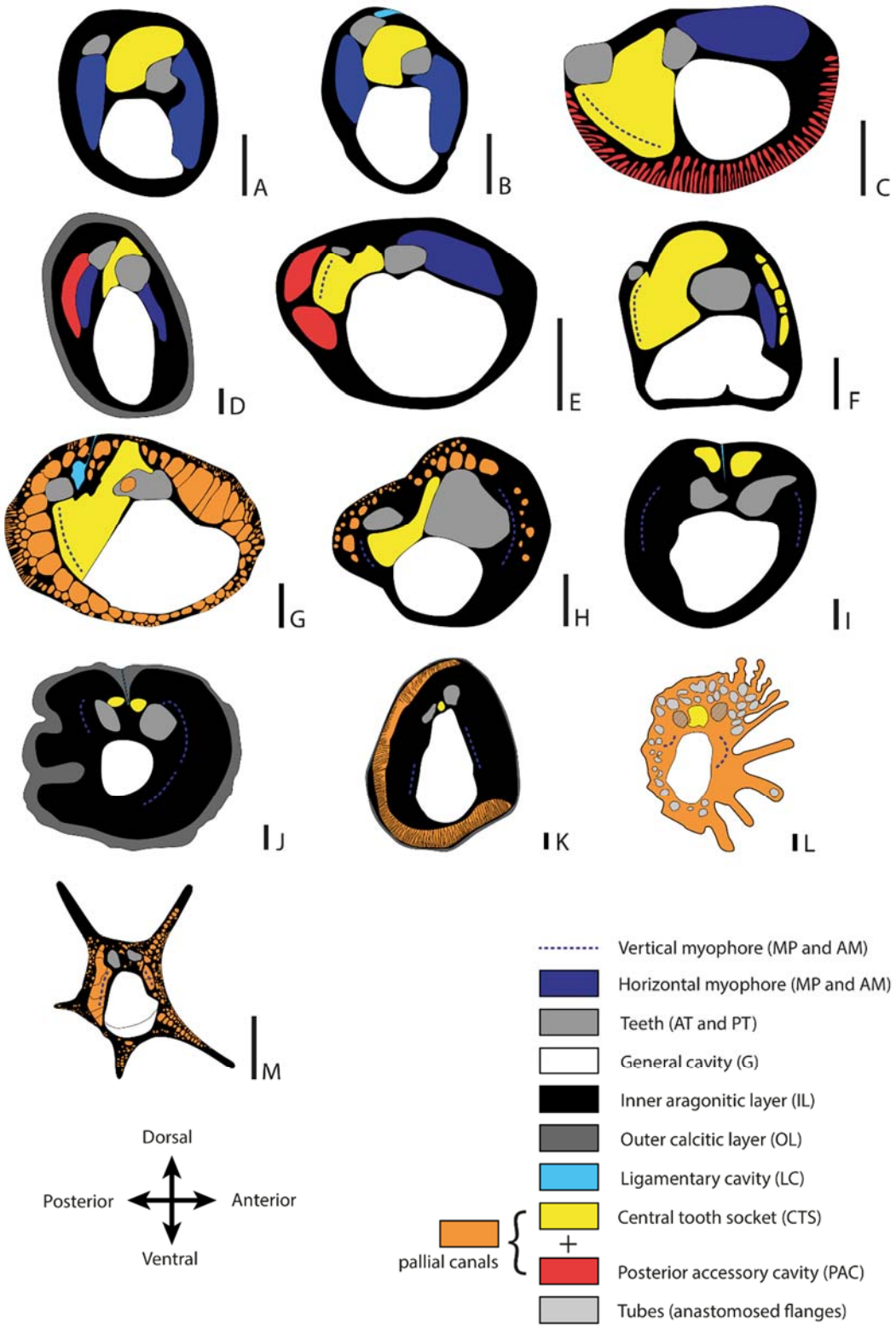


Figure 4. 14. Interpreted cross section into left valves with colors corresponding to homologous features. A – *Valletia germani* (MNHN.F.J05798). B – *Monopleura michaillensis* (UPMC.1272). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). E – *Caprotina striata* (MV 2003.2.5147). F – *Pachytraga paradoxa* (JPMA.17738-1). G – *Caprinula boissyi* (MNHN.6924). H – *Mexicaprina minuta* (JPMA.14097-5). I – *Radiolites radiosus* (MNHN.F.J09021). J – *Hippurites radiosus* (MNHN.F.A58757). K – *Dictyoptychus morgani* (NHM.L-58427). L – *Titanosarcolites giganteus* (UT.47533). M – *Ichthyosarcolites bicarinatus* (IGM.6970). The entire shell of *Titanosarcolites giganteus* is invaded by very small capillary canals (purple), including the teeth (purple hatches). All rudists possess an outer calcitic layer, which is not is usually not visible on fossils when very thin. Scale bars, 10 mm.

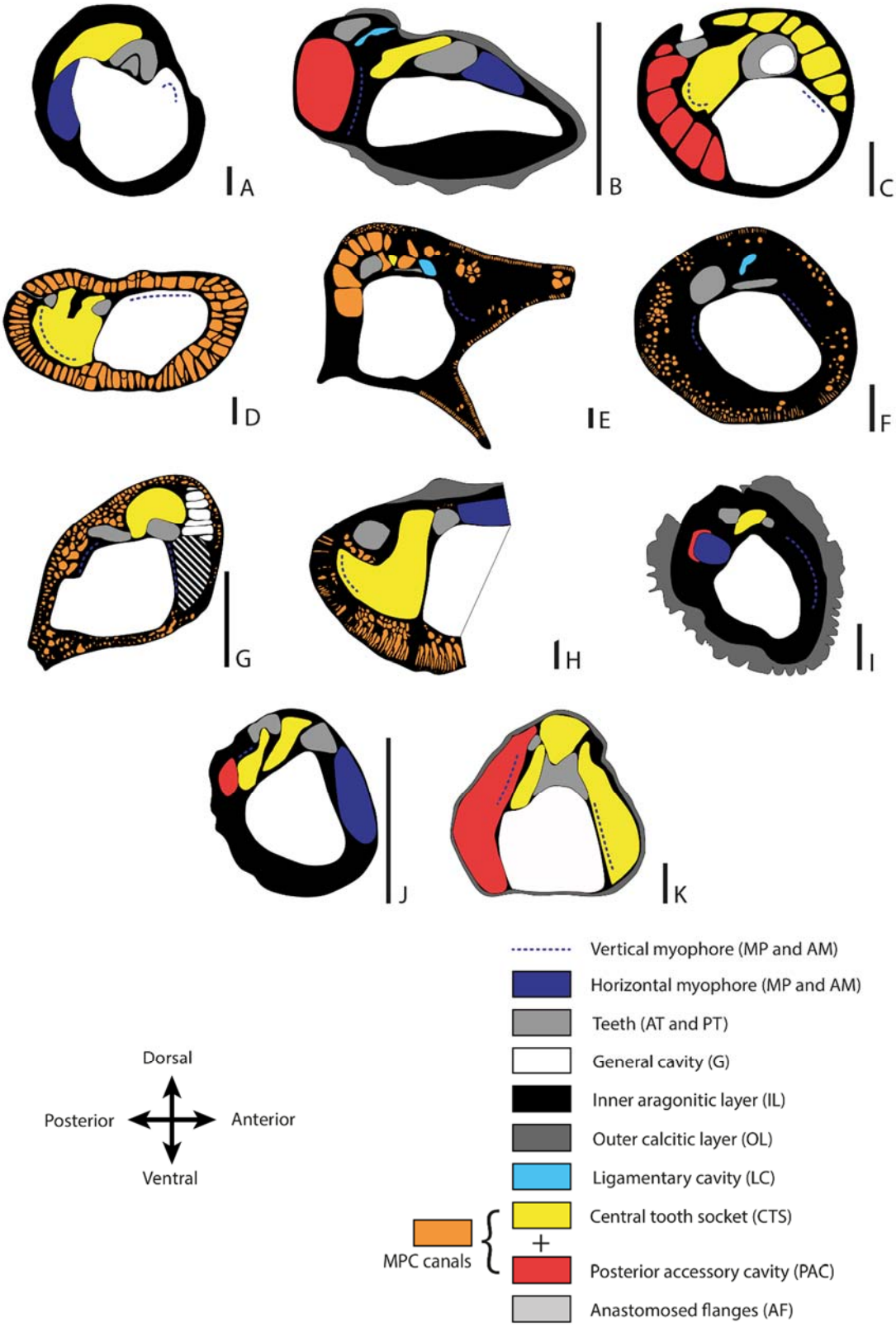


Figure 4. 15. Interpreted cross section into left valves with colors corresponding to homologous features. A – *Diceras arietinum* (MNHN). B – *Glossomyophorus costatus* (MNHN.F.R06779). C – *Amphitriscoelus waringi*. D – *Offneria rhodanica* (UCBL.R275). E – *Semailia smithi* (BM.LL41931). F – *Clinocaprina lowentsami* (USNM.245242). G – *Curtocaprina clabaughikinsorum* (UT.10937). H – *Coralliochama orcutti* (USNM.186662). I – *Tepeyacia corrugata* (CAS.66657.01). J – *Chaperia costata* (MNHN.F.R07950). K – '*Sellaea' sp* (WSA3058). All rudists possess an outer calcitic layer, which is not is usually not visible on fossils when very thin. Scale bars, 10 mm.

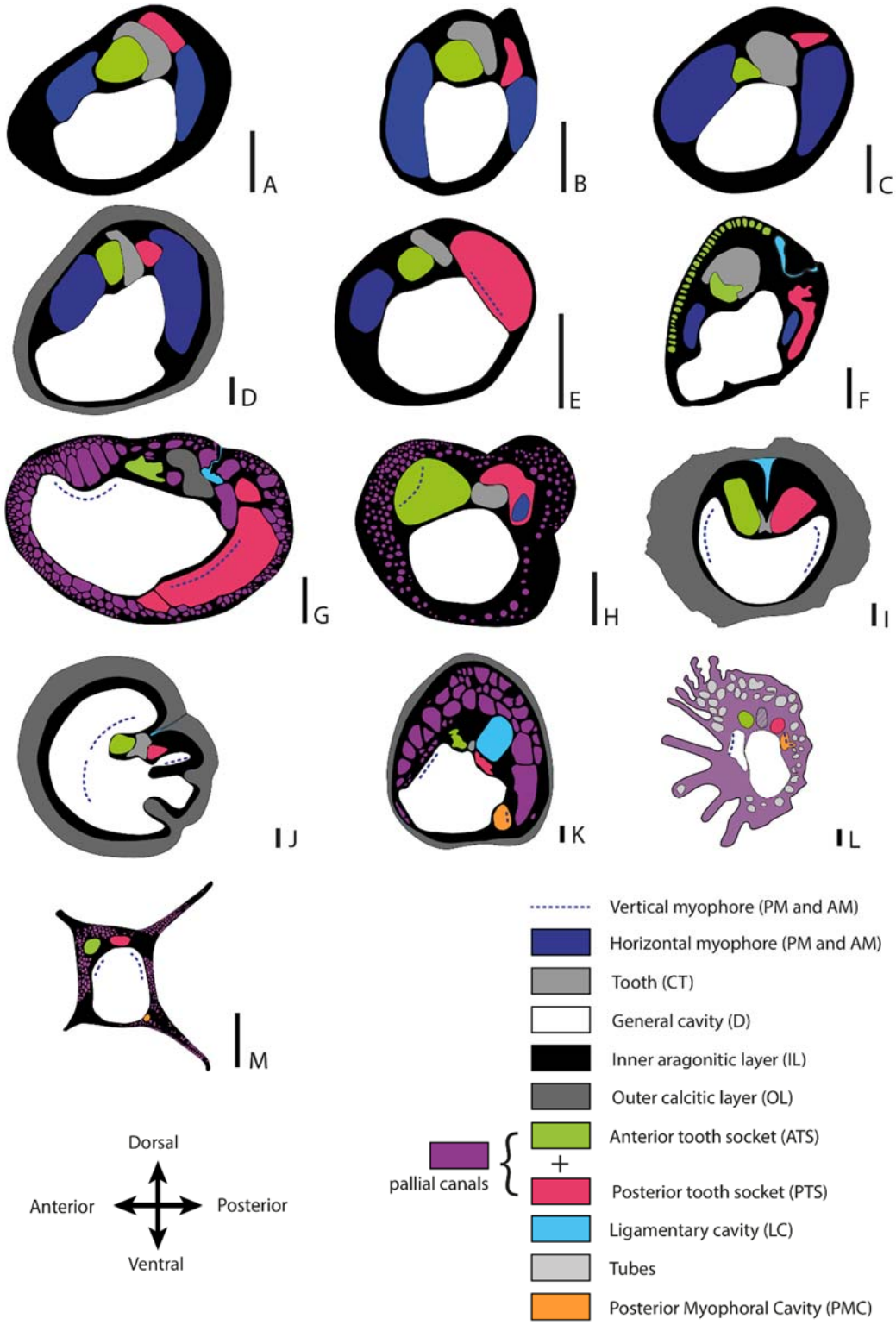


Figure 4. 16. Interpreted cross section into right valves with colors corresponding to homologous features. A – *Valletia germani* (MNHN.F.J05798). B – *Monopleura michaillensis* (UPMC.1272). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). E – *Caprotina striata* (MV 2003.2.5147). F – *Pachytraga paradoxa* (JPMA.17738-2). G – *Caprinula boissyi* (MNHN.6924). H – *Mexicaprina minuta* (JPMA.14097-1). I – *Radiolites radiosus* (MNHN.F.J09021). J – *Hippurites radiosus* (MNHN.F.A58757). K – *Dictyoptychus morgani* (NHM.L-58424). L – *Titanosarcolites giganteus* (UT.47533). M – *Ichthyosarcolites bicarinatus* (IGM.6972). The entire shell of *Titanosarcolites giganteus* is invaded by very small capillary canals (purple), including the teeth (purple hatches). All rudists possess an outer calcitic layer, which is not is usually not visible on fossils when very thin. Scale bars, 10 mm.

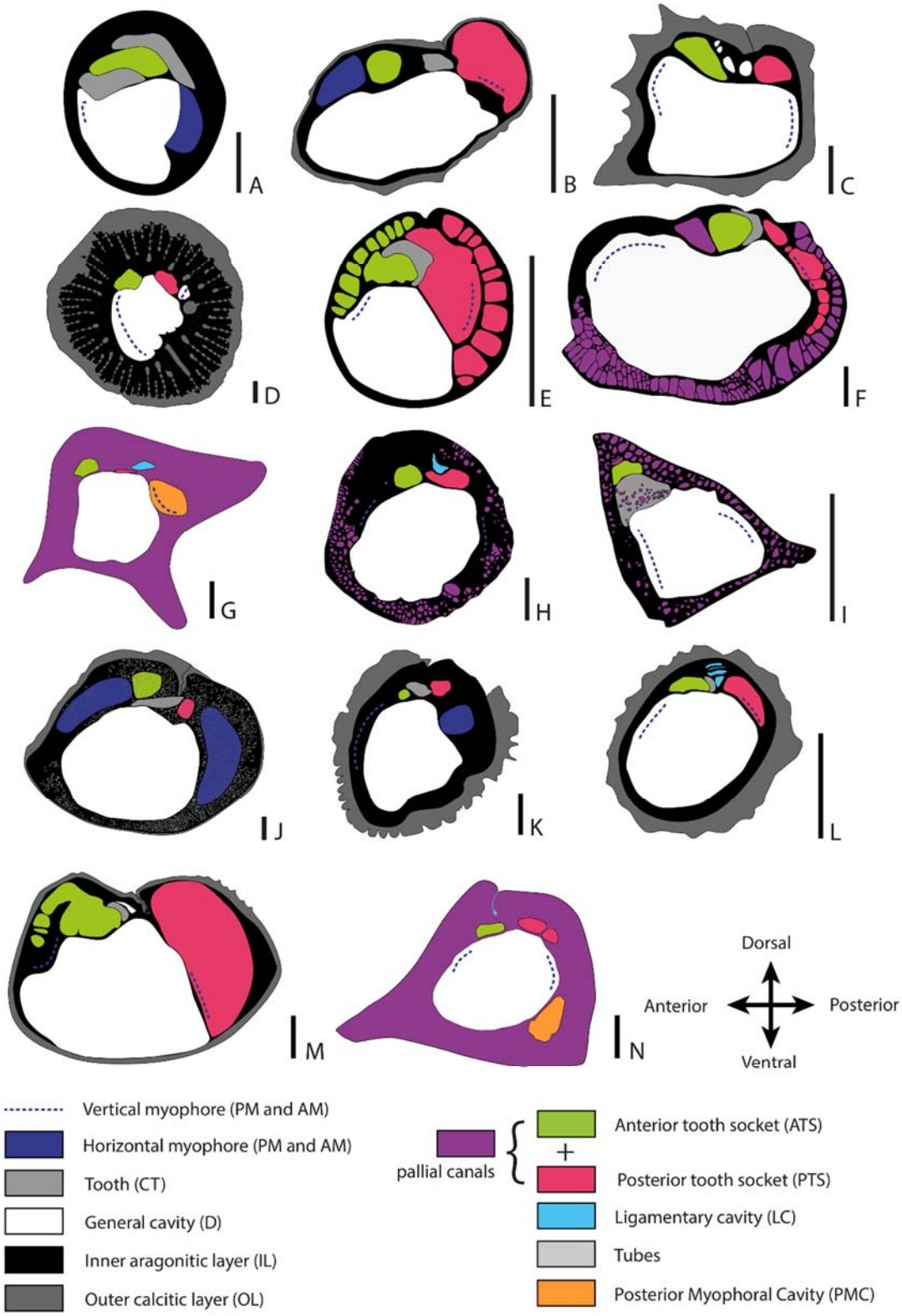


Figure 4. 17. Interpreted cross section into right valves with colors corresponding to homologous features. A – *Diceras arietinum* (MNHN.F.A52439). B – *Glossomyophorus costatus* (MNHN.F.A27280). C – *Agriopleura marticensis* (JPMA.15172-2). D – *Barrettia monilifera* (fig. 6B in Mitchell 2010; UWIGM.RUD.2006.223). E – *Amphitriscoelus waringi*. F – *Offneria rhodanica* (UCBL.R781). G – *Semailia smithi* (BM.LL41931). H – *Clinocaprina lowentsami* (USNM.245242). I – *Curtocaprina clabaughikinsorum* (UT.10922). J – *Coralliochama orcutti* (USNM.186665). K – *Tepeyacia corrugata* (CAS.66657.01). L – *Chaperia costata* (MNHN.F.R07942). M – '*Sellaea' sp* (WSA3058). N – *Oryxia sulcata* (BJ.1901-HR2). The entire shell of *Semailia smithi and Oryxia sulcata* is invaded by very small capillary canals (purple), including the teeth (purple hatches). All rudists possess an outer calcitic layer, which is not is usually not visible on fossils when very thin. Scale bars, 10 mm.

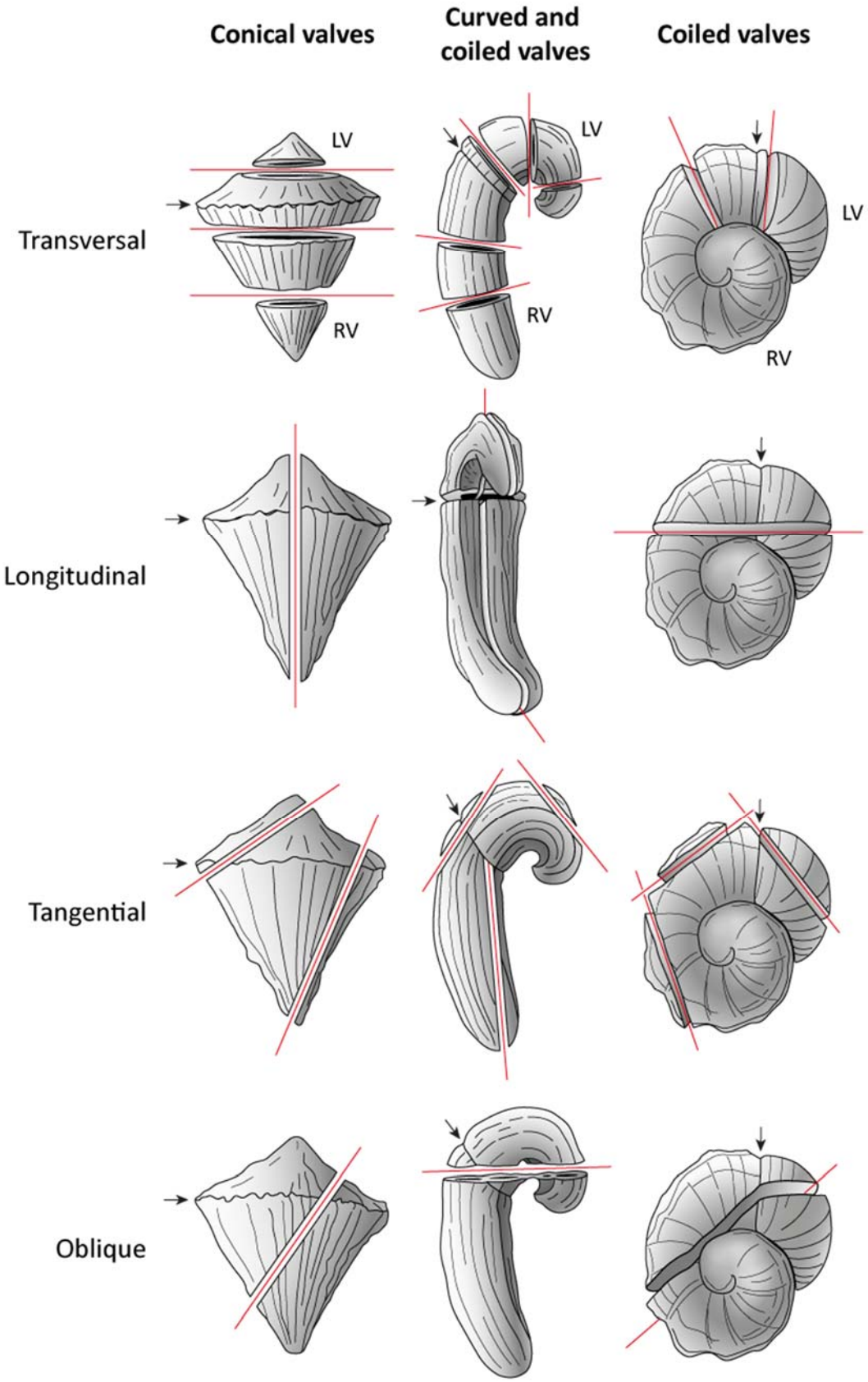


Figure 4. 18. Principal sections (red lines; transversal, longitudinal, tangential) on rudist fossils with various shell shapes. Modified from Dechaseaux (1969). Arrows: commissural plane.

morphology of the cavities formed by the radial blades is extremely diverse, ranging from half the size of the central cavity to tiny canals of less than a millimeter in diameter (capillary canals). The blades can be simple, develop dichotomies or make anastomosis between them. Cavities of moderate size aligned along the outer shell margin and localized in the aragonitic part of the shell are called *pallial canals*. The pallial canals must always be related to their valve as LV pallial canals do not possess the same origin as RV pallial canals. The rudist pallial canals originate from cavities, with similar series of canals potentially related to distinct cavities (#6.3, 11, 28, 29). Some structures found in Antillocaprinidae similar in shape to prominent pallial canals must however have a different origin. Trechmann (1924) described large cavities in *Titanosarcolites* in addition to standard pallial canals. The large canals (hereafter called tubes) are covered by a calcitic layer in the inside (Trechmann 1924), which is characteristic of the external layer of the shell in rudists (the internal layer being aragonitic). The tubes likely appear through anastomosis of the external shell flanges of some Antillocaprinids or of *Ichthyosarcolites bicarinatus*. Because there is no equivalence of rudist canals in extant bivalves, various theories attempt to explain their functional significance.

Any cavity of the shell can also be partitioned by horizontal, oblique to hourglass blades (tabulae; #17, 18, 19, 33, 34). Tabulae occur most often in the largest taxa (*Titanosarcolites giganteus*) but not only (*Offneria rhodanica*). Some taxa such as *Ichthyosarcolites* can show highly curved tabulae.

External features comprise all structures associated to the outer, calcitic layer of the shell. The external features unique to the rudists are mostly linked to the uncoiling of the shell (Skelton 1985; Stone and Telford 1999). The ligament is primitively amphidetic (i. e. below the beaks; Yonge 1967; #37.0). It develops vertically between the extremity of the shell (beak) to the margin in a majority of rudists (#37.1, 37.2), dorsally to the PT or the CT (#38). The change in ligament orientation relaxes structural constraints which allowed the high diversity of coiling patterns of rudist shells. Essential for the valve openings in almost all bivalves, the ligament becomes close to functionless when vertical (Skelton 1979). A system of canals and pores allowing the water to penetrate in the shell can be found in the calcitic layer of Hippuritidae. The canal/pore system is composed of calcitic

canals (#20) and of pores connecting the canals to the exterior (#21). The calcitic layer can be smooth or costulated (costae perpendicular to the commissure, developing from the umbo to the commissure; #39), slightly as in *Monopleura michaillensis* or strongly as in *Tepeyacia corrugata*, and showing small growth lines parallel to the commissure. The external shell layer can fold in at two particular locations named pillars, distorting the entire shell (#35). The pillar P1 is localized posterodorsally and the pillar P2 localized posteroventrally. The number of pillars increases in some taxa, such as *Barrettia*. The folds could support excretory pores for faeces and pseudofaeces (Skelton 1979) for organisms that cannot open wide their valves. In other taxa, at the localization of pillars can be found two radial bands, zones with a qualitatively different ornamentation than the rest of the shell (generally zones with a specific costulation morphology).

3.2.2.1. Teeth

The ancestral conditions of dentition is based on comparision with Requienidina, assumed as sister group of the studied group Hippuritidina. *Diceras* is the only genus of Hippuritidina showing the same dentition as the sister group. *Diceras* possess a single AT on the LV (see Malchus 1995 for previous notations) and two teeth on the RV, a DT on the posterior side and a CT on the anterior side. In all other Hippuritidina, the teeth number and positions seems to be a constant feature (Douvillé 1896). They keep both the AT and the CT. The position where should lie the DT is flat in all of these species, which can be understood as a loss of the tooth (#23). On the other valve, the PM has changed; it may have produce a little bulge of shell material posterior to it (as seen in *Valletia germani*) that becomes the PT in derived taxa (#1). This PT can vary in size relatively to the AT. The AT can change in shape: its shape can be simple or showing to apex as in *Diceras*, the AT being divided in a big protrusion surrounding a small one (#2). We also note the existence of a single round cavity or canal in the AT of *Amphitriscoelus* and *Caprinula* (#15). Finally, the CT can protrude straight into the LV, or be curved posteriorly as in *Monopleura* (#25), where the CT also partially surround the ATS (#24). Some taxa can show supernumerary teeth, such as *Monopleura aliena* (Počta 1890), certainly autapomorphic, or the two additional toothlets 'pat' and 'aat' (Mitchell and Gunter 2006; Mitchell 2010) that are synapomorphic of the subfamily Parasarcolitinae (Mitchell 2013b). Teeth can also be

ornamented, those of the LV of *Radiolites* and *Agriopleura* being slightly costulated dorsally (#3).

Mitchell (2013a) described an additional 'ichthyosarcolitid toothlet' (IT) in *Curtocaprina clabaughinkinsorum*, a tooth postulated to be a diagnostic feature for the family Ichthyosarcolitidae. We discuss here Mitchell's hypothesis and suggest an alternative interpretation (#10; fig. 4.19). *Curtocaprina* shares common features with the Caprinuloideidae, and Mitchell proposed that rudists primitively had an AT and a PT on the left valve (fig. 4.19 right, state 0) and evolved with the apparition of the IT into the Ichthyosarcolitidae (fig. 4.19 right, state 1). However, if *Curtocaprina* bear the three AT, PT and IT (fig. 4.19 right, state 1, down sketch), *Ichthyosarcolites* species has only two teeth that Mitchell interpreted as an AT and an IT (fig. 4.19 right, state 1, top sketch), the PT disappearing. If *Ichthyosarcolites* would have an AT plus an IT and no PT, then it would become impossible to differentiate the IT from the PT found in Radiolitidae, Hippuritidae, Antillocaprinidae, or Trechmanellidae (fig. 4.19 right, dotted arrow). The evolutionary hypothesis of an IT diagnostic of Ichthyosarcolitidae is difficult to support because it grounded on evolutionary interpretation of a single shell of *Curtocaprina* with the anterior part of its IT broken (fig. 4.19-C). We propose here an alternative evolutionary scenario based on the conservation of AT and PT without newly formed tooth (fig. 4.19 left). Assuming that teeth are structurally homologous in all the inverse rudists, we need to explain the differences of teeth position expressed among taxa. The configuration found in *Valletia* or in many Monopleurids is an AT in connection with both G and the CTS where the PT remains eccentric, isolated in the posterodorsal side and in connection with the CTS only (fig. 4.19 left, state 0, down sketch). In these taxa, the CTS/G wall ends posteriorly in the middle of the myophoral shelf. In other taxa such as Plagioptychidae, the CTS/G wall is even more ventral, the PM being surrounded by the CTS (fig. 4.19 left, state 0, top and down sketches, fig. 4.5A). In Radiolitidae or in *Ichthyosarcolites*, the AT occupies a similar position, but the PT is connected to both G and CTS (fig. 4.19 left, state 1). This can be explained by a movement of the CTS/G wall that separates the two cavities G and the CTS. Ancestrally, the CTS/G wall is connected anteriorly to the AT, and posteriorly to the ventral end of the PM (fig. 4.19 left, state 0) or in the middle of the anterior side of the PM. In *Ichthyosarcolites* and *Radiolites*, the posterior insertion of the CTS/G wall shifts dorsally,

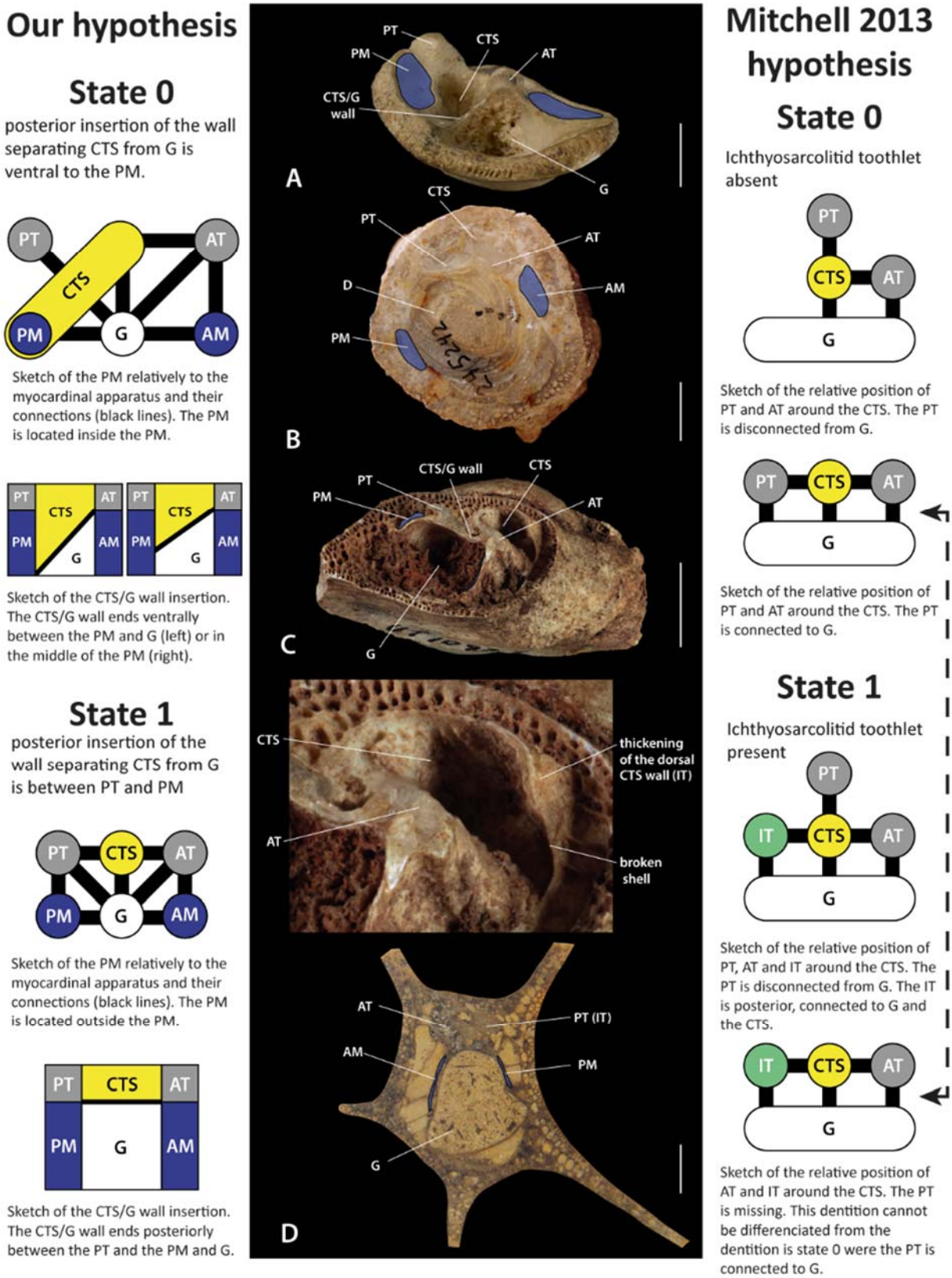


Figure 4. 19. Confronting hypotheses about the CTS and the 'ichthyosarcolitid toothlet'. Left: our hypothesis, illustrating the two states of the character 10. Right: Mitchell (2013) hypothesis about the 'ichthyosarcolitid toothlet', illustrating its character 13. The dotted arrow show two manifestations of the states 1 and 2 that are indistinguishable. Scale bars, 10 mm. Middle: photographies illustrating the two diverging character states. A – *Plagioptychus arnaudi* (UPMC.1273), B – *Clinocaprina lowentsami* (USNM.245242), C – *Curtocaprina clabaughikinsorum* (UT.10937), D – *Ichthyosarcolites bicarinatus* (IGM.6970); picture provided by Javier Aguilar-Pérez.

between the PM and the PT (fig. 4.19 left, state 1). This hypothesis is only suported by a movement of this unique CTS/G wall the diverse morphologies of *Curtocaprina*, *Ichthyosarcolites* and other rudists taxa. Mitchell's hypothesis is less parsimonious because it necessitates the apparition of an IT and the explanation of the diverse localizations of the PM-LV. The move of the CTS/G wall lead to what MacGillavry (1937) called a 'symmetrization' of the myocardinal apparatus. What Mitchell presumed to be the PT in *Curtocaprina* is considered here as an autapomorphic feature, understood as a thickening of the dorsal part of the CTS wall because it does not protrude into the commissure as other rudist teeth do.

3.2.2.2. Myophores

The myophores are of prime interest for the rudist taxonomy (Chartrousse 1998; Skelton 2014). They carry the animal muscles and can be recognized from their muscles scars and their specific microstructures (Kennedy & Taylor 1968; Skelton 1978). Rudist myophores evolved through a high development, strong morphological modifications compared to other bivalves, and a differentiation of in three-dimensional structures that often strongly imbricate from one valve into the other. All rudists are dimyarian (having both an anterior and a posterior muscle), with the only exception of Hippuritidae that develop a tertiary adductor muscle originated from the division of the AM (#6.5). The two homomyarian and heteromyarian conditions can be found in rudists. The two anterior and posterior myophores look globally similar in *Valletia*. The PM and the AM are highly differentiated in *Hippurites*, with strong differences on the two myophores of each valve, and even more differences between the left and right valves. A modification in a myophore of a left valve can drive several responses in its associated myophore of the right valve. Each of the four myophores having proper evolutionary rates and histories, they are good examples of character heterobathmy (Hennig 1966) and must be considered independently. We displayed all myophoral organizations in figure 4.20 with corresponding character states in table 4.3 encountered in our taxonomic sampling. This figure illustrates the great diversity of rudist myophores, and the complete independence between the left and the right valves.

Figure 4. 20. Myophoral arrangements encountered in our taxonomic sampling in transversal views. Positions and orientations of the anterior and posterior muscles are displayed in blue. Corresponding character states are given in table 4.3.

	LV		RV		
	PM (#4)	AM (#6)	PM (#26)	AM (#27)	Morphotypes
Diceras arietinum	0	0	0	0	1
Valletia tombecki	1	1	2	1	2
Monopleura michaillensis	1	1	2	1	2
Plagioptychus arnaudi	2	2	2	1	3
Coralliochama orcutti	2	2	2	1	3
Offneria rhodanica	2	3	4	1	4
Caprinula boissyi	2	?	4	?	4
Caprotina striata	3	1	3	0	5
Glossomyophorus costatus	3	1	3	1	6
Sellaea americana	3	3	3	1	7
Pachytraga paradoxa	3	3	4	0	8
Chaperia costata	3	4	3	0	9
Polyconites verneuili	4	1	1	1	10
Tepeyacia corrugata	4	4	1	0	11
Curtocaprina clabaughikinsorum	5	3	1	0	12
Radiolites radiosus	5	4	1	0	13
Agriopleura marticensis	5	4	1	0	13
Ichthyosarcolites bicarinatus	5	4	6	0	13
Oryxia sulcata	?	?	6	0	13
Clinocaprina lowentsami	5	4	5	0	14
Mexicaprina minuta	5	4	5	2	15
Amphitriscoelus waringi	5	?	5	?	?
Dictyoptichus morgani	5	4	6	0	16
Titanosarcolites giganteus	5	4	6	2	17
Semailia smithi	6	4	6	?	18
Hippurites radiosus	6	5	5	0	19
Barrettia monilifera	6	5	5	0	19

Table 4. 3. Overview of the general myophoral organization (PM-LV, AM-LV, PM-RV, AM-RV) of rudists from our sampling. For each species are displayed the myophoral characters. The last column is the morphotype number of the species. There is a number for each combination of myophoral states that are displayed in Figure 4.20.

Caprinula and *Plagioptychus* possess both a PM-LV in the form of an oblique plate surrounded by the CTS (#4.3). However, even if this structure is identical on the left valve, the associated AM-RV is a horizontal shelf in *Plagioptychus* (#26.2), whereas it is completely vertical with an associated protruding blade in *Caprinula* (#26.4). Each myophore can also vary in position relatively to teeth. *Radiolites* has it PM-LV separated from the PT by a gap (#5), contrary to *Mexicaprina*. The AM-LV can also be in contact with the AT or not (#7).

Primitively, the rudists exhibit anterior and posterior myophores lying on the inner wall of the GC (#4.0, 6.0, 26.0, 27.0), the PM being attached on a thin myophoral blade. The early inverse rudist *Valletia* exhibits shelf-like anterior and posterior myophores, flush with the plane of the commissure, on both valves. The PM-RV lies ventrally to the PTS. The evolution of vertical myophores in derived groups can generally be explained by the high development of cavities, which extend toward the myophoral shelf, that resorbes consequently. In the shells possessing myophoral cavities, the shape of the cavity is always highly similar to the myophore outline (#4.1 and #4.2; #6.1 and #6.3; #26.2 and #26.3). The small PTS and myophoral shelf of the PM-RV in *Plagioptychus* are homologous to the greatly extended PTS that encloses the PM-RV in *Caprotina* (while the AM-RV is still horizontal), which is an example of myophoral verticalization (#26.2 and 26.3). A similar evolution is documented by the vertical AM-RV of *Mexicaprina minuta* with the ATS hosting the myophore (#27.1 and 27.2). In *Sellaea*, these cavities which modify the arrangement and orientation of myophores allow to observe a connection between the CTS and their posterior cavity (#6.3), as well as between the ATS and the anterior cavity divided by blades giving it the appearance of pallial canals (#28.1). The history of myophores and of their verticalization is thus structurally correlated to the evolution of cavities. In the example of Radiolitidae RV, myophores fall directly into D (#26.1, 27.0). Myophores can also protrude strongly into the opposite valve as in Titanosarcolitinae or Radiolitidae, where the AM-LV and PM-LV extend directly to the shell wall of D (#4.5, 6.4). A blade can also protrude in the RV (#26.4), and is actually understood as synapomorphic of the family Caprinidae (represented by *Offneria* and *Pachytraga* in our analysis), while the protrusion

of the PM-LV is synapomorphic of the Caprinuloideidae (*Amphitriscoelus* and *Mexicarpina*) (#4.3). The trend toward blade-like myophore evolution was considered diagnostic feature of Caprinidae plus Caprinuloideidae (corresponding to the previous family Caprinidae and its subfamilies Caprininae and Coalcomaninae differentiated one to each other by the muscle position relatively to the blade carrying the myophore (Chartrousse 1998). However, the myocardinal organization of the Caprinidae plus Caprinuloideidae has to be decompound into characters for the left valve blade and for the right valve blade (#4, PM-LV; #6, AM-LV; #26, PM-RV; #27, AM-RV).

3.2.2.3. Cavities and pallial canals

Accessory cavities and pallial canals derive from the partition of the shell by vertical blades, considering that canals are small cavities between duplicated walls. The large cavities open in rudist shells that are not homologous to the general cavity of other bivalves that are usually named accessory cavities. Among accessory cavities, the 'ectomyophoral' and 'endomyophoral' cavities (Skelton *et al.* 2013) or 'posterior accessory cavity' and 'posterior myophoral cavity' (Chartrousse 1998) are pertinent for description of some taxa but cannot be used for a search of homologies among the rudists. Our approach of homologies leads to a novel evolutionary hypothesis in which cavities can extend over other spaces of the shell (fig. 4.21; #11-16 for the LV and #28-32 for the RV). Thus 11 characters and 27 characters states are necessary to properly transcribe Skelton and Smith's unique character 'Pallial canals regularly developed in aragonitic inner shell layers. 0, no; 1, yes, in LV only; 2, yes, in both valves' (character 32 in Skelton and Smith, 2000), and the characters 11 and 12 of Sano *et al.* (2014) 'LV: 0, none (or sparsely present in some specimens only); 1, around both anterior and posterior valve margins but not continuing around entire ventral margin; 2, continuous around anterior, ventral and posterior valve margins' and 'RV: 0, none (or sparsely present in some specimens only); 1, around both anterior and posterior valve margins but not continuing around entire ventral margin'.

3.2.2.3.1. Left valve

3.2.2.3.1.1. Canals

On the LV, two cavities are present in almost all rudist taxa: G and the CTS. In *Diceras* and *Valletia*, the CTS is small, and fulfills the function of socket for the PT. Tooth and associated socket fits and are approximately of the same size as in Requienids, which is hypothesized to be the ancestral condition. Our hypotheses about the development of pallial canals follows a rationale of cavity development (fig. 4.21). First, a new cavity appears posterior to the PM-LV, and it can be observed in Polyconitids and Caprotinids (#11.1). Then, the CTS extends itself anteriorly and surrounds the AM-LV (#6.4). The character 6.4 represents both the CTS extension and the AM-LV tilting that are structurally interrelated homologs. The raise of two characters, one for the CTS and one for the AM-LV would generate over-ponderation by logical dependence between them. This hypothesis is supported by observations on *Sellaea, Retha* and *Pachytraga* shells where it seems to be a unique cavity that develops from the CTS and extend dorsally to the AT (fig. 4.22). In *Amphitriscoelus* numerous radial walls grow in in both the PAC and the extended CTS (#13.1). The radial walls built canals that can subsequently invade myophores (#16). We identify a row of pallial canals that demarcates themselves from other canals by their position and size (#14). When looking the *Kimbleia capacis* shells illustrated in Aguilar-Pérez thesis (2008), we see this row as an extension of the CTS. Our hypothesis is a two-steps CTS extension: a first step where the cavity extends so much that it ends in merging itself with the PAC (#12.1), and a second step where the CTS extends itself another time to surrounds all G in the aspect of a single row of large rounded canals.

Plagioptychus diverges from other taxa. Because the AM-LV is attached to a horizontal thick shelf, it has not been invaded by the CTS that is kept dorsally. However, *Plagioptychus* possess canals. Our hypothesis is that *Plagioptychus* possess a PAC widely developed ventraly and filled with numerous radial walls. In Caprinuloideids, Caprinids, Trechmanellids, Antillocaprinids and Ichthyosarcolitids, canals are found everywhere in the shell. This morphology results from the ventral contact and coalescence between the PAC and the CTS both filled with radial walls. Then, these canals are probably not homologous to Plagioptychid canals.

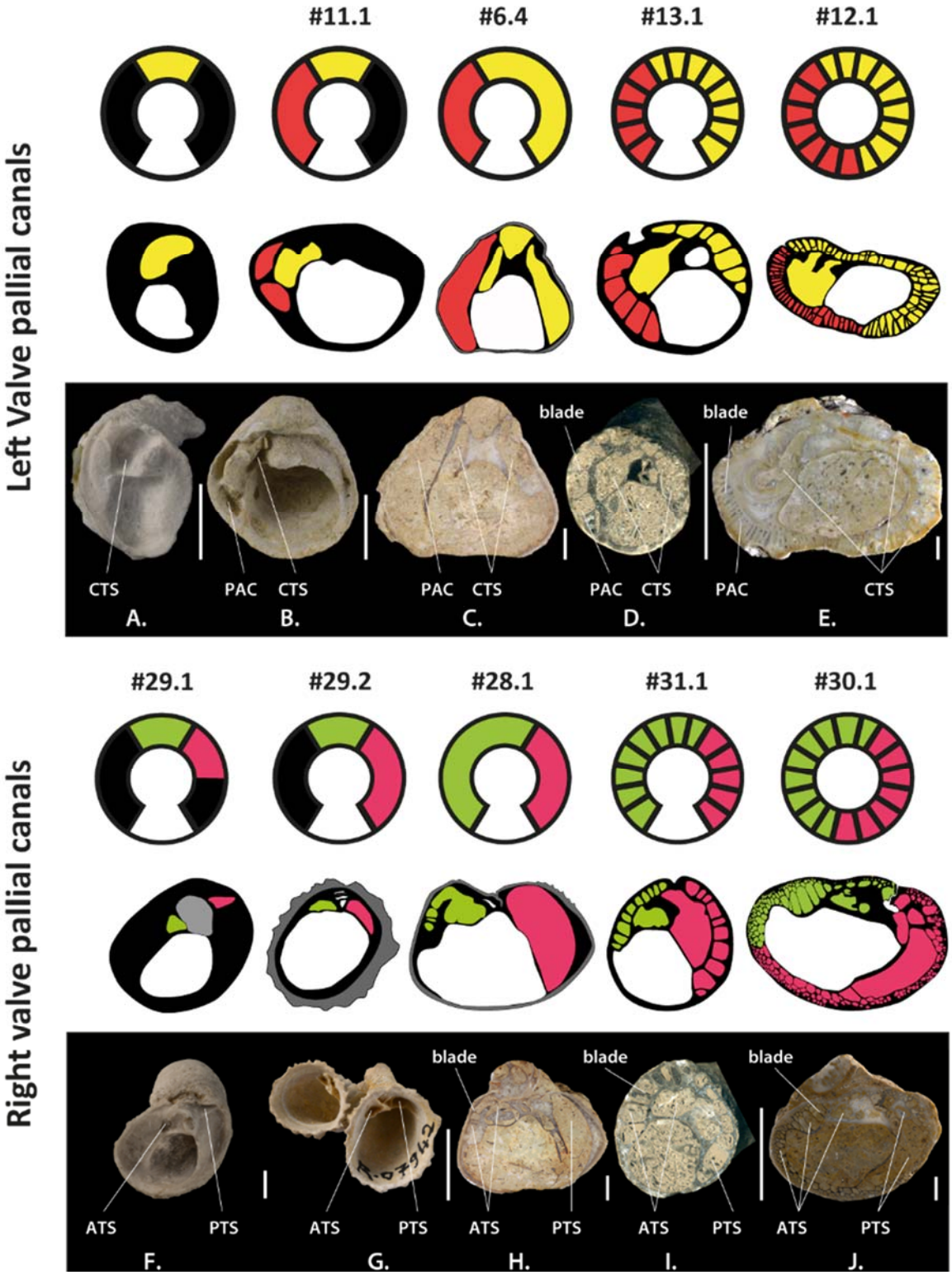


Figure 4. 21. Hypotheses about the origins of pallial canals. Nine character states are implied in their evolution: four on LV (#6.4, 11.1, 12.1, 13.1) and five on RV (#28.1, 29.1, 29.2, 30.1, 31.1). Left valve: digging of a PAC (#11.1), anterior expansion of the CTS (#6.4), cavities invaded with blades (#13.1), ventral connection between PAC and CTS (#12.1). Right valve: digging of a PTS (#29.1), posterior expansion of the PTS (#29.2), anterior expansion of the ATS (#28.1), cavities invaded with blades (#31.1), ventral connection between PTS and ATS (#30.1). Left valves: A – *Monopleura michaillensis* (UPMC.1272), B – *Caprotina striata* (MV 2003.2.5147), C – '*Sellaea' sp*. (WSA3058), D – *Amphitriscoelus waringi* (picture provided by Jean-Pierre Masse), E – *Offneria rhodanica* (UCBL.R275; picture provided by Emmanuel Robert). Right valves: F – *Plagioptychus arnaudi* (UPMC.1273), G – *Caprotina striata* (MV 2003.2.5147), H –'*Sellaea*' *sp*. (WSA3058), I – *Amphitriscoelus waringi* (picture provided by Jean-Pierre Masse), J – *Caprinula boissyi* (MNHN.6924). Scale bars, 10 mm.

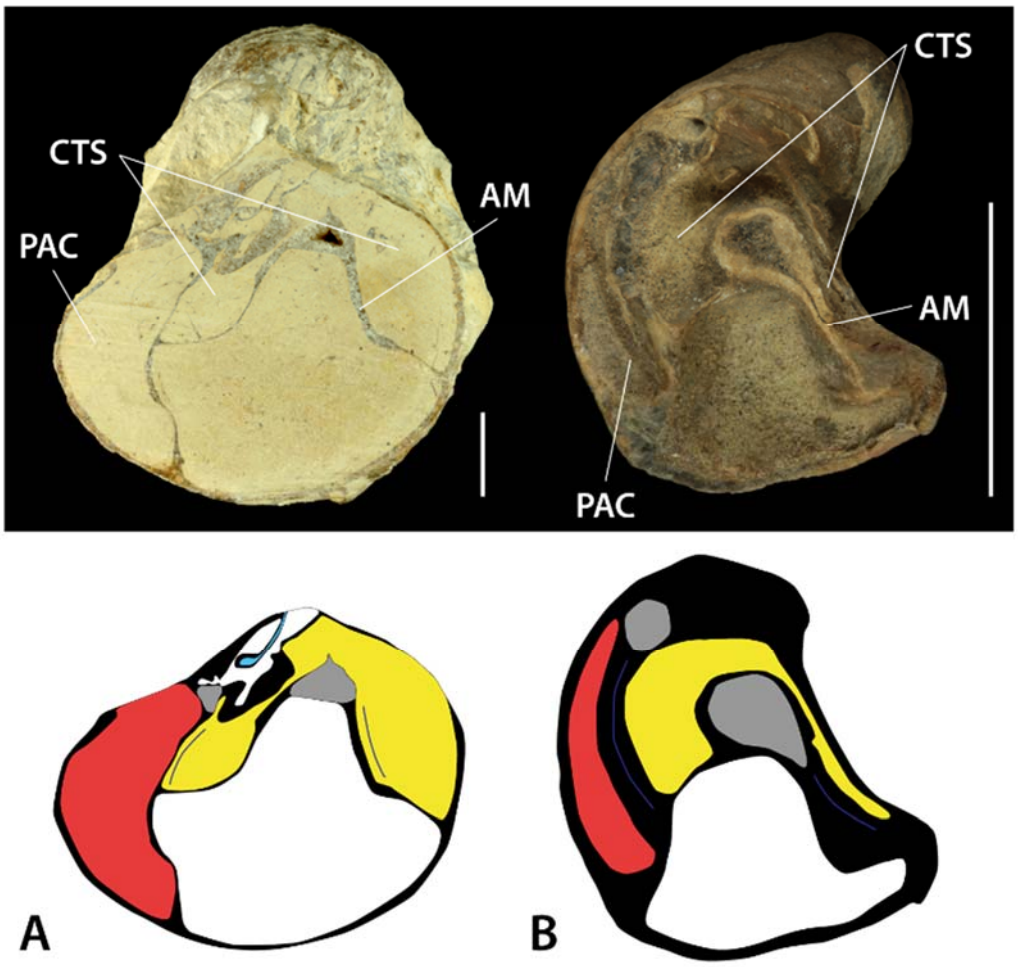


Figure 4. 22. Illustration of the CTS anterior development. A – '*Sellaea' sp* (UT.48167), sliced left valve. B – *Retha sp*. (WSA3831.4), left valve. Scale bars, 10 mm.

3.2.2.3.1.2. Central tooth socket

The CTS varies in several ways. It can invade the shell posteroventrally to surround the PM or be small and disconnected from myophores (#10). It can also vary in orientation (#8). Completely horizontal in *Diceras*, it is tilted in *Monopleura* and it becomes completely vertical relatively to the commissure plane in *Plagioptychus*. Previous authors recognize two cavities, the CTS, and a neoformed cavity supporting the PM (Chartrousse 1998; Skelton and Smith 2000) on descriptive and functional basis. This hypothesis is based on the distinction between the part of the cavity receiving the tooth from the part of the cavity receiving the myophore (fig. 4.23). We also note the presence in some taxa of a blade separating the CTS from the vertical MP. In our taxonomic sampling, only *Chaperia* has a blade between CTS and MP, and maybe a residue of this blade in *Offneria* where the ventral

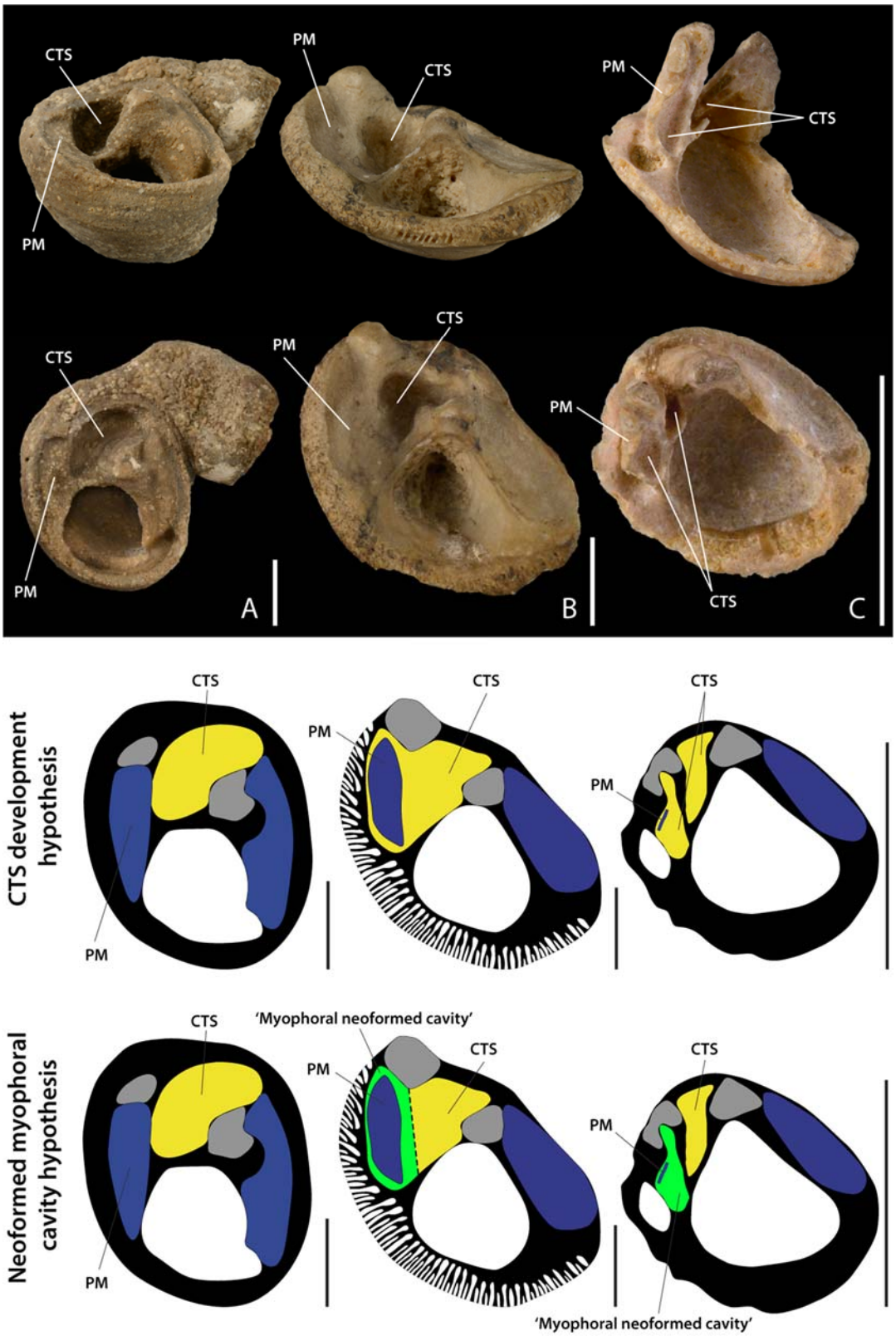


Figure 4. 23. Illustration of confronting hypotheses on CTS homology. A – *Valletia germani* (MNHN.F.J05798); B – *Plagioptychus arnaudi* (UPMC 1273); C – *Chaperia costata* (MNHN.F.R07950). CTS development hypothesis supported herein: the CTS in *Valletia germani* is primitively stranded between the AT and the PM (myophoral shelf): in *Plagioptychus arnaudi*, the posterior myophoral plate becomes oblique, causing the expansion of the CTS; in *Chaperia costata*, the expanded CTS is divided by a thin blade. Myophoral cavity hypothesis: the CTS in *Valletia germani* is primitively stranded between the AT and the PM (myophoral shelf); in *Chaperia costata*, a myophoral cavity fills the position of the posterior myophoral shelf – myophoral cavity and CTS are separated by a blade; in *Plagioptychus arnaudi*, the CTS/myophoral cavity blade disappear. Scale bars, 10 mm.

part of the wall disappeared, marking a connection between the two cavities (CTS and the cavity were is inserted the PM). We go back to the previous hypothesis of a single cavity that can extend itself to surround the PM (#4.2), and we call this cavity CTS to highlight its origins, even if the function in numerous taxa as Caprinids can vary (the CT or the PM-RV blade can be inserted in it). This hypothesis is not new and can be found in Douvillé (1887) were the CTS, noted n, corresponds to a 'fossette cardinal, incomplètement remplie par la dent correspondante N (CT) et jouant le role d'une grande cavité accessoire' (Douvillé 1887: 782). When the PM-LV become tilted on its myophoral shelf, it generates a single huge cavity that can reach the diameter of G as in *Plagioptychus*, and that usually connect to the CTS. This huge CTS can be seen in most Caprinidae, Caprinuloideidae, Caprotinidae, and Caprinulidae. In *Chaperia*, the vertical PM-LV becomes so prominent that it distorts the CTS and raises up to the commissure its ventral part that becomes much less deep than the dorsal part where the CT is inserted.

3.2.2.3.2. Right valve

The same pattern of cavities development occurs on the right and left valves, but at different evolutionary rate and with distinct homologies (fig. 4.21). Following the same argumentation than for the CTS, the PTS is understood as the small cavity, as observed *Plagioptychus* or *Monopleura* (fig. 4.4-C to F), where is inserted the PT of the LV (#29.1). The PTS can extends itself by tilting of the PM-RV shelf (#29.2). When extended (e.g, in *Caprotina*), the PTS occupies exactly the same position as the set PT+PM-RV shelf of *Monopleura*. An anterior extension of the ATS can be seen in *Amphitriscoelus* or *Sellaea* (#28). The cavities can be filled with radial blades (#31.1). Intially well separated, the two cavities can merge into a single undifferentiated 'pallial canal zone', as in *Caprinula*, Antillocaprinidae, Trechmanellidae, some Caprinidae species and most Caprinuloideidae (#30.1). Pallial canals can invade the entire shell, including the CT as in *Titanosarcolites* or *Curtocaprina* (#32). Even if their morphology is strongly modified, we hypothesize these pallial canals being homologous to pallial canals originating from partitioning of the sockets. Tooth pallial canals have the same origins. More generally, the walls involved in the formation of pallial canals can have a great diversity of morphology, (Damestoy 1970; Chartrousse 1994), with walls that can divide themselves dichotomously as in

Plagioptychids, or can be densely packed to form capillary canals as in *Titanosarcolites* (#13). *Coralliochama* exhibit very distinctive canals that look more cellular and angulose than in any other canaliculated species. Their origin cannot be the same as in other canaliculated shell taxa, because its sockets are not developed and never invade the RV myophoral shelves. *Coralliochama* is understood as having structures that cannot be homolog to pallial canals. In our analysis, the *Coralliochama* RV 'canal-like' structure is autapomorphic and cannot support any supra-generic taxa.

3.2.2.3.3. Tabulae

Tabulae are usually constructed in deep cavities of various taxa in the living world, and is generally considered as a convergent feature. However, in rudists, when situated in several cavities, they are found to be generated independently, at the contrary of the cellular microstructure of the calcitic layer in Radiolitids (Pons and Vicens 2008). The occurrence and shape of tabulae are quite independent from the size of the shell, some of them can be retrieved in small *Ichthyosarcolites* pallial canals, sockets and GC whereas there is no trace of tabulae in some big shells of Radiolitids and Hippuritids. We consider the tabulae of the various shell cavities as evolutionary significant. Tabulae occur in G (#17) or in D (#33), in the pallial canals of the LV (#18) or the RV (#34), and in sockets (#19), and they evolved at distinct timings among taxa.

3.2.2.4. Calcitic outer shell layer

Most features are internal and localized on the inner aragonitic shell layer. The external features are those related to the external calcitic layer of the shell (Kennedy and Taylor 1968). Aragonitic and calcitic layer can vary in thickness relatively to each other (#36). Shell ornaments are common in rudists, represented mostly by growth lines and simple ribs of diverse sizes (*Monopleura michaillensis*, *Rotacaprina trilobata*; #39). We define rudist costulations as costae formed on the outside of the calcitic material. Flanges are found mainly in Ichthyosarcolitids and Antillocaprinids, and are ribs involving mainly the aragonitic layer (#39). They distort the conical shape of the shell to increase shell robustness. Flanges and costulations can both appear associated on a single shell. Radial bands are folding of the outer layer that mark costulations and the growth lines and that

might be related to faecal and pseudofaecal streams (Skelton 1976). Hippuritidae and some Radiolitidae develop two pillars produced by infolding of the outer layer in similar position to the radial bands (#35). Pillar is the term used in the literature to describe Hippuritid folds, those of Radiolitids being considered as pseudopillars. All of these features that are pillars, pseudo-pillars and radial bands are considered homologous in virtue of their similar position.

In the family Radiolitidae, shells exhibit a complex cellular calcitic layer (Amico 1978, Pons and Vicens 2008). In Hippuritidae, a very specialized anatomical novelty, the canal/pore system, develops within the calcitic layer of the LV. The surface of the opercular LV is filled with numerous tiny pores (#20) that are highly variable in shape amongst species. The pores communicate with calcitic canals (#21) that radiate from the umbo to the commissural plane. The calcitic canals are not homologous to pallial canals because of their localization in a different shell layer. This system allowed the animal to bring water inside the general cavity without necessitating the opening of the valve. The water enters by the pores and is trasfered to GC though the calcitic canals.

Traces of color patterns were reported in the outer shell layer of rudists. Douvillé (1888) described a left valve of *Plagioptychus* from Gosau (Austria): 'une série de bandes étroites en forme de lignes brisées qui se détachent en clair sur la couleur sombre de la coquille'. The pattern is independent from growth lines of the shell. Unfortunately, colors are reported in too few taxa for being considered in phylogenetic analyses.

3.2.2.5. Ligament

The rudist ligament is known only from few exceptionally preserved specimens (Skelton 1979; Mitchell 2013b). However, the ligament location on the shell is frequently marked by a ligamentary ridge that, in some taxa, can be connected to an internal ligamentary cavity. The ligament of Diceratids is typically parallel to the commissural plane, lying along the dorsal margin of the shell (#37). It becomes perpendicular to the shell margin (vertical) in *Monopleura* and in all other rudists, or even folded inside the shell, marking an internal cavity (Yonge 1967; Dechaseaux 1969). The ligament disappears in some taxa such as *Titanosarcolites* or *Curtocaprina*, where no ligamentary structure

remains (no ridge nor cavity). The ligamentary ridge can fold into the shell margin (#38), even if always located dorsally. The ligamental infloding in Radiolitids and Hippuritids becomes similar in shape with a pillar and can develop itself through the CTS (#9). One of the most obvious attribute of rudists is the loss of valves symmetry, leading to the characterization of rudists as 'bizarre', 'aberrant' or 'baroque' bivalves. Each of the two valves can be coiled or uncoiled, highly developed (*Titanosarcolites*) or regressed to an opercula (Hippuritidae). The ligament evolution is related to the shell uncoiling, and may contribute to the high diversity of rudist morphologies by relaxing functional constraints (Skelton 1985; Stone and Telford 1999).

The valve coiling is a continuous feature, and delineation of states is difficult. We choose to consider as uncoiled only completely capuloid shells, or shells in which the umbo shows no torsion. Any other slight coiling is rejected from an 'uncoiled stage'. For this reason, only the LV can be represented in the characters (#22), the RV being always coiled in its juvenile stage. Note that vertical ligament and a shell uncoiling can be observed in other bivalve taxa, such as *Saccostrea*.

3.2.3. Characters

Our coding approach introduces new hypotheses of homology. The codings of left and right valves are completely independent, except for the ornaments and the ligament that connects the valves. Teeth development is considered conservative, with anterior, central and posterior teeth retrieved in most taxa. The hypothesis of an ichthyosarcolitid toothlet is rejected (fig. 4.19), to the benefit of a movement of the CTS/G wall. Myophores are assumed present in all taxa, but can be of two types: horizontal (myophoral shelf) or vertical (blade-like, tooth-like). The vertical myophores may originate by invasion of a cavity resulting in the tilting of myophoral shelves. Myophore orientations are also considered. Pallial canals benefit of a radically different coding than previous studies by highlighting multiple origins (fig. 4.21). Definition and delineation of cavities based on functional aspects and descriptive terms are rejected to the profit of homologies of phylogenetic significance. The hypothesis of a primitive amphidetic ligament is retained; the ligament becoming vertical in Monopleurids and invaginating or disappearing later on.

40 characters are built for our phylogenetic analysis: 22 for the LV, 13 for the RV and 5 involving both valves. All characters and associated taxa are shown on the cladistic matrix in table 4.4. Here are presented in detail each character, coded as hierarchical hypotheses for their subsequent analysis in 3ta (following Williams and Ebach 2006, Cao *et al.* 2007). For each character are associated a number, a name, its position (LV, RV or both), the supporting structure (teeth, myophores, vertical blades, pallial canals, horizontal blades, canal/pore system, pillars, outer calcitic shell layer, ligament, and ornamentation), its hierarchical relations between states, the states definitions (excepting the state 0, uninformative, for which we give sometimes a description of what a shell must look like when bearing no derived feature, or the term [root] by default; the term absence being avoided), the result of the phylogenetic test character state by character state, the global retention index of the character and remarks, including its relation to characters of previous cladistic analyses (Chartrousse 1994, Smith 1994, Skelton and Masse 1998, Steuber 1999, Stone and Telford 1999, Skelton and Smith 2000, Mitchell 2010, Sano *et al.* 2014). Some characters however are coded in a different way and are not easily comparable (e.g. a new proposal for homologies on myophores).

3.2.4. Parameters of the phylogenetic analyses

27 species from the 13 Hippuritidina families are used as terminal taxa. Each family is represented by two species to test their monophyly. The two species from a family are chosen as the most morphologically distant to sample a maximum of the rudist diversity. This includes *Diceras arietinum* and *Valletia germani* (Diceratidae), *Monopleura michaillensis* and *Glossomyophorus costatus* (Monopleuridae), *Polyconites verneuili* and *Tepeyacia corrugata* (Polyconitidae), *Plagioptychus arnaudi* and *Coralliochama orcutti* (Plagioptychidae), *Caprotina semistriata* and *Chaperia costata* (Caprotinidae), *Pachytraga paradoxa* and *Offneria rhodanica* (Caprinidae), *Sellaea sp*. and *Caprinula boissyi* (Caprinulidae), *Amphitriscoelus waringi* and *Mexicaprina minuta* (Caprinuloideidae), *Semailia smithi* and *Dictyoptychus morgani* (Trechmanellidae), *Clinocaprina lowentsami* and *Titanosarcolites giganteus* (Antillocaprinidae), *Curtocaprina clabaughikinsorum* and *Ichthyosarcolites bicarinatus* (Ichthyosarcolitidae; see chapter 3), *Agriopleura marticensis* and *Radiolites radiosus* (Radiolitidae), *Hippurites radiosus* and *Barrettia monilifera*

Character number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Character hierarchy	(0(1(2)(3)))	(0(1))	(0(1))	(0(1(2(3(4)(5(6))))))	(0(1))	(0(1(2)(3(4(5))))	(0(1))	(0(1(2)))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1(2)(3)))	(0(1(2)))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1)(2(3(4)(5(6)))))	(0(1(2))	(0(1))	(0(1(2)))	(0(1))	(0(1(2)))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1))	(0(1(2(3))))	(0(1))	(0(1))	((1)0)
Diceras arietinum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Valletia tombecki	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Monopleura michaillensis	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	1	0	1	0	0	0	0	0	0	0	1	0	1	0
Glossomyophorus costatus	1	1	0	3	0	1	0	?	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	?	3	1	0	1	0	0	0	0	0	0	0	1	0	1	0
Radiolites radiosus	2	1	1	5	1	4	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	2	0	0	0	1	0	0	1	1	1	1	0
Agriopleura marticensis	2	1	1	5	0	4	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	1	0	1	0	0	0	0	1	1	0	1	0
Hippurites radiosus	2	1	0	6	1	5	0	2	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	5	0	0	1	0	0	0	1	0	1	1	1	1	1	0
Barrettia monilifera	2	1	0	6	1	5	0	2	?	1	0	0	0	0	0	0	0	0	?	1	1	1	1	1	1	5	0	0	1	0	0	0	1	0	1	0	1	0	1	0
Mexicaprina minuta	1	1	0	5	0	4	0	2	0	0	1	1	2	1	0	0	0	0	0	0	0	0	1	1	1	5	2	1	2	1	2	0	0	0	0	0	1	0	1	0
Amphitriscoelus waringi	1	1	0	5	0	?	0	2	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	?	5	?	1	2	0	1	0	0	0	0	0	1	0	1	0
Pachytraga paradoxa	1	1	0	3	0	3	0	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	4	0	1	2	0	1	0	0	0	0	0	1	0	1	0
Offneria rhodanica	1	1	0	2	1	3	0	2	0	0	1	1	3	0	0	0	0	1	1	0	0	0	1	1	1	4	1	1	2	1	0	0	0	1	0	0	1	0	1	0
Dictyoptichus morgani	3	1	0	5	1	4	0	2	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	6	0	1	2	1	1	0	1	1	0	1	2	0	1	0
Semailia smithi	3	1	0	6	1	4	?	2	0	1	1	1	2	2	0	1	0	0	0	0	0	0	1	1	1	6	?	1	2	1	2	0	0	0	0	0	2	0	1	1
Titanosarcolites giganteus	2	1	0	5	1	4	1	2	0	1	1	1	2	0	0	1	1	1	0	0	0	0	1	1	1	6	2	1	2	1	2	1	1	1	?	0	3	?	1	1
Clinocaprina lowentsami	3	1	0	5	0	4	0	2	0	1	1	1	2	0	0	1	0	1	0	0	0	0	1	1	1	5	0	1	2	1	2	0	1	1	0	0	2	0	0	0
Ichthyosarcolites bicarinatus	2	1	0	5	0	4	0	2	0	1	1	1	2	2	0	0	1	1	?	0	0	0	1	1	1	6	0	1	2	1	2	0	1	1	0	0	3	0	0	1
Curtocaprina clabaughikinsorum	1	1	0	5	0	3	0	2	0	1	1	1	2	2	0	0	0	0	0	0	0	0	1	1	1	1	0	1	2	1	2	1	0	0	0	0	3	0	0	1
Plagioptychus arnaudi	2	1	0	2	0	2	0	2	0	0	1	0	3	0	0	0	0	0	0	0	0	0	1	0	1	2	1	0	1	0	0	0	0	0	0	0	1	0	1	0
Coralliochama orcutti	2	1	0	2	0	2	0	2	0	0	1	0	3	0	0	1	0	0	0	0	0	0	1	1	1	2	1	0	1	0	0	1	0	0	0	0	1	0	1	0
Polyconites verneuili	1	1	0	4	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	1	1	0	1	0
Tepeyacia corrugata	1	1	0	4	0	4	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	?	1	1	1	1	0	?	1	0	0	0	0	0	1	1	1	0	1	0
Caprotina striata	1	1	0	3	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3	0	0	2	0	0	0	0	0	0	0	1	0	1	0
Chaperia costata	1	1	0	3	0	4	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3	0	0	2	0	0	0	0	0	0	0	1	0	1	0
Caprinula boissyi	2	1	0	2	0	?	0	2	0	0	1	1	2	2	1	0	0	0	1	0	0	0	1	1	1	4	?	1	2	1	2	0	0	0	0	0	1	0	?	0
Sellaea americana	2	1	0	3	0	3	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3	1	1	2	0	0	0	0	0	0	0	1	0	1	0
Oryxia sulcata	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	?	?	6	0	1	1	1	2	1	1	0	0	0	1	1	0	1

Table 4. 4. Matrix of 3ta characters with their hierarchical coding and their assignation to OTUs of this analysis.

(Hippuritidae). *Sellaea sp*. has a special status, as only a single specimen was coded. This sliced specimen was chosen from Austin's collections (WSA.3058, labelled '*Sellaea americana* Van de Flied') because it exhibits a clearly original myophoral configuration of importance for the rudist phylogeny.

The taxa were analyzed using three-taxon analysis (3ta) with 40 characters. In 3ta, the characters are understood as hierarchies, in the same way as taxa (Zaragüeta i Bagils and Bourdon 2007). On the 40 characters, 29 are binary (with a state representing the root and a derived character state), 5 are composed of three states and 1 of four states. On these 35 characters, each state is included and/or include a unique other state (only orthologous clades, no paralogy). The 5 other characters are designed with a paralogous clade and can be divided into several sub-characters following the free-paralogy subtree analysis (Nelson and Ladiges 1996). These five characters are #1: (0(1(2)(3))) (teeth), #4: (0(1(2(3(4)(5(6))))))) (myophores), #6: (0(1(2)(3(4(5)))) (myophores), #13: (0(1(2)(3))) (pallial canals), #26: (0(1)(2(3(4))(5(6)))) (myophores). The 40 characters are composed of 105 character-states (out of which 65 derived states). The matrix and the associate hierarchical coding is given in table 4.4. All characters are phylogenetically informative – a minimum of two taxa in the derived state and 1 taxa in the ancestral state are required. The 3ta was conducted on LisBeth 1.0 (Zaragüeta i Bagils *et al.* 2012) for the file formatting (appendix 2) and PAUP 4.0b10 (Swofford 2003) for the heuristic search and 1000 replicates.

All character hierarchies being rooted, they are polarized. Polarization is made *a priori* to the tree reconstruction in the 3ta. The polarization is a process of the character formulation with reference to outgroup criteria. Our analysis considers only the taxa Hippuritidina. As outgroup criterion for polarization, we looked two species from the families Diceratidae and Epidiceratidae. Outgroups are always external to the analysis. 3ta allow to choose a different outgroup for each character, or several outgroups, leading to greater flexibility in the coding process and morphoanatomical comparisons. The term 'absence' is systematically excluded from descriptions and character state definitions. The ancestral character state is the root and define only the 'universe' of the analysis (in a

mathematical sense). Thus, there is no need to define it. When possible, there is a description, and where it is impossible, the term [root] is used.

4. Results

The 3ta produces three cladograms varying only in the relative position of taxa inside the clade gathering *Glossomyophorus*, *Caprotina*, and *Chaperia*. No triplet from the characters allows grouping of two of these taxa from a third. The reduced cladistic consensus give the same results as the strict consensus, giving a single consensus tree with a trichotomy inside this clade. The strict consensus is interpreted as a cladogram and is taken as working hypothesis for Hippuritidina phylogenetic relationships (see chapter 1, part consensus trees, for the justification). The strict consensus cladogram is reported in figure 4.24 with all its synapomorphies. It contains six paralogous clades. The retention index for the 3ta tree is 0.77 (i. e. 77% of the phylogenetic information contained in the characters is retained in the optimal tree). RI by character is detailed in table 4.5, with the result for the homology test. Out of the 65 derived character states, 32 pass the character state testing procedure, and 33 are rejected as homoplastic (chapter 1, p. 91 for the detailed procedure).

Cladistic information (triplets) is located on the phylogeny using the Nodal Retention index (fig. 4.25). The repartition of the phylogenetic information is globally uniform because of the pectinate structure of the tree (fig. 4.25). However, interesting informations can be interpreted from the NRI values. The teeth mainly supports the most inclusive clades of our phylogeny, the ornamentation supports mainly small clades (Caprotinidae, Polyconitidae, Plagioptychidae, Hippuritidae), and horizontal blades supports mainly the clade 15 (fig. 4.25-B). Following Figure 4.28-C, characters about myophores supports mainly the clades 4 and 12, the Polyconitidae, characters about pallial canals supports mainly the clade 15. All other clades (mainly deep nodes) are equally supported by canals and myophores. Finally, when looking at the information repartition by valve, it seems equilibrate on the main ortholog serie (from clade 1 to 24), the small clades 4 and 17 being mainly supported by RV characters, and clades 8 and 12 being mainly supported by LV characters.

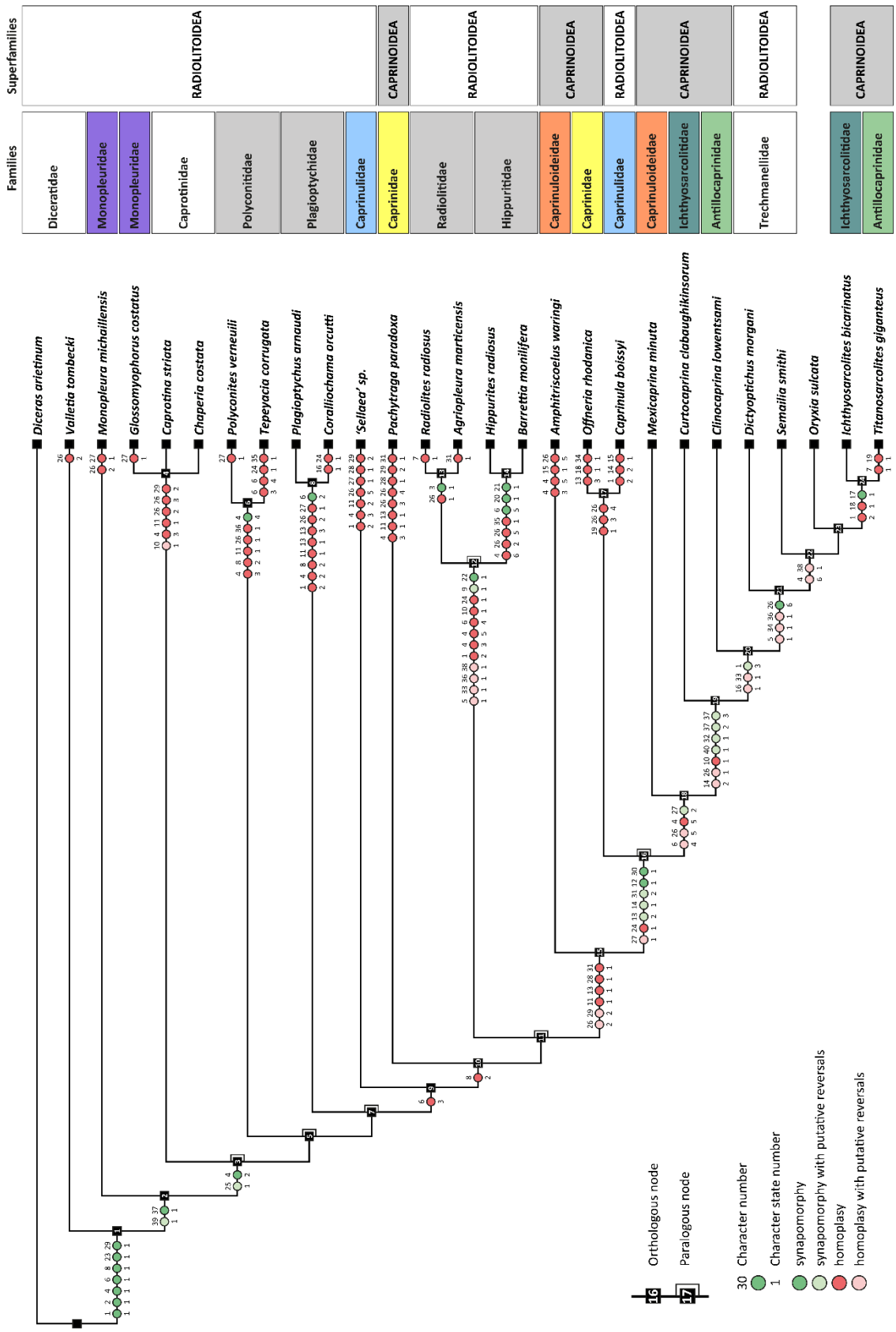


Figure 4. 24. Strict consensus tree of the two optimal trees obtained in 3ta, with corresponding character states. Squares represent the nodes of the tree, and numbers refer to clade labels. Circles represent character states that characterize clades. At the right of the tree, panels correspond to families and superfamilies according to Skelton (2013). Families – white: paraphyletic group, grey: monophyletic group, colored: polyphyletic group. Superfamilies – white: Radiolitoidea, grey: Caprinoidea.

Character localisation	Character number	Kind	Number of informative states	Retention index	Synapomorphic states
LV	1	Teeth	3	0.72	1, 3
	2		1	1.00	2
	3		1	1.00	1
	4	Myophores	6	0.80	1, 4
	5		1	0.74	
	6		5	0.92	1, 2, 5
	7		1	0.52	
	8	Vertical blades	2	0.95	1
	9		1	0.96	1
	10		1	0.74	
	11	Canals	1	0.60	
	12		1	1.00	1
	13		3	0.82	2
	14		2	0.84	1
	15		1	0.67	
	16		1	0.64	
	17	Horizontal blades	1	1.00	1
	18		1	0.83	
	19		1	0.83	
	20	Canal/pore system	1	1.00	1
	21		1	1.00	1
	22	Coiling	1	1.00	1
RV	23	Teeth	1	1.00	1
	24		1	0.86	
	25		1	0.84	1
	26	Myophores	6	0.60	6
	27		2	0.31	2
	28	Vertical blades	1	0.92	
	29		2	0.64	1
	30	Canals	1	1.00	1
	31		2	0.92	2
	32		1	0.94	1
	33	Horizontal blades	1	0.79	
	34		1	0.84	
Both valves	35	Pillars	1	0.51	
	36	Outer layer	1	0.51	
	37	Ligament	3	0.95	1, 2, 3
	38		1	0.65	
	39	Ornamentation	1	0.38	1
	40		1	0.96	1

Table 4. 5. Overview of the results of the character testing procedure using 3ta. List of hierarchical characters with their localization (LV, RV or both), their number in this paper, their kind, their number of informative states (i. e. excluding the root), their retention index, and the character states numbers that pass the congruence test and become synapomorphies on the cladogram.

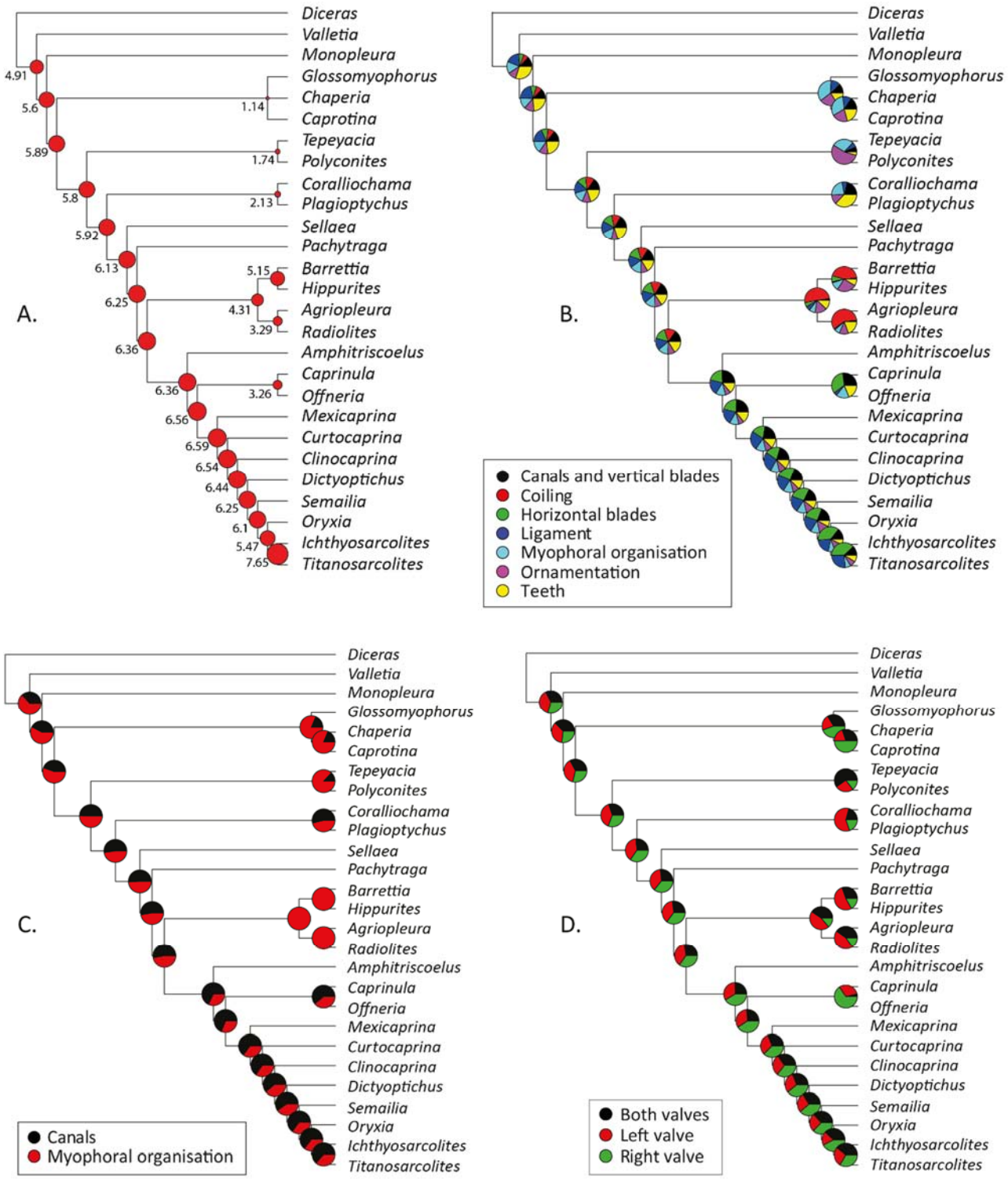


Figure 4. 25. Description of the phylogeny and clades by their *Nodal Retention Index* (NRI) displaying the distribution of cladistic information amongst the tree. A triplet is generally counted in the NRI of several ortholog clades (ortholog clades being dependent). A – Absolute NRI values; B – proportions of NRI by kinds of characters; C – proportions of NRI between canals and myophores; D – proportions of NRI by valve.

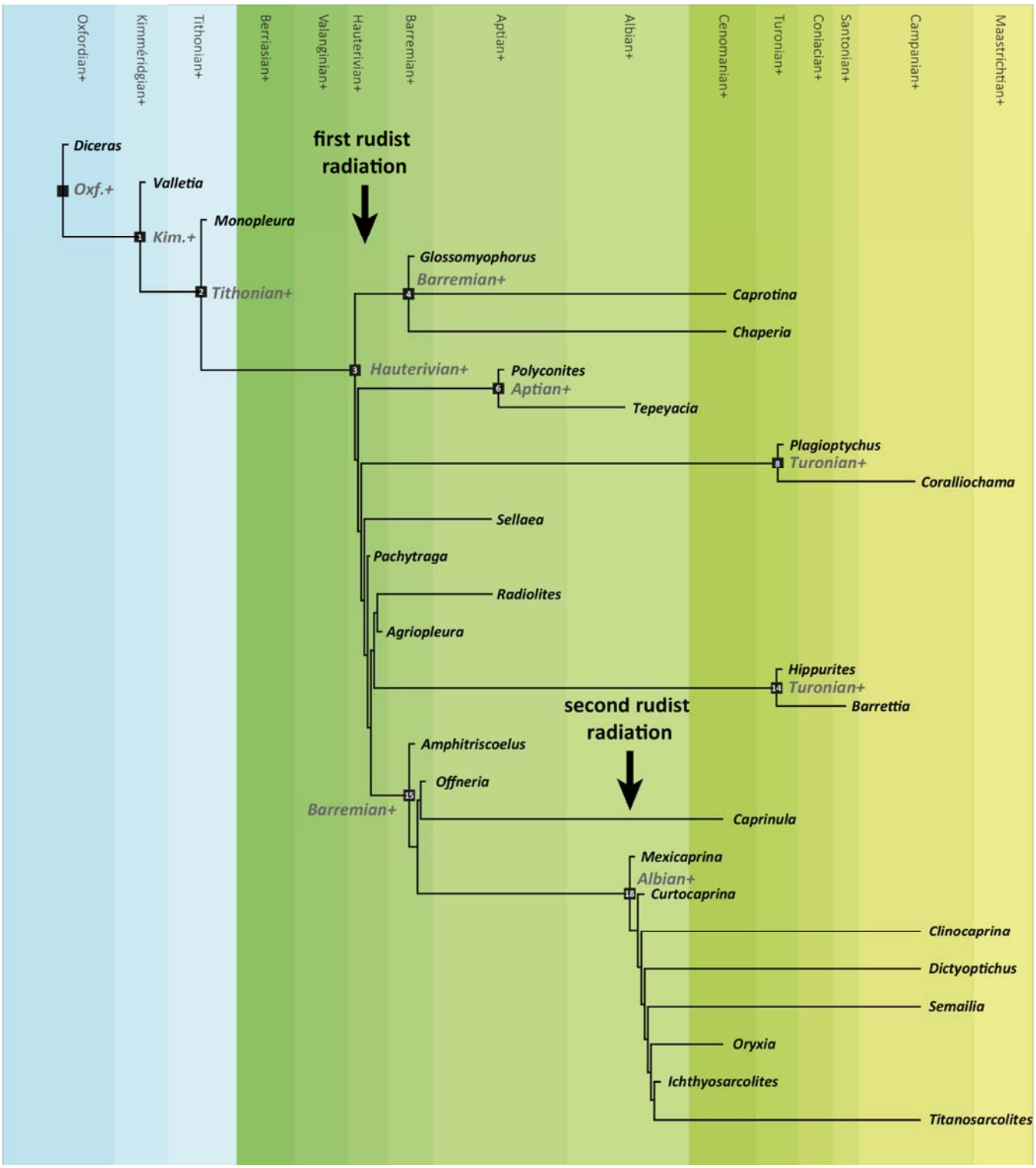


Figure 4. 26. Dated phylogeny highlighting two radiations of clades. Clades are characterized by their age of origin.

Finally, the cladogram is dated following procedure described in Chapter 1, and displayed in Figure 4.26.

5. Discussion

5.1. Cladogram

The cladogram reconstructing rudist evolution exhibits a highly pectinate structure, with six paralogous clades (3, 6, 8, 12, 13, and 17) interpreted as six diversification events. We compare these resulting clades to the most recent and complete update on rudist taxonomy (Skelton 2013a) presented for the revised Bivalvia volumes of the 'Treatise of Invertebrate Paleontology' (table 4.6). This taxonomic classification has never been tested by cladistic analyses, and some families and one superfamily have been erected as paraphyletic taxa. On the 13 families, 5 are retrieved monophyletic (Caprotinidae, Polyconitidae, Plagioptychidae, Radiolitidae, and Hippuritidae). The family Polyconitidae is retrieved monophyletic whereas it was precedently hypothesized as a paraphyletic 'stem group' for Hippuritidae and Plagioptychidae (Skelton and Smith 2000; Skelton 2013a). The hypothesis of *Tepeyacia* as sister group to Hippuritidae in virtue of sharing pillars and infoldings of the outer shell layer is rejected. The paraphyly of Diceratidae makes consensus and is confirmed by our analysis. The Monopleuridae are polyphyletic on our cladogram, *Glossomyophorus* being closer to the Caprotinidae than to *Monopleura*, pledging for the reintegration of *Glossomyophorus* as a Caprotinidae (a former hypothesis of Masse *et al.* 1984). The tree supports a redefinition of the family Caprotinidae which was isolated from other taxa with only two genera in Skelton 2013a because of uncertainties on their phylogenetic position. More surprisingly are the rejection of numerous families previously hypothesized monophyletic: Caprinulidae is found polyphyletic, and Trechmannellidae, Antillocaprinidae, Ichthyosarcolitidae and Caprinidae are found paraphyletic. The Caprinuloideidae were precedently understood as paraphyletic taxa composed of several sister groups of Antillocaprinidae. We found the Caprinuloideidae

	Taxa	Skelton and Smith 2000	Skelton's 2013 presumptions	Our test
Superfamily	Radiolitoidea	paraphyletic	paraphyletic	polyphyletic
	Caprinoidea			polyphyletic
Family	Diceratidae	paraphyletic	paraphyletic	paraphyletic
	Monopleuridae	polyphyletic	paraphyletic	polyphyletic
	Polyconitidae	paraphyletic	paraphyletic	
	Caprotinidae	not tested		paraphyletic
	Plagioptychidae	not tested		
	Radiolitidae			
	Hippuritidae			
	Caprinulidae	not tested		polyphyletic
	Trechmanellidae	not tested		paraphyletic
	Caprinidae			paraphyletic
	Caprinuloideidae		paraphyletic	polyphyletic
	Ichthyosarcolitidae	not tested		paraphyletic
	Antillocaprinidae	not tested		paraphyletic

Table 4. 6. Table of families and superfamilies following Skelton 2013 and their phylogenetic status (monophyletic, paraphyletic, polyphyletic, or not tested) under Skelton and Smith phylogenetic test, under Skelton 2013 presumptions on the rudist taxonomy, and under our phylogenetic analysis. Empty lines imply monophyly.

polyphyletic, *Mexicaprina* being close to Antillocaprinids, Trechmanellid and Ichthyosarcolitid taxa, *Amphitriscoelus* being sister group of a wider clade comprising also *Offneria* and *Caprinula*. We agree on the polyphyly of canals if they are understood as a unique and indivisible feature (Douvillé 1888; Mac Gillavry 1937; Skelton and Smith 2000). Skelton and Smith wrote about Douvillé's thought that 'it was already evident that pallial canals had evolved convergently in a number of unrelated groups' (Skelton and Smith 2000: 100) and, further, 'MacGillavry (1937) emphasized on the polyphyletic origins of canaliculated rudists'. However, our new character hypotheses and the derived tree suggest canals to support a substantial phylogenetic significance (fig. 4.27). All taxa showing pallial canals in both valves and distributed all around the shell are associated in the clade 16, with *Amphitriscoelus* as a sister group. This result is emphasized by the more important weight concerning myophores (19 states in characters #4, #6, #26, #27) than for pallial canals (13 states in characters #11, #12, #13, #14, #28, #29, #30, #31). Although very different in their shell construction, many families showing pallial canals that are Caprinidae, Caprinuloideidae, Trechmanellidae, Ichthyosarcolitidae, Antillocaprinidae, Plagioptychidae, and some canaliculated Radiolitidae as *Sabinia* may heritate this feature from a common ancestor. All families amongst Hippuritidina where gathered into two superfamilies raised by Skelton (2013): the superfamily Caprinoidea, assumed

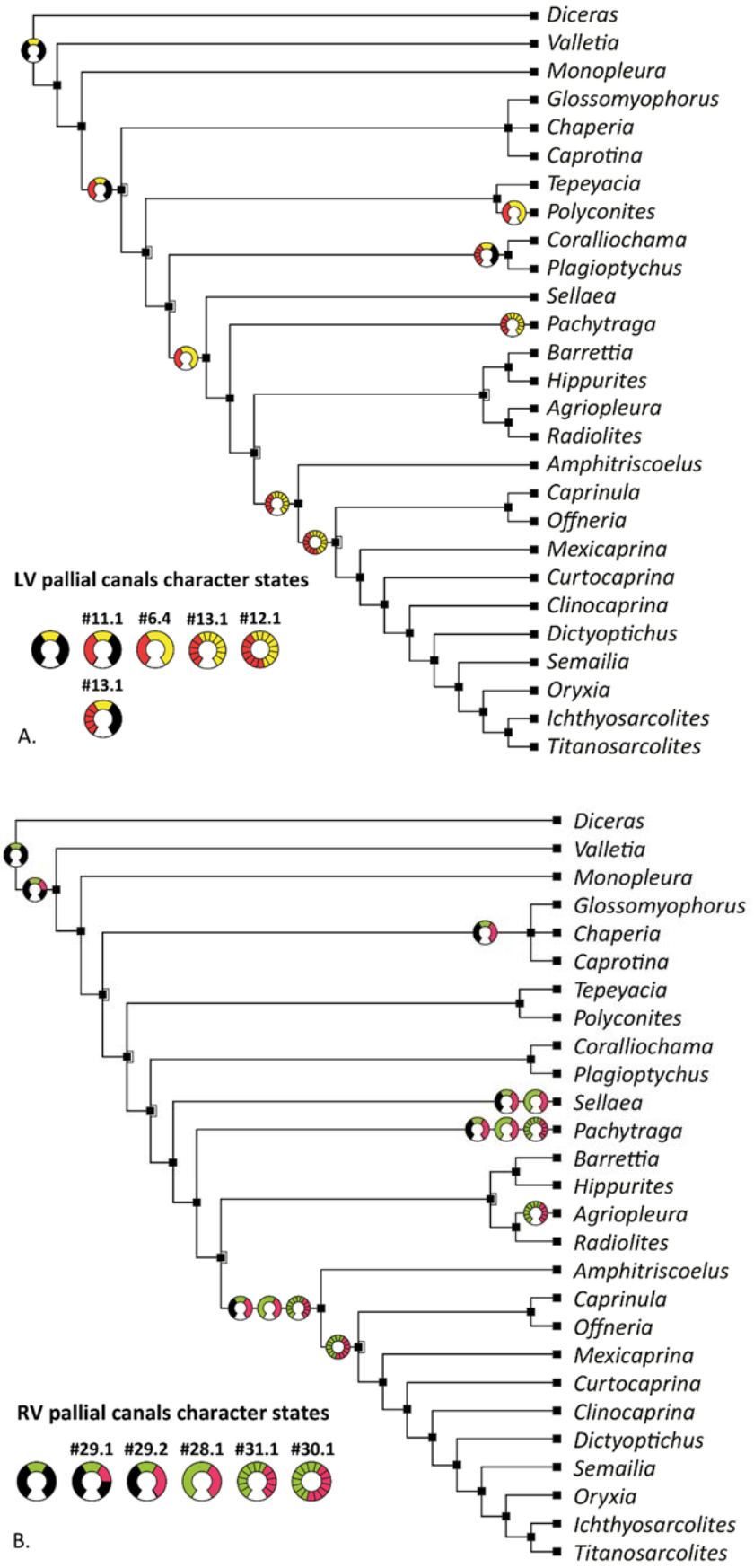


Figure 4. 27. Probable position of pallial canals character states on the phylogeny. A – Left valve pallial canals. B – Right valve pallial canals.

monophyletic, and the superfamily Radiolitoidea, assumed paraphyletic (gathering all non-Caprinoidea families). Our analysis rejects both superfamilies as monophyletic.

The following section discusses the tree clade by clade. It begins with its root dated to the Oxfordian+, representing the suborder Hippuritidina gathering all rudists attached to the substratum by their LV (synapomorphy of the group). *Diceras* is placed as the sister group of all other rudists (as in Skelton and Smith 2000), as is assumed to exhibit only ancestral character states. By tracing the key synapomorphies of the cladogram, a phylogenetic position is suggested for the rudist genera not represented in the tree. The result is displayed in Figure 4.28 and has only an exploratory value since it has not been tested by cladistic analysis.

5.1.1. Clade 1

This first clade dated from Kimmeridgian+ is consistent with all previous phylogenies, with *Diceras* as a sister group of all other rudists attached by their LV (Smith 1994; Steuber 1999; Stone and Telford 1999; Skelton and Smith 2000). It is the most strongly supported clade with seven full synapomorphic states. Three states mark the 'inversion of the dentition', with the presence of a PT (#1.1), the flattening of the 'diceratid tooth' zone (#23.1) and the presence of a PTS (#29.1). Important new states modify drastically the myophoral organization compared to *Diceras*, with the PM-LV and the AM-LV forming a horizontal shelf (respectively #4.1 and #6.1). They are followed by the simple AT (#2.1; *Diceras* having a complex AT with two apex) and a slight tilting of the CTS/G wall (#8.1). All states supporting this clade have a RI=1.

5.1.2. Clade 2

This clade is dated from the Tithonian+. *Valletia* is the sister group of this clade. As the clade 1, this clade is consistent with previous analyses (Smith 1994; Steuber 1999; Skelton and Smith 2000). It gathers all rudists attached by their RV except Diceratidae. The family Diceratidae is paraphyletic in our analysis, *Diceras* and *Valletia* being the only genera of this family according to Skelton (2013). The clade 2 differentiates itself by the ligament that becomes vertical and perpendicular to the commissure of the valves (#37.1,

Figure 4. 28. Stratigraphic extension of 154 rudist genera of the Hippuritidina suborder, from Oxfordian to Maastrichtian, modified from Steuber et *al*. (2016). Genera are grouped according to the clade to which they probably belong (or clades for the probably polyphyletic genus *Offneria* and *Caprina*).

Continuation of Figure 4.28

RI=1), and becomes invaginated in a ligamentary cavity linked to the outside by a ligamentary groove. The small close costulations on the shell external side (#39.1, except *Clinocaprina*, *Oryxia*, *Curtocaprina*, *Ichthyosarcolites*) are also characterizing this clade.

5.1.3. Clade 3

The clade 3 sheds light on two novel synapomorphies: the inward tilting of the PM-LV that becomes surrounded by the CTS which extends itself posteroventrally (#4.2, RI=1), and the straight CT (#25.1, except *Polyconites*, *Pachytraga*). The presence of a PAC (#11.1) is probably also a synapomorphy of this clade, but is rejected by its absence from the clade 12, which may be a loss in a highly derived clade. The age of origin of the cklade 3 is Hauterivian+. It is a paralogous clade from which diversify two lineages that are the clades 4 and 6. *Monopleura* is the sister group of the clade 3. *Glossomyophorus* is the other taxon that belongs to the Monopleuridae. The relative position of *Monopleura* and *Glossomyophorus* implies polyphyly of Monopleuridae. *Glossomyophorus* might be rejected of it to the benefit of the family Caprotinidae (see clade 4). In regard to their morphology, the genus *Debrunia*, *Mathesia*, *Bicornucopina*, *Gyropleura*, *Oedomyophorus*, *Pseudopetalodontia*, *Petalodontia*, *Araeopleura*, *Simacia*, *Arnaudia* and *Artigesia* probably belong to the clade 2 (fig. 4.28) but not to the clade 3. All of these genera are presently included in Monopleuridae, excepted *Oedomyophorus* that is assigned to Caprinuloideidae because of the presence of an 'outward-facing LV posterior myophore, situated directly on the posterior valve margin' (Skelton 2004: 40) hypothesized as synapomorphic to *Oedomyophorus* and *Retha*. However, the PM-LV of *Oedomyophorus* is not included in the CTS by tilting of the myophoral plate (#4.2), and it does not possess the PAC of *Retha* (#11.1). The similarity between *Oedomyophorus* PM-LV and those of Caprinuloideidae is probably not synapomorphic, and may be an inflated myophoral shelf converging with those of Caprinuloideidae, all other characters being those of this clade 2 (Skelton 2004). The phylogentic arguments support the removal of *Oedomyophorus* from the family Caprinuloideidae. Taxa probably belonging to the clade 3 are the Polyconitid genera *Douvillelia*, *Praecaprotina*, *Jerjesia*, and *Magallanesia*, and the Caprinulidae *Himeraelites*, (fig. 4.28).

5.1.4. Clade 4

The clade of age Barremian+ gathers the family Caprotinidae with *Glossomyophorus*. The clade supports the first placement of *Glossomyophorus* within Caprotinidae (Masse *et al.* 1984; Caprotinidae *sensu* Skelton 1978). In the original publication, *Glossomyophorus* was placed in the Caprotinidae because of '- fixation par la valve droite, - ligament invaginé, - absence de *tabulae*, de canaux ou de cellules dans le test'. The last two features are absence, making possibly the family paraphyletic. We were not able to find synapomorphies for the family Caprotinidae sensus Skelton 2013a in our analysis. However, our cladogram allows to redefine the Caprotinidae on a cladistic basis with positive features; thus, the clade 4 is characterized by the PM-LV on a blade projecting into the RV (#4.3), the acquisition of a PAC (#11.1) and the inclusion of the PM-RV in the extended PTS (#26.3). It is also characterized by a PTS extended on the posteroventral side (#29.2), marking a longer cavity than in *Monopleura* for example. The state #10.1 is only possessed by *Glossomyophorus* but is probably a convergence with the clades 13 and 20, marking a more dorsal insertion of the posterior side of the CTS/G wall. There is a clue to understand the difference between this state in *Glossomyophorus* and in taxa from clades 13 and 20: whereas in those clades, the CTS becomes a round cavity, the cavity is crooked in *Glossomyophorus* in an anterodorsal/posteroventral axis, as in Plagioptychidae, Caprotinidae, and Polyconitidae.

We agree with previous thought of Skelton (1976) who recognizes the Caprotinidae *sensu* Dechaseaux (1969: N784) to be a 'polyphyletic assemblage […] merely characterized by possession of accessory cavities', as Dechaseaux included *Caprotina*, *Chaperia*, but also *Horiopleura*, *Pachytraga*, *Polyconites*, *Praecaprotina*, *Retha*, and *Sellaea* in the family Caprotinidae. The Caprotinidae are probably including *Glossomyophorus*, *Praecaprotina*, *Caprotina* and *Chaperia*, the fossil register being continuous from Barremian to Cenomanian. The tilting of the AM-LV shelf of *Chaperia* is a probable synapomorphy of the genus, making a convergence with the clades 13, 19 (except *Curtocaprina*) and *Tepeyacia.*

5.1.5. Clade 5

No comprehensive synapomorphy characterizes the clade 5. This clade would probably not remain stable in further phylogentic analyses. It is a paralogous clade leading to the diversification of the clades 7 and 8. It gathers together the families Polyconitidae, Plagioptychidae, Caprinulidae, Caprinidae, Radiolitidae, Hippuritidae, Caprinuloideidae, Ichthyosarcolitidae, Antillocaprinidae and Trechmanellidae.

5.1.6. Clade 6: Polyconitidae

The Polyconitidae, erected by Mac Gillavry (1937), is retrieved in our analysis as the Aptian+ clade 6 made of *Polyconites* and *Tepeyacia*. *Tepeyacia* was found as a sister group of Hippuritidae by Skelton and Smith (2000: 121) by sharing pillars. We coded the pillars as a common feature to *Tepeyacia*, *Hippurites* and *Barrettia*. However, the myophoral organization of these taxa seems very dissimilar and *Tepeyacia* is found to be Polyconitidae with some particularities (#6.4, #24.1, #35.1). The Hippuritidae do not possess a PAC, contrary to *Polyconites* and *Tepeyacia* (#11.1). Our result is thus conservative in regard to the family Polyconitidae, and challenges the *Tepeyacia*-Hippuritidae hypothesis of Skelton and Smith (2000). The family Polyconitidae could be redefined from a unique character: the synapomorphy #4.4 (RI=1) which is a projecting PM blade outwardly twisted, parallel to the commissure. The Polyconitidae would include only *Polyconites*, *Horiopleura* and *Tepeyacia*. The other typical anatomical features of the Polyconotidae are a vertical CTS/G wall (#8.2; a state rejected as synapomorphy because it is convergent with other taxa sharing a capuloid LV of these shells, marking major modifications upon internal anatomy), the presence of a PAC (#11.1), a PM-RV vertical lying on the inner shell (#26.1), and tabulae in D (#33.1). The state #27.1 on AM-RV morphology (*Polyconites*) is doubtful because of the difficulty to make the difference on some taxa between a tilted shelf and a slightly thickened myophore lying directly on the shell wall. Interestingly, the difference seems to be easier to make on the LV (#4.1, #6.1) than to RV (#26.2, #27.1), the first character states being well accepted, the second being all rejected. The RV is generally subject to stretching because of its important development, modifying the internal parts and bluring the recognition of myophore homologs.

5.1.7. Clade 7

There is no synapomorphy to the clade 7. It is a paralogous clade leading to the diversification of the clades 9 and 10. It gathers together the families Plagioptychidae, Caprinulidae, Caprinidae, Radiolitidae, Hippuritidae, Caprinuloideidae, Ichthyosarcolitidae, Antillocaprinidae and Trechmanellidae.

5.1.8. Clade 8: Plagioptychidae

The clade 8 corroborates the monophyletic status of the family Plagioptychidae. Plagioptychidae has no recorded fossil occurrence earlier than the Cenomanian (Počta 1887; Pleničar *et al.* 1999; figs. 4.1 and 2). The clade is dated from the Turonian+ in our tree data, but it is however rooted in a Hauterivian+ clade 5, which would suggest a long Early Cretaceous history of the clade without recognized fossil occurrence.

Plagioptychus possesses only few derived character states and mainly ancestral ones. Plagioptychidae do not possess canals on their RV, and have a myophoral arrangement close to that of *Monopleura*, except for a vertical MP-LV (#4.2), a synapomorphy of the clade 3. They have one synapomorphy: an easily recognizable large and quadrangular AM-LV shelf (#6.2, RI=1). They are also characterized by numerous states: the AT and PT that are of equal size (#1.2), a vertical PM-LV surrounded by the CTS (#4.2), a vertical CTS/G wall (#8.2), and shelves supporting both myophores on the RV (#26.2, #27.1).

Plagioptychidae are generally grossly understood as rudists with pallial canals that evolved independently from other canaliculated rudists (Skelton 1978). We define pallial canals of the LV as the conjunction of a PAC and a CTS extended anteriorly, both filled with blades, and their ventral contact. Plagioptychidae are taxa which have a PAC (#11.1), located posteriorly to the CTS. The PAC extends on the ventral side of the shell, and is completely filled with numerous radial walls of characteristic dichotomic shape (#13.1, #13.3). The Plagioptychidae have only a PAC, and their ‘canals’ cannot be stated as homologous to those of Caprinuloideidae for example. In *Coralliochama* the PAC-canals are so extended that they invade the AM-LV myophoral shelf. Because it is still a shelf, it has not been

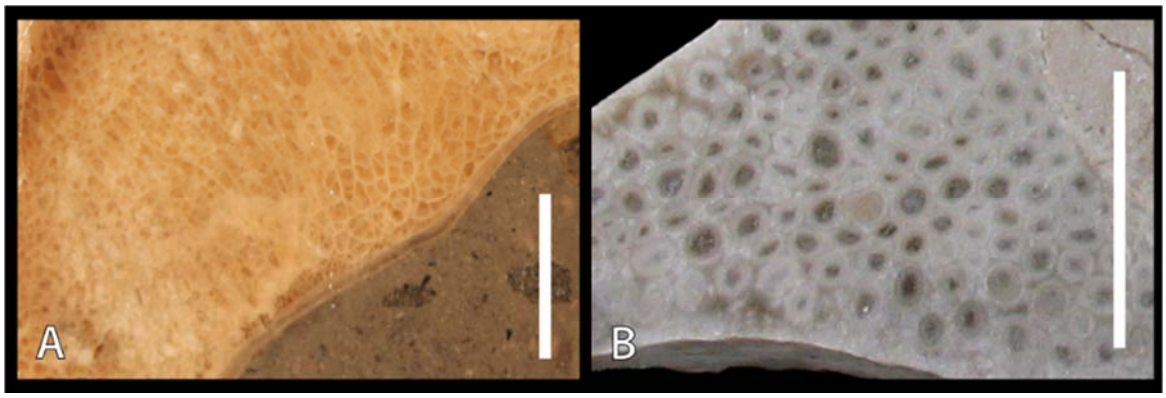


Figure 4. 29. Detail of the structuration of the aragonitic layer in A – *Coralliochama orcutti* (USNM.186665) and B – *Oryxia sulcata* (BJ.1901-HR2; picture provided by Bogdan Jurkovšek) right valves. *Coralliochama* 'pseudocanals' are not homologous to *Oryxia*'s pallial canals. Scale bars, 10 mm.

invaded by a CTS, and does not produce pallial canals *sensu stricto* (fig. 4.29). Caprina *mulleri* and *Caprina choffati* possess an AM-LV shelf and an extended PAC filled with numerous radial blades and the two species could represent plesiomorphic Plagioptychidae. Both Caprina *mulleri* and *Caprina choffati* are of Albian age, and would partially fill the temporal gap implied by an early rooting of the Plagioptychid clade.

The late appearance of Plagioptychids in the fossil record could be explained by incorrect assignment of some Caprinidae that could represent the first diverging branches of the family Plagioptychidae during the lower Cretaceous.

5.1.9. Clade 9

The clade 9 includes all taxa usually assigned to the Caprinoidea (plus other assigned to Radiolitidoidea). The superfamily could be redefined and characterized by the verticalization of the AM-LV (#6.3; convergent with *Tepeyacia*). We hypothesized the verticalization of the AM-LV to be a unique evolutionnary event. The origin of the verticalization of the AM-LV is probably linked to the anterior development of the CTS. This development is no longer visible in Hippuritidae and Radiolitidae. On other taxa (especially taxa from clade 15), the extended CTS become part of the pallial canals, and become highly modified. *Retha* probably belong to this clade, as it possesses an anteriorly extended CTS.

5.1.10. Clade 10

The clade 10, with *Sellaea* as sister group, is characterized by a completely verticalized CTS/G (#8.2). However, the character state is also shared by Polyconitids and Plagioptychids. It is probably a synapomorphy of the clade 5, with exclusion of *Sellaea*.

Pachytraga diverge from the clade 11 that include *Offneria rhodanica*, assuming the Caprinidae polyphyletic. The usual synapomorphy of Caprinidae (Chartrousse 1998) of the projecting PM-RV blade being rejected (clades 18, *Pachytraga*) to the benefit of other characters neglected in modern phylogenies, such as pallial canals evolutionary stages. *Pachytraga* exhibits a first stage of canal developement, with the presence and extension of cavities filled with some radial canals that occupies the main part of the shell on both valves (fig. 4.27). This assemblage of character states makes *Pachytraga* far from *Offneria rhodanica* and to other Caprinidae with complete pallial canals on both valves (clade 15).

5.1.11. Clade 11

The clade 11 is poorly supported, only by an assemblage of triplets that came from the characters, but no exclusive synapomorphy. It is a paralogous clade leading to the diversification of the clades 12 (Hauterivian+, and made of Radiolitidae plus Hippuritidae) and the clade 15 (Barremain+, and including Caprinuloideidae, Ichthyosarcolitidae, Antillocaprinidae and Trechmanellidae, *Caprinula*, *Offneria*). The clade 11 could be related to a clade found in Skelton and Smith (2000) gathering Radiolitidae (*Chiapasella*, *Durania*, *Osculigera*, *Pseudosabinia*, *Sauvagesia*, *Eoradiolites*, and *Agriopleura*), Ichthyosarcolitidae (*Ichthyosarcolites*), Antillocarpinidae (*Antillocaprina*) and Trechmanellidae (*Dictyoptychus*), which is seen as a similarity in their myocardinal arrangement. Our clade is larger, including Hippuritidae, Caprinuloideidae, *Offneria*, and *Caprinula*. The high amount of homoplasy in this area of the cladogram is critical. Clade 12 supports the greatest number of character states, but the clade 17 is responsible for rejection of numerous of these character states. These states are, amongst others, a vertical and projecting PM-LV blade outwardly directed (#4.5, clades 13, 19, and *Amphitriscoelus*), and a PM and PT separated by a gap, giving a tooth-like or blade-like aspect (#5.1, clades 13 and 22), but also a CTS/G wall excluding the PM-LV (#10.1, clades 13 and 20;

equivalent to the 'symetrization' of the myocardinal apparatus in MacGillavry 1937), a conical shaped CT (#24.1, clades 13 and 17), and a ligamentary cavity dorsal from CT and CTS (#38.1, clades 13 and 23).

5.1.12. Clade 12

One of the two branches rooted into the clade 11 is the original association of Radiolitidae and Hippuritidae. However, the clade 12 associates Radiolitidae and Hippuritidae. It is one of the best supported clades of the phylogeny, with 12 character states, two of which being synapomorphic equal PT and AT (#1.2), a conical CT (#24.1), a vertical PM-LV blade, outwardly directed (#4.5), a vertical AM-LV blade, outwardly directed (#6.4; the LV myophoral organization is constant amongst this clade, except for some variability in shape and thickness), a PM situated outside from the CTS (#10.1) and separated from the PT by a gap (#5.1, except *Agriopleura*), tabulae in D (#33.1, except *Agriopleura*), a thick outer calcitic shell layer (#36.1, except *Barrettia*), and a ligamentary insertion dorsal from the CT and CTS (#38.1, except *Agriopleura* and *Barrettia*). The two synapomorphies supporting this clade are the division of the CTS into two cavities by a strong ventral development of the ligament (#9.1), and a completely uncoiled LV (#22.1, RI=1; other taxa may show uncoiled valves, but there is always a slight torsion of the shell that remains as in *Dictyoptychus*). The question of the origins of Radiolitidae and Hippuritidae are of critical importance, as they are the most diverse families counting for half the generic diversity in Hippuritidina (59 genera for Radiolitidae and 17 genera in Hippuritidae). Our results challenge the scenario of the monopleurid ancestry of the Radiolitidae with *Monopleura* as sister group (Skelton and Smith 2000; Masse and Fenerci-Masse 2010) because of derived character states amongst myophores (clade 11).

Hippuritid morphology and most of its specific character states can be explained by the torsion of the Radiolitid shell through infolding of radial bands, and leading to strongly infolded structures called pillars (fig. 4.30). The infoldings P1 and P2 compress the posterior shell wall because of the general cavity bending. The posterior myophore is reduced from a blade-like to a tooth-like myophore because it becomes stranded between the P1 and the PTS. In some Hippuritids, as in *Hippurites radiosus*, the PM-LV seems to

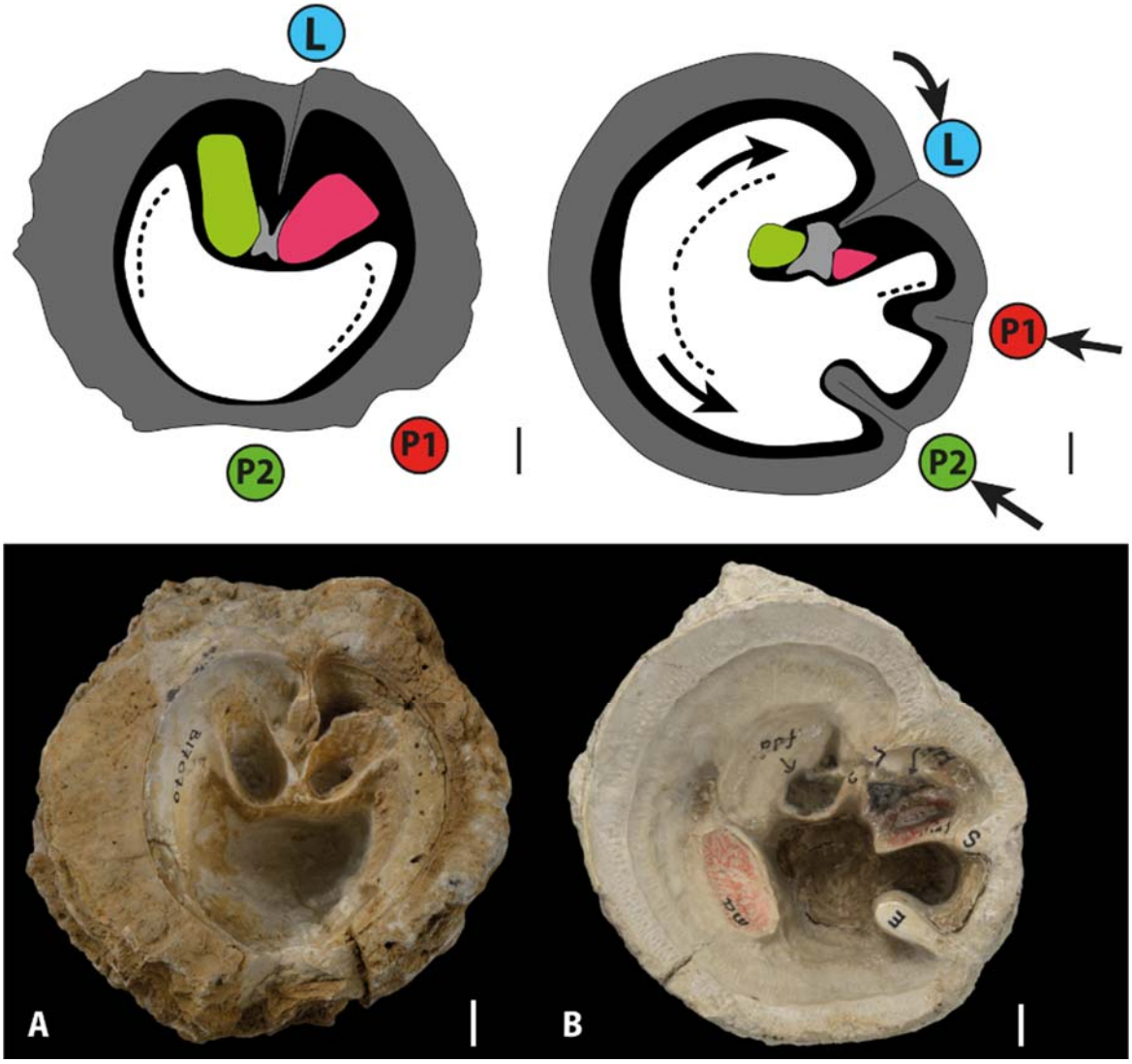


Figure 4. 30. Sketch illustrating *Hippurites* and *Radiolites* kinship (RV). The siphonal bands of *Radiolites* develop into pillars in *Hippurites* (P1 and P2). These infoldings are the source of numerous structural deformations in Hippuritid shells: (i) the ligament moves posteriorly, (ii) the vertical posterior muscle becomes stranded between the pillar P1 and the cardinal apparatus, and (iii) the anterior muscle stretch itself and become divided into two parts. A – *Radiolites radiosus* (MNHN.F.J09021). B – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm.

be inserted in the RV as a blade in a scabbard that in fact is part of the distorted general cavity D (the hypothesis #26.5 of a cavity homologous to *Dictyoptychus*, *Titanosarcolites* or *Ichthyosarcolites* PMC is rejected, as the hypotheses mady by Götz following observations on juvenile Hippuritids; Götz 2003). The pillar P1 is developed up to connect with the hinge extremity forming a cavity that looks like a socket. The wall separating this neoformed socket from D is also generally smaller than PTS and ATS. While the shell is

compressed on the posterior side, it is stretched on the anterior side. The anterior muscle stretching leads to elongation of the AM that is cut in two parts. The ligament moves toward a posterodorsal position through the shell torsion. The ligament movement opens an empty space between itself and the AM, a space that was always described as an 'accessory cavity' near to the ligament in Hippuritidae. It results that the anterior and posterior myophore shapes, the ligament position and the differentiation of new cavities can all be explained by the single and simple change in the shell conformation: the formation of the pillars. Thus, Hippuritidae are probably highly modified Radiolitids with pillars and a loss of the shell cellular microstructure The emblematic family Radiolitidae would be paraphyletic, the monophyly found in our analysis being those of a minor Radiolitid clade (see next part, clade 13).

Hippuritids are Radiolitids with deformed shell due to the development of pillars, which implied the reduction of the PM, the stretch of the AM that divides in two parts, and the posterior shift of the ligament crest. The pillars are not homologous to the infolded radial bands of *Tepeyacia* because of numerous other contradictory states (see clade 6). Pillars are probably differentiated from Radiolitid 'pseudopillars' as seen in *Lapeirousia* (Coniacian-Maastrichtian) and *Osculigera* (Campanian). The complex celluloprismatic mesostructure found in the outer calcitic shell layer is generally used as synapomorphy of Radiolitidae. However, it is present is some Radiolitidae but not all (i. e. *Agriopleura*), and may be present is few Hippuritidae species (i. e. *Praebarrettia corrali* in MacGillavry 1937, *Carribea muellerriedi* and *Praebarrettia sparcilirata* illustrated in Pons *et al.* 2013; a homology rejected by the authors, but that is worth further examination), supporting our hypothesis of a Radiolitid origin of Hippuritidae.

5.1.13. Clade 13: Radiolitidae

The clade 13 corresponds to the family Radiolitidae, found monophyletic herein. However, the synapomorphy of dorsally costulated or infolded teeth (#3.1) does not apply to every Radiolitidae, and the complex structure of the celluloprismatic calcitic layer cannot be a character state in our analysis, because *Radiolites radiosus* is the only taxa to have it

in our taxonomic sampling. Further analyses should consider a wider diversity of Radiolitidae for a formal testing of their synapomorphies.

In the literature, *Agriopleura* is hypothesized to be the sister group of all other Radiolitidae (Skelton and Smith 2000; Masse and Fenerci-Masse 2014) because of its thick outer calcitic shell layer. *Agriopleura* also shares with Radiolitidae a PM-RV vertical and inserted directly in D (#26.1), which is a probable state supporting the clade 12, the PM-RV of Hippuritids being probably also inserted on the posterior wall of D, and this wall being strongly infolded thereafter.

5.1.14. Clade 14: Hippuritidae

The Hippuritidae is found monophyletic in our analysis, corresponding to clade 14 and dated to Turonian+ (with a Radiolitid taxa as probable sister group). The Hippuritidae is supported by an important number of derived states that makes it easily distinguishable. Three synapomorphies with RI=1 supports the clade 14: a stretchered AM divided in two parts (#6.5, consequence of infoldings), making the Hippuritidae the only trimyarian rudist taxa, and the complex system of calcitic canals and pores (#20.1 and #21.1; see Schumann 2000). The phylogenetic position of *Praetorreites* and *Torreites* remains to resolve. They could either be a sister group of all other Hippuritidae or nested within the Hippuritidae following a secondary loss of the canal/pore system. Other states support the Hippuritid clade: the presence of two pillars (#35.1), a vertical PM-LV blade facing the PT and stranded perpendicular to the shell wall (#4.6), and the PM-LV projecting in a scabbard cavity in the RV (#26.5). *Vaccinites fontalbensis* was hypothesized as sister group of all other Hippuritidae (Philip 1978).

5.1.15. Clade 15

The clade 15 (Barremian+) is moslty supported by pallial canal character states. It gathers all taxa having pallial canals in both valves except on their ventral side, traducing the presence of a PAC (#11.1), an extended CTS (#6.3, clade 9), PTS (#29.2, except *Oryxia*) and ATS (#28.1), all being filled with radial blades (#13.1 and #31.1; figs. 4.21, 4.27): Caprinuloideidae, Ichthyosarcolitidae, Antillocaprinidae, Trechmanellidae, *Caprinula* and

Offneria rhodanica. The character states are shared by other genera: three Caprinuloideidae (*Conchemipora*, *Pantojaloria*, *Planocaprina*), three Caprinidae (*Offneria*, *Pseudocaprina*, *Praecaprina*), and a Caprinulidae (*Neocaprina*; fig. 4.28). There is also a PM-RV on a myophoral shelf (#26.2), except *Curtocaprina* in which both myophores are inserted directly on D. This clade is surprisingly similar to the 'Caprinidae' sensu Dechaseaux (1969) and Damestoy (1971), except the placements of *Plagioptychus*, *Coralliochama* and *Mitrocaprina*. This clade was assumed polyphyletic in Skelton and Smith analysis (2000) with reason because they rejected the synapomorphy 'pallial canals'. However, they neglected the phylogenetic significance of rudist cavities, for which we provide refinement, and show them to be significant for definition of several clades, including well-established groups. The myophores (#4, 6, 26, and 27) support the clade 1, 4, 7, 9, 13, 14, 15, 18, 19, 20, 23. Our cladogram can be seen as a trade way between Dechaseaux (1969) that overemphasize the role of cavities and canals on rudist classification, and Skelton and Smith (2000) that neglected it to the benefit of myophoral characters.

5.1.16. Clade 16

The clade 16, with *Amphitriscoelus* as sister group, mark the transition to completely canaliculated shells without interruption on any valve (#12.1, #30.1, both synapomorphic with RI=1). On this basis, the clade should also include the Caprinuloideidae (*Huetamia*, *Coalcomana*, *Caprinuloideidea*, *Texicaprina*, *Youngicaprina*, *Neokimbleia*, *Kimbleia*, *Jalpania*, *Mexicaprina*, *Cobbanicaprina*) and the Caprinulidae *Caprinula*. It is a paralogous clade from which the clades 18 and 20 differentiate.

The grouping is also supported by: the pallial canals from the LV specialized with two morphologies (#13.2, except *Offneria*, *Dictyoptychus*) as in the RV (#31.2, except *Offneria*, *Curtocaprina*, *Dictyoptychus*), and a large row of polygonal canals originating from the CTS (#14.1, except *Offneria*, *Dictyoptychus*, *Clinocaprina*, *Titanosarcolites*). Other supporting states are a horizontal shelf for the AM-RV (#27.1, except *Oryxia*, *Dictyoptychus*, *Curtocaprina*, *Clinocaprina*, *and Ichthyosarcolites*) and a conical CT (#24.1).

5.1.17. Clade 17

This small clade links *Caprinula* with *Offneria rhodanica* (we specify that according to characters #12.1 and #30.1, the genus *Offneria* is suspected of paraphyly). The tabulae in the CTS (#19.1) are probably converging with *Titanosarcolites*. Clade 17 is also supported by a projecting blade bearing the PM-RV (#26.4). This state is used in the current classification of rudists as synapomorphy of the Caprinidae family (Chartrousse 1998; Skelton and Masse 1998; Steuber 1999; Skelton and Smith 2000; Sano *et al.* 2014). Our analysis shows that this character state is informative only in conjunction with the characters on pallial canals. A possible phylogenetic redefinition of the Caprinidae family will probably have to take into account the possession of pallial canals all around the RV and LV. The myophoral organization of *Pachytraga* is a convergence according to this hypothesis. According to this definition, the Caprinidae includes species of the genus *Offneria* which do not possess a ventral interruption of the pallial canals (i. e. *Offneria arabica*, but not *Offneria interrupta*), the genus *Caprina* (except *Caprina mulleri* and *Caprina choffati* which probably belong to the family of Plagioptychidae), *Sphaerucaprina*, *Orthoptychus*. The state #15.1 (circular canal in the AT) is rejected, the grouping of *Amphitriscoelus* with *Caprinula* being impossible with regard to their PM. These two taxa both carrying a posterior myophoral blade, but not homologous, with Amphitriscoelus carrying the LV (#4) and Caprinula carrying the RV (#26).

5.1.18. Clade 18

The clade 18 (Albian+) gathers Antillocaprinids, Trechmanellids, Ichthyosarcolitids and *Mexicaprina* (Caprinuloideid). It is the sister group of the clade 17, marking the longest serie of ortholog clades, from 20 to 25. The clade 18 can be compared with the clade found by Mitchell (2013) linking Caprinuloideidae, Ichthyosarcolitidae and Antillocaprinidae on pallial canal characters. The main difference is due to the lack of Trechmanellidae in the 2013 Mitchell's analysis. The only synapomorphy of clade 18 is the tilting of the AM-RV in the ATS (#27.2), that is however only present in *Mexicaprina* and *Titanosarcolites*. The AM-RV of other taxa from this clade lies on the inner shell in D. The clade is also characterized by an outwardly directed PM-LV blade (#4.5) as observed in *Amphitriscoelus*

and taxa from the clade 12, by an outwardly directed AM-LV blade (#6.4, except *Curtocaprina*), and a PM-RV inwardly oriented included in the PTS (#26.5, except *Curtocaprina* and *Oryxia*).

5.1.19. Clade 19

The clade 19 groups together Antillocaprinidae, Ichthyosarcolitidae and Trechmanellidae, and is similar to a clade found in Skelton and Smith (2000), although they included the Radiolitidae on the basis of myophoral character states. Our analysis group found a Caprinuloideidae (*Mexicaprina*) as sister group because it shares similar pallial canals (Mitchell 2013b). The presence of flanges (#40.1, except *Dictyoptychus*, *Clinocaprina*; RI=0.96) and capillary canals that invade the CT (#32.1, except *Dictyoptychus*, *Semailia*, *Clinocaprina*) are synapomorphies of the clade 19. There are two other ambiguous synapomorphies related to the loss of ligamentary structure. The loss of a ligamentary groove (#37.2) and of the ligamentary cavity (#37.3). The first change excludes *Oryxia*, and the second excludes the taxa *Clinocaprina*, *Dictyoptychus*, *Semailia,* and *Oryxia*. Inconstancy in such critical character evolution is probably due to the problematic taxa *Oryxia* and *Curtocaprina*. *Oryxia* possess a ligament included in a cavity linked to the outside by a groove. Its position as sister group of the clade 24 is due to the highly similar myophores. However, *Oryxia* is a taxa assigned to the minimum of characters, because its LV is unknown. Thus, only 16 characters are coding *Oryxia* on a total of 40. While its position away from the clade 19 is highly unlikely, a revised phylogenetic position could lead to reassess the history of the ligament. The other problematic taxa, *Curtocaprina*, has however all its states assigned without any missing element. We found no problem to understand the morphology of its LV. However, the RV of its taxa is extraordinary different from any other rudist taxa. *Curtocaprina* bears autapomorphies of critical importance: the PTS and D are directly connected, the PTS is reduced to a little notch on the posterior wall of D, the ATS is completely disconnected from D, and a crescentiform CT partially surround D. All those elements are certainly of phylogenetically significant but the understanding of the evolutionary pathways leading to such a character combination remains unclear. The loss of the ligamentary groove and cavity that certainly arise more than once – at least two times, in this clade and in the Radiolitidae, as the genus *Durania*.

A comprehensive history of the ligament would be a verticalization in the clade 1, a loss in Radiolitidae and *Curtocaprina*, an invagination in the clade 20 and a third loss in clade 24 (*Ichthyosarcolites* plus *Titanosarcolites*). The assignement of *Immanitas anahuacensis* to the clade 19 is possible because of a loss of ligament, and several arguments suggest close relationship to Antillocaprinidae (Aguilar-Pérez 2008). The morphology of *I. anahuacensis* is quite unique with the presence of a cordiform PTS, a shared feature with the Caprinuloideidae *Cobbanicaprina* and *Mexicaprina cornuta*, and the Ichthyosarcolitidae *Ichthyosarcolites coraloidea*.

The clade 19 is also characterized by a CTS-G wall connected dorsally to the PM-LV that is not included in the CTS (#10.1), leading to a small round CTS that may disappear in taxa like *Ichthyosarcolites*, a conflation of the pallial canals of the LV in contact with the myophores (#14.2, except *Clinocaprina*, *Dictyoptychus*, *Titanosarcolites*) that cannot be homologous to those of Caprinuloideidae and may be not originating from a latte extension of the CTS into first pallial canals, but probably by an enlargement of previously existing canals. Taxa from the clade 19 also have a vertical PM-RV (#26.1, RI=0.59), a rejected grouping hypothesis retrieved also in the clades 7 and 14 that is only bear by *Curtocaprina* and may be another autapomorphy of this species.

5.1.20. Clade 20

This clade has *Curtocaprina* as sister group, implying paraphyly of the family Ichthyosarcolitidae. It is supported by the presence of capillary canals in *Semailia*, *Clinocaprina* and *Titanosarcolites* (#16.1), and of tabulae in D (#33.1). No tabulae are reported so far in *Semailia*. The most original character state found for clade 20 is the synapomorphy #1.3 that gathers rudists having a PT squeezed by the ligamentary cavity. *Clinocaprina*, *Dictyoptychus* and *Semailia* have this state that may be lost in further taxa (*Titanosarcolites* and *Ichthyosarcolites*). The association of Trechmanellidae and Antillocaprinidae was assumed by Mac Gillavry (1937).

5.1.21. Clade 21

The clade 21 is the sister group of *Clinocaprina*, and includes *Titanosarcolites*, which makes the Antillocaprinidae a paraphyletic family. The Antillocaprininae *Antillocaprina* and *Rotacaprina* (*Rudicaprina* is too poorly known) are likely to form a clade with *Clinocaprina*, with a CT displaced dorsally as synapomorphy (Mitchell 2013b, character 13), probably also with the Trechmanellidae *Eodictyoptychus* that also have a squeezed PT, and have the same flattened RV myopohores. The presence of a PMC in the RV where is inserted a myophoral blade from the LV is a synapomorphy of the clade 21 (#26.1, RI=1). The character states PM and PT separated by a gap (#5.1, except *Ichthyosarcolites*), the tabulae in pallial canals of the RV (#34.1 except *Oryxia*, *Semailia,* and arising in a convergent way to *Offneria*), also support this clade. A thick calcitic outer shell layer (#36.1) is found only in *Dictyoptychus* and does not participate to the clade support, assuming indepent evolution in Polyconitidae and clade 12.

5.1.22. Clade 22

The clade 22 gathers *Semailia*, *Oryxia*, *Ichthyosarcolites* and *Titanosarcolites*. With *Dictyoptychus* as sister group, it makes the Trechmanellidae family paraphyletic. The vertical and projecting PM-LV stranded perpendicular to the shell wall and facing the PT (#4.6) found in *Semailia* also occurs in *Alencasteria*, *Stellacaprina* or *Oligosarcolites* that are members of the Parasarcolitinae. The placement of *Oryxia* is doubtful according to its ligamentary groove and cavity (see clade 19).

5.1.23. Clade 23

The clade 23 is supported only by an assemblage of triplets that came from the characters, but no comprehensive synapomorphy is characterizing the clade. This clade is probably not stable. It gathers *Oryxia*, *Ichthyosarcolites* and *Titanosarcolites*.

5.1.24. Clade 24

The clade 10 offers a response to the phylogenetic position of the problematic rudist taxon: *Ichthyosarcolites*. Several alternative hypotheses attempted to trace origins the genus (Damestoy 1971; Skelton and Smith 2000), and a family was erected to stand with its unusual morphology (Douvillé 1887; fig. 4.2). Here, *Ichthyosarcolites* is placed with *Titanosarcolites*. This clade is supported by the presence of tabulae in G (#37.1, RI=1), a feature less common than tabulae in D. The ligament evolution reaches an end, with a complete loss of both external groove and ligamentary cavity (#37.3). The grouping is also supported by tabulae in pallial canals (#18.1) and AT and PT of equal size (#1.2). The character state combination places the subfamilies Parasarcolitinae and Titanosarcolitinae in this clade: including *Stellacaprina*, *Alencasteria*, *Antillosarcolites*, *Parantillosarcolites*, *Oligosarcolites*, *Polysarcolites*, *Parasarcolites*, but also *Sawkinsonia*, *Caenosarcolites* and *Eosarcolites* that may form a clade with *Titanosarcolites* by sharing tubes, amongst other states (Mitchell 2013b).

5.2. Time calibration of the tree and historical trends in rudit diversity

Assignement of an age to the clades (fig. 4.26), shows two major diversification events in rudist history that can be related to the carbonate platform development and an increasing taxonomic richness (fig. 4.31). The first radiation (clades 3, 6, 8, 10, 11, 12, 14) is localized in the Hauterivian+, and take place during a high enhancement of carbonate platforms. The second radiation (clades 19, 20, 21, 22, 23, 24, 25) took place in Albian+, and may also be related to an important development of carbonate platforms.

Our cladogram is only a preliminary step to understand the global evolutionary history of rudists. Only 26 genera are sampled (assuming their species to be representative) on 147 in Skelton's review (2013a). We tried to use the tested synapomorphies to make a first reorganization of these genera, grouping them into clades (figure 4.28). We use the oldest known fossil occurrence to date each clade (*e.g. Amphitriscoelus* for clade 10, *Agriopleura* for clade 13, *Offneria* for clade 11, *Curtocaprina* for clade 19 or

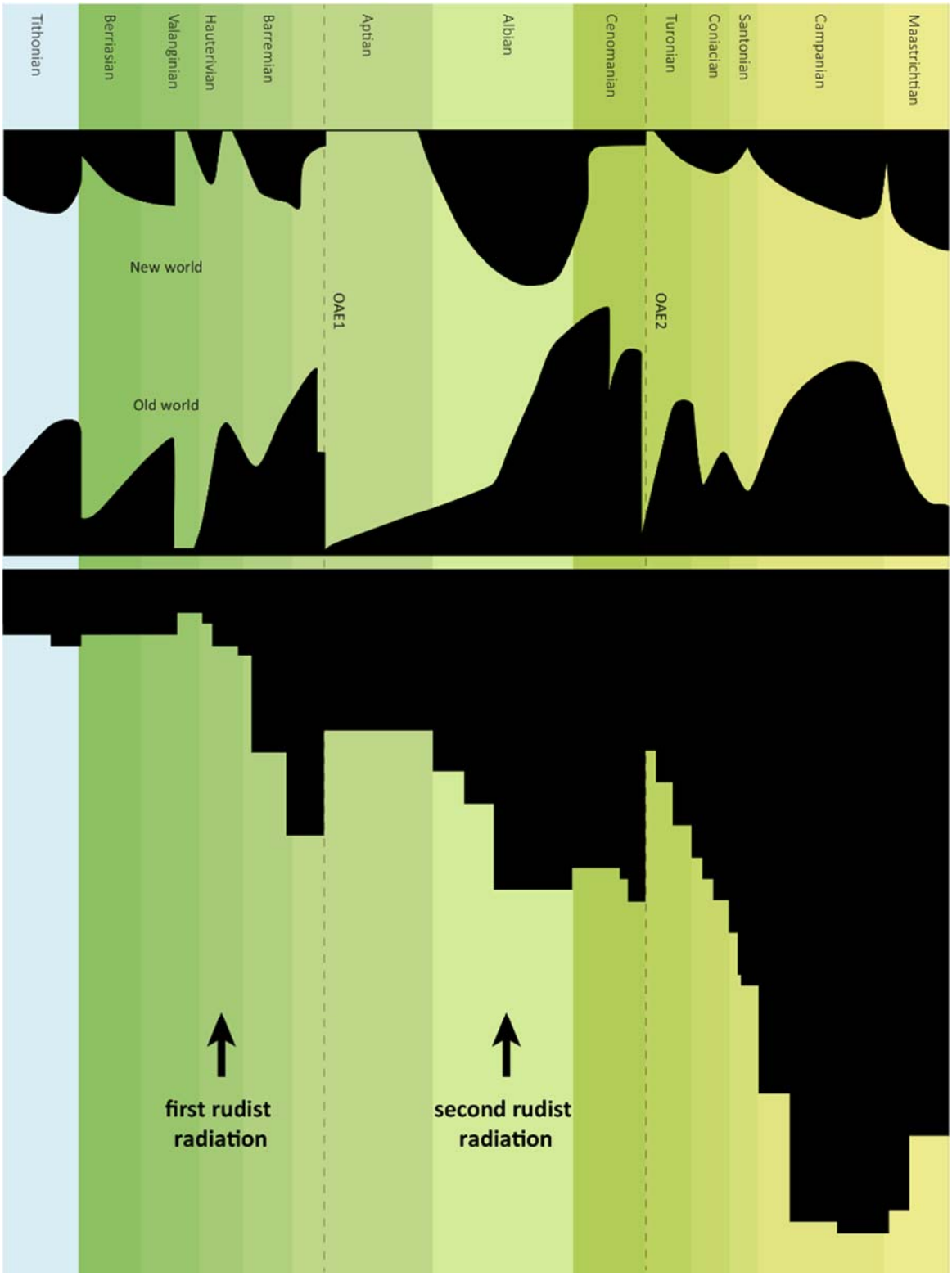


Figure 4. 31. Number of rudist genera from Oxfordian to Maastrichtian compared to the spatial extent of carbonate plateforms of the New world (North and South America) and the old World (Europe, Africa, and Arabia; adapted from Skelton 2003 and Steuber et *al*. 2016). The arrows represent the two Hauterivian and Albian rudist radiations.

Ichthyosarcolites for clade 23). We suggest that the origin of Plagioptychidae need to be studied, as a huge temporal gap separates it from the clade 5. Probable sister groups have to be found in the family Caprinidae (*Caprina mulleri* and *Caprina choffati*, see clade 8). The family Hippuritidae is probably included in the family Radiolitidae and originates from it. Refinements in character analysis and taxonomic sampling would be necessary to assess precisely where in the Radiolitidae the Hippuritidae have their roots. Finally, considering the clade 15, there is no trace of taxa between the Turonian and the Coniacian. This other huge temporal gap might be investigated more precisely in further studies. The endemicity of previous families (Antillocaprinidae: Caribbean region; Trechmannellidae: Afro-Arabian plate) might reflect a poor knowledge more than real endemism.

6. Chapter conclusion

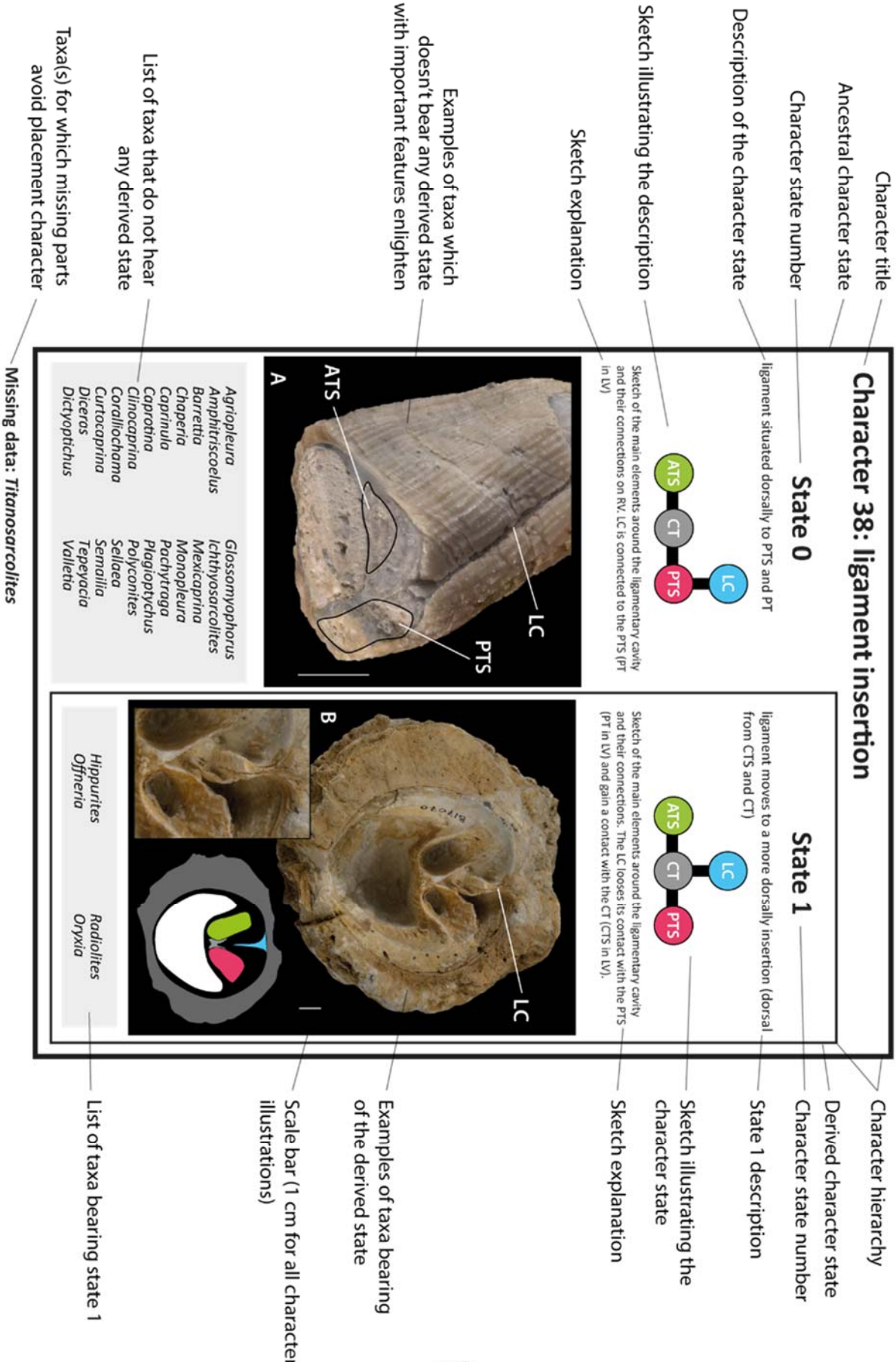
The current view of rudist systematics and evolution is grounded on the seminal cladistic analysis produced by Skelton and Smith (2000). However, many relationships are still debated and the taxonomic scheme is a mix of established monophyletic, paraphyletic and polyphyletic groups. Our main goal was to reassess rudist phylogeny, using a renovated approach of comparative anatomy designed for improvement of phylogenetic resolution. We suggest new, complex characters of critical importance for the reconstruction of the evolutionary tree of rudists, focusing mainly on cavities and myophores. The morphological comparisons between rudist families and the following phylogeny lead to a surprisingly complex evolution of the myophores. The evolutionary events in muscles history are also coeval to evolution of cavities and sockets. Our study highlights one of the main difficulties of the rudist anatomy, which is the correspondence between cavities, and their relation to pallial canals. We propose here a new way to see canals with several independent origins.

New phylogenetic relationships emerge from our cladistic analysis, and are only understandable by focusing on the new characters. The new phylogenetic scheme breaks the tenacious idea of a Monopleuridae stem group from which many families would have evolved independently, and drastically reduces the frequency of morphological convergences. The phylogeny changes to a more pectinate tree, and reintegrates the pallial

canals by refining hypothesis associated to them. We found monophyletic only the families Caprotinidae, Polyconitidae, Plagioptychidae, Hippuritidae and Radiolitidae, all other are paraphyletic, as the Diceratidae, Trechmanellidae, Caprinidae, Ichthyosarcolitidae, and Antillocaprinidae, or polyphyletic for Monopleurdiae, Caprinulidae, and Caprinuloideidae. The two largest rudist families Radiolitidae and Hippuritidae are found to be sister clades, which is an original association. The Hippuritid morphology originated from a Radiolitid form through infolding of siphonal bands into internal pillars. Another clade resolves one of the main problems of Skelton and Smith (2000), by gathering in a unique clade Trechmanellidae, Ichthyosarcolitidae and Antillocaprinidae. The main dichotomy among rudists attached by their RV, between the superfamilies Radiolitoidea and Caprinoidea, is completely rejected, both being polyphyletic.

7. Character list

Figure 4. 32. Example of hierarchical character with associated captions. Scale bars, 10 mm.


Character 38: ligament insertion
State 0
ligament situated dorsally to PTS and PT
ATS
CT
PTS
LC
Sketch of the main elements around the ligamentary cavity and their connections on RV. LC is connected to the PTS (PT in LV)
LC
PTS
ATS
A
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Pachytraga
Plagioptychus
Polyconites
Sellaea
Semailia
Tepeyacia
Valletia
State 1
ligament moves to a more dorsally insertion (dorsal from CTS and CT)
ATS
CT
PTS
LC
Sketch of the main elements around the ligamentary cavity and their connections. The LC looses its contact with the PTS (PT in LV) and gain a contact with the CT (CTS in LV).
LC
B
Hippurites
Offneria
Radiolites
Oryxia
Character title
Ancestral character state
Character state number
Description of the character state
Sketch illustrating the description
Sketch explanation
Examples of taxa which doesn't bear any derived state with important features enlighten
List of taxa that do not hear any derived state
Taxa(s) for which missing parts avoid placement character
Missing data: Titanosarcolites
Character hierarchy
Derived character state
Character state number
State 1 description
Sketch illustrating the character state
Sketch explanation
Examples of taxa bearing of the derived state
Scale bar (1 cm for all character illustrations)
List of taxa bearing state 1

7.1. Character 1

Name: posterior tooth morphology

Position: left valve

Kind: teeth

Coding: (0(1(2)(3)))

States definitions

0 dorsal zone of the PT flat

1 presence of a PT dorsally to the PM

2 PT of equal size than AT

3 PT squeezed by the ligament and elongated anteroposteriorly

States tests

1 passed – clade 1

2 rejected – clades 9, 13, 25, *Sellaea*, *Caprinula*

3 passed – clade 20 (except *Titanosarcolites*, *Ichthyosarcolites*)

Retention index

0.72

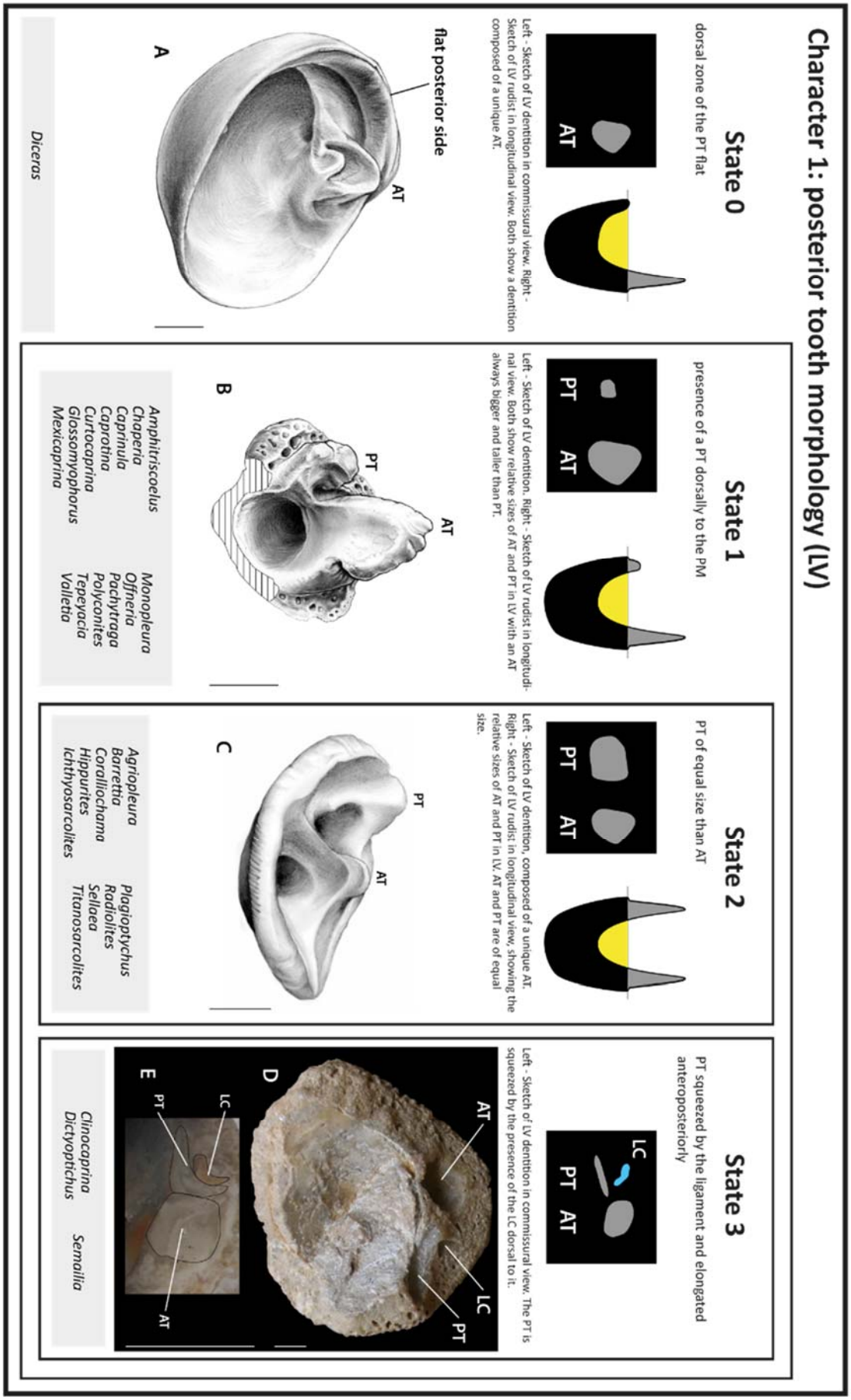
Remarks

Related to character 3b of Skelton and Masse (1998; #1.1, #1.2). Related to characters 10 (#1.3), 11 (#1.2) and 12 (#1.1) of Mitchell (2013a). Related to character 3 of Sano *et al.* (2014).

This character includes a paralogous clade and can be understood as three independent characters after free-paralogy subtree analysis (Nelson and Ladiges 1996)

following procedure described in chapter 1. The character (0(1(2)(3))) can be divided into three independent characters (0(1), (1(2)), and (1(3)).

Figure 4. 33. Hierarchical character #1. A – *Diceras arietinum* (MNHN). B – *Mexicaprina minuta* (JPMA.14097-5). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Clinocaprina lowentsami* (USNM.245242). E – *Clinocaprina lowentsami* (USNM.245241). Scale bars, 10 mm.


Character 1: posterior tooth morphology (LV)
State 0
dorsal zone of the PT flat
AT
Left - Sketch of LV dentition in commissural view. Right - Sketch of LV rudist in longitudinal view. Both show a dentition composed of a unique AT.
flat posterior side
AT
A
Diceras
State 1
presence of a PT dorsally to the PM
PT AT
Left - Sketch of LV dentition. Right - Sketch of LV rudist in longitudinal view. Both show relative sizes of AT and PT in LV with an AT always bigger and taller than PT.
AT
PT
B
Amphitriscoelus
Chaperia
Caprinula
Caprotina
Curtocaprina
Glossomyophorus
Mexicaprina
Monopleura
Offneria
Pachytraga
Polyconites
Tepeyacia
Valletia
State 2
PT of equal size than AT
PT AT
Left - Sketch of LV dentition, composed of a unique AT. Right - Sketch of LV rudist in longitudinal view, showing the relative sizes of AT and PT in LV. AT and PT are of equal size.
PT
AT
C
Agriopleura
Barrettia
Coralliochama
Hippurites
Ichthyosarcolites
Plagioptychus
Radiolites
Sellaea
Titanosarcolites
State 3
PT squeezed by the ligament and elongated anteroposteriorly
LC
PT AT
Left - Sketch of LV dentition in commissural view. The PT is squeezed by the presence of the LC dorsal to it.
AT
LC
PT
D
LC
PT
AT
E
Clinocaprina
Dictyoptichus
Semailia
Missing data: Oryxia

7.2. Character 2

Name: anterior tooth morphology

Position: left valve

Kind: teeth

Coding: (0(1))

States definitions

0 two apex AT

1 simple AT

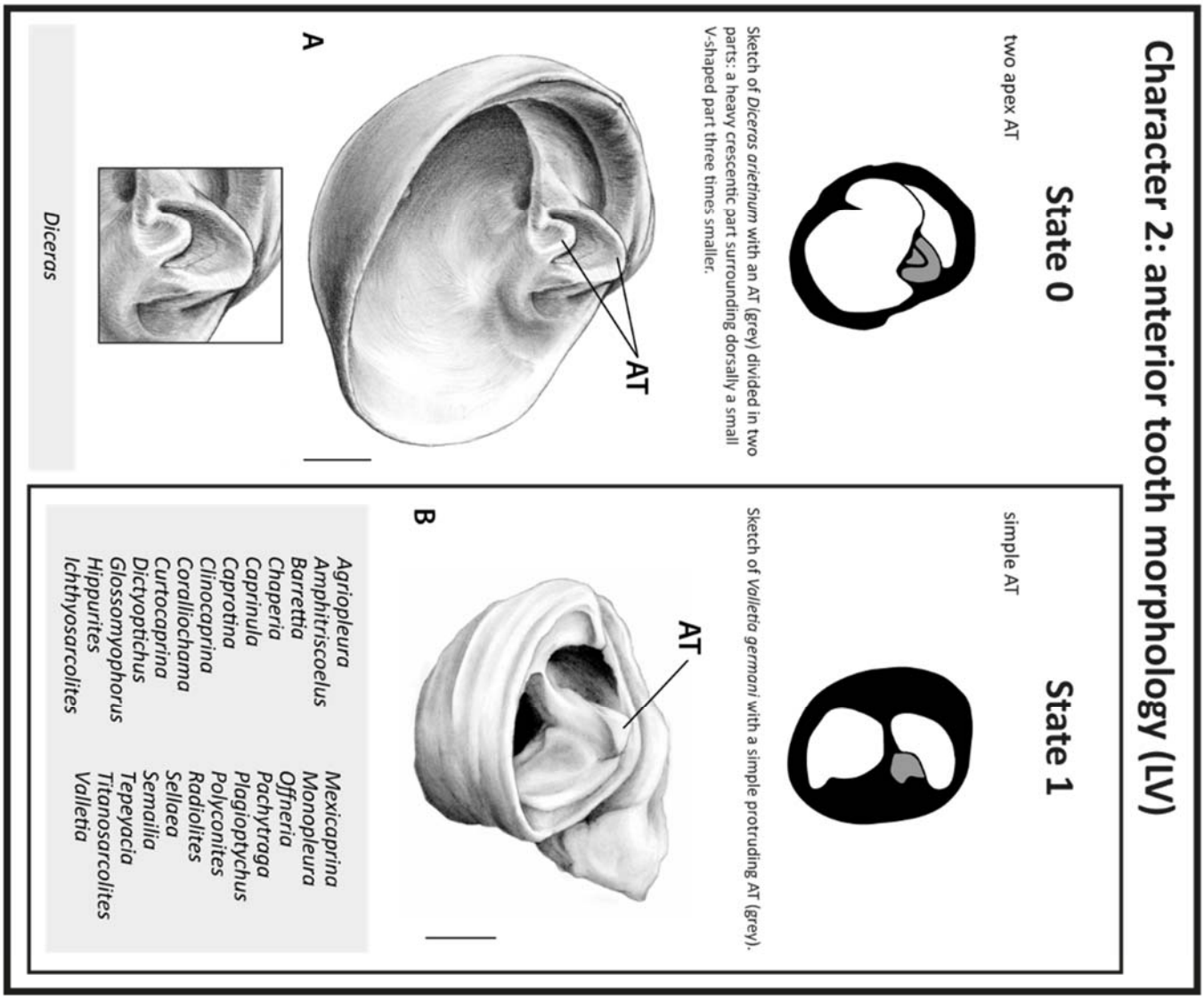
States tests

1 passed – clade 1

Retention index

1.00

Figure 4. 34. Hierarchical character #2. A – *Diceras arietinum* (MNHN). B – *Valletia germani* (MNHN.F.J05798). Scale bars, 10 mm.


Character 2: anterior tooth morphology (LV)
State 0
two apex AT
Sketch of Diceras arietinum with an AT (grey) divided in two parts: a heavy crescentic part surrounding dorsally a small V-shaped part three times smaller.
AT
A
Diceras
State 1
simple AT
Sketch of Valletia germani with a simple protruding AT (grey).
AT
B
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia
Missing data: Oryxia

7.3. Character 3

Name: teeth ornamentation

Position: left valve

Kind: teeth

Coding: (0(1))

States definitions

0 smooth teeth

1 dorsally costulated teeth

States tests

1 passed – clade 13

Retention index

1.00

Figure 4. 35. Hierarchical character #3. A – *Valletia germani* (MNHN.F.J05798). B – *Agriopleura marticensis*. C – *Radiolites radiosus* (MNHN.F.J09021; picture provided by Jean-Pierre Masse). Scale bars, 10 mm.

Character 3: teeth ornamentation (LV)
State 0
smooth teeth
PT
AT
Sketch of LV dentition in commissural view. Both AT and PT are smooth.
AT
PT
A
State 1
dorsally costulated teeth
PT
AT
Sketch of LV dentition in commissural view. Both AT and PT have small close costulations on their dorsal side.
PT
PT
B
C
PT
AT
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia
Agriopleura
Radiolites
Missing data: Oryxia

7.4. Character 4

Name: posterior myophore morphology

Position: left valve

Kind: myophores

Coding: (0(1(2(3(4)(5(6))))))

States definitions

0 PM on a blade inside G, on its posterior side

1 PM forming a horizontal shelf alongside the shell commissure

2 PM shelf included in the CTS, vertical and inserting on the posterior side of the CTS. Inwardly directed

3 vertical PM blade projecting straight into the RV

4 projecting PM blade twisted, parallel to the commissure

5 vertical and projecting PM blade becoming outwardly directed ('backflips')

6 vertical and projecting PM blade stranded perpendicular to the shell wall, facing the PT

States tests

1 passed – clade 1

2 passed – clade 3

3 rejected – clades 4, 7, 13, 19, *Amphitriscoelus*, *Pachytraga*, *Sellaea*

4 passed – clade 6

5 rejected – clades 13, 19, *Amphitriscoelus*

6 rejected – clades 15, 23 (except *Titanosarcolites*, *Ichthyosarcolites*)

Retention index

0.80

Remarks

Related partially to characters 1 and 6 of Chartrousse (1994; #4.2, 4.3, 4.5). Related to character 7 of Skelton and Masse (1998; #4.1, #4.5). Related partially to character 5 of Sano *et al.* (2014).

This character includes a paralogous clade and can be understood as three independent characters after free-paralogy subtree analysis (Nelson and Ladiges 1996) following procedure described in chapter 1.

We propose here a complex character with the recognition of seven possible variants on the PM-LV. The ordering reflects our evolution hypothesis. By comparing to *Diceras* and *Epidiceras*, we state that the ancestral condition to the character is the presence of a blade where inserts the PM-LV. The accumulation of shell material to fill this zone, elevate the PM that becomes parallel to the commissure, supported by a strong shell thickening developed from the myophoral blade called myophoral shelf. The state 2 is believed to be a modification of the state 1, where the myophore supported by a shelf become tilted inwardly (fig. 4.23). Its tilting causes a spill of the PM-LV in the CTS. The argument for their state 2 to be included in the state 1 is that the extended CTS from state 2 occupies the same shell surface as the myophore on its myophoral shelf plus the reduced CTS in the state 1). This feature can be seen in the schemes of the figure 2 from Masse *et al.* (1984: 727). Then, the state 3 is logically differentiated from the state 2 by the development of the dorsal CTS wall supporting the PM. This state has been retained as synapomorphic of the family Caprinuloideidae by Chartrousse (1998). The PM in states 2 and 3 is inwardly directed (in direction of G). This PM-LV blade can be tilted outwardly in some taxa, marking a right angle and the myophore returning to a position parallel to the commissure. Another hypothesis would propose a rise of the myophoral shelf and a cavity excavating the posterior side of the shell. However, on most Polyconitids, the myophore has clearly a blade-like aspect and never an altered protruding shelf. The state 5 is difficult to root in the character, as it presents a conformation where there is still a blade, the myophore being at the contrary outwardly directed. Two possibilities remain: (i) this blade

is convergent with the other blade, and the outward blade develop itself from the myophoral shelf in state 1, and (ii) the myophoral blade is the same, the myophore undergoing what we name a 'myophoral backflip' were the myophore moving from a side to another of the blade (state 5 rooted in state 3). This hypothesis is supported by two arguments: a single explanation for the blade development, and the observation on *Mexicaprina minuta* shell of the PM-LV on its protruding blade on the apex of the blade (fig. 4.39, state 5). On this shell, a muscle scar can be observed at a transitional stage between the inward and the outward orientation. For these reasons, we choose to include the state 5 in the state 4. The state 6 is explained by the torsion of the myophoral blade at 90 degrees, facing the ventral side of the shell.

Figure 4. 36. Hierarchical character #4, its character states intensions, and its treatment with paralogy-free subtree analysis.

Character 4: posterior myophore morphology (LV)

State 0

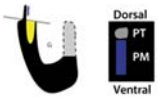
PM on a blade inside G, on its posterior side

State 1

PM forming a horizontal shelf onto the shell commissure

State 2

PM shelf included in the CTS, vertical and inserting on the posterior side of the CTS. Inwardly directed

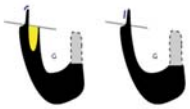


State 3

vertical PM blade projecting straight into the RV

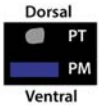
State 4

projecting PM blade twisted, parallel to the commissure



State 5

vertical and projecting PM blade becoming outwardly directed ("backflips")



State 6

vertical and projecting PM blade stranded perpendicular to the shell wall, facing the PT

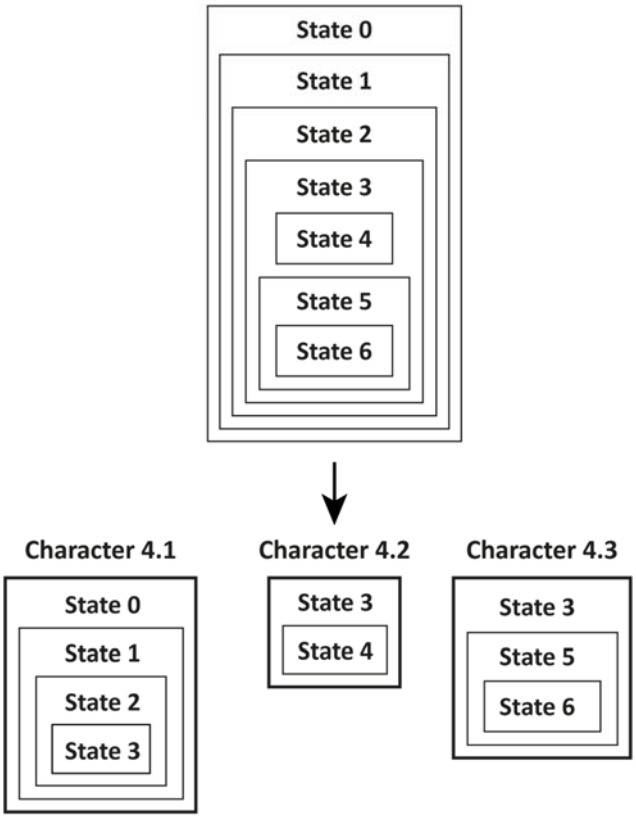


Figure 4. 37. Hierarchical character #4.1. A – *Diceras arietinum* (MNHN). B – *Valletia germani* (MNHN.F.J05798). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Caprotina costata* (MV 2003.2.5147). Scale bars, 10 mm.

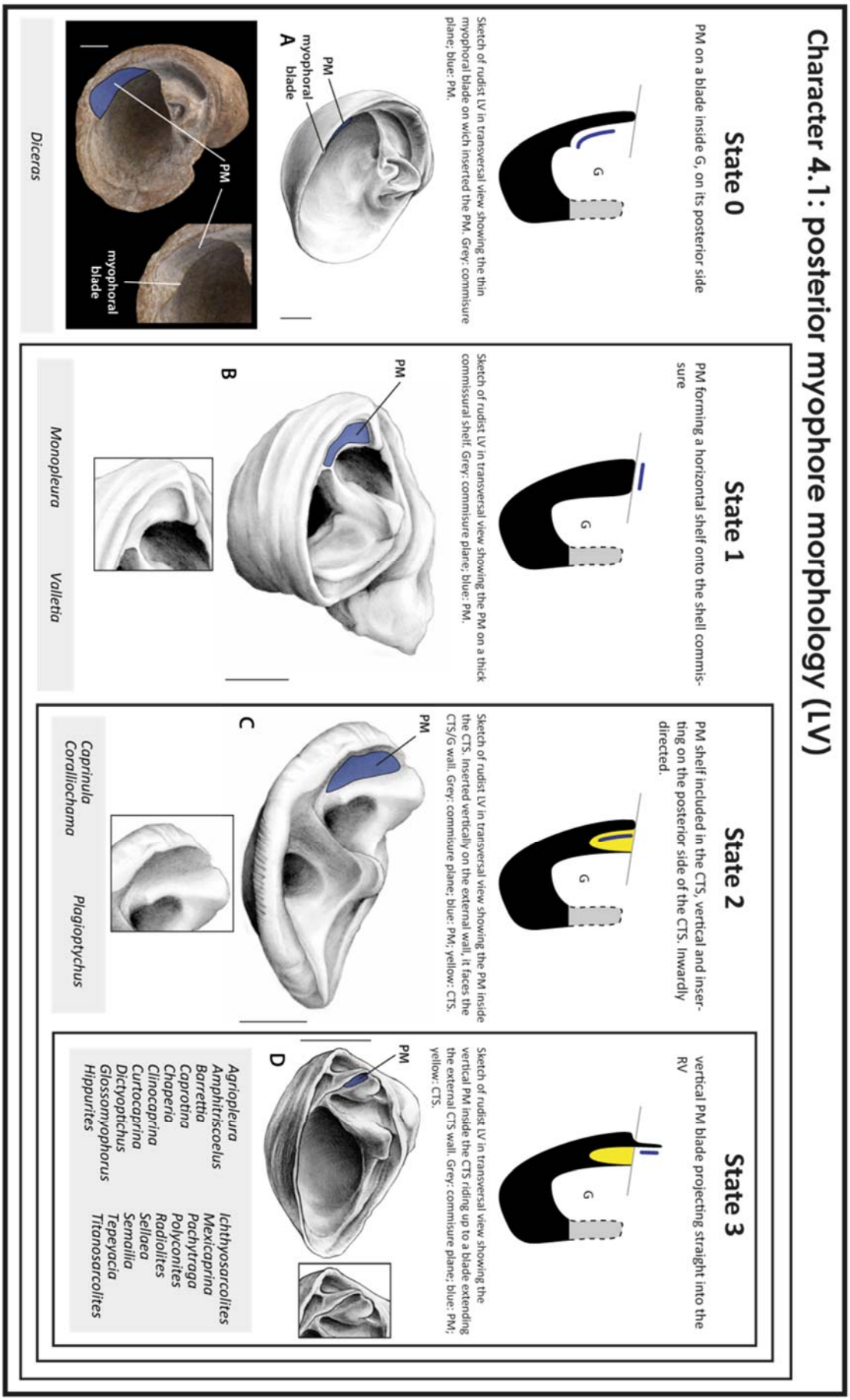

Character 4.1: posterior myophore morphology (LV)
State 0
PM on a blade inside G, on its posterior side
G
Sketch of rudist LV in transversal view showing the thin myophoral blade on wich inserted the PM. Grey: commisure plane; blue: PM.
PM
myophoral blade
A
PM
myophoral blade
Diceras
State 1
PM forming a horizontal shelf onto the shell commissure
G
Sketch of rudist LV in transversal view showing the PM on a thick commissural shelf. Grey: commisure plane; blue: PM.
PM
B
Monopleura
Valletia
State 2
PM shelf included in the CTS, vertical and inserting on the posterior side of the CTS. Inwardly directed.
G
Sketch of rudist LV in transversal view showing the PM inside the CTS. Inserted vertically on the external wall, it faces the CTS/G wall. Grey: commisure plane; blue: PM; yellow: CTS.
PM
C
Caprinula
Coralliochama
Plagioptychus
State 3
vertical PM blade projecting straight into the RV
G
Sketch of rudist LV in transversal view showing the vertical PM inside the CTS riding up to a blade extending the external CTS wall. Grey: commisure plane; blue: PM; yellow: CTS.
PM
D
Agriopleura
Amphitriscoelus
Barrettia
Caprotina
Chaperia
Clinocaprina
Curtocaprina
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Pachytraga
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Missing data: Oryxia

Figure 4. 38. Hierarchical character #4.2. A – *Caprotina costata* (MV 2003.2.5147). B – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). Scale bars, 10 mm.

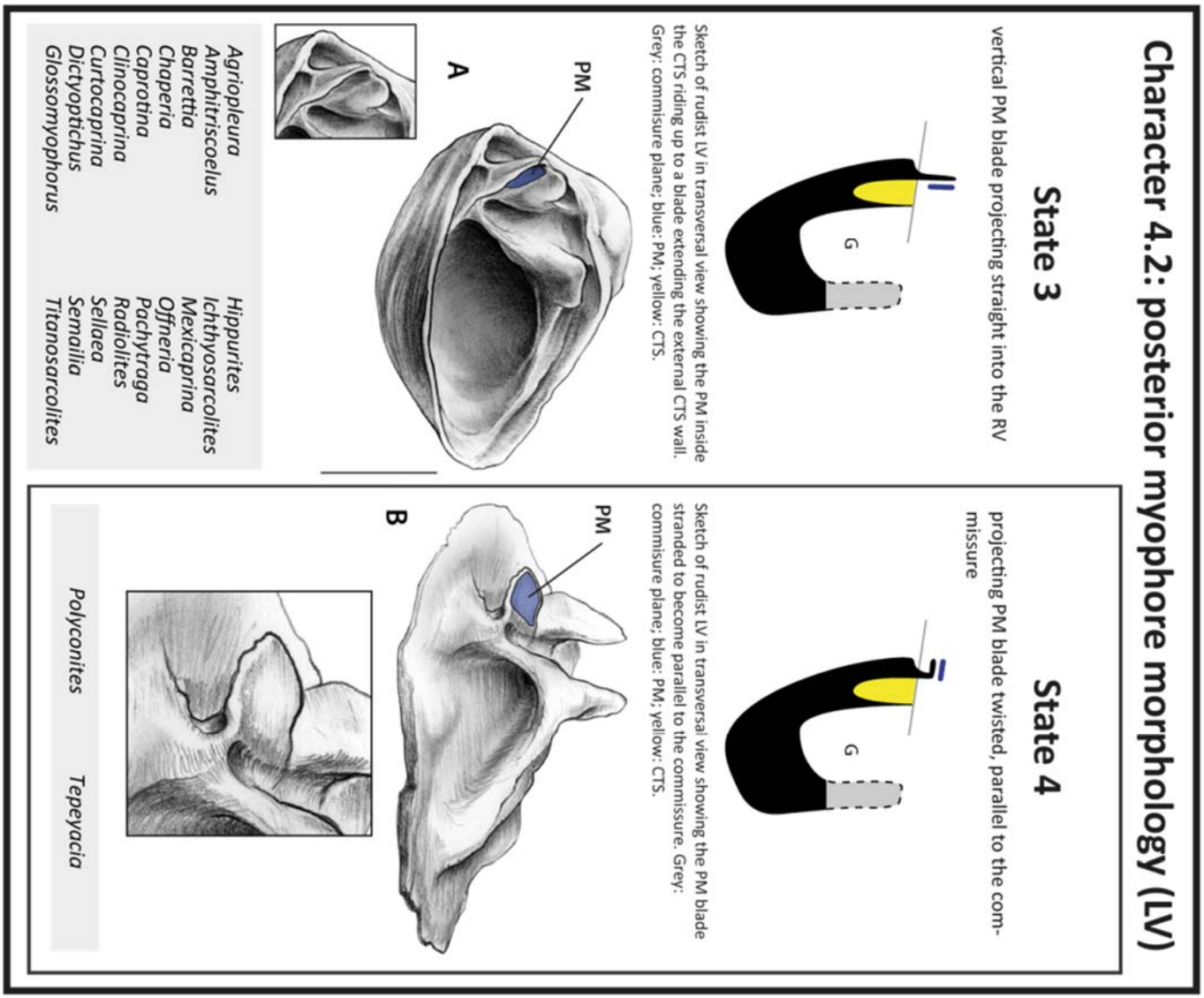
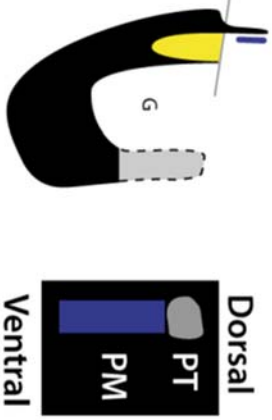

Character 4.2: posterior myophore morphology (LV)
State 3
vertical PM blade projecting straight into the RV
G
Sketch of rudist LV in transversal view showing the PM inside the CTS riding up to a blade extending the external CTS wall. Grey: commisure plane; blue: PM; yellow: CTS.
PM
A
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprotina
Clinocaprina
Curtocaprina
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Offneria
Pachytraga
Radiolites
Sellaea
Semailia
Titanosarcolites
State 4
projecting PM blade twisted, parallel to the commissure
G
Sketch of rudist LV in transversal view showing the PM blade stranded to become parallel to the commissure. Grey: commisure plane; blue: PM; yellow: CTS.
PM
B
Polyconites
Tepeyacia
Missing data: Oryxia

Figure 4. 39. Hierarchical character #4.3. A – *Caprotina costata* (MV 2003.2.5147). B – *Mexicaprina minuta* (JPMA.14097-5). C – *Radiolites radiosus* (MNHN.F.J09021). D – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm.

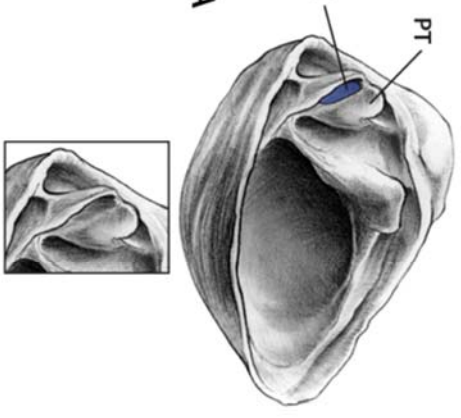
Character 4.3: posterior myophore morphology (LV)

State 3

vertical PM blade projecting straight into the RV



Left - Sketch of rudist LV in transversal view showing the PM on a blade extending the external CTS wall. Grey: commisure plan; blue: PM; yellow: CTS. Right - sketch of relative position between PT and PM. The PM is elongated in the dorso-ventral axis and has a dorsal contact with the PT.

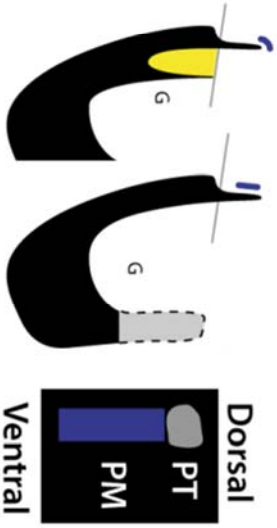


Chaperia
Caprotina
Glossomyophorus
Pachytraga

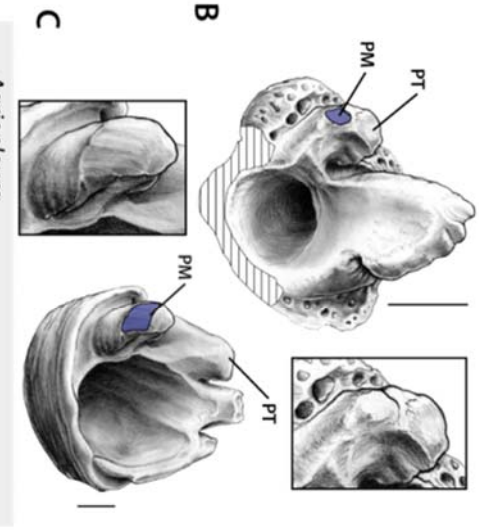
Polyconites
Sellaea
Tepeyacia

State 5

vertical and projecting PM blade becoming outwardly directed ("backflips")



Sketch of rudist LV in transversal view showing the PM on a myophoral blade. Left: the PM is at the apex of the myophoral blade (B). Right: the PM is vertical and outwardly directed, on the external side of the blade (C). The blade departs from the CTS. Grey: commisure plan; blue: PM; yellow: CTS.

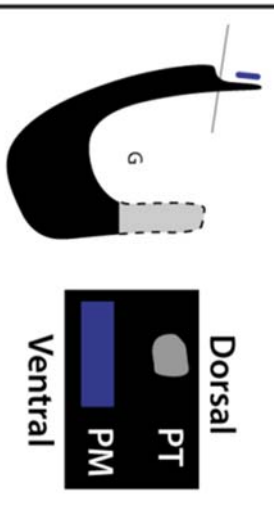


Agriopleura
Amphitriscoelus
Clinocaprina
Curtocaprina
Dictyoptichus

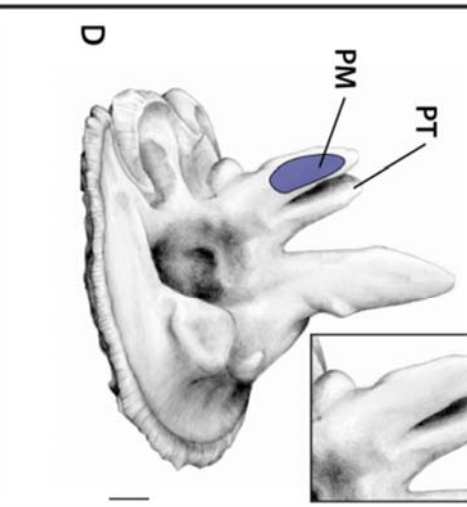
Ichthyosarcolites
Mexicaprina
Radiolites
Titanosarcolites

State 6

vertical and projecting PM blade stranded perpendicular to the shell wall, facing the PT



Sketch of relative position between PT and PM. The PM rotates of 90 degrees, and becomes elongated in the antero-posterior axis. By rotating, it looses its contact with the PT.



Barrettia
Hippurites

Semailia

Missing data: ***Oryxia***

7.5. Character 5

Name: posterior myophore position

Position: left valve

Kind: myophores

Coding: (0(1))

States definitions

0 contact between PM and PT

1 PM and PT separated by a gap

States tests

1 rejected – clades 13 (except *Agriopleura*), 22 (except *Ichthyosarcolites*)

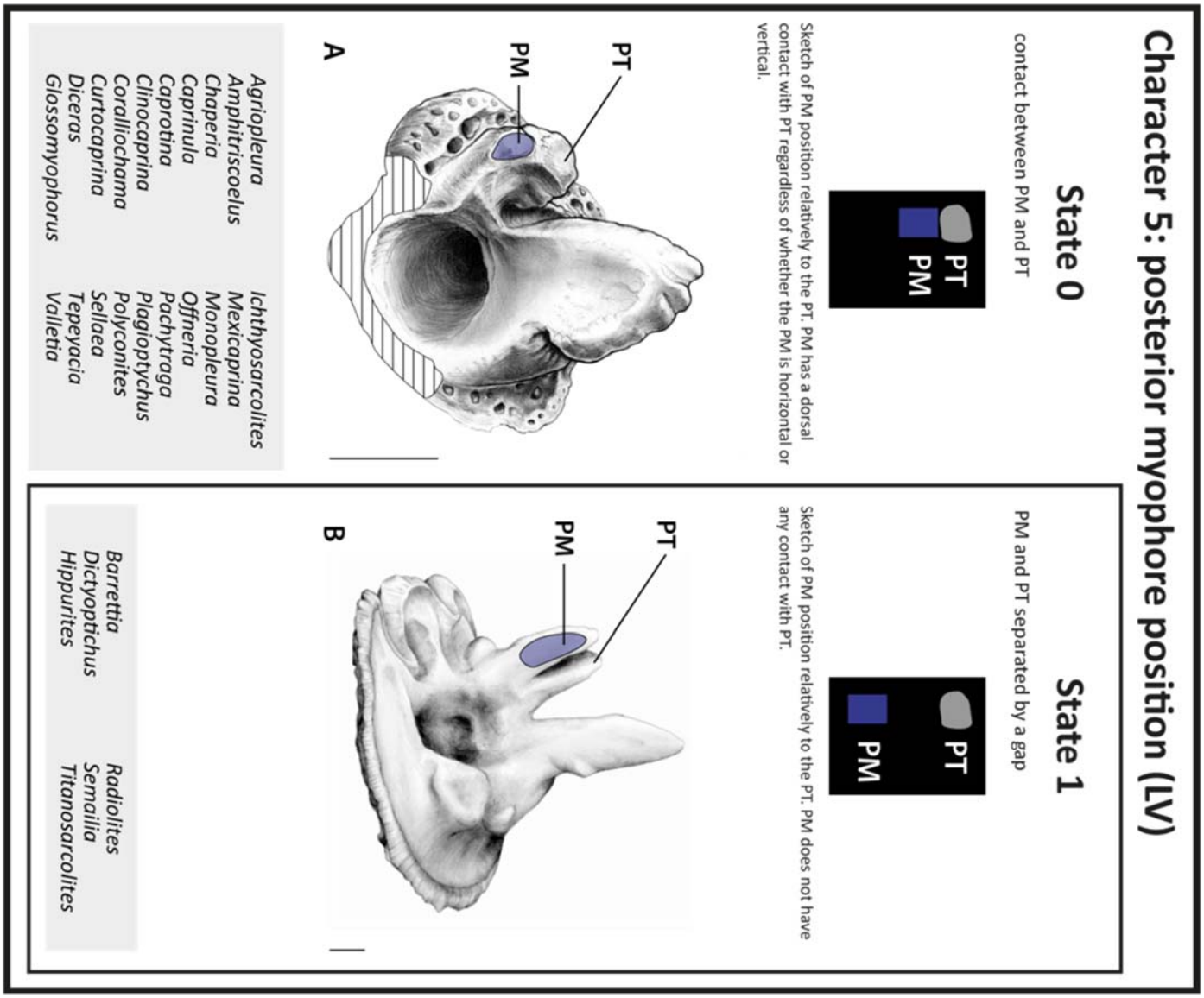
Retention index

0.74

Remarks

If protruding sharply in a thick blade, the PM-LV separated by a gap from its associated PT can lead to a 'tooth-like myophore' (see character 7 of Mitchell 2013b).

Figure 4. 40. Hierarchical character #5. A – *Mexicaprina minuta* (JPMA.14097-5). B – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm.


Character 5: posterior myophore position (LV)
State 0
contact between PM and PT
PT
PM
Sketch of PM position relatively to the PT. PM has a dorsal contact with PT regardless of whether the PM is horizontal or vertical.
PT
PM
A
Agriopleura
Amphitriscoelus
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Sellaea
Tepeyacia
Valletia
State 1
PM and PT separated by a gap
PT
PM
Sketch of PM position relatively to the PT. PM does not have any contact with PT.
PT
PM
B
Barrettia
Dictyoptichus
Hippurites
Radiolites
Semailia
Titanosarcolites
Missing data: Oryxia

7.6. Character 6

Name: anterior myophore morphology

Position: left valve

Kind: myophores

Coding: (0(1(2)(3(4(5))))

States definitions

0 AM vertical inserted directly on the inner shell surface

1 AM on the shell to the commissural plane, forming a horizontal shelf

2 AM shelf large and quadrangular

3 AM vertical and outwardly directed by extension of the CTS anteriorly

4 AM outwardly directed projecting itself in the RV by the development of the AMC/G wall (may be rooted to the thick shelf)

5 AM shelf stretched and divided in two contiguous parts

States tests

1 passed – clade 1

2 passed – clade 8

3 rejected – clade 9, *Tepeyacia*

4 rejected – clades 13, 19 (except *Curtocaprina*), *Tepeyacia*

5 rejected – clades 15

Retention index

0.92

Remarks

This character includes a paralogous clade (independence instancy) and can be understood as three independent characters after free-paralogy subtree analysis (Nelson and Ladiges 1996) following procedure described in chapter 1.

The AM-LV show a great complexity of varying forms. The ancestral state is a direct insertion of the myophore on the anterior wall of G. There is no differenciated structure to carry the vertical myophore. On state 1, the myophoral zone is thickened. However, this myophoral shelf has a different morphology than the PM-LV (#4.1) probably due to the absence of an ancestral myophoral blade. The comparison between the two myophoral shelves of the same valves offers the best conditions to understand the differences, as *Monopleura* or *Valletia* (Fig. X; see also Schneider and Skupien 2015). While the PM is well delineating in its contact with G, marking an acute angle because of the ancestral presence of the myophoral blade, the AM looks like more as an accreted amount of aragonite on the anterior side that can seems to spill out in G. The AM-LV is generally not as well delineating as PM-LV. In state 2, the AM shelf become slightly better delineated and takes a form easily recognizable with a globally quadrangular shape, increasing in surface and completely parallel to the commissural plane. In some Plagioptychid taxa having state 2 as *Plagioptychus coquandi*, the AM-LV can heavily protrude while staying horizontal. The state 1 is a paralog clade from which two independent states arise: state 2 and state 3. In the state 3, the myophoral shelf can be invaded by the CTS which extends itself anteriorly by surrounding dorsally the AT, and become inserted vertically on the wall separating the CTS extension from G. It results in a vertical outwardly directed myophore. This hypothesis is made by the position of the myophoral shelf corresponding to this cavity in shells having a vertical AM-LV. On some *Pachytraga, Retha, Caprotina,* and *Sellaea* species, this anterior cavity can be seen in connection with the CTS, allowing to hypothesize an origin of the first in the second (fig. 4.22). The vertical and outwardly directed myophore can develop itself to protrude out of the commissural plane (state 4). Hippuritids show this blade, even if it is thick and modified by the capuloid LV. In this family, the myophore become highly extended in the dorsoventral axis that it became divided in two parts (state 5).

Figure 4. 41. Hierarchical character #6, its character states intensions, and its treatment with paralogy-free subtree analysis.

Character 6: anterior myophore morphology (LV)

State 0

AM vertical inserted directly on the inner shell surface

State 1

AM on the shell to the commissural plane, forming a horizontal shelf

State 2

AM shelf large and quadrangular

State 3

AM vertical and outwardly directed by extension of the CTS anteriorly

State 4

AM outwardly directed projecting itself in the RV by the development of the AMC/G wall

State 5

AM shelf stretched and divided in two contiguous parts

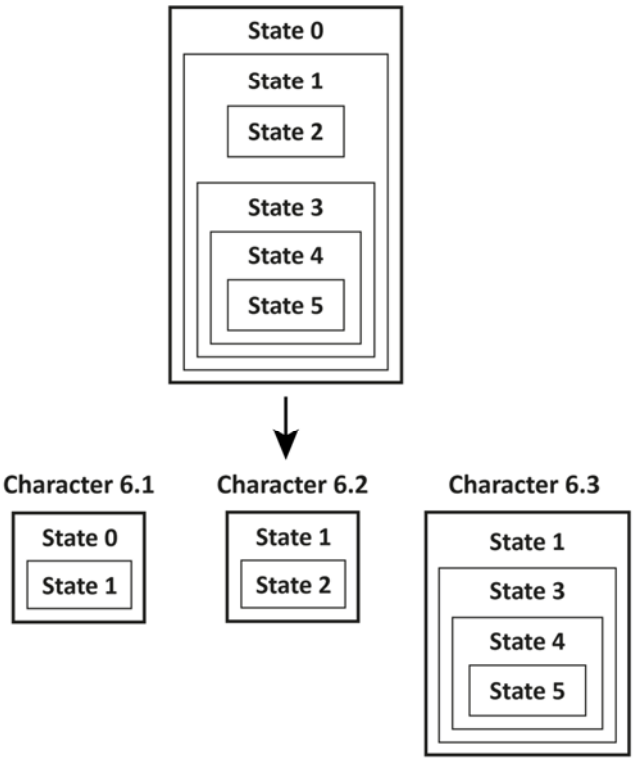


Figure 4. 42. Hierarchical character #6.1. A – *Diceras arietinum* (MNHN). B – *Caprotina costata* (MV 2003.2.5147). Scale bars, 10 mm.

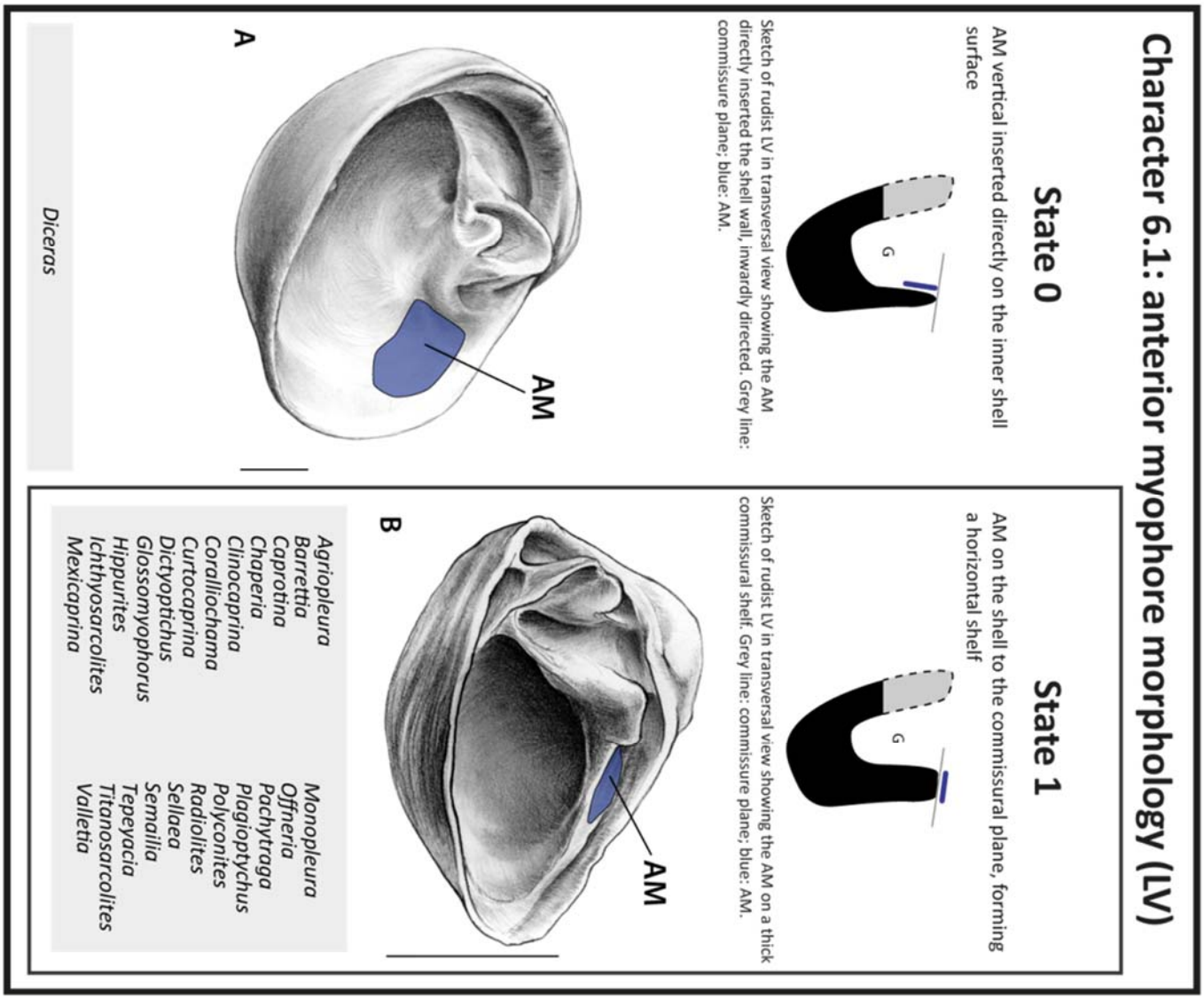

Character 6.1: anterior myophore morphology (LV)
State 0
AM vertical inserted directly on the inner shell surface
G
Sketch of rudist LV in transversal view showing the AM directly inserted the shell wall, inwardly directed. Grey line: commissure plane; blue: AM.
AM
A
State 1
AM on the shell to the commissural plane, forming a horizontal shelf
G
Sketch of rudist LV in transversal view showing the AM on a thick commissural shelf. Grey line: commissure plane; blue: AM.
AM
B
Agriopleura
Barrettia
Caprotina
Chaperia
Clinocaprina
Coralliochama
Curtocaprina
Dictyopticlus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia
Diceras
Missing data: Amphitriscoelus, Caprinula, Oryxia

Figure 4. 43. Hierarchical character #6.2. A – *Caprotina costata* (MV 2003.2.5147). B – *Coralliochama orcutti* (USNM.186662). C – *Plagioptychus arnaudi* (UPMC.1273) Scale bars, 10 mm.

Character 6.2: anterior myophore morphology (LV)

State 1

AM on the shell to the commissural plane, forming a horizontal shelf

Agriopleura
Barrettia
Caprotina
Chaperia
Clinocaprina
Curtocaprina
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia

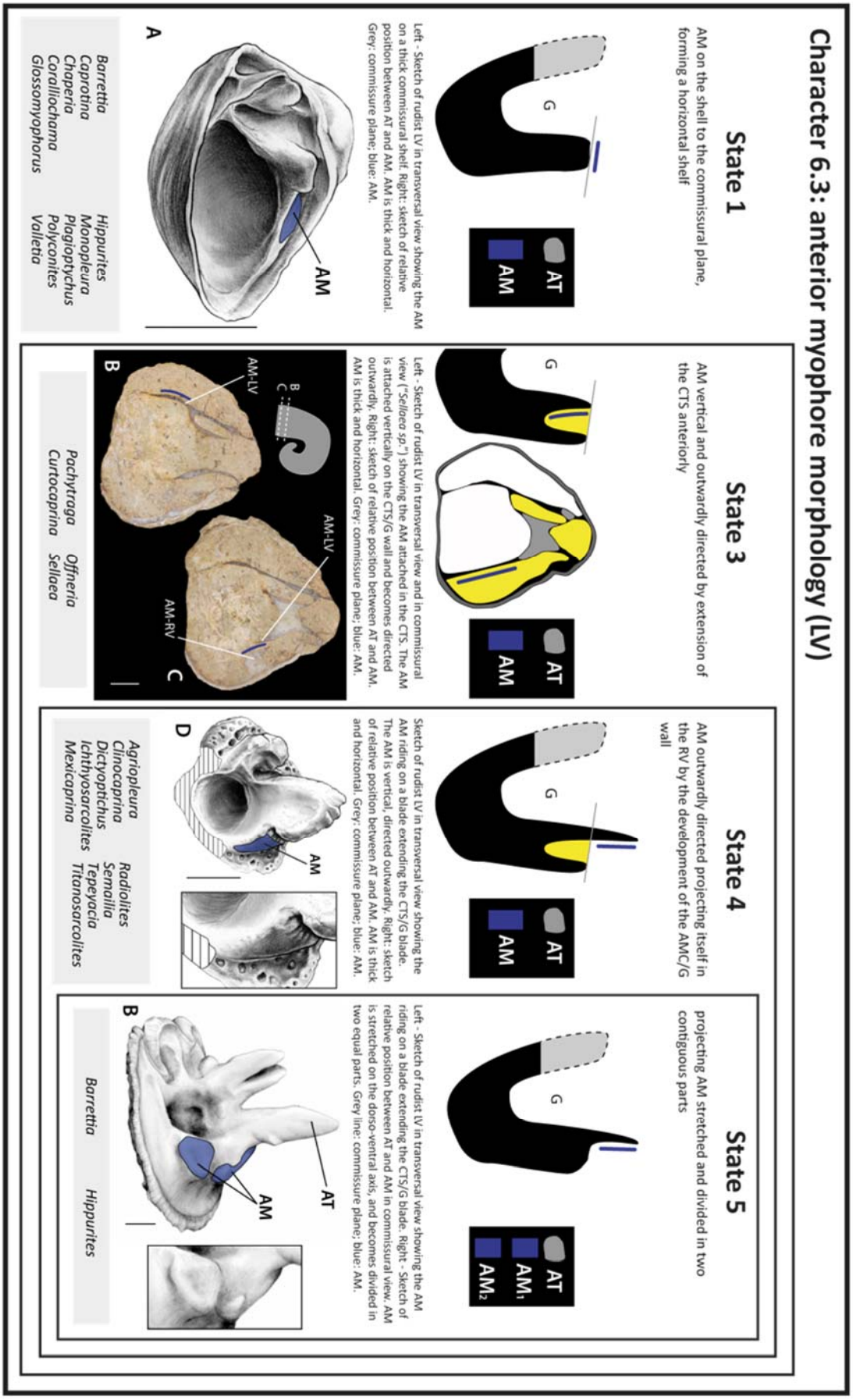
State 2

AM shelf large and quadrangular

Coralliochama
Plagioptychus

Missing data: ***Amphitriscoelus, Caprinula, Oryxia***

Figure 4. 44. Hierarchical character #6.3. A – *Caprotina costata* (MV 2003.2.5147). B – '*Sellaea' sp.* (WSA3058). C – *Mexicaprina minuta* (JPMA.14097-5). D – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm.


Character 6.3: anterior myophore morphology (LV)
State 1
AM on the shell to the commissural plane, forming a horizontal shelf
G
AT
AM
Left - Sketch of rudist LV in transversal view showing the AM on a thick commissural shelf. Right: sketch of relative position between AT and AM. AM is thick and horizontal. Grey: commissure plane; blue: AM.
AM
A
Barrettia
Caprotina
Chaperia
Coralliochama
Glossomyophorus
Hippurites
Monopleura
Plagioptychus
Polyconites
Valletia
State 3
AM vertical and outwardly directed by extension of the CTS anteriorly
G
AT
AM
Left - Sketch of rudist LV in transversal view and in commissural view ("Sellaea sp.") showing the AM attached in the CTS. The AM is attached vertically on the CTS/G wall and becomes directed outwardly. Right: sketch of relative position between AT and AM. AM is thick and horizontal. Grey: commissure plane; blue: AM.
B
C
AM-LV
AM-LV
AM-RV
C
B
Pachytraga
Curtocaprina
Offneria
Sellaea
State 4
AM outwardly directed projecting itself in the RV by the development of the AMC/G wall
G
AT
AM
Sketch of rudist LV in transversal view showing the AM riding on a blade extending the CTS/G blade. The AM is vertical, directed outwardly. Right: sketch of relative position between AT and AM. AM is thick and horizontal. Grey: commissure plane; blue: AM.
AM
D
Agriopleura
Clinocaprina
Dictyoptichus
Ichthyosarcolites
Mexicaprina
Radiolites
Semailia
Tepeyacia
Titanosarcolites
State 5
projecting AM stretched and divided in two contiguous parts
G
AT
AM1
AM2
Left - Sketch of rudist LV in transversal view showing the AM riding on a blade extending the CTS/G blade. Right - Sketch of relative position between AT and AM in commissural view. AM is stretched on the dorso-ventral axis, and becomes divided in two equal parts. Grey line: commissure plane; blue: AM.
AT
AM
B
Barrettia
Hippurites
Missing data: Amphitriscoelus, Caprinula, Oryxia

7.7. Character 7

Name: anterior myophore position

Position: left valve

Kind: myophores

Coding: (0(1))

States definitions

0 contact between AM and AT

1 AM and AT separated by a gap

States tests

1 rejected – *Radiolites*, *Titanosarcolites*

Retention index

0.52

Remarks

If protruding sharply as a thick blade, the PM-LV separated by a gap from its associated PT can lead to a 'tooth-like myophore'. See character 7 of Mitchell (2013), whoever apply only on the PM, its taxonomic sampling does not allow description of an equivalent character state for the AM-LV (see character 8 of Mitchell 2013b).

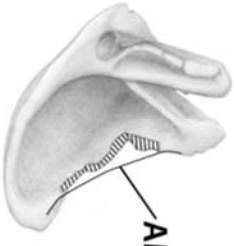
Figure 4. 45. Hierarchical character #7. A – *Chaperia costata* (MNHN.F.R07950). B – *Radiolites radiosus* (MNHN.F.J09021). Scale bars, 10 mm.

Character 7: anterior myophore position (LV)

State 0

contact between AM and AT

Sketch of AM position relatively to the AT. AM has a dorsal contact with AT regardless of whether the AM is horizontal or vertical.



Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Sellaea
Tepeyacia
Valletia

State 1

AM and AT separated by a gap

Left - Sketch of AM position relatively to the AT. Right - Commissural view of *Radiolites radiosus*. AM doesn't have any contact with AT.

Titanosarcolites
Radiolites

Missing data: ***Oryxia, Semailia***

7.8. Character 8

Name: orientation of the central tooth socket/general cavity wall

Position: left valve

Kind: vertical blades

Coding: (0(1(2)))

States definitions

0 wall separating CTS from the G horizontal

1 tilting of the wall to the dorsal side of the shell

2 verticalization of the wall to the dorsal side of the shell

States tests

1 passed – clade 1

2 rejected – clades 7, 9, 11

Retention index

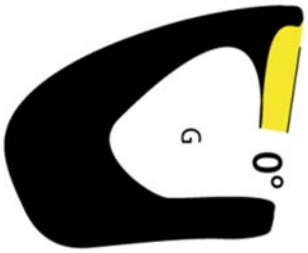
0.95

Figure 4. 46. Hierarchical character #8. A – *Diceras arietinum* (MNHN). B – *Monopleura michaillensis* (UPMC.1272). C – *Plagioptychus arnaudi* (UPMC.1273). Scale bars, 10 mm.

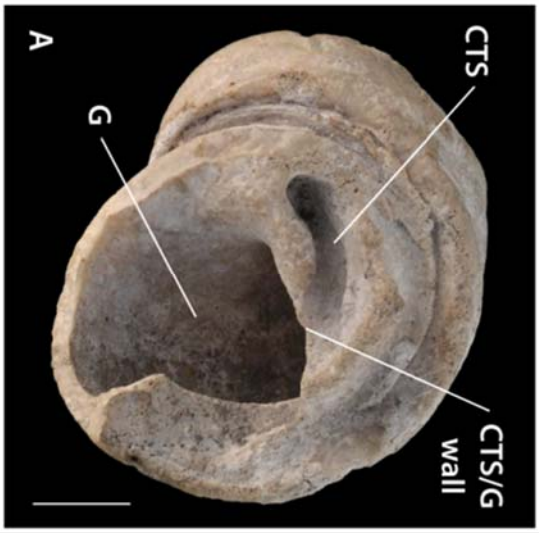
Character 8: orientation of the central tooth socket/general cavity wall

State 0

wall separating CTS from the G horizontal



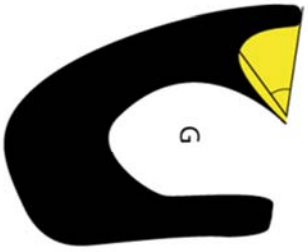
Sketch of rudist LV in transversal view. The wall separating the CTS (yellow) from G is parallel to the commissure.



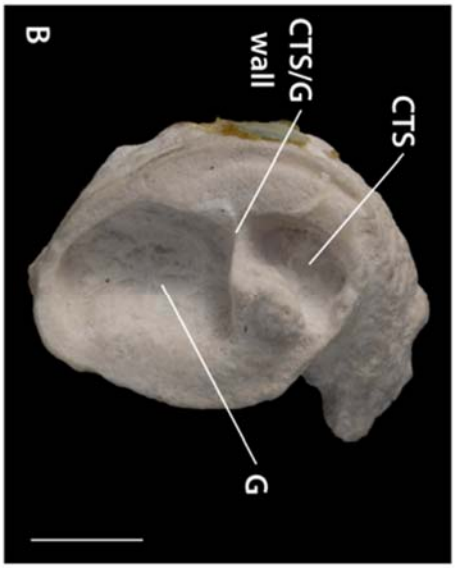
Diceras

State 1

tilting of the wall to the dorsal side of the shell



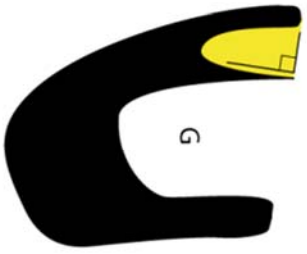
Sketch of rudist LV in transversal view. The wall separating the CTS from G is tilted and makes an acute angle relatively to the commisure plan.



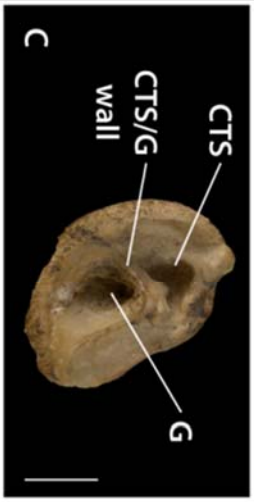
Caprotina
Chaperia
Monopleura
Sellaea
Valletia

State 2

verticalization of the wall to the dorsal side of the shell



Sketch of rudist LV in transversal view. The wall separating the CTS from G stands vertical. It makes a right angle relatively to the commisure plan.



Agriopleura
Amphitriscoelus
Barrettia
Caprinula
Clinocaprina
Coralliochama
Curtocaprina
Dictyoptichus
Hippurites
Ichthyosarcolites
Mexicaprina
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Semailia
Tepeyacia
Titanosarcolites

Missing data: ***Glossomyophorus, Oryxia***

7.9. Character 9

Name: central tooth socket morphology

Position: left valve

Kind: vertical blades

Coding: (0(1))

States definitions

0 CTS located between ligament and G

1 CTS divided in two daughter cavities by the development of an internally oriented ligamentary ridge

States tests

1 passed – clade 12 (except *Agriopleura*)

Retention index

0.96

Remarks

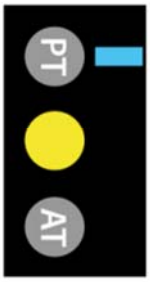
The CTS is generally not recognized in Hippuritids and Radiolitids, in which it is assumed to be lost, the space being too small between teeth. Cavities however can be found dorsally to the teeth that are generally associated to accesory cavities or to shell deformation when the LV is capuloid. In shells where the AT and the PT are close from each other, the CTS can be retreived in a more dorsal position. Example of dorsal CTS can be found in *Rotacaprina*. The dorsal CTS can be divided in two parts when the ligamentary structures extend ventrally. This character state can be seen in shellls of the genus *Hippurites*, *Agriopleura* and *Radiolites*. It is more parsimonious to postulate that this cavity is the CTS than a disappearance of the CTS and the excavation of a neoformed accessory cavity. This cavity is also unlikely a structure resulting from the modification of the LV onto a capuloid valve, a cavity excavating the shell material being distinctly visible.

Figure 4. 47. Hierarchical character #9. A – *Monopleura michaillensis* (UPMC.1272). B – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm.

Character 9: central tooth socket morphology (LV)

State 0

CTS located between ligament and G

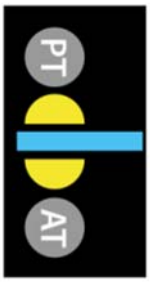


Position of the CTS position (yellow) relatively to PT, AT and ligament (blue). The CTS is stranded between the two teeth. The ligament has a single contact, with the PT.

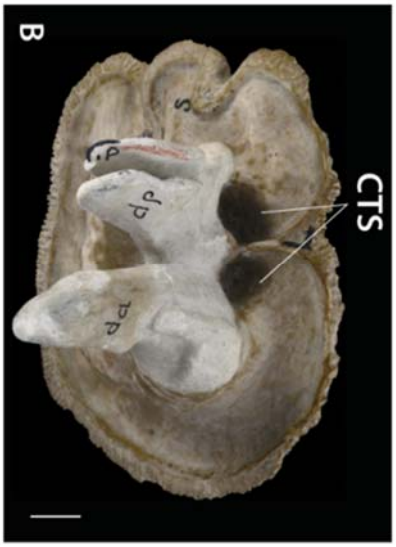
Agriopleura
Amphitriscoelus
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia

State 1

CTS divided in two daughter cavities by the development of an internaly oriented ligamentary ridge



Position of the CTS (yellow) relatively to PT, AT and ligament (blue). The ligament moves dorsally to the CTS and looses its contact with the PT. By developing itself ventraly, the ligamentary ridge divide the CTS in two parts.



Hippurites
Radiolites

Missing data: ***Barretia*, *Oryxia***

7.10. Character 10

Name: central tooth socket/general cavity wall position

Position: left valve

Kind: vertical blades

Coding: (0(1))

States definitions

0 posterior insertion of the wall separating CTS from G is ventral to the PM

1 posterior insertion of the wall separating CTS from G is between PT and PM

States tests

1 rejected – clade 4 (except *Caprotina*, *Chaperia*), 13, 20

Retention index

0.74

Remarks

Related to character 30 of Skelton and Smith (2000) with diverging interpretations (see fig. 4.23).

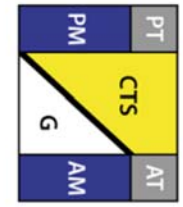
Related to what MacGillavry (1937) call 'symetrization'. A discussion of this character linked to the ichthyosarcolitid toothlet hypothesis can be found in Figure 4.19.

Figure 4. 48. Hierarchical character #10. A – *Plagioptychus arnaudi* (UPMC.1273). B – *Titanosarcolites giganteus* (UT.47533). C – *Curtocaprina clabaughikinsorum* (UT.10937). Scale bars, 10 mm.

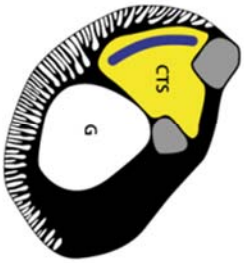
Character 10: wall separating the central tooth socket from the general cavity (CTS/G wall)

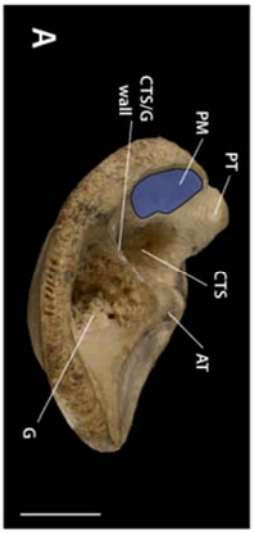
State 0

posterior insertion of the wall separating CTS from G is ventral to the PM.



Left - sketch of the PM relatively to the myocardinal apparatus and their connections (black lines). The PM is located inside the PM. Right - sketch of the CTS/G wall insertion. The CTS/G wall ends ventrally between the PM and G.

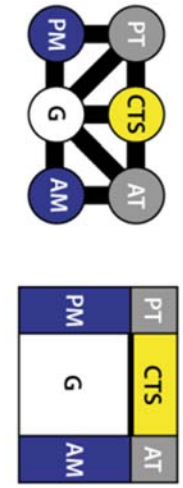




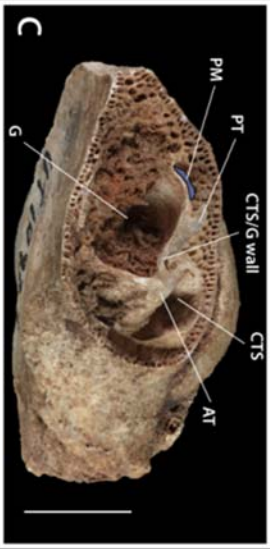
Amphitriscoelus
Chaperia
Caprinula
Caprotina
Coralliochama
Diceras
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Sellaea
Tepeyacia
Valletia

State 1

posterior insertion of the wall separating CTS from G is between PT and PM



Left - sketch of the PM relatively to the myocardinal apparatus and their connections (black lines). The PM is located outside the PM. Right - sketch of the CTS/G wall insertion. The CTS/G wall ends posteriorly between the PT and the PM and G.



Agriopleura
Barrettia
Clinocaprina
Curtocaprina
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Radiolites
Semailia
Titanosarcolites

Missing data: ***Oryxia***

7.11. Character 11

Name: posterior accessory cavity

Position: left valve

Kind: canals

Coding: (0(1))

States definitions

0 root

1 presence of a posterior accessory cavity

States tests

1 rejected – clades 4, 7, 9, 16, *Sellaea*, *Pachytraga*

Retention index

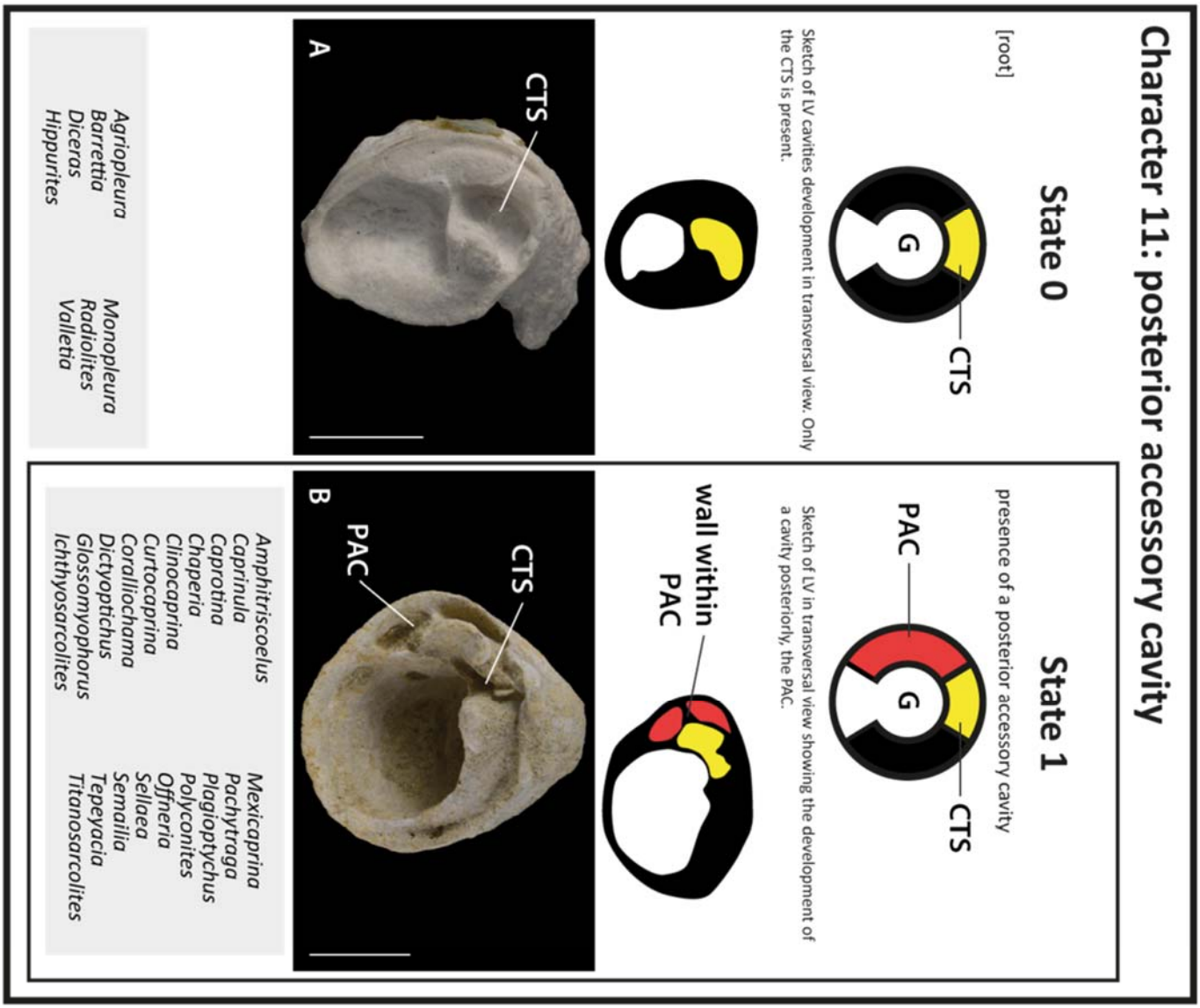
0.60

Remarks

Related partially to character 31 of Skelton and Smith (2014).

The posterior accessory cavity can be found as a small ovoid cavity stranded between the posterior side of the shell and the PM-LV in Polyconitidae. In Caprotinidae, a blade develops to form a ventral and a dorsal cavities. The posterior accessory can be highly modified by a wide ventral development and the multiplication of radial walls giving to it an aspect of pallial canals as in *Plagioptychus* and *Coralliochama*. We however reserve the term pallial canals on the LV for structures resulting from the merge between the posterior accessory cavity and the CTS extended anteriorly (#4.2). We call Plagioptychid-like canals the canals deriving from bifurcating walls and developped PAC.

Figure 4. 49. Hierarchical character #11. A – *Monopleura michaillensis* (UPMC.1272). B – *Caprotina costata* (MV 2003.2.5147). Scale bars, 10 mm.


Character 11: posterior accessory cavity
State 0
[root]
CTS
G
Sketch of LV cavities development in transversal view. Only the CTS is present.
CTS
A
State 1
presence of a posterior accessory cavity
CTS
PAC
G
Sketch of LV in transversal view showing the development of a cavity posteriorly, the PAC.
wall within PAC
CTS
PAC
B
Agriopleura
Barrettia
Diceras
Hippurites
Monopleura
Radiolites
Valletia
Amphitriscoelus
Caprinula
Caprotina
Chaperia
Clinocaprina
Curtocaprina
Coralliochama
Dictyoptichus
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Pachytraga
Plagioptychus
Polyconites
Offneria
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Missing data: Oryxia

7.12. Character 12

Name: posterior accessory cavity and central tooth socket position

Position: left valve

Kind: canals

Coding: (0(1))

States definitions

0 LV ventral side thick, devoid of canals

1 ventral contact between the CTS and the PAC

States tests

1 passed – clade 16

Retention index

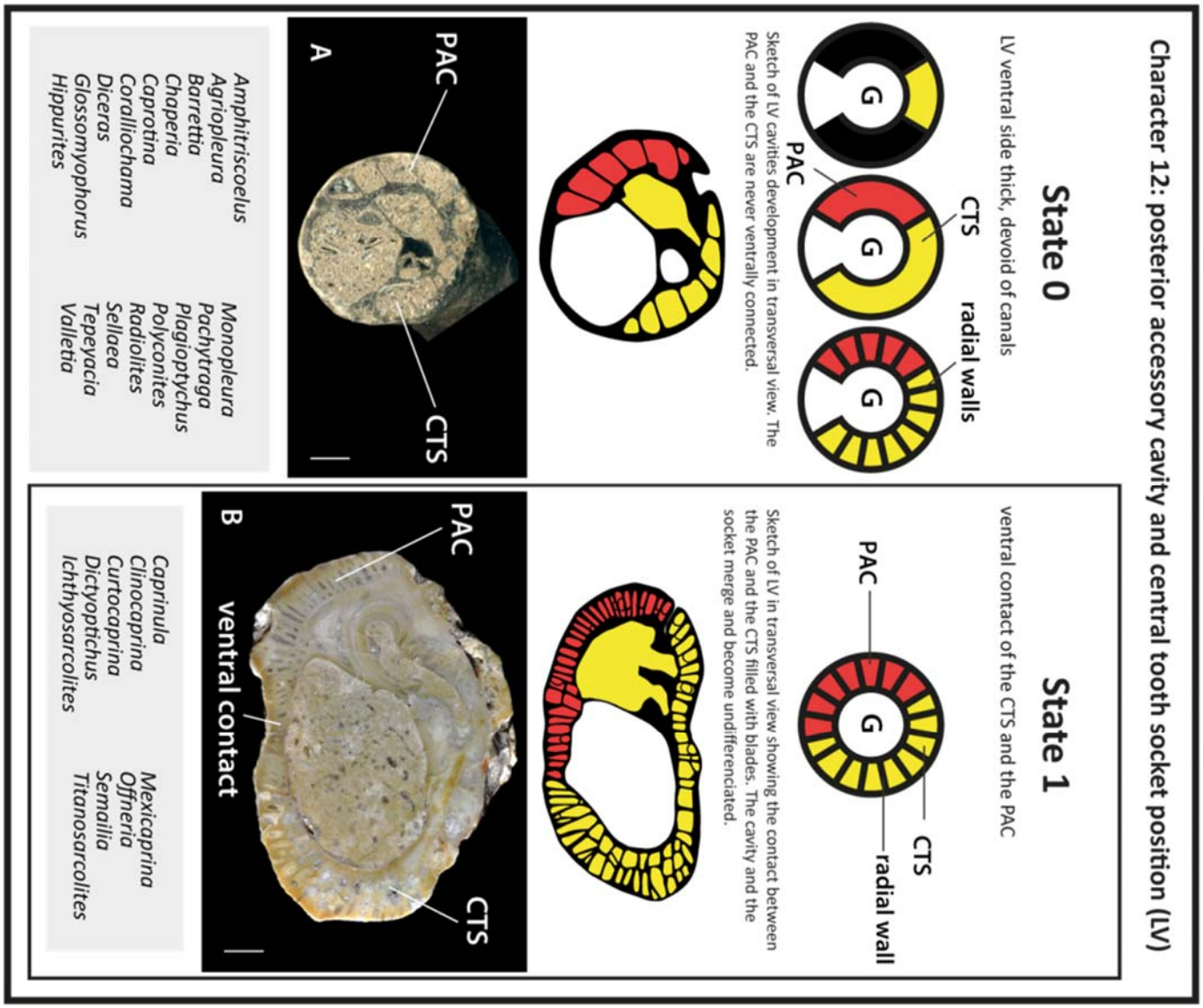
1.00

Remarks

Related to character 10 of Chartrousse (1994). Related partially to character 19 of Mitchell (2013a) (with #30). Related partially to character 11 of Sano *et al.* (2014).

In other analyses, this character is only described as a ventral interruption of the pallial canals. We associate this morphological aspect of the pallial to be of critical importance to understand the construction of pallial canals in rudist shells. The derived state is reached when both PAC and CTS extended anteriorly (#4.2) invade the ventral side of the shell. When the cavities are in contact, they are generally filled with radial blades and it becomes virtually impossible to differentiate the two cavities. The G become entirely surrounded by pallial canals that originate from the PAC and the CTS filled with blades.

Figure 4. 50. Hierarchical character #12. A – *Amphitriscoelus waringi*. B – *Offneria rhodanica* (UCBL.R275). Scale bars, 10 mm.


Character 12: posterior accessory cavity and central tooth socket position (LV)
State 0
LV ventral side thick, devoid of canals
CTS
radial walls
G
G
G
PAC
Sketch of LV cavities development in transversal view. The PAC and the CTS are never ventrally connected.
PAC
CTS
A
Amphitriscoelus
Agriopleura
Barrettia
Chaperia
Caprotina
Coralliochama
Diceras
Glossomyophorus
Hippurites
Monopleura
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Tepeyacia
Valletia
State 1
ventral contact of the CTS and the PAC
CTS
radial wall
PAC
G
Sketch of LV in transversal view showing the contact between the PAC and the CTS filled with blades. The cavity and the socket merge and become undifferenciated.
PAC
CTS
B
ventral contact
Caprinula
Clinocaprina
Curtocaprina
Dictyoptichus
Ichthyosarcolites
Mexicaprina
Offneria
Semailia
Titanosarcolites
Missing data: Oryxia

7.13. Character 13

Name: radial walls morphology

Position: left valve

Kind: canals

Coding: (0(1(2)(3)))

States definitions

0 root

1 numerous radial walls invading cavities (and forming canals with the MPC)

2 walls specialized in two morphologies leading to two morphologies of canals: a row of rectangular canals along the external side of the shell wall, and capillary canals everywhere else

3 walls bifurcating (plagioptychid-like canals)

States tests

1 rejected – clades 9, 16, *Pachytraga*

2 passed – clade 16 (except *Offneria*, *Dictyoptychus*)

3 rejected – clades 9, *Offneria*

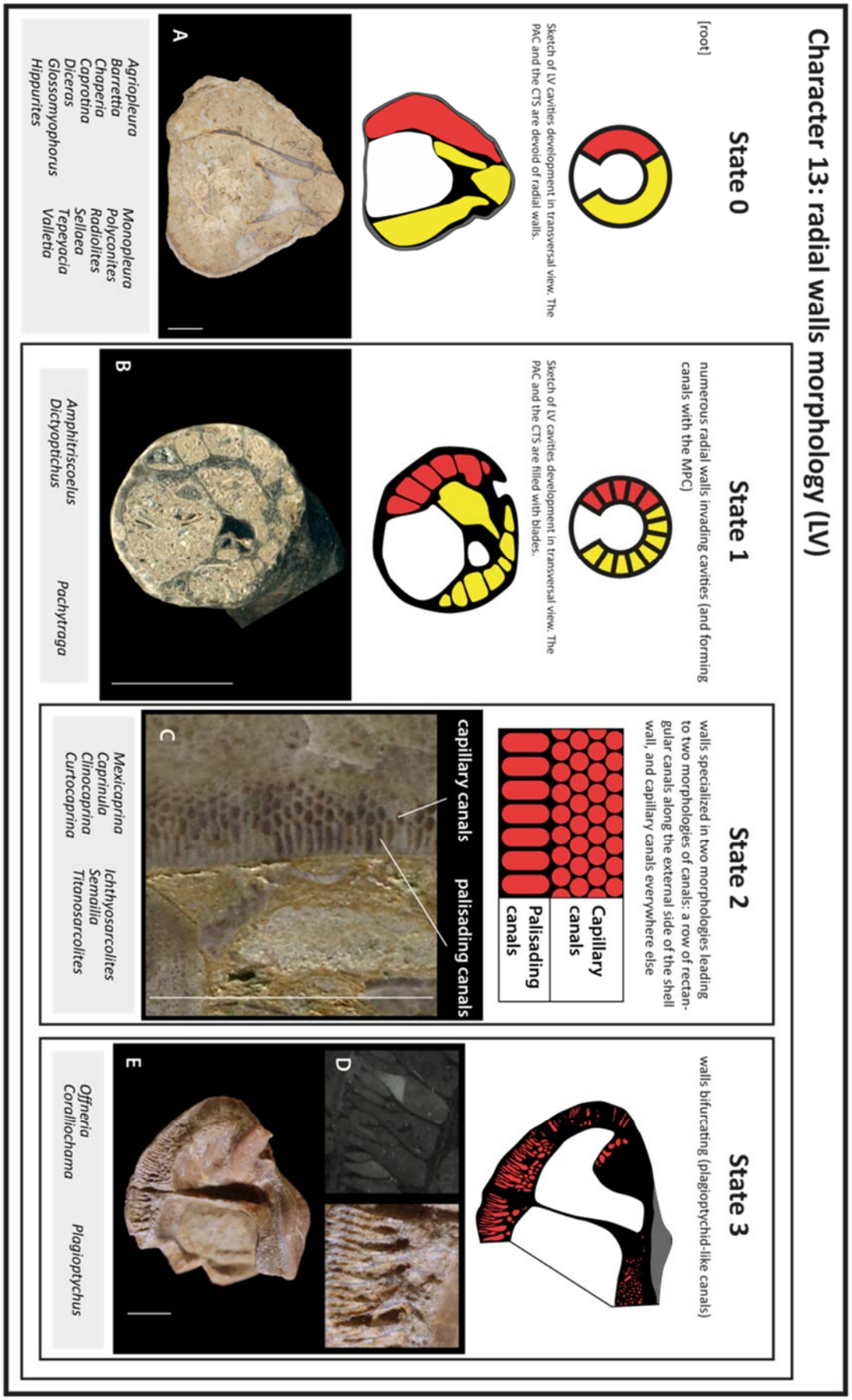
Retention index

0.82

Remarks

Related to characters 12 and 16 of Chartrousse (1994). Related partially to character 10 of Smith (1994) (#13.1). Related to character 9 of Stone and Telford (1999; #13.1).

Figure 4. 51. Hierarchical character #13. A – *'Sellaea' sp.* (WSA3058). B – *Amphitriscoelus waringi*. C – *Titanosarcolites giganteus* (UT.47533). D – *Schiosa ramosa* (R485, Université Claude Bernard Lyon 1; Coalcoman, Mexico). E – *Coralliochama orcutti* (USNM.186662). Scale bars, 10 mm.


Character 13: radial walls morphology (LV)
State 0
[root]
Sketch of LV cavities development in transversal view. The PAC and the CTS are devoid of radial walls.
A
Agriopleura
Barrettia
Chaperia
Caprotina
Diceras
Glossomyophorus
Hippurites
Monopleura
Polyconites
Radiolites
Sellaea
Tepeyacia
Valletia
State 1
numerous radial walls invading cavities (and forming canals with the MPC)
Sketch of LV cavities development in transversal view. The PAC and the CTS are filled with blades.
B
Amphitriscoelus
Dictyoptichus
Pachytraga
State 2
walls specialized in two morphologies leading to two morphologies of canals: a row of rectangular canals along the external side of the shell wall, and capillary canals everywhere else
Capillary canals
Palisading canals
capillary canals
palisading canals
C
Mexicaprina
Caprinula
Clinocaprina
Curtocaprina
Ichthyosarcolites
Semailia
Titanosarcolites
State 3
walls bifurcating (plagioptychid-like canals)
D
E
Offneria
Coralliochama
Plagioptychus
Missing data: Oryxia

7.14. Character 14

Name: large pallial canals row morphology

Position: left valve

Kind: canals

Coding: (0(1(2)))

States definitions

0 root

1 one row of large polygonal canals bordering G

2 the large polygonal canals increase in size between the two myophores and the shell/exterior wall

States tests

1 passed – clade 16 (except *Offneria*, *Dictyoptychus*, *Clinocaprina*, *Titanosarcolites*)

2 rejected – clade 19 (except *Clinocaprina*, *Dictyoptychus*, *Titanosarcolites*), *Caprinula*

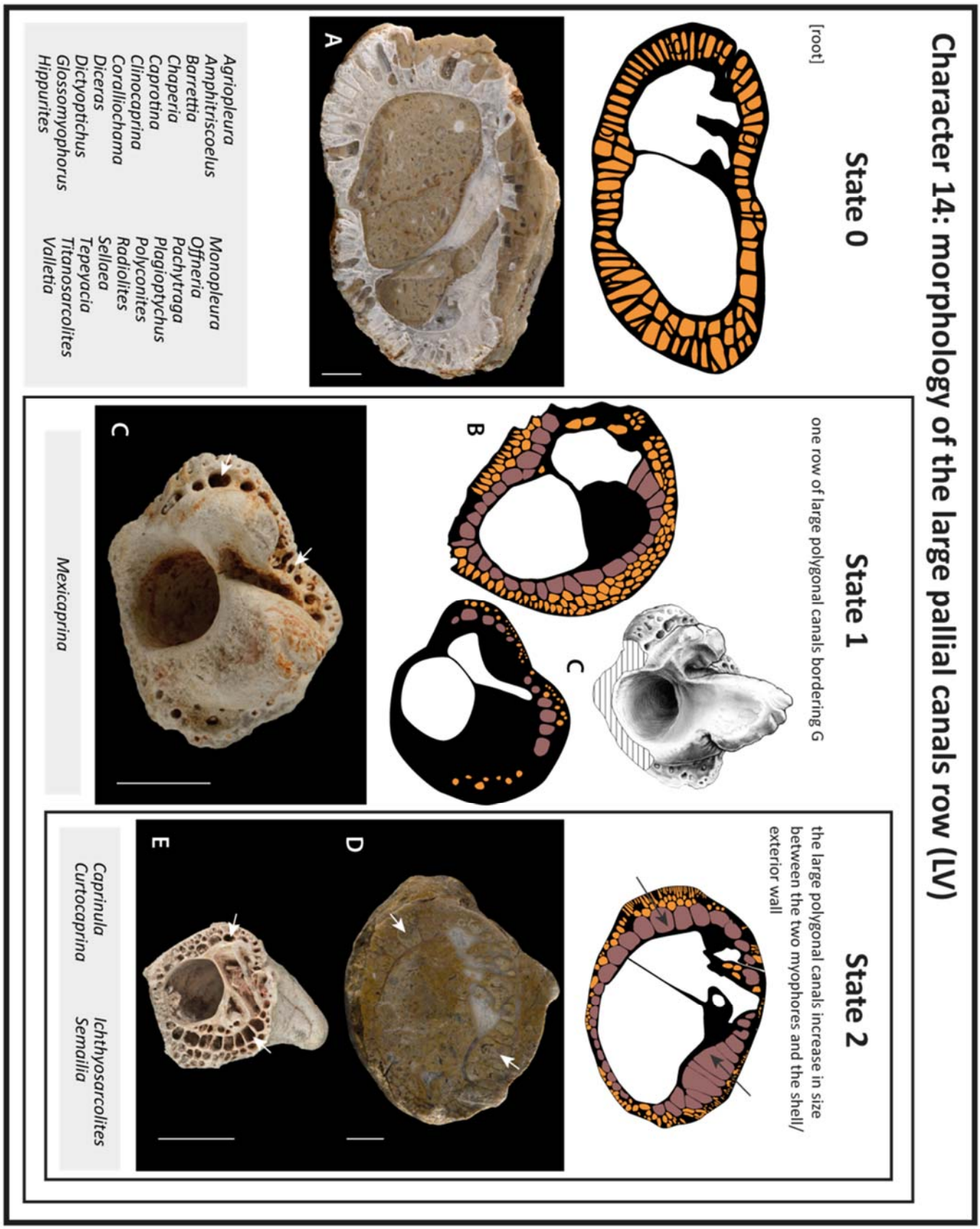
Retention index

0.84

Remarks

The postulated row of large polygonal canals may be originated from the CTS partitioning following the observations on *Kimbleia capacis* (Aguilar-Perez 2008) bordering G (see *Caprinuloidea romeri* as another example of state 2).

Figure 4. 52. Hierarchical character #14. A – *Offneria rhodanica* (JPMA.16494). B – *Kimbleia capacis* (PUAB-81521). C – *Mexicaprina minuta* (JPMA.14097-5). D – *Caprinula boissyi* (MNHN.6924). E – *Caprinuloidea sp.* (UT.10922-2). Scale bars, 10 mm.


Character 14: morphology of the large pallial canals row (LV)
State 0
[root]
A
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprotina
Clinocaprina
Coralliochama
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Tepeyacia
Titanosarcolites
Valletia
State 1
one row of large polygonal canals bordering G
B
C
Mexicaprina
State 2
the large polygonal canals increase in size between the two myophores and the shell/exterior wall
D
E
Caprinula
Curtocaprina
Ichthyosarcolites
Semailia
Missing data: Oryxia

7.15. Character 15

Name: anterior tooth canal

Position: left valve

Kind: canals

Coding: (0(1))

States definitions

0 thick AT

1 circular canal in the AT

States tests

1 rejected – *Amphitriscoelus*, *Caprinula*

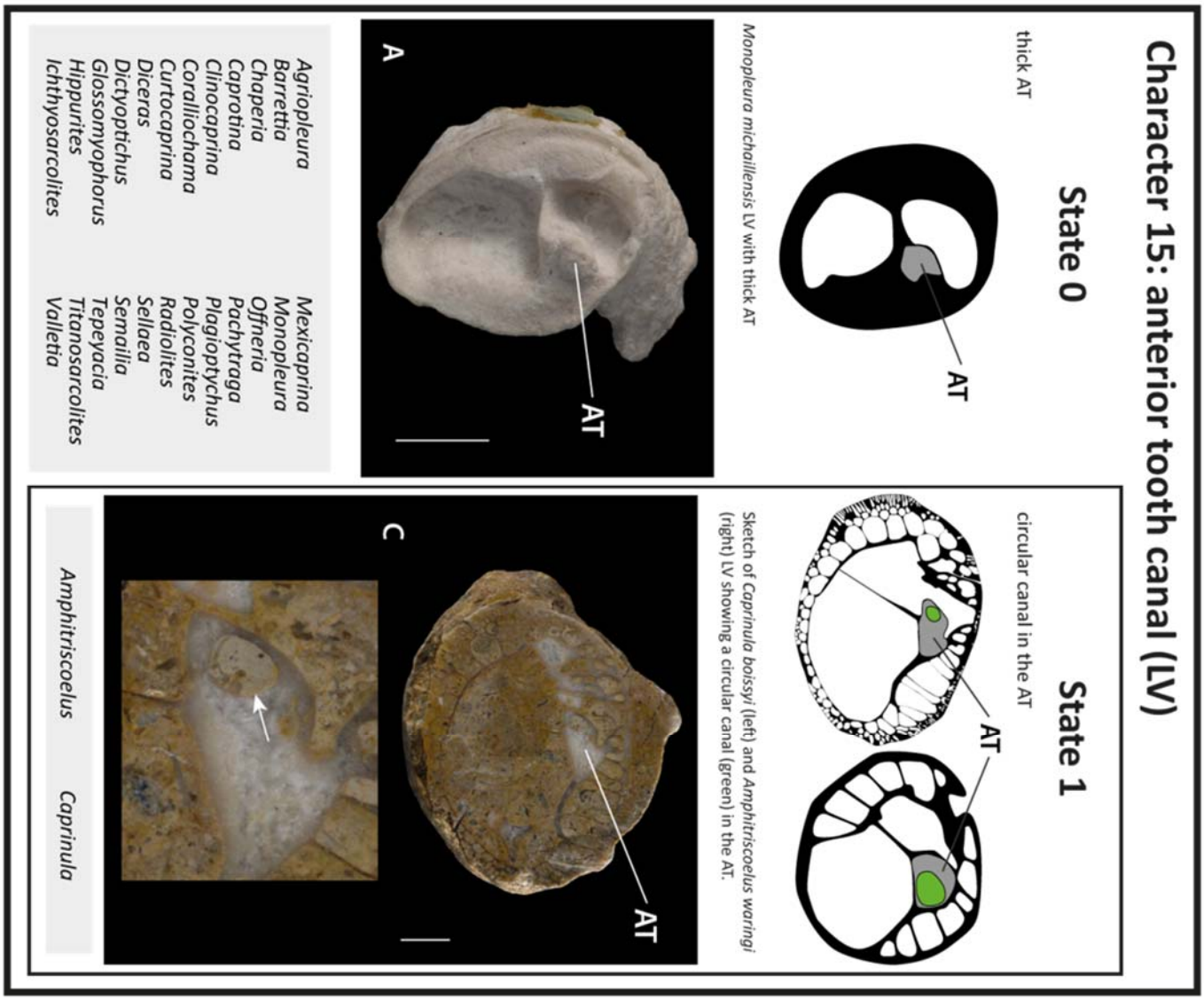
Retention index

0.67

Remarks

The anterior tooth canal is probably unrelated to pallial canals of the LV, and most likely to be a neoformation. Other tooth canals can be seen in *Alencasteria*, *Stellacaprina*.

Figure 4. 53. Hierarchical character #15. A – *Monopleura michaillensis* (UPMC.1272). B – *Caprinula boissyi* (MNHN.6924). Scale bars, 10 mm.


Character 15: anterior tooth canal (LV)
State 0
thick AT
AT
Monopleura michaillensis LV with thick AT
AT
A
Agriopleura
Barrettia
Chaperia
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia
Missing data: Oryxia
State 1
circular canal in the AT
AT
Sketch of Caprinula boissyi (left) and Amphitriscoelus waringi (right) LV showing a circular canal (green) in the AT.
AT
C
Amphitriscoelus
Caprinula

7.16. Character 16

Name: capillary canals in the myophore

Position: left valve

Kind: canals

Coding: (0(1))

States definitions

0 thick anterior and posterior myophores

1 capillary canals in AM and PM

States tests

1 rejected – *Coralliochama*, clade 20 (except *Dictyoptychus*, *Ichthyosarcolites*)

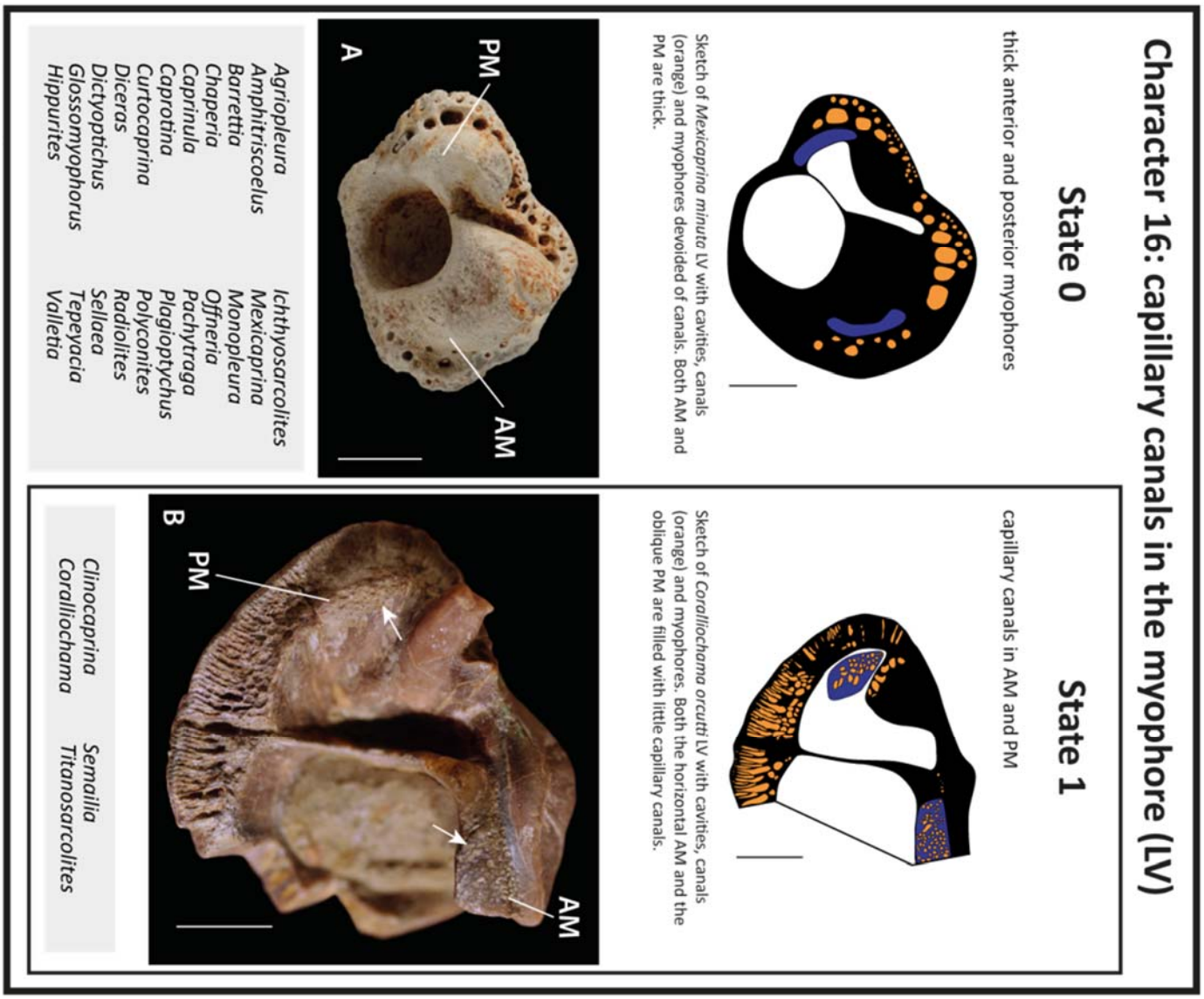
Retention index

0.64

Remarks

These canals are likely to be derived pallial canals, the morphology of myophoral canals being approximately the same as other pallial canals in the shell.

Figure 4. 54. Hierarchical character #16. A – *Mexicaprina minuta* (JPMA.14097-5). B – *Coralliochama orcutti* (USNM.186662). Scale bars, 10 mm.


Character 16: capillary canals in the myophore (LV)
State 0
thick anterior and posterior myophores
State 1
capillary canals in AM and PM
Sketch of Mexicaprina minuta LV with cavities, canals (orange) and myophores devoided of canals. Both AM and PM are thick.
Sketch of Coralliochama orcutti LV with cavities, canals (orange) and myophores. Both the horizontal AM and the oblique PM are filled with little capillary canals.
PM
AM
A
AM
B
PM
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Tepeyacia
Valletia
Clinocaprina
Coralliochama
Semailia
Titanosarcolites
Missing data: Oryxia

7.17. Character 17

Name: tabulae in the general cavity G

Position: left valve

Kind: horizontal blades

Coding: (0(1))

States definitions

0 empty G

1 tabulae in G

States tests

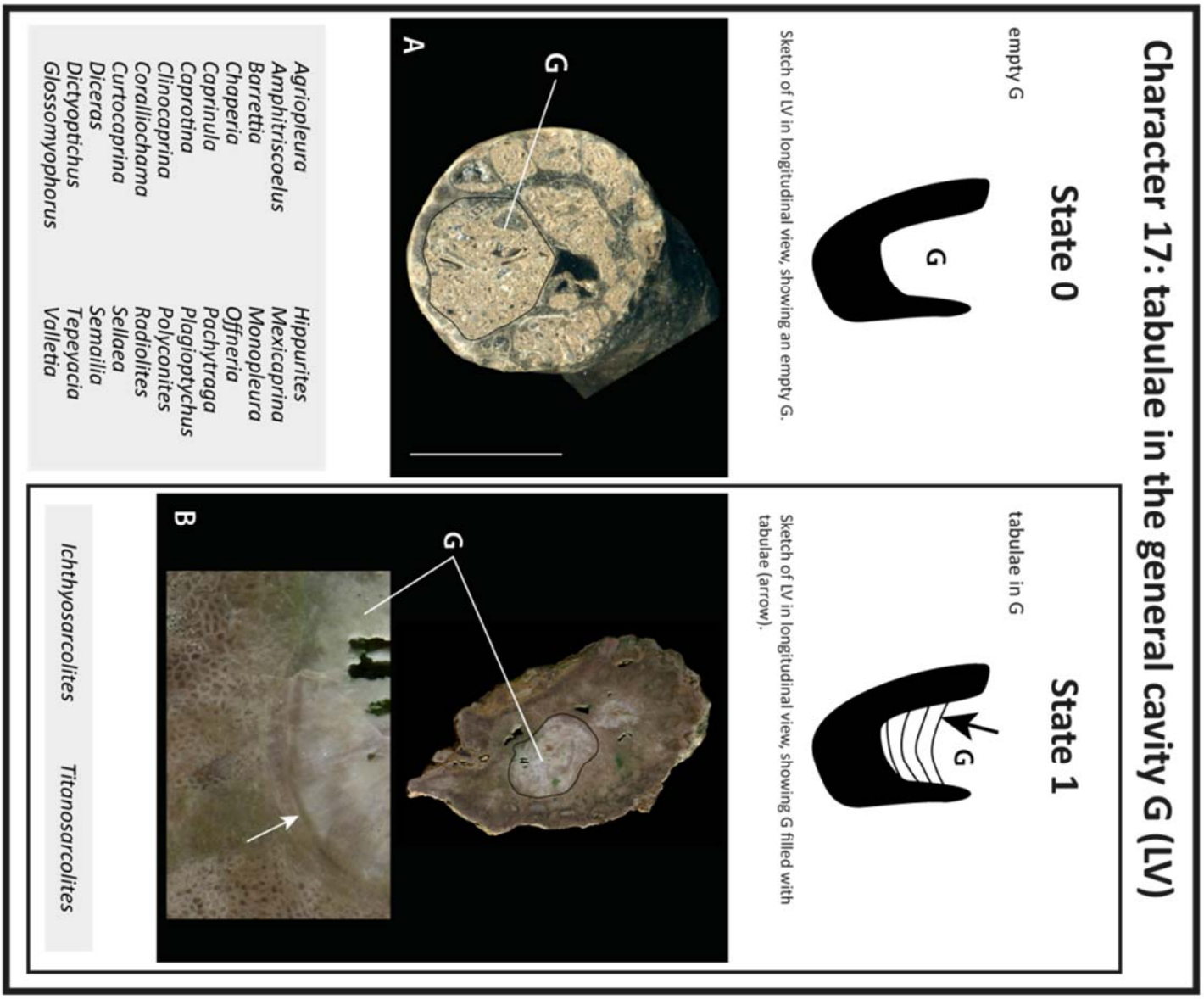
1 passed – clade 24

Retention index

1.00

Remarks

Figure 4. 55. Hierarchical character #17. A – *Amphitriscoelus waringi*. B – *Titanosarcolites giganteus* (UT.47533). Scale bars, 10 mm.



Missing data: ***Oryxia***

7.18. Character 18

Name: tabulae in pallial canals

Position: left valve

Kind: horizontal blades

Coding: (0(1))

States definitions

0 root

1 tabulae in pallial canals

States tests

1 rejected – clade 24, *Offneria*

Retention index

0.83

Remarks

Related to character 8 of Chartrousse (1994; with #34). Related partially to character 21 of Mitchell (2013a) (with #34).

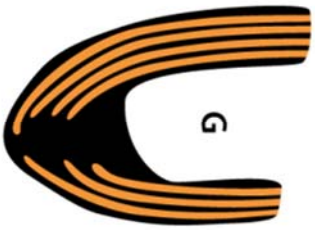
Generation of tabulae in pallial canals are not ryhtmic in rudists, the generation of tabulae being asynchronous with canal growth.

Figure 4. 56. Hierarchical character #18. A – *Caprinula boissyi* (MNHN.6924). B – *Offneria rhodanica* (JPMA.16494). Scale bars, 10 mm.

Character 18: tabulae in the pallial canals (LV)

State 0

[root]

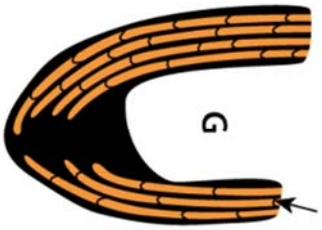


Sketch of canaliculated rudist LV in longitudinal view. Canals are shown in orange.

Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Mexicaprina
Monopleura
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Valletia

State 1

tabulae in pallial canals



Sketch of canaliculated rudist LV in longitudinal view. Canals shown in orange are filled with tabulae.

Ichthyosarcolites
Offneria
Titanosarcolites

Missing data: ***Oryxia***

7.19. Character 19

Name: tabulae in the central tooth socket

Position: left valve

Kind: horizontal blades

Coding: (0(1))

States definitions

0 root

1 tabulae in CTS

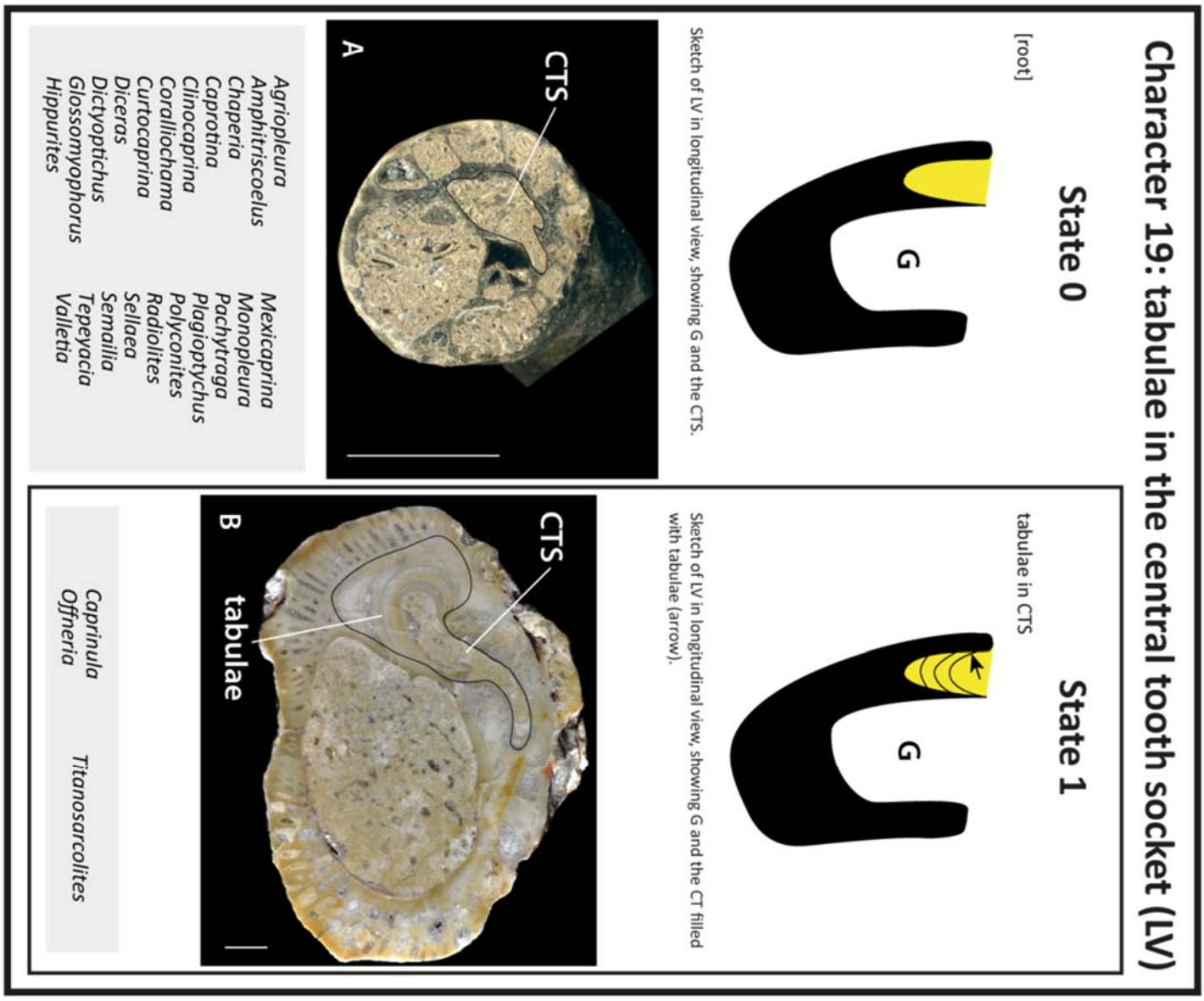
States tests

1 rejected – clade 17, *Titanosarcolites*

Retention index

0.83

Figure 4. 57. Hierarchical character #19. A – *Amphitriscoelus waringi*. B – *Offneria rhodanica* (UCBL.R275). Scale bars, 10 mm.


Character 19: tabulae in the central tooth socket (LV)
State 0
[root]
G
Sketch of LV in longitudinal view, showing G and the CTS.
CTS
A
Agriopleura
Amphitriscoelus
Chaperia
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Mexicaprina
Monopleura
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Valletia
State 1
tabulae in CTS
G
Sketch of LV in longitudinal view, showing G and the CT filled with tabulae (arrow).
CTS
B
tabulae
Caprinula
Offneria
Titanosarcolites
Missing data: Barrettia, Oryxia, Ichthyosarcolites

7.20. Character 20

Name: calcitic canals of the canal/pore system

Position: left valve

Kind: canal/pore system

Coding: (0(1))

States definitions

0 compact calcitic outer shell layer

1 radial canals in the outer calcitic shell layer

States tests

1 passed – clade 14

Retention index

1.00

Remarks

Related to character 13 of Steuber (1999; with #21). Related to character 4 of Skelton and Smith (2014) (with #21).

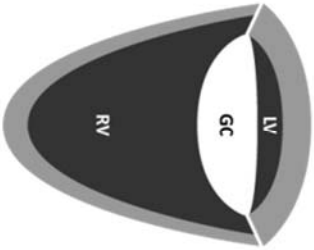
Pores and canals are coded as separated homologs herein. These structures are considered independent, their morphologies being dissociated in Hippuritids with different evolutionary rates.

Figure 4. 58. Hierarchical character #20. A – *Radiolites radiosus* (MNHN.F.J09021; umbonal view). B – *Hippurites radiosus* (MNHN.F.A58757; umbonal view). Scale bars, 10 mm.

Character 20: calcitic canals of the canal/pore system (LV)

State 0

compact calcitic outer shell layer

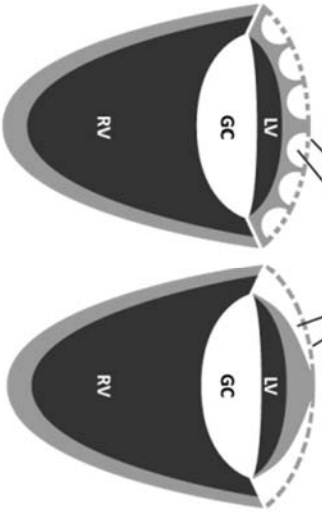


Sketch of rudist bivalve with thick valves. Upper valve: LV; lower valve: RV; grey: calcitic outer shell layer; black: aragonitic inner shell layer; GC: general cavity (G+D).

Agriopleura
Amphitriscoelus
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia

State 1

radial canals in the outer calcitic shell layer



Sketch of Hippuritid pore system. The pore system situated in the calcitic outer shell layer is constituted of canals arranged radially and covered by a thin layer filled with pores. Left: tangential view; right: transversal view. Upper valve: LV; lower valve: RV; grey: calcitic outer shell layer; black: aragonitic inner shell layer; GC: general cavity (G+D).

Barrettia
Hippurites

Missing data: ***Oryxia***

7.21. Character 21

Name: calcitic pores of the canal/pore system

Position: left valve

Kind: canal/pore system

Coding: (0(1))

States definitions

0 compact outer shell layer

1 pores in the outer calcitic shell layer

States tests

1 passed – clade 14

Retention index

1.00

Remarks

Related to character 13 of Steuber (1999) (with #20). Related to character 4 of Skelton and Smith (2014) (with #20).

Pores and canals are coded as separated homologs herein. These structures are considered as independent, their morphologies being dissociated in Hippuritids with different evolutionary rates.

Figure 4. 59. Hierarchical character #21. A – *Radiolites radiosus* (MNHN.F.J09021; umbonal view). B – *Hippurites radiosus* (MNHN.F.A58757; umbonal view). C – *Barrettia monilifera* (fig. 3D in Mitchell 2010; UWIGM.2010.01.0003). Scale bars, 10 mm.

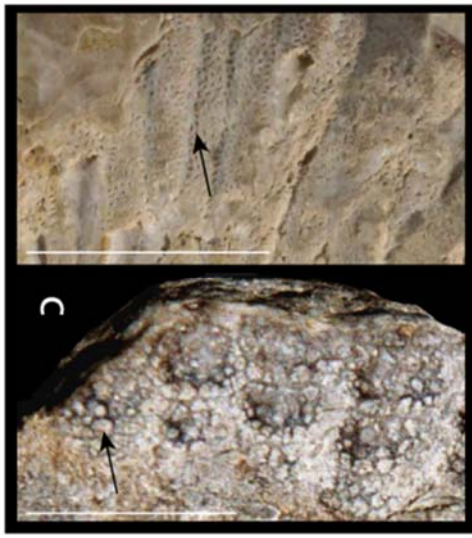
Character 21: calcitic pores of the canal/pore system (LV)

State 0

compact calcitic outer shell layer

State 1

pores in the outer calcitic shell layer



Agriopleura
Amphitriscoelus
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia

Barrettia
Hippurites

Missing data: ***Oryxia***

7.22. Character 22

Name: left valve coiling

Position: left valve

Kind: coiling

Coding: (0(1))

States definitions

0 coiled left valve

1 complete uncoiling of the left valve

States tests

1 passed – clade 12

Retention index

1.00

Remarks

Related partially to character 2 of Smith (1994). Related to character 15 of Skelton and Smith (2000). Related to character 2 of Sano *et al.* (2014).

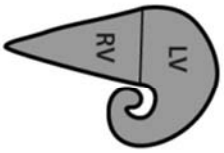
Only shells with the umbo at the center of the shell in umbonal view are considered as fully uncoiled.

Figure 4. 60. Hierarchical character #22. A – *Glossomyophorus costatus* (MNHN.F.R06779). B – *Hippurites radiosus* (MNHN.F.A58757). C – *Radiolites radiosus* (MNHN.F.J09021). Scale bars, 10 mm.

Character 22: left valve coiling (LV)

State 0

coiled left valve

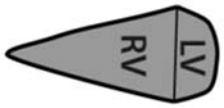


Sketch of valves differential coiling. The left valve is coiled whereas the right valve is uncoiled.

Amphitriscoelus
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Sellaea
Semailia
Titanosarcolites
Valletia

State 1

complete uncoiling of the left valve



Sketch of valves coiling, showing a LV completely uncoiled that can be concave(left) or convex (right).

Agriopleura
Barrettia
Hippurites
Radiolites

Missing data: Oryxia, *Tepeyacia*

7.23. Character 23

Name: diceratid tooth

Position: right valve

Kind: teeth

Coding: (0(1))

States definitions

0 presence of an anterior crescentic Diceratid Tooth (DT) along D

1 loss of the tooth

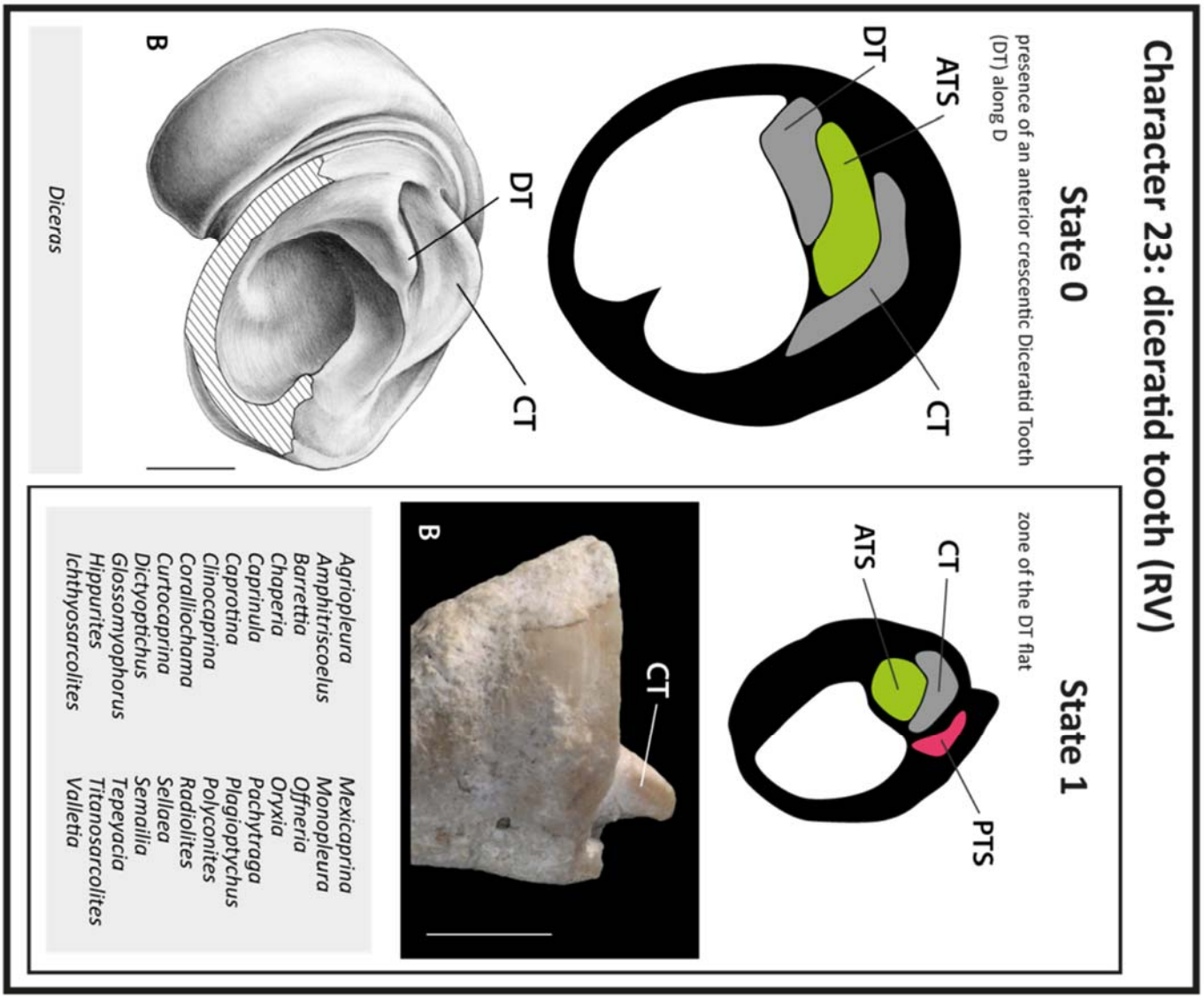
States tests

1 passed – clade 1

Retention index

1.00

Figure 4. 61. Hierarchical character #23. A – *Diceras arietinum* (MNHN.F.A52439). B – *Monopleura michaillensis* (UPMC.1272). Scale bars, 10 mm.


Character 23: diceratid tooth (RV)
State 0
presence of an anterior crescentic Diceratid Tooth (DT) along D
ATS
CT
DT
DT
CT
B
Diceras
State 1
zone of the DT flat
CT
ATS
PTS
CT
B
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Oryxia
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Valletia

7.24. Character 24

Name: central tooth morphology

Position: right valve

Kind: teeth

Coding: (0(1))

States definitions

0 crescent-shaped CT surrounding the ATS

1 conical-shaped CT between ATS and PTS

States tests

1 rejected – clade 12, 17, *Tepeyacia*, *Coralliochama*

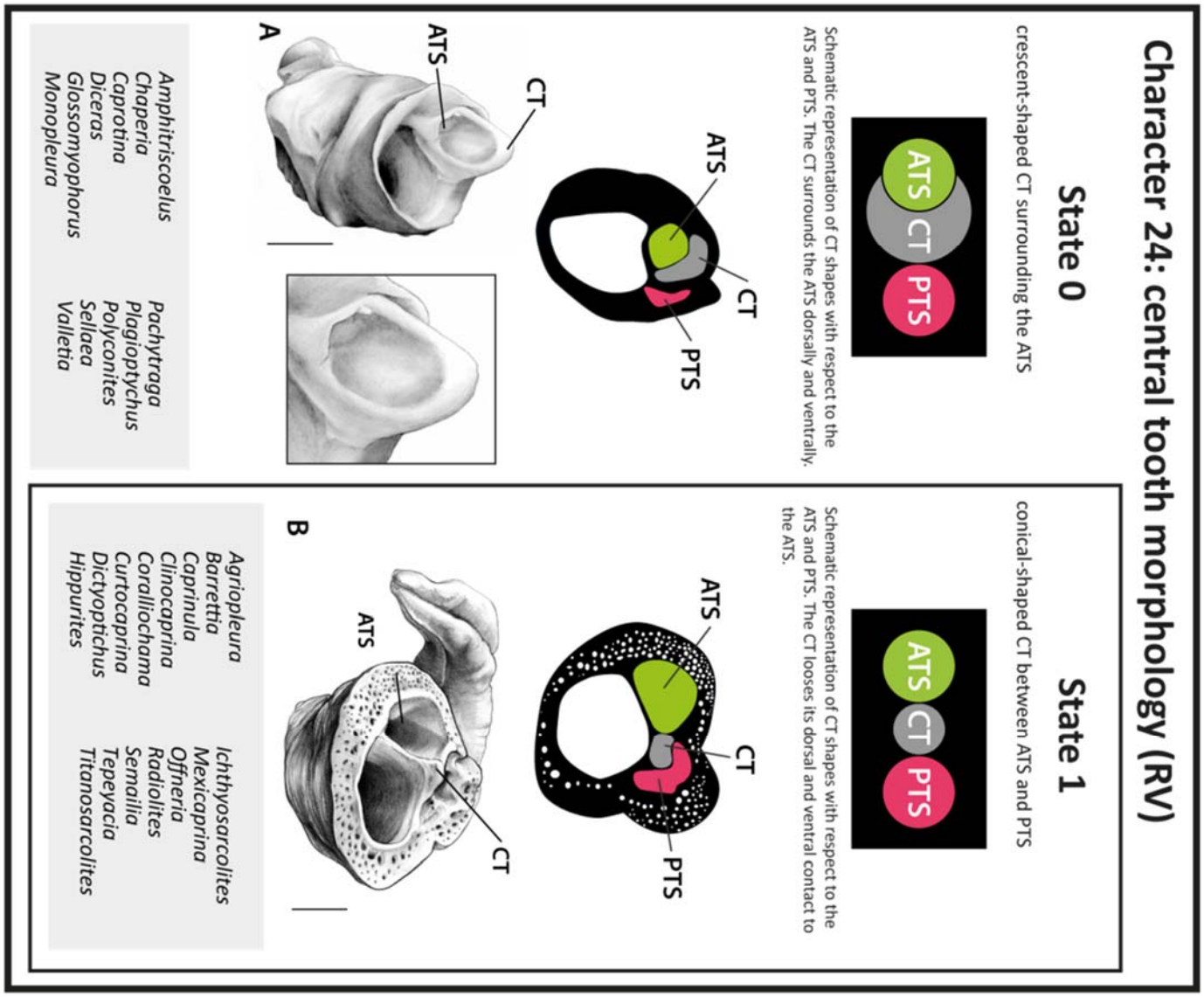
Retention index

0.86

Remarks

Related to character 18 of Skelton and Smith (2000). Related to character 4 of Sano *et al.* (2014; with #25).

Figure 4. 62. Hierarchical character #24. A – *Monopleura michaillensis* (UPMC.1272). B – *Mexicaprina minuta* (JPMA.14097-1). Scale bars, 10 mm.


Character 24: central tooth morphology (RV)
State 0
crescent-shaped CT surrounding the ATS
ATS
CT
PTS
Schematic representation of CT shapes with respect to the ATS and PTS. The CT surrounds the ATS dorsally and ventrally.
ATS
CT
PTS
CT
ATS
A
Amphitriscoelus
Chaperia
Caprotina
Diceras
Glossomyophorus
Monopleura
Pachytraga
Plagioptychus
Polyconites
Sellaea
Valletia
State 1
conical-shaped CT between ATS and PTS
ATS
CT
PTS
Schematic representation of CT shapes with respect to the ATS and PTS. The CT looses its dorsal and ventral contact to the ATS.
ATS
CT
PTS
CT
ATS
B
Agriopleura
Barrettia
Caprinula
Clinocaprina
Coralliochama
Curtocaprina
Dictyoptichus
Hippurites
Ichthyosarcolites
Mexicaprina
Offneria
Radiolites
Semalia
Tepeyacia
Titanosarcolites
Missing data: Oryxia

7.25. Character 25

Name: central tooth orientation

Position: right valve

Kind: teeth

Coding: (0(1))

States definitions

0 CT curved to the posterior side

1 CT straight

States tests

1 passed – clade 3 (except *Polyconites*, *Pachytraga*)

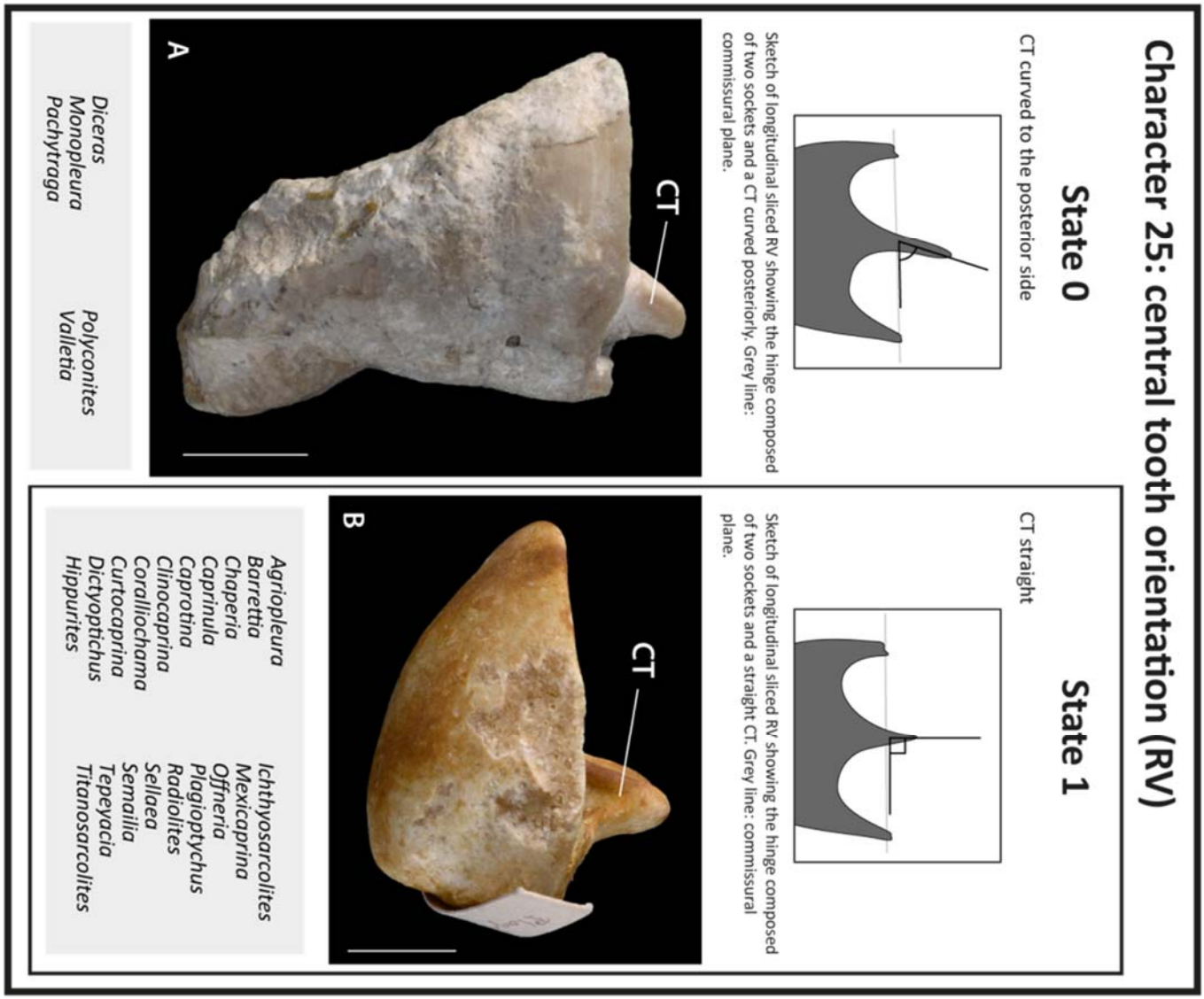
Retention index

0.84

Remarks

Related to character 4 of Sano *et al.* (2014; with #24).

Figure 4. 63. Hierarchical character #25. A – *Monopleura michaillensis* (UPMC.1272). B – *Plagioptychus arnaudi* (UPMC without number). Scale bars, 10 mm.


Character 25: central tooth orientation (RV)
State 0
CT curved to the posterior side
Sketch of longitudinal sliced RV showing the hinge composed of two sockets and a CT curved posteriorly. Grey line: commissural plane.
CT
A
Diceras
Monopleura
Pachytraga
Polyconites
Valletia
State 1
CT straight
Sketch of longitudinal sliced RV showing the hinge composed of two sockets and a straight CT. Grey line: commissural plane.
CT
B
Agriopleura
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Dictyopticlus
Hippurites
Ichthyosarcolites
Mexicaprina
Offneria
Plagioptychus
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Missing data: Amphitriscoelus, Glossomyophorus, Oryxia

7.26. Character 26

Name: posterior myophore morphology

Position: right valve

Kind: myophores

Coding: (0(1)(2(3(4))(5(6))))

States definitions

0 PM on a blade inserted behind the CT in D

1 PM vertical in D

2 PM on the commissure and forming a horizontal shelf posterior to the CT

3 PM included in the PTS, vertical and outwardly oriented

4 outwardly verticalized PM projecting into the LV by PTS/D wall development

5 PM included in the PTS, vertical and inwardly oriented

6 posterior myophoral cavity distinct from the PTS

States tests

1 rejected – clade 6, 14, 20 (except *Clinocaprina*, *Dictyoptychus*, *Semailia*, *Titanosarcolites*, *Ichthyosarcolites*)

2 rejected – clades 6, 9, 14, 15 (except *Curtocaprina*), *Valletia*, *Monopleura, Sellaea*

3 rejected – clades 4, 18, *Sellaea*, *Pachytraga*

4 rejected – clades 18, *Pachytraga*

5 rejected – clades 15, 19 (except *Curtocaprina, Oryxia*), *Amphitriscoelus, Sellaea*

6 passed – clade 21 (except *Oryxia*)

Retention index

0.59

Remarks

Related partially to character 1 of Chartrousse (1994; #26.2, #26.3, #26.4). Related to character 5 of Skelton and Masse (1998; #26.2, #26.5). Related partially to character 10 of Steuber (1999; #26.1, #26.2).

This character includes a paralogous clade (independence instancy) and can be understood as four independent characters after free-paralogy subtree analysis (Nelson and Ladiges 1996) following procedure described in chapter 1.

The posterior myophore morphology, as LV myophores, is highly complex and show seven different states. The ancestral state is a thin myophoral blade visible in Diceratidae as the PM-LV ancestral state. Two states are quite similar to it: the PM being simply inserted in the posterior wall of D without any particular structure (state 1) and the PM-RV horizontal on a thick and well delineated myophoral shelf (state 2). No evidence of relation can be found between state 1 and 2. We choose to include both of them in the state 0 that became a paralogous clade. The state 1 is understood as resulting from a flattening of the myophoral blade, and the state 2 is seen oppositely as a strong development of the myophoral blade by aragonite accretion in a myophoral shelf. As other myophores, the myophore on its shelf is in some taxa tilted to contact the PTS and extends it (state 3). The tiltening can results in an inwardly oriented (state 3) or an outwardly oriented (state 5) vertical myophore, making the state 2 paralogous. When outwardly orientated, the myophore can rise out of the commissural plane by a blade (state 4). This state is considered by Chartrousse (1998) as the synapomorphy of the family Caprinidae. The state 6 is included in the state 5, where the PM-RV is surrounded by the PTS and inwardly oriented. We hypothesize this cavity to be separated in two parts in some taxa: the dorsal part collects the PT whereas the ventral part collects the myophoral blade of the PM-LV. *Ichthyosarcolites* show the two cavities (the new PTS and the PMC). It thus belongs to the state 6. However, its PM-LV does not insert in the PMC. The two features are not anymore aligned; this is an autapomorphic feature of no interest for clade support in our analysis.

Figure 4. 64. Hierarchical character #26, its character states intensions, and its treatment with paralogy-free subtree analysis.

Character 26: posterior myophore morphology (RV)

State 0

PM on a blade inserted behind the CT in D

State 1

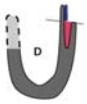
PM vertical in D

State 2

PM on the commissure and forming a horizontal shelf posterior to the CT

State 3

PM included in the PTS, vertical and outwardly oriented

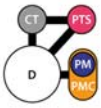


State 4

outwardly verticalized PM projecting into the LV by PTS/D wall development

State 5

PM included in the PTS, vertical and inwardly oriented



State 6

Posterior myophoral cavity distinct from the PTS

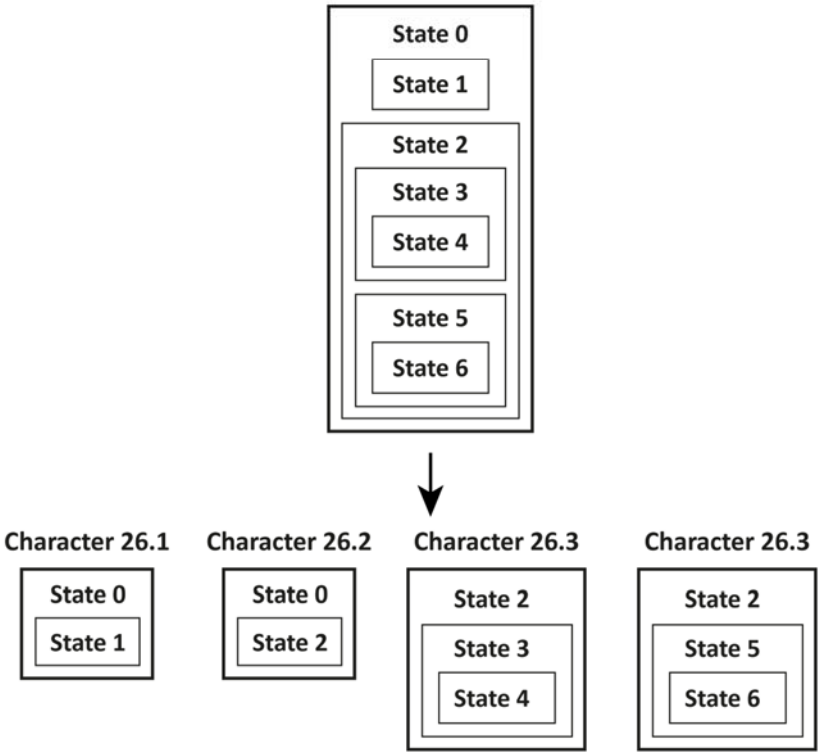


Figure 4. 65. Hierarchical character #26.1. A – *Diceras arietinum* (MNHN.F.A52439). B – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). Scale bars, 10 mm.

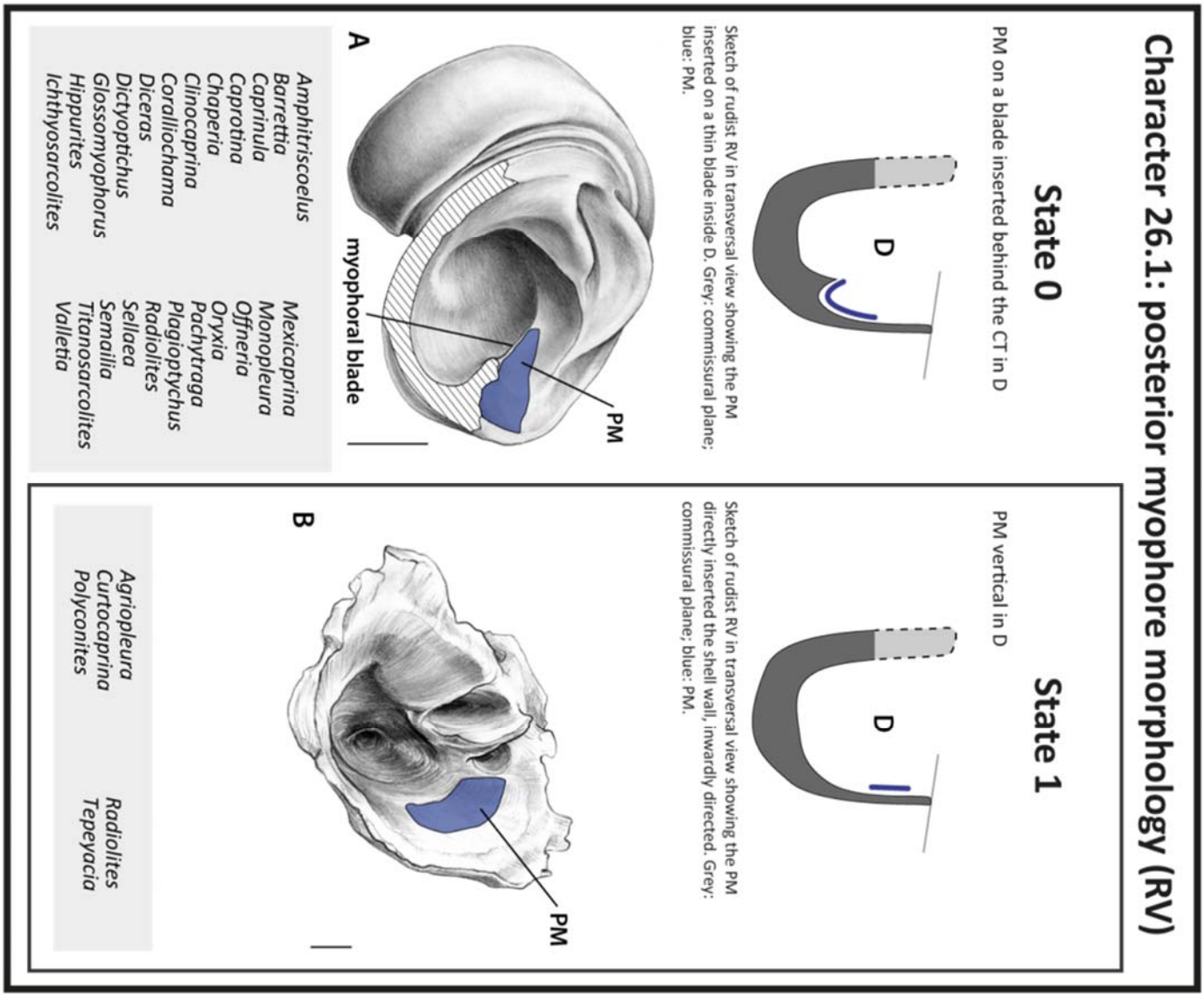

Character 26.1: posterior myophore morphology (RV)
State 0
PM on a blade inserted behind the CT in D
D
Sketch of rudist RV in transversal view showing the PM inserted on a thin blade inside D. Grey: commissural plane; blue: PM.
PM
A
myophoral blade
Amphitriscoelus
Barrettia
Caprinula
Caprotina
Chaperia
Clinocaprina
Coralliochama
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Oryxia
Pachytraga
Plagioptychus
Radiolites
Sellaea
Semailia
Titanosarcolites
Valletia
State 1
PM vertical in D
D
Sketch of rudist RV in transversal view showing the PM directly inserted the shell wall, inwardly directed. Grey: commissural plane; blue: PM.
PM
B
Agriopleura
Curtocaprina
Polyconites
Radiolites
Tepeyacia

Figure 4. 66. Hierarchical character #26.2. A – *Diceras arietinum* (MNHN.F.A52439). B – *Plagioptychus arnaudi* (UPMC.1273). Scale bars, 10 mm.

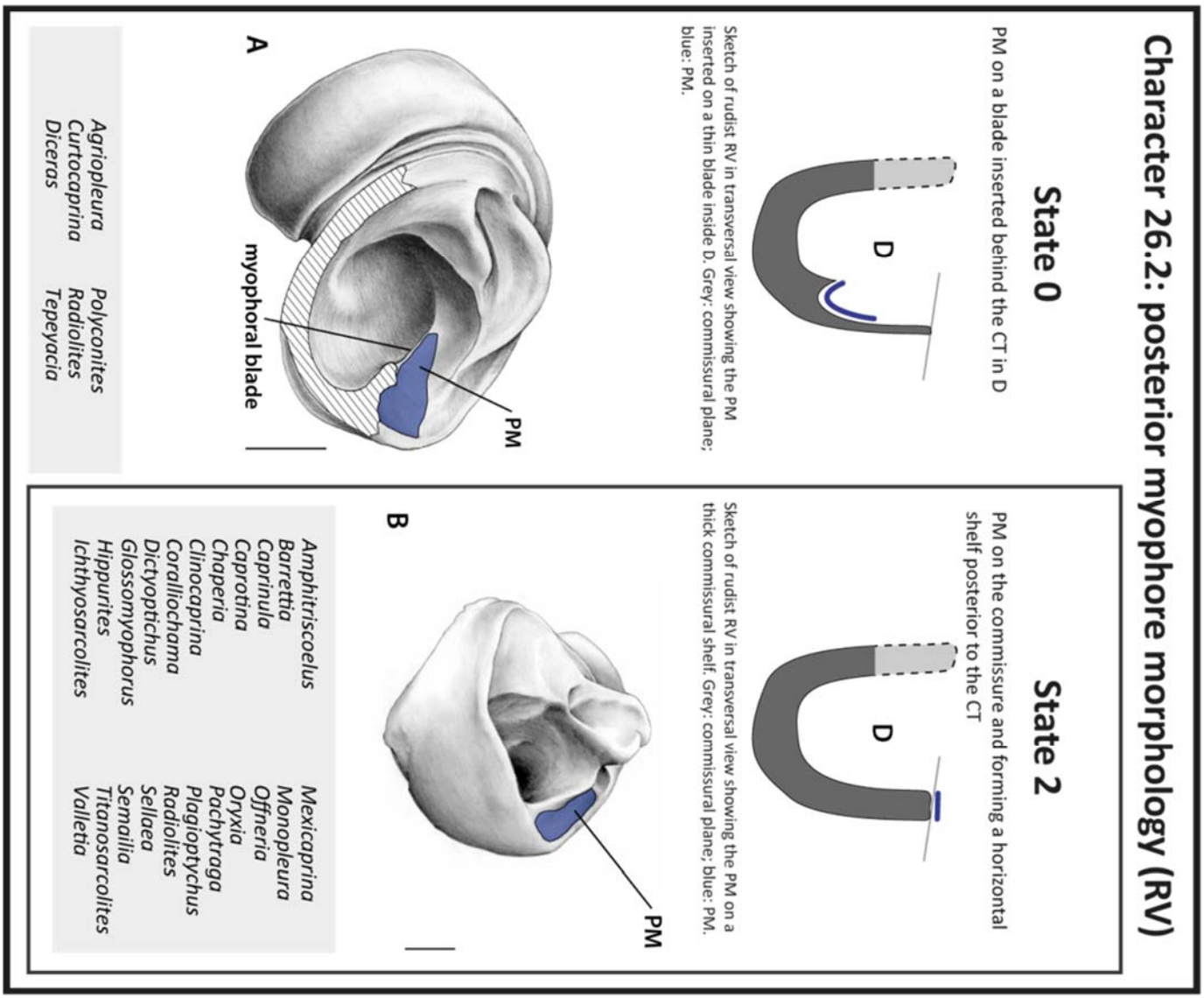
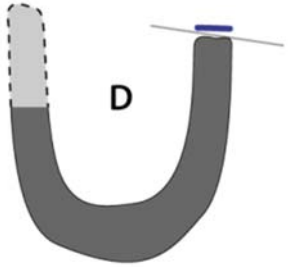

Character 26.2: posterior myophore morphology (RV)
State 0
PM on a blade inserted behind the CT in D
D
Sketch of rudist RV in transversal view showing the PM inserted on a thin blade inside D. Grey: commissural plane; blue: PM.
PM
myophoral blade
A
Agriopleura
Curtocaprina
Diceras
Polyconites
Radiolites
Tepeyacia
State 2
PM on the commissure and forming a horizontal shelf posterior to the CT
D
Sketch of rudist RV in transversal view showing the PM on a thick commissural shelf. Grey: commissural plane; blue: PM.
PM
B
Amphitriscoelus
Barrettia
Caprinula
Caprotina
Chaperia
Clinocaprina
Coralliochama
Dictyoptichus
Glossomyophorus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Oryxia
Pachytraga
Plagioptychus
Radiolites
Sellaea
Semailia
Titanosarcolites
Valletia

Figure 4. 67. Hierarchical character #26.3. A – *Plagioptychus arnaudi* (UPMC.1273). B – *Caprotina costata* (MV 2003.2.5147). C – *Offneria rhodanica* (UCBL.R781). Scale bars, 10 mm.

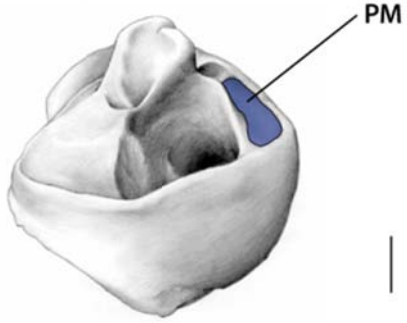
Character 26.3: posterior myophore morphology (RV)

State 2

PM on the commissure and forming a horizontal shelf posterior to the CT



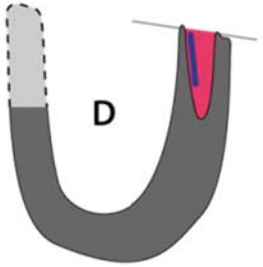
Sketch of rudist RV in transversal view showing the PM on a thick commissural shelf. Grey: commissural plane; blue: PM.



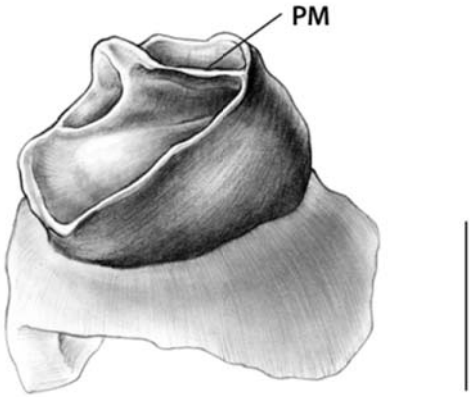
Amphitriscoelus
Barrettia
Clinocaprina
Coralliochama
Dictyoptichus
Hippurites
Ichthyosarcolites
Mexicaprina
Monopleura
Oryxia
Plagioptychus
Radiolites
Sellaea
Semailia
Titanosarcolites
Valletia

State 3

PM included in the PTS, vertical and outwardly oriented



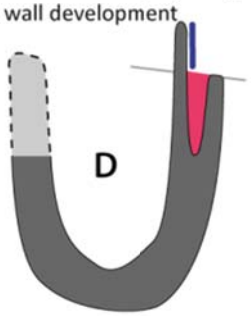
Sketch of rudist RV in transversal view showing the PM in the PTS. The PM is attached vertically on the PTS/D wall and becomes directed outwardly. Grey: commissural plane; blue: PM; red: PTS.



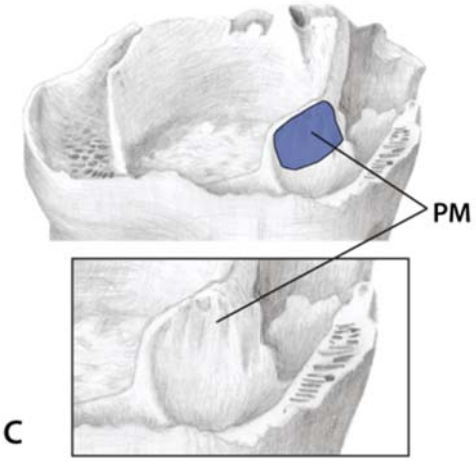
Caprotina
Chaperia
Glossomyophorus

State 4

outwardly verticalized PM projecting into the LV by PTS/D wall development



Sketch of rudist RV in transversal view showing the PM on the PTS/D wall extended beyond the commissure. The PM is vertical, directed outwardly. Grey: commissural plane; blue: PM; red: PTS.



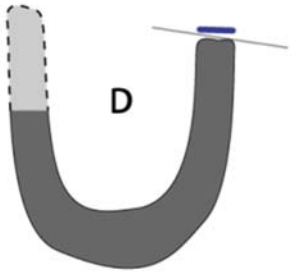
Caprinula
Offneria
Pachytraga

Figure 4. 68. Hierarchical character #26.4. A – *Plagioptychus arnaudi* (UPMC.1273). B – *Amphitriscoelus waringi*. C – *Titanosarcolites giganteus* (UT.47533). Scale bars, 10 mm.

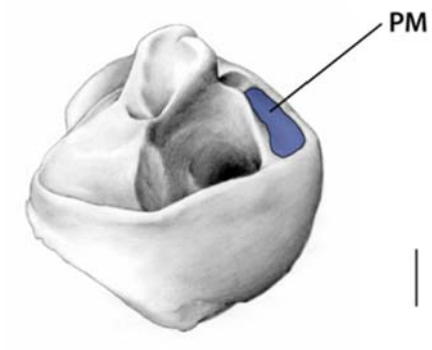
Character 26.4: posterior myophore morphology (RV)

State 2

PM on the commissure and forming a horizontal shelf posterior to the CT



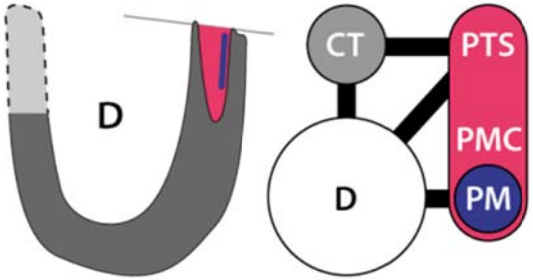
Sketch of rudist RV in transversal view showing the PM on a thick commissural shelf. Grey: commissural plane; blue: PM.



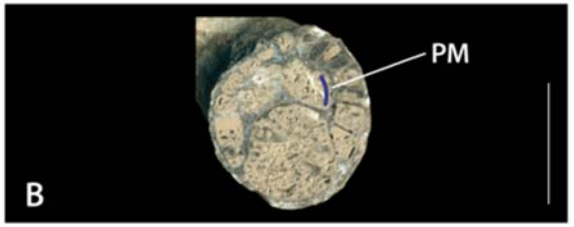
Caprinula
Caprotina
Chaperia
Coralliochama
Glossomyophorus
Monopleura
Offneria
Oryxia
Pachytraga
Plagioptychus
Radiolites
Valletia

State 5

PM included in the PTS, vertical and inwardly oriented



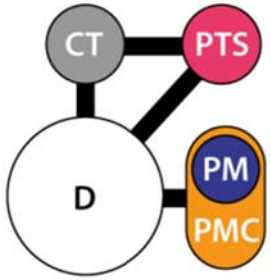
Left: Sketch of rudist RV in transversal view showing the PM falling in the PTS. The PM is attached vertically on the external wall of the PTS and becomes directed inwardly. Right: sketch of the main elements around the PM and their connections. The PM is inserted inside the PTS/PMC. Grey line: commissural plane; blue: PM; red: PTS.



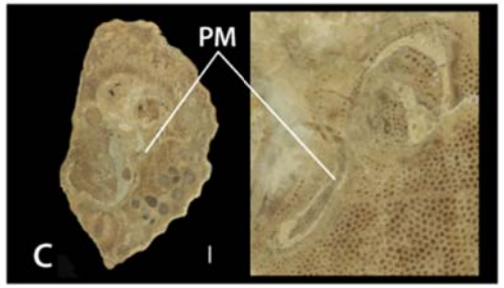
Amphitriscoelus
Barrettia
Clinocaprina
Hippurites
Mexicaprina
Sellaea

State 6

Posterior myophoral cavity distinct from the PTS



Sketch of the main elements around the PM and their connections. The PTS/PMC is divided in two daughter cavities: the PTS and the PMC. The PM is inserted inside the PMC. Grey line: commissural plane; blue: PM; red: PTS.



Dictyoptichus
Ichthyosarcolites
Semailia
Titanosarcolites

7.27. Character 27

Name: anterior myophore morphology

Position: right valve

Kind: myophores

Coding: (0(1(2))

States definitions

0 AM on the inner shell in D

1 AM on a horizontal shelf

2 AM vertical inwardly oriented by shelf digging because of the ATS development anteriorly on the dorsal side of the shell

States tests

1 rejected – clades 9, clades 17 (except *Oryxia*, *Dictyoptychus*, *Curtocaprina*, *Clinocaprina*, *Ichthyosarcolites*), *Monopleura*, *Glossomyophorus*, Polyconites, *Sellaea*

2 passed – clades 19 (except *Oryxia*, *Dictyoptychus*, *Curtocaprina*, *Clinocaprina*, *Ichthyosarcolites*)

Retention index

0.31

Remarks

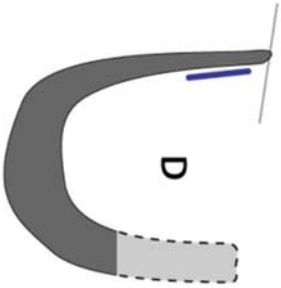
The ancestral state of the AM-RV is a vertical insertion of the inner shell. It can become inserted on a myophoral shelf that can be not delineated as well as PM-RV (see remarks of character #6). By inward tiltening, the myophore can fall into the ATS.

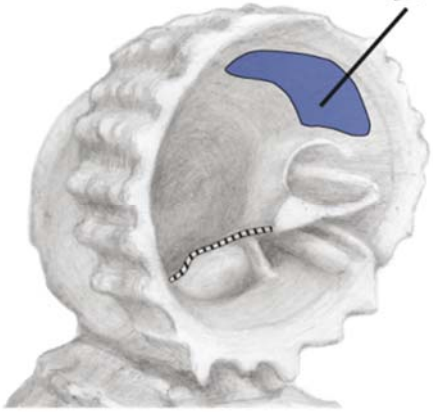
Figure 4. 69. Hierarchical character #27. A – *Chaperia costata* (MNHN.F.R07942). B – *Plagioptychus arnaudi* (UPMC.1273). C – *Mexicaprina minuta* (JPMA.14097-1). Scale bars, 10 mm.

Character 27: anterior myophore morphology (RV)

State 0

AM on the inner shell in D

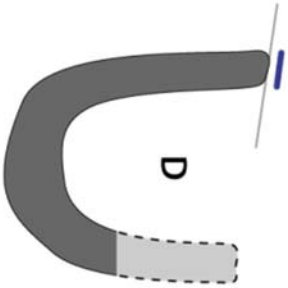


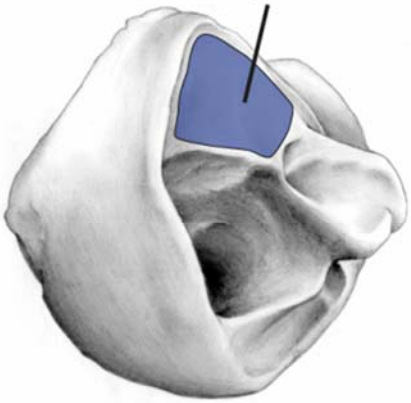


Agriopleura
Barrettia
Chaperia
Caprotina
Clinocaprina
Curtocaprina
Diceras
Dictyoptichus
Hippurites
Ichthyosarcolites
Oryxia
Pachytraga
Radiolites
Tepeyacia

State 1

AM on a horizontal shelf

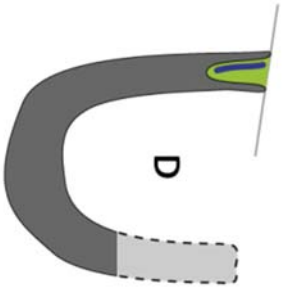


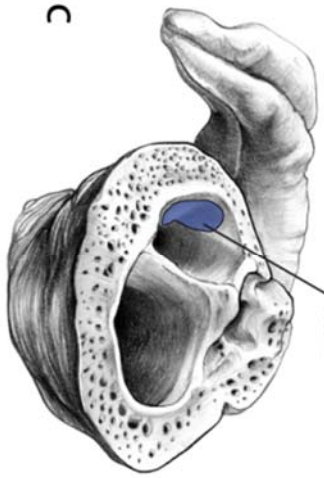


Coralliochama
Glossomyophorus
Monopleura
Offneria
Sellaea
Plagioptychus
Polyconites
Valletia

State 2

AM vertical inwardly oriented by shelf digging because of the ATS development anteriorly on the dorsal side of the shell





Mexicaprina
Titanosarcolites

Missing data: ***Amphitriscoelus, Caprinula, Semailia***

7.28. Character 28

Name: Position of the anterior tooth socket

Position: vertical blades

Kind: myophores

Coding: (0(1))

States definitions

0 ATS between CT and AM

1 ATS extended anteriorly to the AM

States tests

1 rejected – clades 16, *Sellaea, Pachytraga*

Retention index

0.92

Remarks

The characters 27 and 28 are independent. The anterior development of the ATS is compatible both with an inner shell insertion of the AM-RV, a surrounding of the myophore by the extended ATS or even with the development of a myophoral shelf.

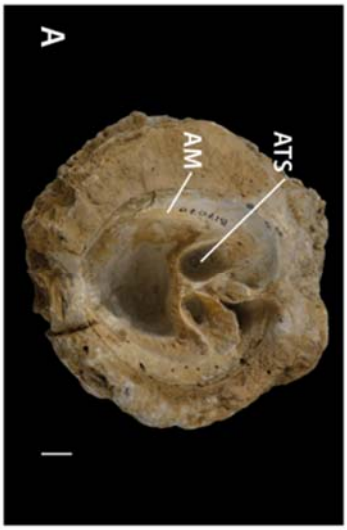
Figure 4. 70. Hierarchical character #28. A – *Radiolites radiosus* (MNHN.F.J09021). B – '*Sellaea' sp.* (WSA3058). Scale bars, 10 mm.

Character 28: Position of the anterior tooth socket

State 0

ATS between CT and AM

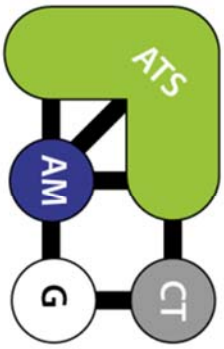
Sketch of ATS position with respect to the elements in connection with it. The ATS is dorsal to the AM and anterior to the CT



Agriopleura
Barrettia
Chaperia
Caprotina
Coralliochama
Diceras
Glossomyophorus
Hippurites
Monopleura
Plagioptychus
Polyconites
Radiolites
Tepeyacia
Valletia

State 1

ATS extended anteriorly to the AM



Sketch of ATS position with respect to the elements in connection with it. The ATS extend itself to surround dorsally and anteriorly the AM.

Amphitriscoelus
Caprinula
Clinocaprina
Curtocaprina
Dictyoptichus
Ichthyosarcolites
Mexicaprina
Oryxia
Offneria
Pachytraga
Sellaea
Semailia
Titanosarcolites

7.29. Character 29

Name: posterior tooth socket morphology

Position: right valve

Kind: vertical blades

Coding: (0(1(2)))

States definitions

0 root

1 presence of a Posterior Tooth Socket between the CT and the PM

2 PTS highly developed and including the PM

States tests

1 passed – clade 1

2 rejected – clades 5, 16 (except *Oryxia*), *Sellaea*, *Pachytraga*

Retention index

0.64

Remarks

Related partially to character 12 of Steuber (1999; #29.2).

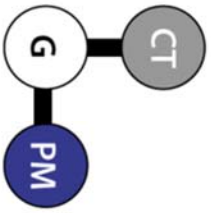
See remarks of character #28.

Figure 4. 71. Hierarchical character #29. A – *Diceras arietinum* (MNHN.F.A52439). B – *Radiolites radiosus* (MNHN.F.J09021). C – *Chaperia costata* (MNHN.F.R07942). Scale bars, 10 mm.

Character 29: posterior tooth socket morphology

State 0

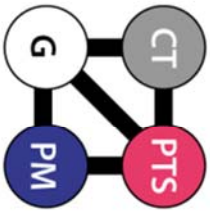
[root]



Sketch of the main elements of the RV.

State 1

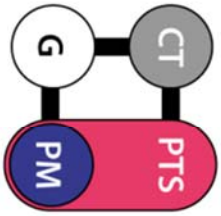
presence of a posterior tooth socket between the CT and the MP



Sketch of the main elements around the PTS and their connections on RV. The PTS appears between the CT on its anterior side, G on its antero-ventral side and the PM on its ventral side. Black line: contact.

State 2

PTS highly developed and including the PM



Sketch of the main elements around the PTS and their connections on RV. The PTS surrounds the PM and includes it. Black line: contact.

Missing data: ***Oryxia***

Diceras		*Agriopleura*	*Radiolites*
		Barrettia	*Plagioptychus*
		Coralliochama	*Polyconites*
		Glossomyophorus	*Tepeyacia*
		Hippurites	*Valletia*
		Monopleura	

Amphitriscoelus	*Mexicaprina*
Chaperia	*Offneria*
Caprinula	*Oryxia*
Caprotina	*Pachytraga*
Clinocaprina	*Sellaea*
Curtocaprina	*Semailia*
Dictyoptichus	*Titanosarcolites*
Ichthyosarcolites	

7.30. Character 30

Name: posterior and anterior sockets position

Position: right valve

Kind: canals

Coding: (0(1))

States definitions

0 root

1 ventral contact between the PTS and ATS

States tests

1 passed – clade 16

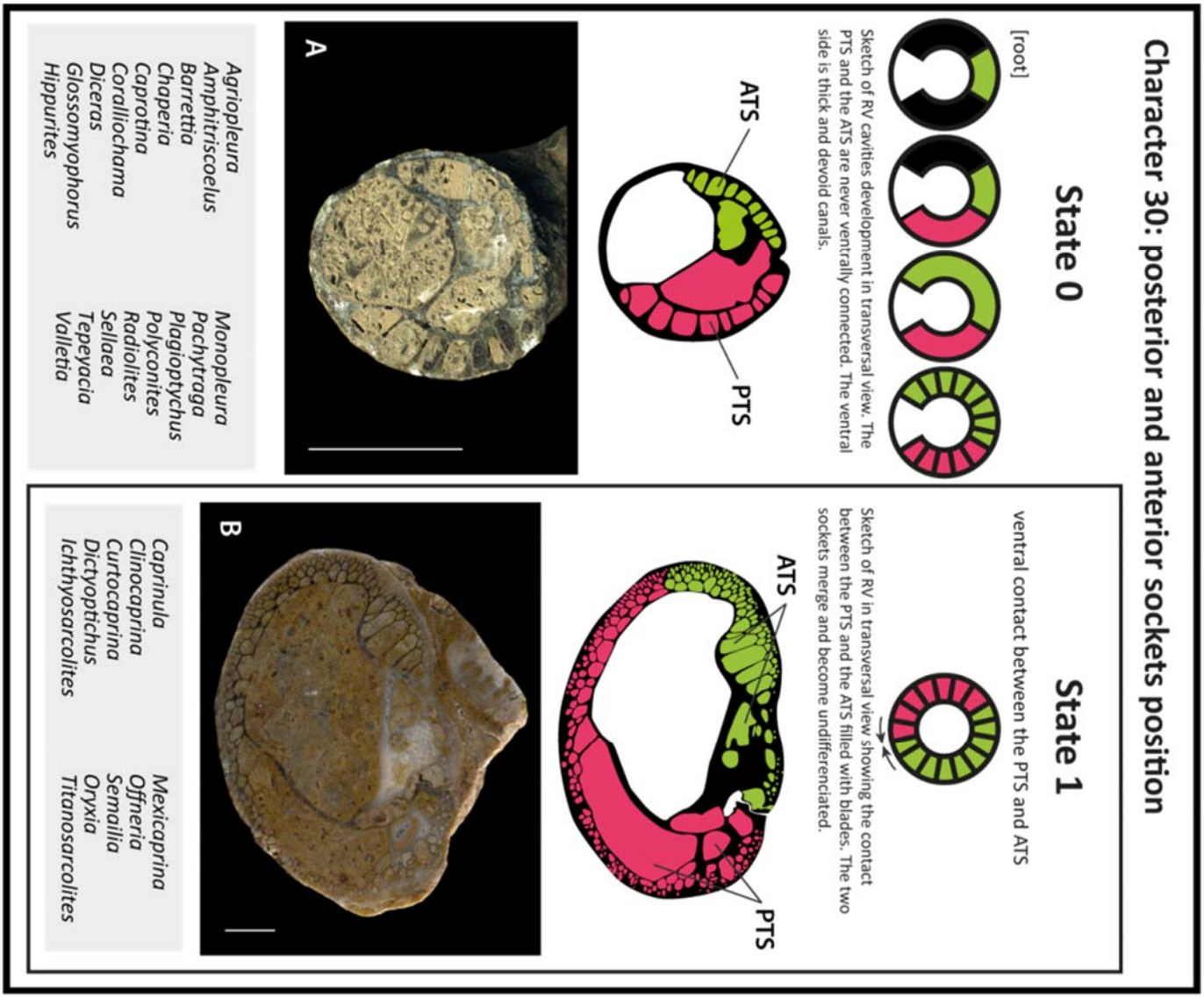
Retention index

1.00

Remarks

Related to character 9 of Chartrousse (1994). Related partially to character 19 of Mitchell (2013a; with #12).

Figure 4. 72. Hierarchical character #30. A – *Amphitriscoelus waringi*. B – *Caprinula boissyi* (MNHN.6924). Scale bars, 10 mm.


Character 30: posterior and anterior sockets position
State 0
[root]
Sketch of RV cavities development in transversal view. The PTS and the ATS are never ventrally connected. The ventral side is thick and devoid canals.
ATS
PTS
A
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprotina
Coralliochama
Diceras
Glossomyophorus
Hippurites
Monopleura
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Tepeyacia
Valletia
State 1
ventral contact between the PTS and ATS
Sketch of RV in transversal view showing the contact between the PTS and the ATS filled with blades. The two sockets merge and become undifferenciated.
ATS
PTS
B
Caprinula
Clinocaprina
Curtocaprina
Dictyoptichus
Ichthyosarcolites
Mexicaprina
Offneria
Semailia
Oryxia
Titanosarcolites

7.31. Character 31

Name: radial walls morphology

Position: right valve

Kind: canals

Coding: (0(1(2)))

States definitions

0 root

1 radial walls invading cavities and forming canals

2 walls differentiated in two morphologies leading to two morphologies of canals: a row of palisading canals along the external side of the shell wall, and capillary canals everywhere else

States tests

1 rejected – clade 15, *Pachytraga*, *Agriopleura*

2 passed – clade 16 (except *Offneria*, *Curtocaprina*, *Dictyoptychus*)

Retention index

0.92

Remarks

Related to character 11 of Chartrousse (1994). Related partially to character 11 in Smith (1994; #31.1). Related to character 10 of Stone and Telford (1999; #31.1).

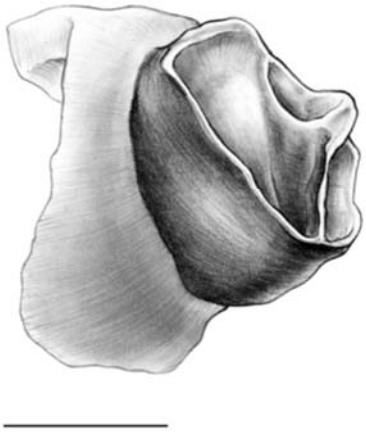
See remarks of character #12.

Figure 4. 73. Hierarchical character #31. A – *Caprotina costata* (MV 2003.2.5147). B – *Amphitriscoelus waringi*. C – *Ichthyosarcolites bicarinatus* (Upper Cenomanian, D559a road, south exit of La Bédoule, Var, France; picture provided by Marc Floquet). Scale bars, 10 mm.

Character 31: radial walls morphology (RV)

State 0

Sketch of RV cavities development in transversal view. The sockets devoid of walls.



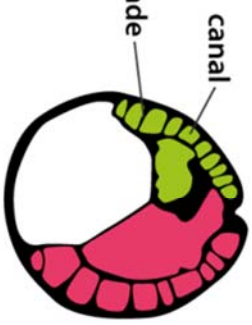
A

Barrettia	*Monopleura*
Chaperia	*Plagioptychus*
Caprotina	*Polyconites*
Coralliochama	*Radiolites*
Curtocaprina	*Sellaea*
Diceras	*Tepeyacia*
Glossomyophorus	*Valletia*
Hippurites	

State 1

radial walls invading cavities and forming canals

Sketch of RV cavities development in transversal view. The sockets (left: ATS; right: PTS) are filled with blades.

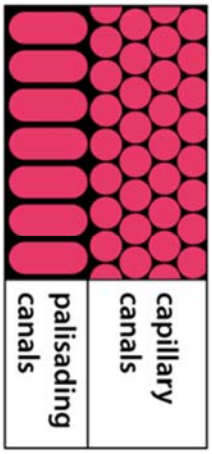


B

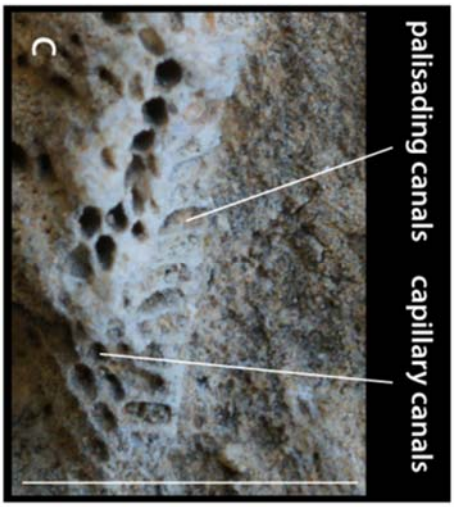
Agriopleura	*Offneria*
Amphitriscoelus	*Pachytraga*
Dictyoptichus	

State 2

walls diffrenciated in two morphologies leading to two morphologies of canals: a row of palisading canals along the external side of the shell wall, and capillary canals everywhere else.



Sketch of the two canal morpologies.



C

Caprinula	*Oryxia*
Clinocaprina	*Semailia*
Ichthyosarcolites	*Titanosarcolites*
Mexicaprina	

7.32. Character 32

Name: capillary canals in the central tooth

Position: right valve

Kind: canals

Coding: (0(1))

States definitions

0 compact shell in the CT

1 capillary canals in the CT

States tests

1 passed – clade 19 (except *Clinocaprina*, *Dictyoptychus*, *Semailia*)

Retention index

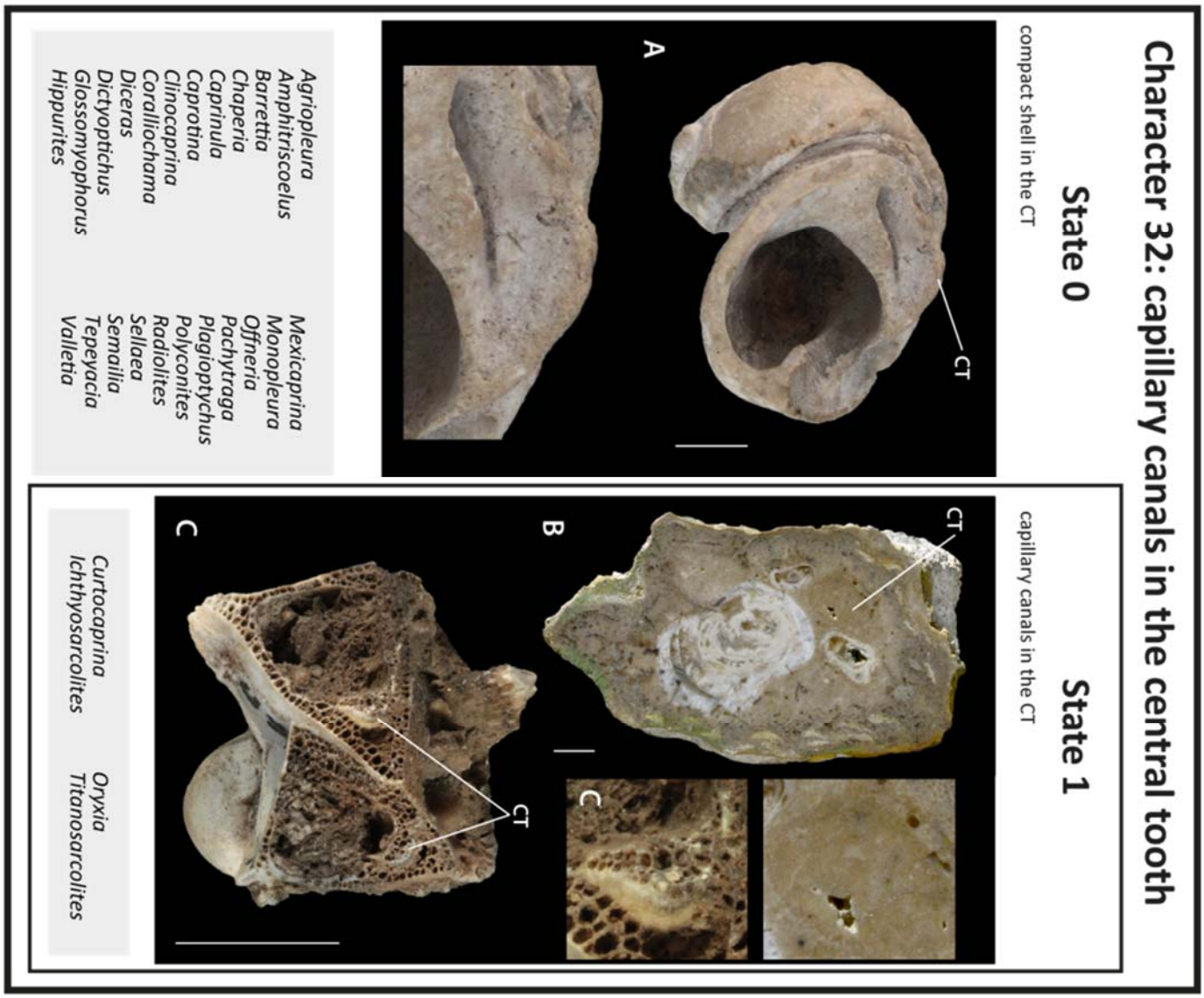
0.94

Remarks

Related to character 21 of Mitchell (2013a).

Because canals of the central tooth have always the same morphology as other pallial canals, they are hypothesized homologous and sharing the same origins.

Figure 4. 74. Hierarchical character #32. A – *Diceras arietinum* (MNHN.F.A52439). B – *Titanosarcolites giganteus* (UT.47533). C – *Curtocaprina clabaughikinsorum* (UT.10922). Scale bars, 10 mm.


Character 32: capillary canals in the central tooth
State 0
compact shell in the CT
State 1
capillary canals in the CT
CT
A
B
C
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Valletia
Curtocaprina
Ichthyosarcolites
Oryxia
Titanosarcolites

7.33. Character 33

Name: tabulae in the general cavity

Position: right valve

Kind: horizontal blades

Coding: (0(1))

States definitions

0 root

1 tabulae in D

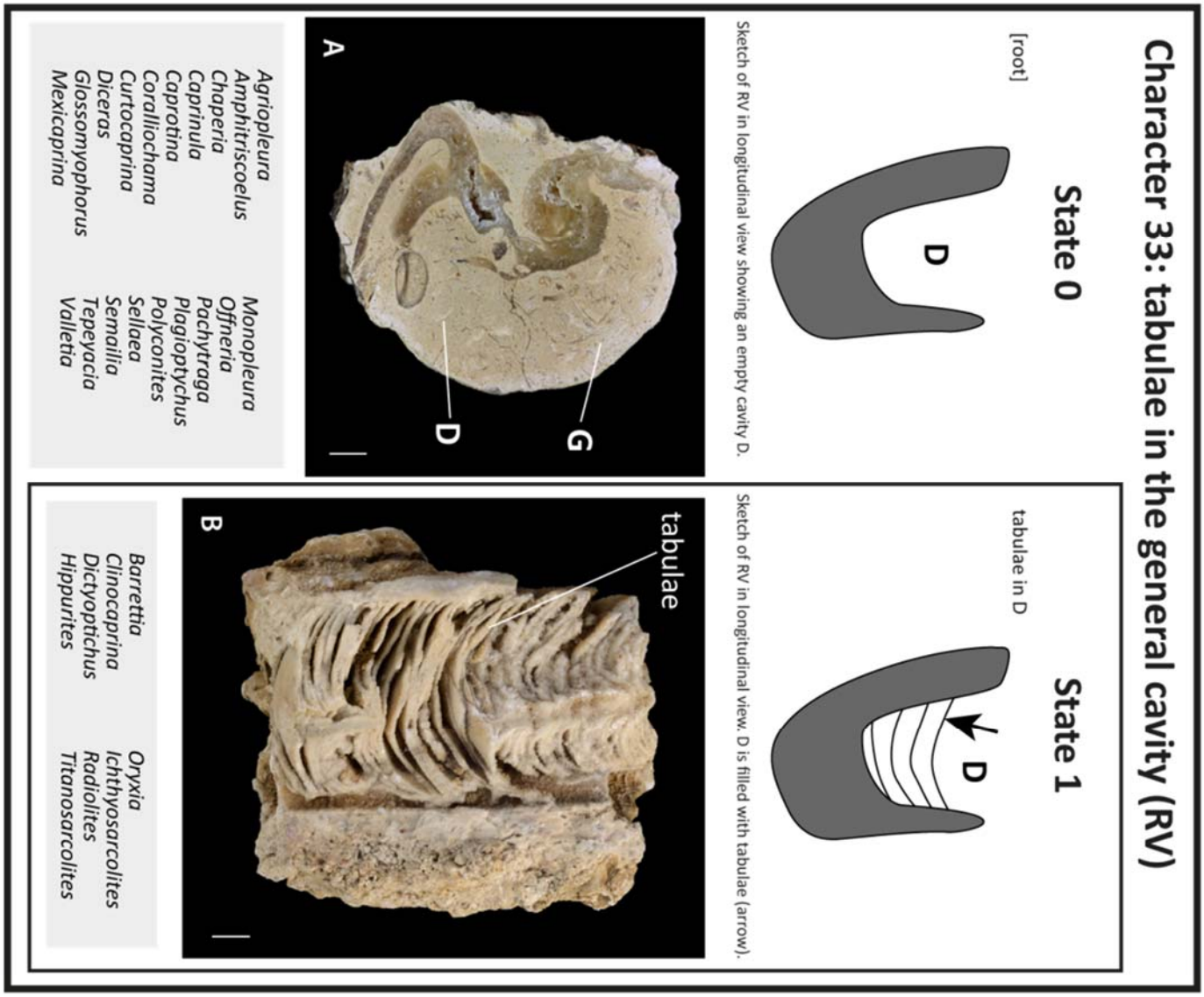
States tests

1 rejected – clades 13 (except *Agriopleura*), 21 (except *Semailia*)

Retention index

0.79

Figure 4. 75. Hierarchical character #33. A – '*Sellaea*' *sp.* (UT.48167). B – *Hippurites sp.* (UPMC without number). Scale bars, 10 mm.


Character 33: tabulae in the general cavity (RV)
State 0
[root]
D
Sketch of RV in longitudinal view showing an empty cavity D.
G
D
A
Agriopleura
Amphitriscoelus
Chaperia
Caprinula
Caprotina
Coralliochama
Curtocaprina
Diceras
Glossomyophorus
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Sellaea
Semailia
Tepeyacia
Valletia
State 1
tabulae in D
D
Sketch of RV in longitudinal view. D is filled with tabulae (arrow).
tabulae
B
Barrettia
Clinocaprina
Dictyoptichus
Hippurites
Oryxia
Ichthyosarcolites
Radiolites
Titanosarcolites

7.34. Character 34

Name: tabulae in pallial canals

Position: right valve

Kind: horizontal blades

Coding: (0(1))

States definitions

0 root

1 tabulae in pallial canals

States tests

1 rejected – clade 21 (except *Oryxia*, *Semailia*), *Offneria*

Retention index

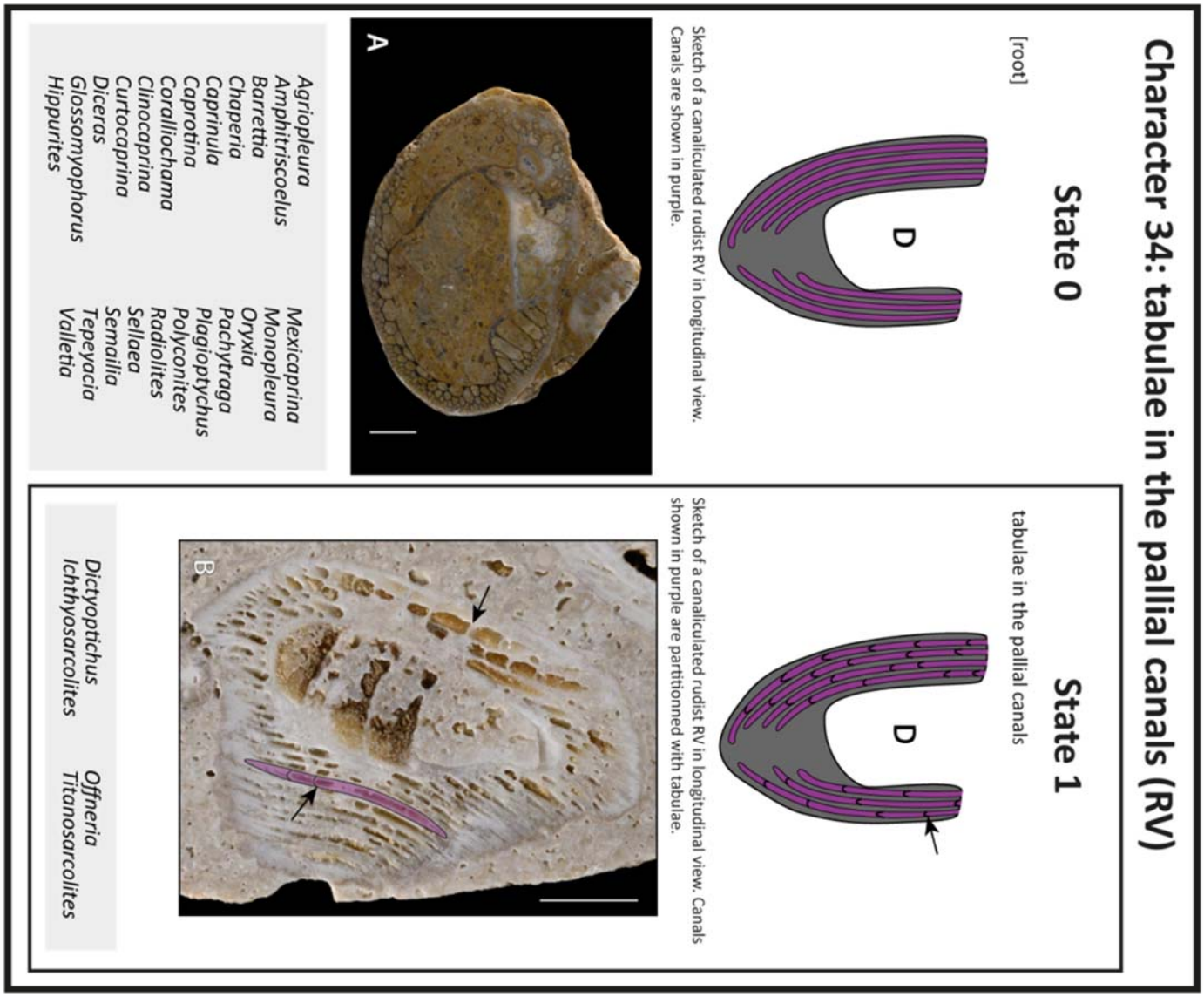
0.84

Remarks

Related to character 8 of Chartrousse (1994; with #18). Related partially to character 21 of Mitchell (2013a; with #18).

See remarks of character #18.

Figure 4. 76. Hierarchical character #34. A – *Caprinula boissyi* (MNHN.6924). B – *Caprinuloidea sp.* (JPMA.without number). Scale bars, 10 mm.


Character 34: tabulae in the pallial canals (RV)
State 0
[root]
D
Sketch of a canaliculated rudist RV in longitudinal view. Canals are shown in purple.
A
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Coralliochama
Clinocaprina
Curtocaprina
Diceras
Glossomyophorus
Hippurites
Mexicaprina
Monopleura
Oryxia
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Valletia
State 1
tabulae in the pallial canals
D
Sketch of a canaliculated rudist RV in longitudinal view. Canals shown in purple are partitionned with tabulae.
B
Dictyoptichus
Ichthyosarcolites
Offneria
Titanosarcolites

7.35. Character 35

Name: pillars

Position: both valves

Kind: pillars

Coding: (0(1))

States definitions

0 flat siphonal bands

1 infolding of the siphonal bands

States tests

1 rejected – clade 14, *Tepeyacia*

Retention index

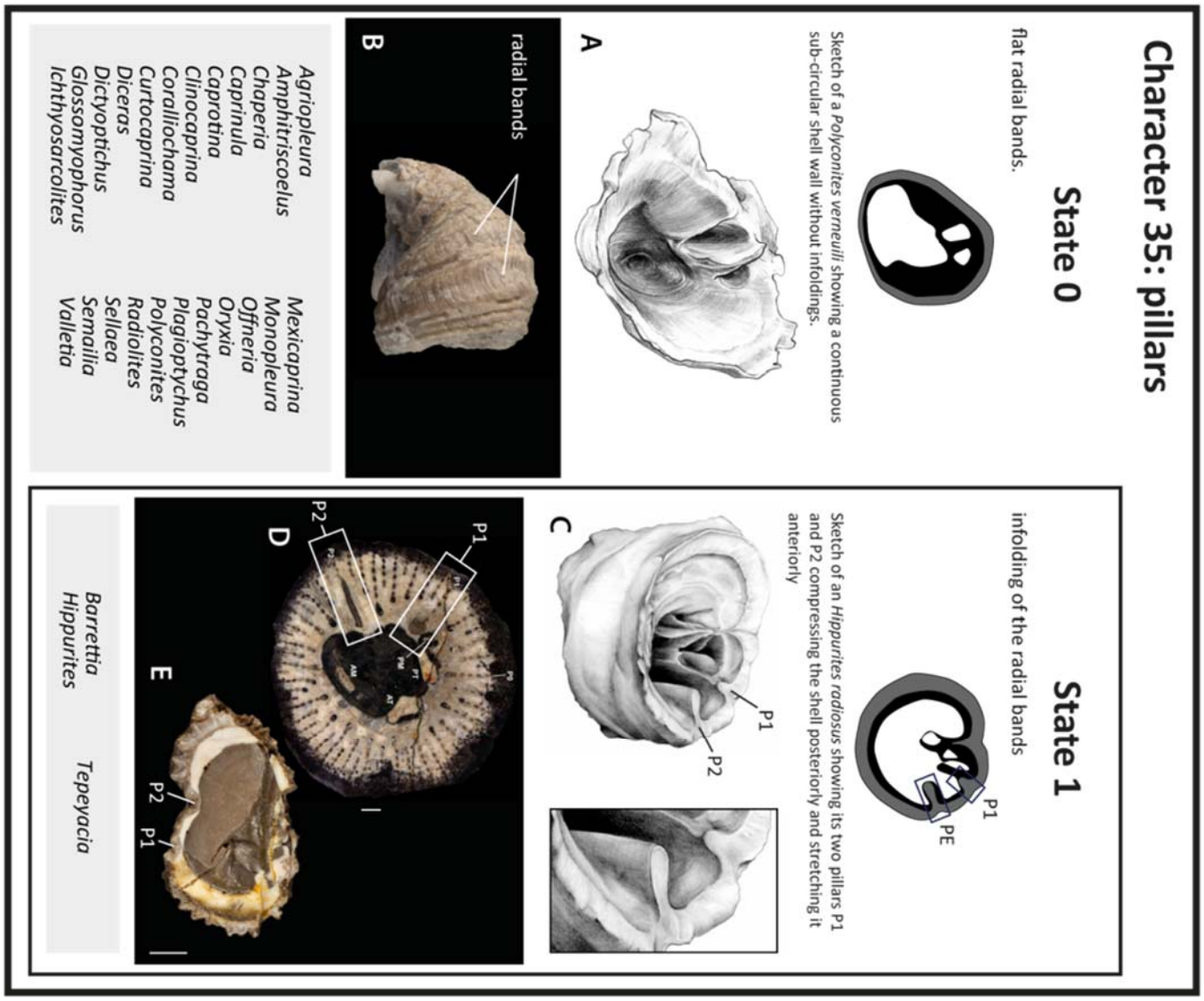
0.51

Remarks

Related to character 8 of Skelton and Smith (2000).

Pillars are postulated homologous to radial bands because they are both by pair and situated in the same posterior side of the shell. Pillars are understood as strong infoldings of the radial bands. Radiolitid pseudopillars may also be folded radial bands, but are uninformative with our taxonomic sampling.

Figure 4. 77. Hierarchical character #35. A – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). B – *Glossomyophorus costatus* (MNHN.F.R06779). C – *Hippurites radiosus* (MNHN.F.A58757). D – *Barrettia monilifera* (UWIGM.RUD.2006.223). E – *Tepeyacia corrugata* (CAS.66657.01). Scale bars, 10 mm.


Character 35: pillars
State 0
flat radial bands.
Sketch of a Polyconites verneuili showing a continuous sub-circular shell wall without infoldings.
A
radial bands
B
Agriopleura
Amphitriscoelus
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Oryxia
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Valletia
State 1
infolding of the radial bands
P1
PE
Sketch of an Hippurites radiosus showing its two pillars P1 and P2 compressing the shell posteriorly and stretching it anteriorly
P1
P2
C
P1
P2
D
E
P2
P1
Barrettia
Hippurites
Tepeyacia
Missing data: Titanosarcolites

7.36. Character 36

Name: outer calcitic shell layer

Position: both valves

Kind: outer layer

Coding: (0(1))

States definitions

0 thin calcitic outer shell layer (less than a tenth of the size of the shell)

1 thick calcitic outer shell layer

States tests

1 rejected – clades 7, 13 (except *Barrettia*), 22 (except *Semailia*, *Oryxia*, *Titanosarcolites*, *Ichthyosarcolites*)

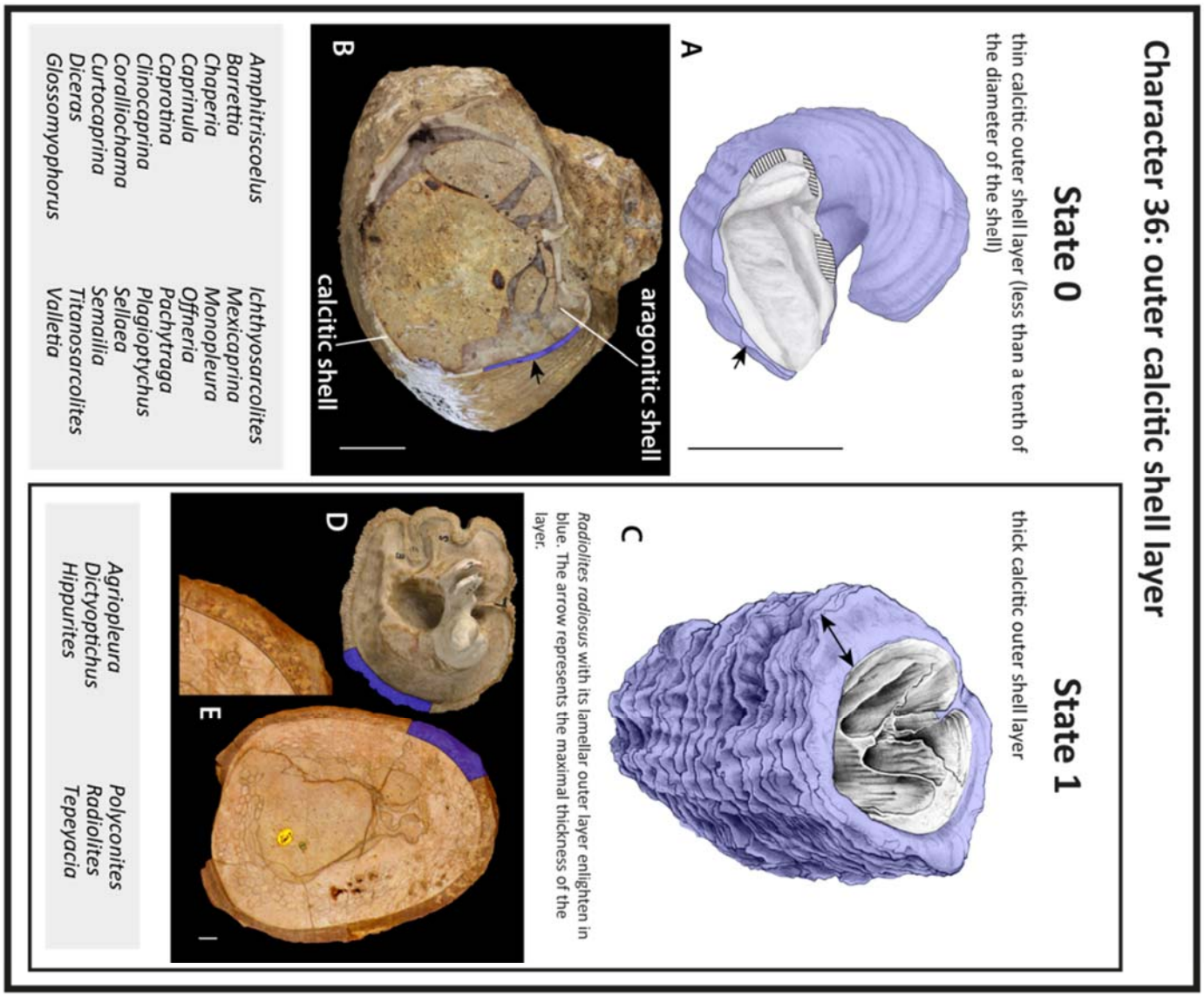
Retention index

0.51

Remarks

Related to characters 5 of Chartrousse (1994). Related to character 3c in Skelton and Masse (1998). Related partially to characters 2 and 9 in Steuber (1999). Related to character 1 in Sano *et al.*, (2014) with a more relative definition in order to reduce shell size effect.

Figure 4. 78. Hierarchical character #36. A – *Glossomyophorus costatus* (MNHN.F.R06779). B – *'Sellaea' sp.* (WSA3058). C – *Radiolites radiosus* (MNHN.F.J09021). D – *Hippurites radiosus* (MNHN.F.A58757). E – *Dictyoptychus morgani* (NHM.L-58427; picture provided by Peter W. Skelton). Scale bars, 10 mm.


Character 36: outer calcitic shell layer
State 0
thin calcitic outer shell layer (less than a tenth of the diameter of the shell)
State 1
thick calcitic outer shell layer
A
aragonitic shell
B
calcitic shell
C
Radiolites radiosus with its lamellar outer layer enlighten in blue. The arrow represents the maximal thickness of the layer.
D
E
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Sellaea
Semaiia
Titanosarcolites
Valletia
Agriopleura
Dictyoptichus
Hippurites
Polyconites
Radiolites
Tepeyacia

7.37. Character 37

Name: ligamentary structures

Position: both valves

Kind: ligament

Coding: (0(1(2(3))))

States definitions

0 external ligamentary groove parallel to the commissure

1 invagination of the ligament, forming a ligamentary cavity perpendicular to the commissure and connected to the exterior by a groove

2 ligamentary groove loss and complete internalization of the ligamentary cavity

3 complete loss of the groove and of the ligamentary cavity

States tests

1 passed – clade 2

2 passed – clade 19 (except *Oryxia*)

3 passed – clade 19 (except *Clinocaprina*, *Dictyoptychus*, *Semailia, Oryxia*)

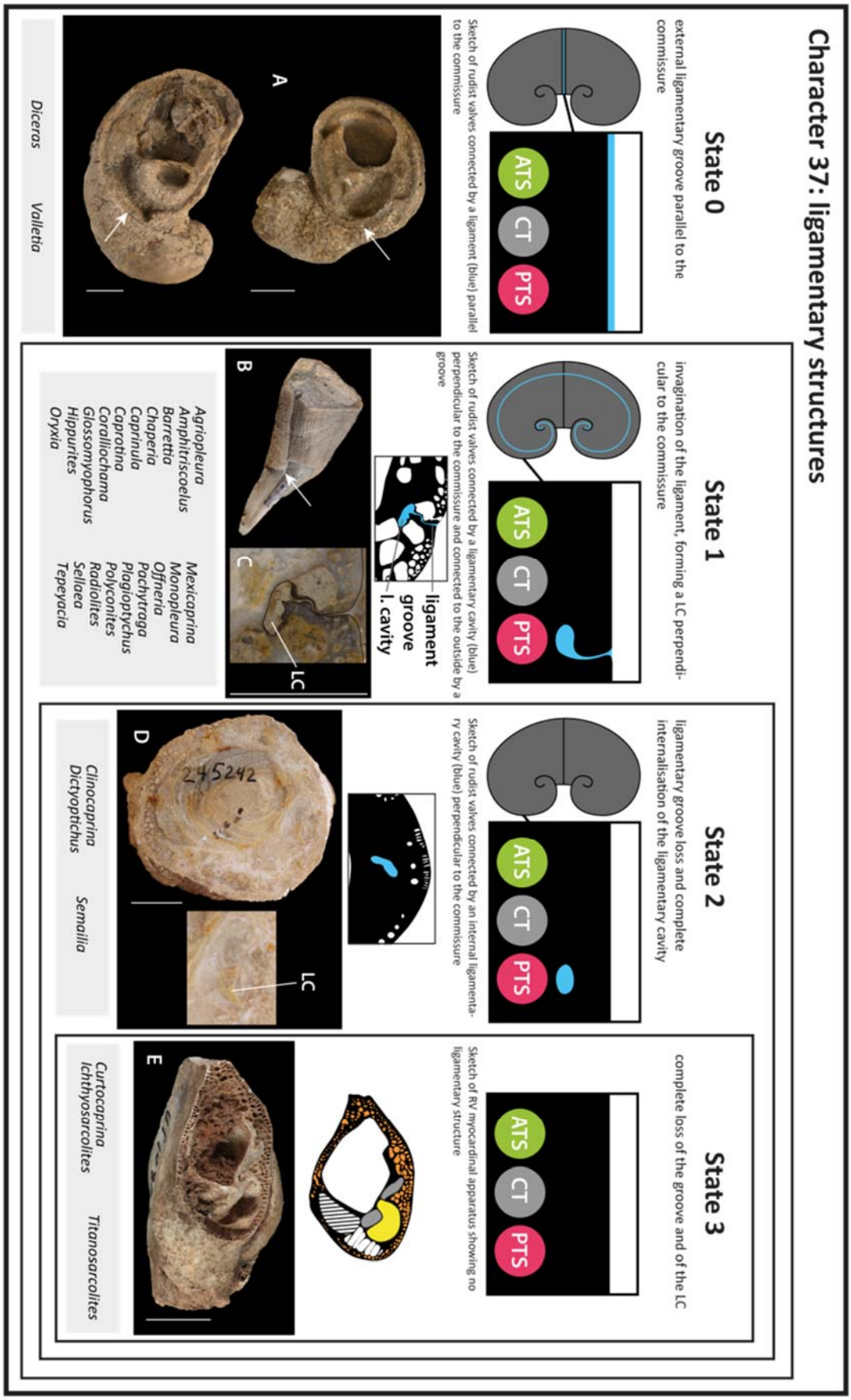
Retention index

0.95

Remarks

Related partially to character 5 of Smith (1994; #37.0, #37.1). Related partially to character 1 in Chartrousse (1994; #37.0, #37.1). Related to character 2 of Skelton and Masse (1998; #37.0, #37.1). Related partially to character 5 of Steuber (1999; #37.0, #37.1). Related to characters 11 and 12 of Skelton and Smith (2000), ordered herein. Related partially to character 6 of Mitchell (2013a; #37.1, #37.2). See Yonge (1967) for a complete discussion on ligament verticalization.

Figure 4. 79. Hierarchical character #37. A – *Valletia germani* (MNHN.F.J05798). B – *Glossomyophorus costatus* (MNHN.F.A27280). C – *Caprinula boissyi* (MNHN.6924). D – *Clinocaprina lowentsami* (USNM.245241). E – *Curtocaprina clabaughikinsorum* (UT.10937). Scale bars, 10 mm.


Character 37: ligamentary structures
State 0
external ligamentary groove parallel to the commissure
ATS
CT
PTS
Sketch of rudist valves connected by a ligament (blue) parallel to the commissure
A
Diceras
Valletia
State 1
invagination of the ligament, forming a LC perpendicular to the commissure
ATS
CT
PTS
Sketch of rudist valves connected by a ligamentary cavity (blue) perpendicular to the commissure and connected to the outside by a groove
ligament groove
l. cavity
B
C
LC
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Coralliochama
Glossomyophorus
Hippurites
Oryxia
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Tepeyacia
State 2
ligamentary groove loss and complete internalisation of the ligamentary cavity
ATS
CT
PTS
Sketch of rudist valves connected by an internal ligamentary cavity (blue) perpendicular to the commissure
D
245242
LC
Clinocaprina
Dictyoptichus
Semailia
State 3
complete loss of the groove and of the LC
ATS
CT
PTS
Sketch of RV myocardinal apparatus showing no ligamentary structure
E
Curtocaprina
Ichthyosarcolites
Titanosarcolites

7.38. Character 38

Name: ligament insertion

Position: both valves

Kind: ligament

Coding: (0(1))

States definitions

0 ligamentary cavity dorsally to PT/PTS

1 ligamentary cavity dorsally to CT/CTS

States tests

1 rejected – clade 12 (except *Agriopleura*, *Barrettia*), 24 (except *Ichthyosarcolites*)

Retention index

0.65

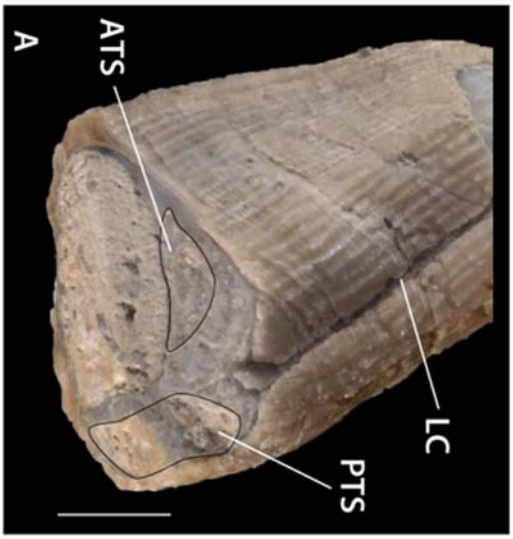
Figure 4. 80. Hierarchical character #38. A – *Glossomyophorus costatus* (MNHN.F.A27280). B – *Radiolites radiosus* (MNHN.F.J09021). Scale bars, 10 mm.

Character 38: ligament insertion

State 0

ligamentary cavity dorsally to PTS/PT

Sketch of the main elements around the ligamentary cavity and their connections on RV. LC is in contact with the PTS (PT in LV)

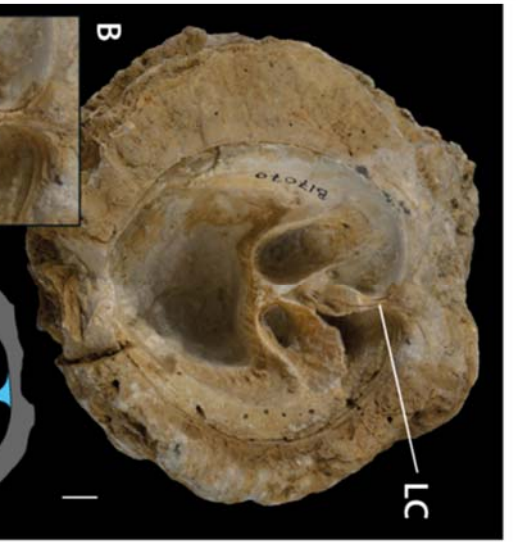


Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Curtocaprina
Diceras
Dictyoptichus
Glossomyophorus
Ichthyosarcolites
Mexicaprina
Monopleura
Pachytraga
Plagioptychus
Polyconites
Sellaea
Semailia
Tepeyacia
Valletia

State 1

ligamentary cavity dorsally to CTS/CT

Sketch of the main elements around the ligamentary cavity and their connections. The LC looses its contact with the PTS (PT in LV) and gain a contact with the CT (CTS in LV).



Hippurites
Offneria
Radiolites
Oryxia

Missing data: ***Titanosarcolites***

7.39. Character 39

Name: shell costulations

Position: both valves

Kind: ornamentation

Coding: (0(1))

States definitions

0 smooth shell

1 small close costulations

States tests

1 passed – clade 2 (except *Clinocaprina*, *Oryxia*, *Curtocaprina*, *Ichthyosarcolites*)

Retention index

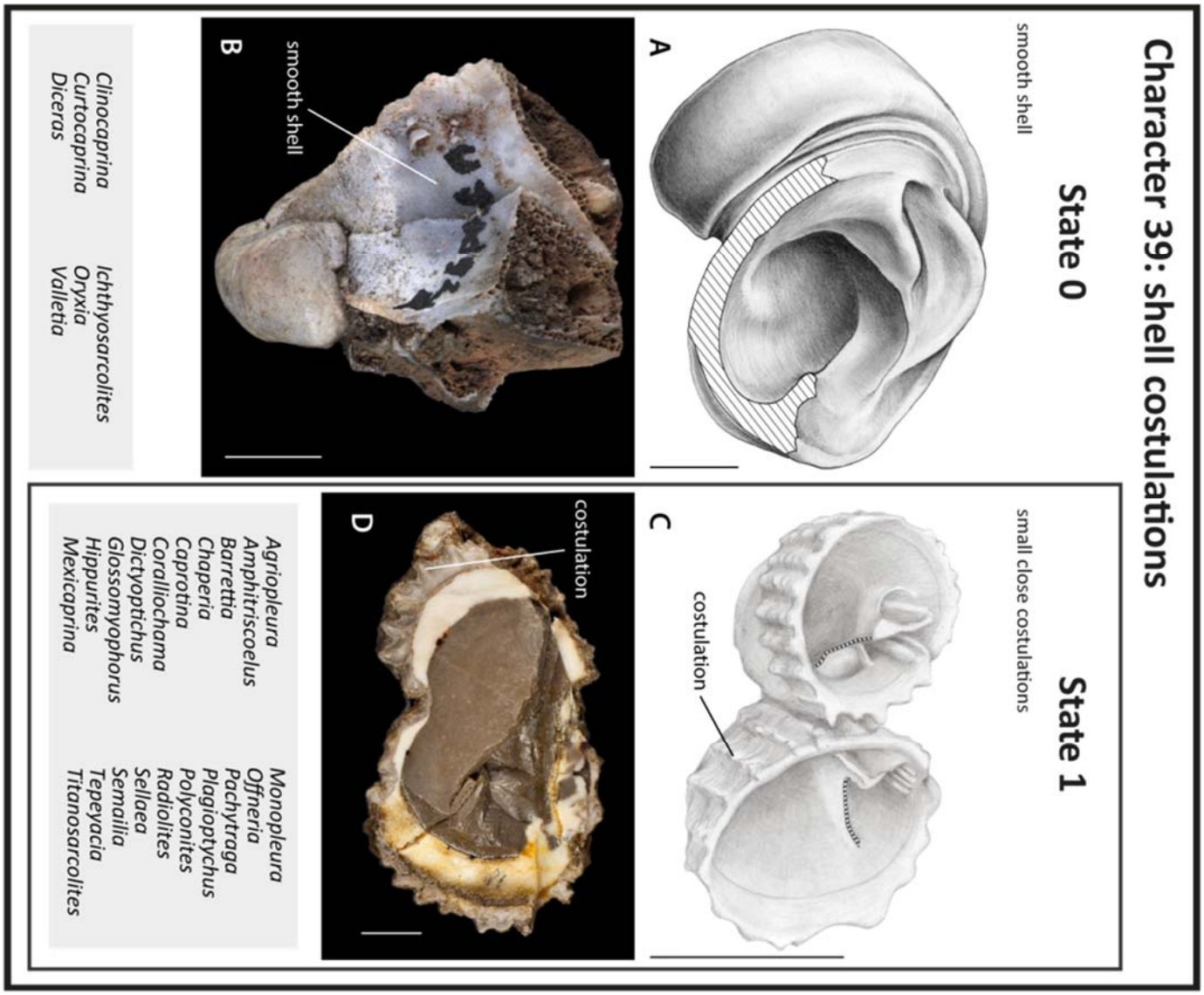
0.38

Remarks

Related to character 3 in Skelton and Smith (2000).

Costulations are defined as ornamental costae of the outer calcitic shell layer that do not impact the course of the inner aragonitic shell layer.

Figure 4. 81. Hierarchical character #39. A – *Diceras arietinum* (MNHN.F.A52439). B – *Curtocaprina clabaughikinsorum* (UT.10922). C – *Chaperia costata* (MNHN.F.R07942). D – *Tepeyacia corrugata* (CAS.66657.01). Scale bars, 10 mm.


Character 39: shell costulations
State 0
smooth shell
State 1
small close costulations
A
B
smooth shell
C
costulation
costulation
D
Clinocaprina
Curtocaprina
Diceras
Ichthyosarcolites
Oryxia
Valletia
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprotina
Coralliochama
Dictyoptichus
Glossomyophorus
Hippurites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Semailia
Tepeyacia
Titanosarcolites
Missing data: Caprinula

7.40. Character 40

Name: flanges

Position: both valves

Kind: ornamentation

Coding: (0(1))

States definitions

0 circular to sub-circular shape

1 presence of elongated narrow flanges

States tests

1 passed – clade 19 (except *Dictyoptychus*, *Clinocaprina*)

Retention index

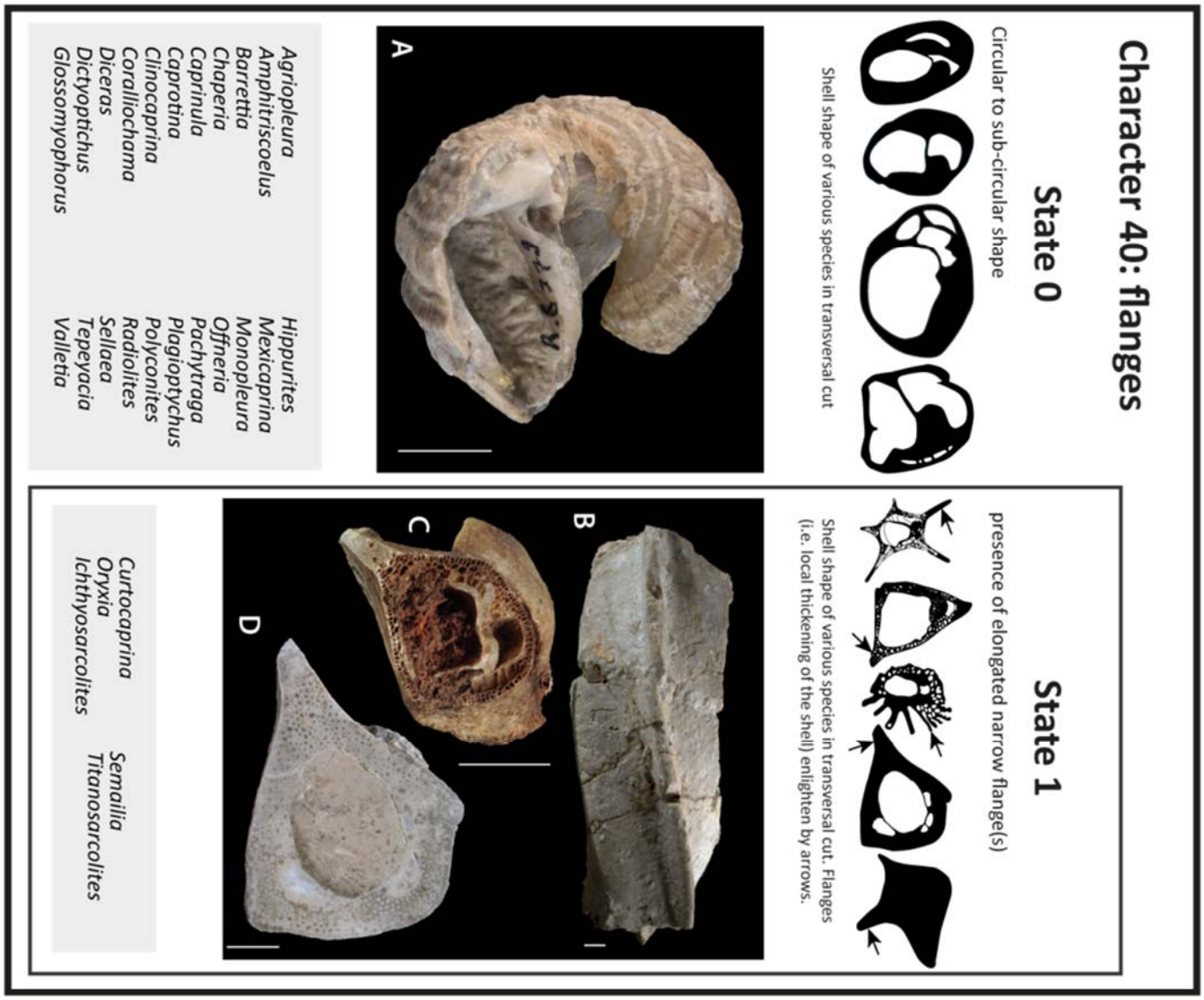
0.96

Remarks

Related to character 4 of Mitchell (2013a).

Contrary to costulations, flanges involves the inner aragonitic shell layer. We code a wide variety of flange morphologies and number as a single derived state. Antillocaprinids especially can reach an important number of flanges.

Figure 4. 82. Hierarchical character #40. A – *Glossomyophorus costatus* (MNHN.F.R06779). B – *Ichthyosarcolites bicarinatus* (MGUP.021.2-110; picture provided by Carolina d'Arpa and Tony Costaglia). C – *Curtocaprina clabaughikinsorum* (UT.10937). D – *Oryxia sulcata* (BJ.1901-HR2). Scale bars, 10 mm.


Character 40: flanges
State 0
Circular to sub-circular shape
Shell shape of various species in transversal cut
A
Agriopleura
Amphitriscoelus
Barrettia
Chaperia
Caprinula
Caprotina
Clinocaprina
Coralliochama
Diceras
Dictyoptichus
Glossomyophorus
Hippurites
Mexicaprina
Monopleura
Offneria
Pachytraga
Plagioptychus
Polyconites
Radiolites
Sellaea
Tepeyacia
Valletia
State 1
presence of elongated narrow flange(s)
Shell shape of various species in transversal cut. Flanges (i.e. local thickening of the shell) enlighten by arrows.
B
C
D
Curtocaprina
Oryxia
Ichthyosarcolites
Semailia
Titanosarcolites

Chapitre 5. Paléoécologie, paléoenvironnements et événements biosédimentaires enregistrés dans la succession de communautés à rudistes du Bassin Sud-Provençal au Cénomanien supérieur

Ce chapitre a été écrit en collaboration avec M. Floquet et L. Villier.

1. Introduction 461
2. Cadre géologique 462
3. Matériel et méthodes 464
4. Résultats 470
5. Discussion 496
6. Conclusion du chapitre 504

1. Introduction

Durant le Crétacé, les rudistes sont les principaux taxons de bivalves vivant dans des eaux chaudes et peu profondes des plates-formes carbonatées des marges de la Téthys (Masse *et al.* 1981; Ross et Skelton 1993; Gili *et al.* 1995; Skelton et Masse. 2000; Steuber et Löser 2000). Leur diversité et leur abondance en font les principaux producteurs de carbonates de l'époque, conditionnant la croissance des plates-formes sur lesquels ils vivent (Carannante *et al.* 1993, 1997, 2000; Scott 1995; Floquet, 1982; Ruberti 1997; Gili et Skelton 2000). Ce développement est ponctué par plusieurs épisodes de diversification tout au long du Crétacé et par des adaptations à de nouveaux modes de vie et à la dynamique sédimentaire de leur environnement (Masse et Montaggioni 2001; Philip 1981). Le nombre de taxons de rudistes est directement corrélé à l'étendue des plates-formes. Il est également influencé par des épisodes de crise, associés aux deux événements anoxiques (Steuber et Löser 2000). Il en résulte une croissance par paliers du nombre de taxons, avec l'un des maxima de diversité au Cénomanien terminal, avec 32 genres. Dès le début du Turonien, juste après l'EAO 2, on ne compte guère plus que 17 genres. Les familles Polyconitidae, Caprinidae, Caprinuloideidae, Caprotinidae et Ichthyosarcolitidae (Skelton 2013a) disparaissent alors définitivement du registre fossile (Steuber *et al.* 2016).

Une communauté biotique ou biocénose se définit par un ensemble d'organismes vivant en interdépendance (Möbius 1877). Elle se caractérise notamment par un assemblage de taxons présents dans des proportions stables. L'étude des communautés à rudistes connues en Provence au cours du Crétacé inférieur a permis de montrer que leur histoire est perturbée lors des épisodes de crise. Ces communautés possèdent cependant une dynamique évolutive qui leur est propre liée à une spécialisation écologique des taxons (Fenerci-Masse *et al.* 2004, 2005; Masse et Fenerci-Masse 2006). Au Cénomanien, les communautés à rudistes sont abondantes et diversifiées, notamment autour de la Téthys. Le passage Cénomanien-Turonien est donc une période clé pour comprendre l'évolution des rudistes dans une perspective synécologique.

Le Bassin Sud-Provençal fait partie du Domaine Téthysien, et permet de comprendre l'évolution des communautés à rudistes dans un cadre paléogéographique restreint et bien étudié, où les séquences sédimentaires offrent un enregistrement des

successions de communautés jusqu'à la toute fin du Cénomanien (Philip, 1967, 1970, 1978; Masse et Philip, 1976; Philip, Airaud, et Tronchetti, 1989; Hennuy et Floquet, 2001; Hennuy, 2003; Masse, Fenerci, et Pernarcic, 2003; Floquet *et al.*, 2006; Hairabian, 2007). La définition des communautés du Cénomanien supérieur du Bassin Sud-Provençal permet de mieux appréhender le milieu de vie de ces taxons, leur comportement au sein d'un écosystème, leurs relations avec leur environnement, et leur lien avec la macroévolution des rudistes. La série de calcaires à rudistes d'âge Cénomanien supérieur du Bassin Sud-Provençal est donc notre objet d'étude pour ce chapitre, par l'étude de deux coupes : celle de la Bastide d'Orves et la Barre des Aiguilles. Ces deux coupes permettent d'observer les fréquences de taxons en continu sur 14 mètres pour la Bastide d'Orves et 50 mètres pour la Barre des Aiguilles. Le comptage des tests fossilisés dans ces séries permet de définir des communautés en regroupant des assemblages homogènes de taxons. Ces communautés sont ensuite décrites dans leur environnement. La représentativité des assemblages est discutée en analysant les possibles biais de préservation (Tomasovych et Schlogl 2008) : processus écologiques et taphonomiques (Kidwell et Holland 2002; Smith et Nelson 2003; Villier 2009). La définition de ces communautés permet d'inférer le mode de vie des taxons et leur relation avec la production carbonatée *in situ* et le paléoenvironnement de l'époque.

2. Cadre géologique

Au Cénomanien supérieur, le Bassin Sud-Provençal fait partie de la marge nord-ouest téthysienne. Les plates-formes carbonatées sont à cette époque à leur extension maximale (Skelton et Gili 2012). Les mouvements de transtension entre la Plaque Ibérique et la Plaque Eurasienne entraînent la formation du Rift Pyrénéo-Provençal (Floquet et Hennuy 2001, 2003; Hennuy et Floquet 2001; Hennuy 2003; Floquet *et al.* 2005; fig. 5.1). Le Bassin Sud-Provençal est bordé au Nord par le Bombement Durancien (Philip 1970; Philip 1993; Masse et Philip 1976) et au Sud par le Massif Méridional (Redondo 1986; Floquet et Hennuy 2003; Hennuy 2003). Ce bassin est marginal au Rift Pyrénéo-Provençal et sa largeur n'excédait alors pas 15 km.

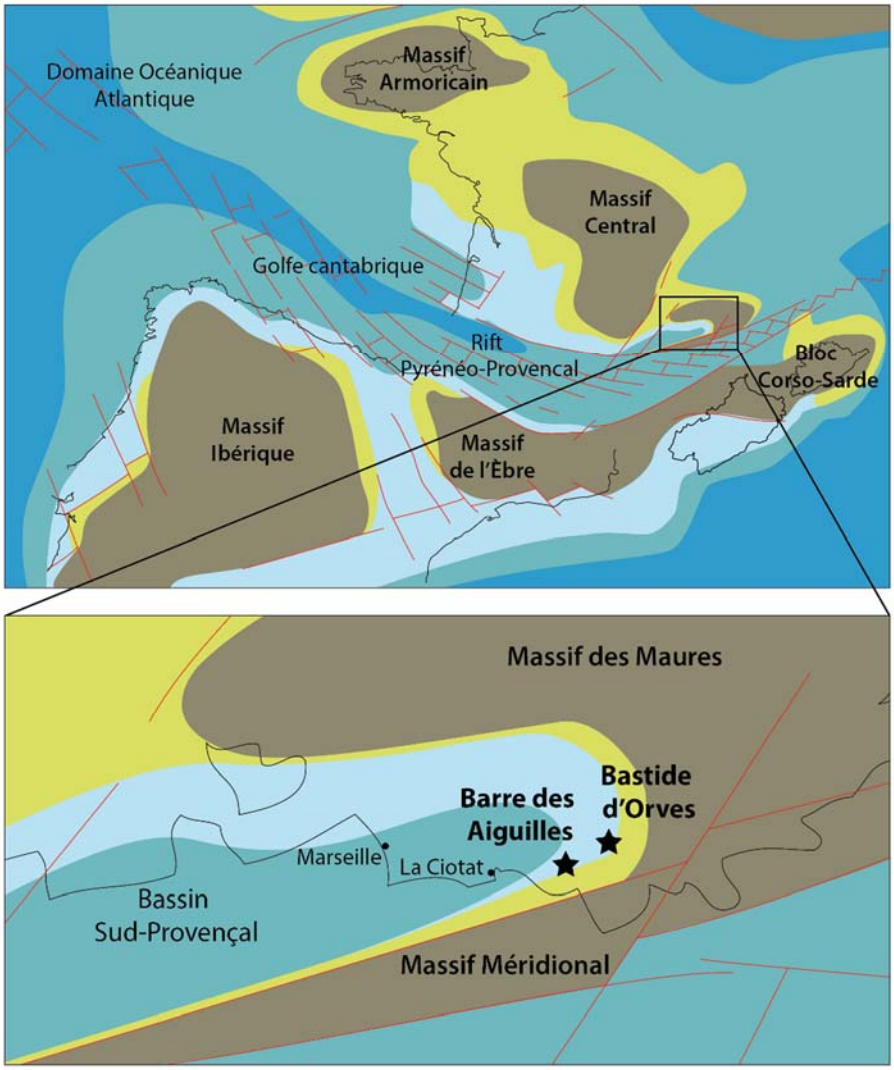


Figure 5. 1. Carte paléogéographique de l'Europe de l'Ouest au Cénomanien supérieur - Turonien, modifiée à partir de Floquet et Hennuy 2001. Marron : terre émergée ; jaune : côtier ; bleu clair : plate-forme carbonatée ; vert : plate-forme externe ; bleu foncé : marin profond.

Une plate-forme carbonatée à rudistes bordait ce bassin au Cénomanien moyen et supérieur, plate-forme que l'on retrouve aujourd'hui sous la forme de deux Barres à Rudistes (fig. 5.2). La première Barre à Rudistes, ou Barre de la Marcouline, est datée du Cénomanien moyen. La seconde Barre à Rudistes, ou Barre de Fontblanche, est datée du Cénomanien supérieur au Turonien basal (Philip 1970). La Barre de Fontblanche se découpe elle-même en deux séquences (Floquet 2005, 2006) : Fontblanche 1 (Cénomanien supérieur) et Fontblanche 2 (Cénomanien supérieur à Turonien basal). Les deux coupes étudiées concernent la séquence Fontblanche 1 de la Barre de Fontblanche. Elles sont situées dans la commune d'Evenos (Var, France) : coupe de la Barre des Aiguilles (latitude

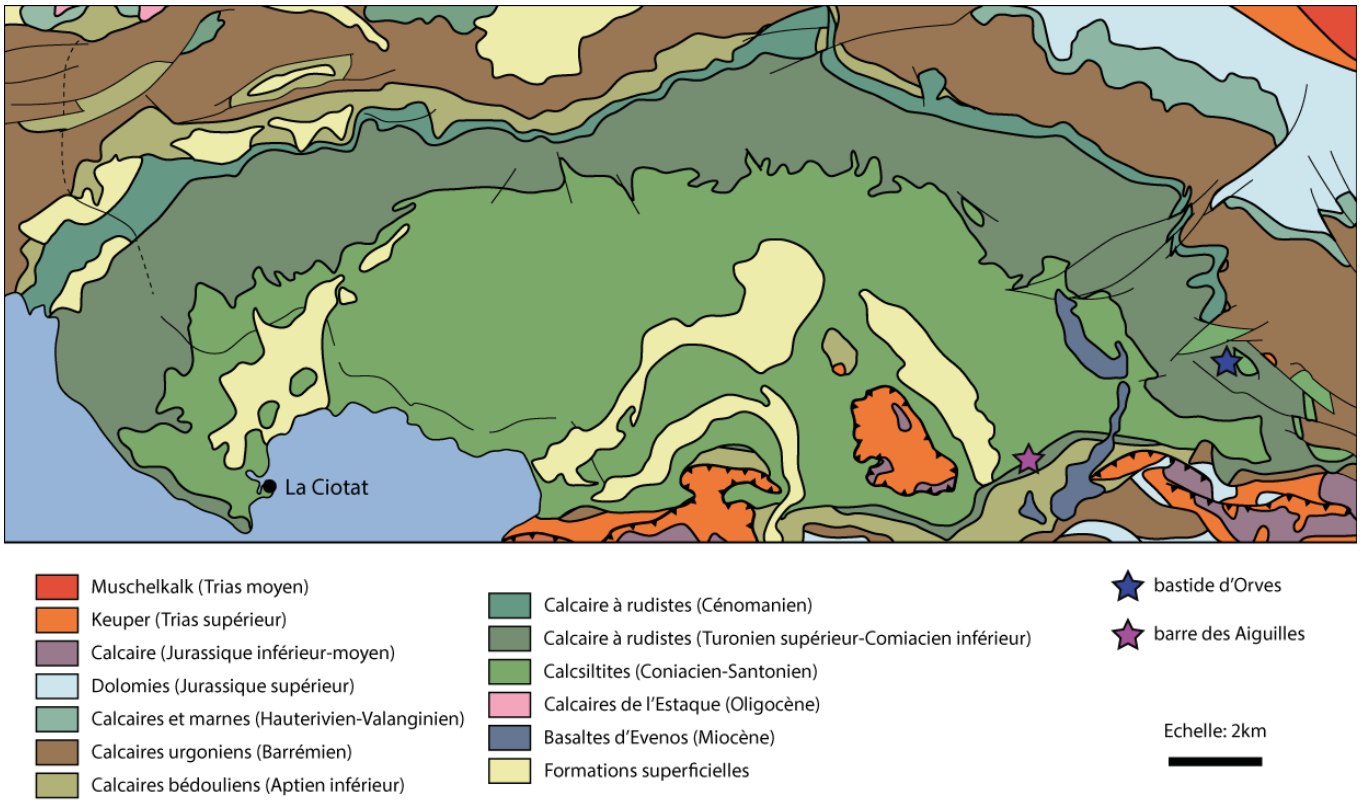


Figure 5. 2. Carte géologique de la région du Bassin Sud-Provençal (actuel synclinal du Beausset), Sud-Est de la France. La ville de La Ciotat est indiquée en point de repère, ainsi que les deux localités étudiées de la Barre des Aiguilles et de la Bastide d'Orves. Echelle : 2 km. Modifié à partir des cartes géologiques à 1.50000 du BRGM Aubagne-Marseille et La Ciotat.

43°10'30.13''N, longitude : 5°50'23.41''E) et coupe de la carrière de la Bastide d'Orves (latitude 43°12'5.59''N, longitude : 5°54'2.12''E. fig. 5.1). L'exploitation des calcaires de ces carrières se faisant par découpage au fil, les surfaces étudiées sont lisses et permettent une identification aisée des taxons ainsi que leur comptage directement sur le terrain. La coupe de la Barre des Aiguilles a été levée sur une épaisseur de 50 m, dont 32 m en surface lisse, celle de la Bastide d'Orves sur une épaisseur de 14,5 m, totalement en surface lisse.

3. Matériel et méthodes

Les assemblages de rudistes sont étudiés par comptage des individus dans des quadras. Les figures 5.3 et 5.4 présentent les localisations des comptages sur les coupes de la Bastide d'Orves et de la Barre des Aiguilles. Chacune des deux coupes est découpée en séquences (S dans figs. 5.3 et 5.4) bordées par des discontinuités d'émersion (D dans figs.

5.3 et 5.4; Blénet, 2010; Ackouala, 2012). Un point de comptage correspond à un quadra de 0,25/1 m. Seuls les individus entiers et les morceaux reconnaissables de plus de 2 cm de taille sont décomptés. Au total, 18 niveaux stratigraphiques sont analysés à la Bastide d'Orves et 20 autres à la Barre des Aiguilles. Les comptages sont ensuite standardisés en fréquences d'occurrence des taxons. Des courbes de dominance sont produites à partir de ces fréquences (fig. 5.3 et 5.4). Une matrice de dissimilarité est construite suivant l'indice de dissimilarité de Bray-Curtis, permettant d'évaluer la dissimilarité entre deux points de comptage. Une Classification Ascendante Hiérarchique (CAH) des points de comptage est calculée à partir de cette matrice, permettant de rassembler les points en assemblages aux propriétés similaires en communautés. Ces communautés biotiques sont définies par la fréquence d'occurrence relative de taxons. La méthode de Positionnement Multidimensionnel Non-Métrique (PMNM; Non-Metric Multidimensionnal Scaling en anglais) est ensuite utilisée pour visualiser les similarités entre les points de comptage dans un espace en deux dimensions, et de montrer la cohérence de la reconstitution des communautés biotiques proposées. Les données de comptage sont synthétisées en courbes de richesse spécifique.

Les environnements sédimentaires sont caractérisés à chaque point de comptage par une analyse des textures des calcaires en lames minces, suivant les nomenclatures de Dunham (1962) et d'Embry et Klovan (1971). 34 échantillons en lames minces permettent d'étudier les microfaciès de la coupe de la Bastide d'Orves (fig. 5.3) et 42 pour la coupe de la Barre des Aiguilles (fig. 5.4). Chaque lame mince est ensuite décrite et interprétée en termes de paléoenvironnements d'après la texture et le contenu micropaléontologique. Les fréquences d'éléments présents dans les lames minces sont estimées à l'aide du logiciel jmicrovision 1.2.7 (Roduit 2008) par le comptage de 250 points disposés aléatoirement sur chacune d'elles. La corrélation entre communautés biotiques et cadres sédimentaires, tels que le taux de bioclastes par rapport au taux de matrice (microbioclastique, microsparitique ou micritique) ou de ciment (sparitique), est ensuite testée statistiquement par ANOVA. On comptabilise comme bioclastes les grains reconnaissables au microscope de taille inférieure à 2 cm. On teste également la corrélation entre le taux de bioclastes et la richesse spécifique des environnements avec un test de corrélation de Pearson sur les valeurs absolues

Figure 5. 3. Coupe de la Bastide d'Orves avec les unités stratigraphiques Fontblanche 1 (Cénomanien supérieur) et Fontblanche 2 (Cénomanien terminal/Turonien inférieur). Les séquences S1 à S6 et les discontinuités D1 à D6 (matériel et méthodes) sont reportées ainsi que les macrocommunautés (1 à 4), la texture, la richesse spécifique, le taux de bioclastes et les valeurs de dominance pour les macrocommunautés.

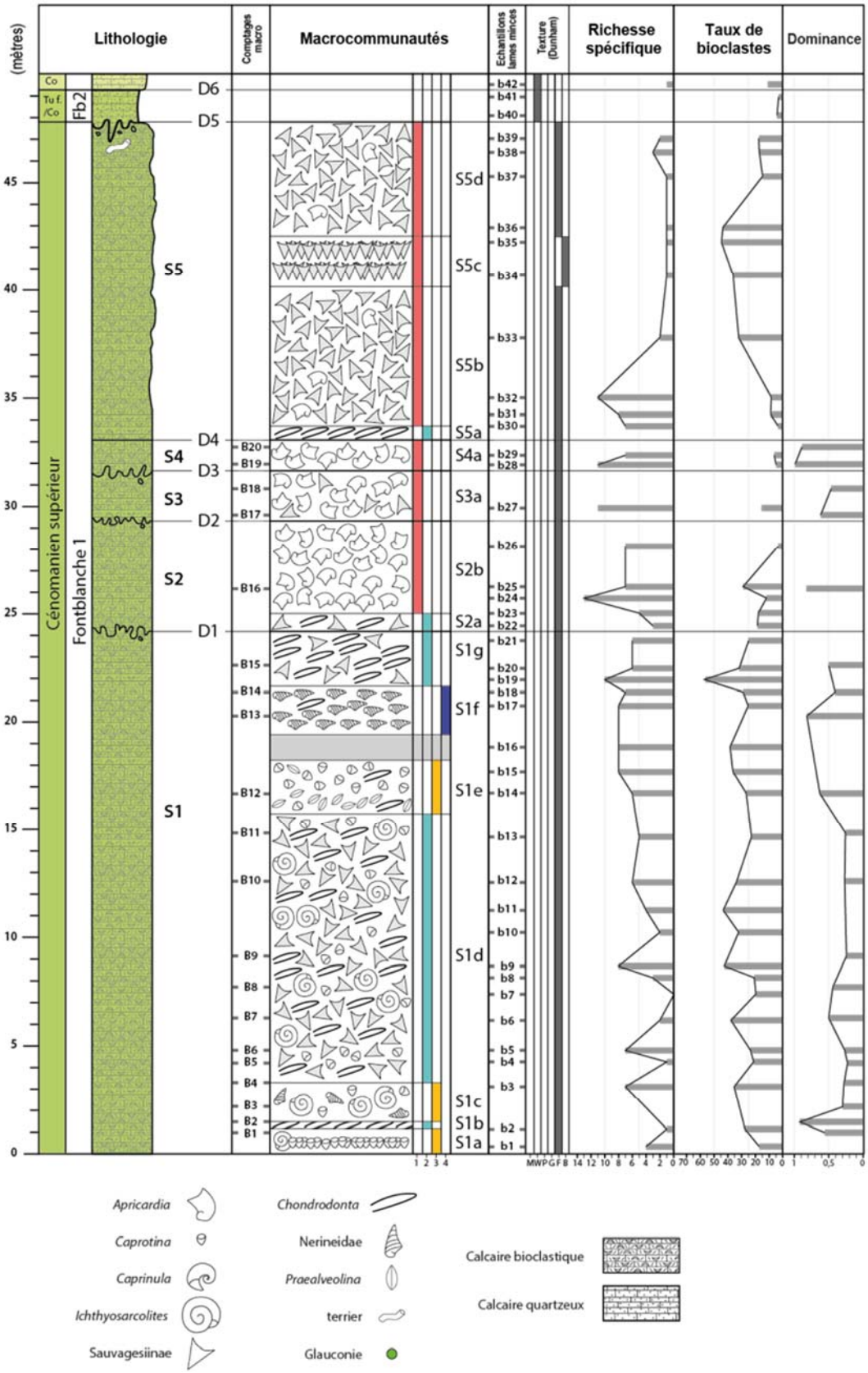


Figure 5. 4. Coupe de la Barre des Aiguilles avec les unités stratigraphiques Fontblanche 1 (Cénomanien supérieur) et Fontblanche 2 (Fb2 ; Cénomanien terminal/Turonien inférieur). Les séquences S1 à S5 et les discontinuités D1 à D5 sont reportées ainsi que les macrocommunautés (1 à 4), la texture, la richesse spécifique, le taux de bioclastes et les valeurs de dominance pour les macrocommunautés.

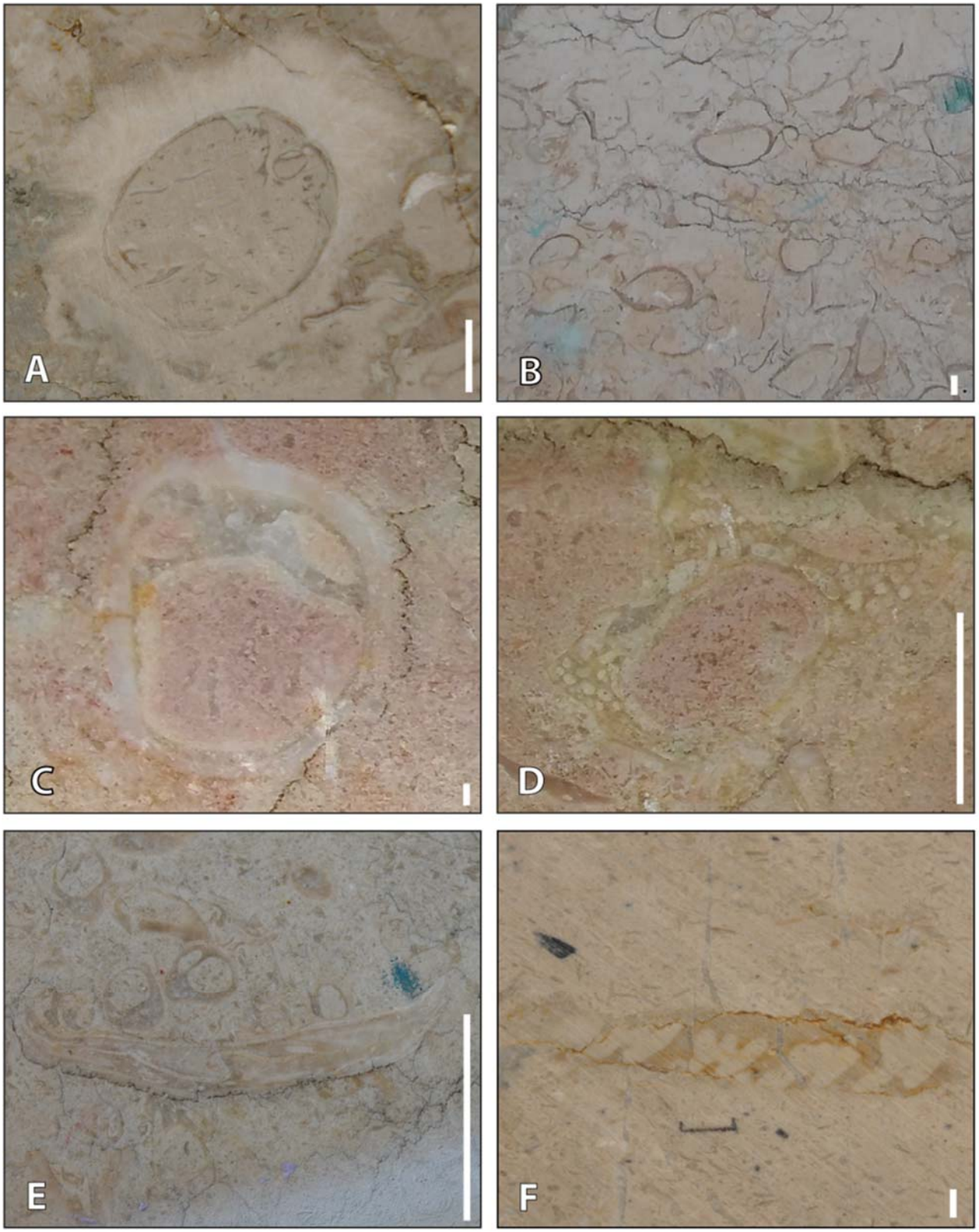


Figure 5. 5. Taxons rencontrés à la Bastide d'Orves et à la Barre des Aiguilles. A – *Sauvagesia sharpei* Bayle 1857 (Radiolitidae, rudiste); B – *Apricardia* sp. (Requienidae, rudiste), C – *Caprotina* sp. (Caprotinidae, rudiste); D – *Ichthyosarcolites bicarinatus* Gemmellaro 1865 (Ichthyosarcolitidae, rudiste); E – *Chondrodonta joannae* Choffat 1886 (Chondrodontidae); F – Nerinea sp. (Nerineidae). Crédit photographique (photo F) : Marc Floquet et Aurélien Blénet. Echelle: 1 cm.

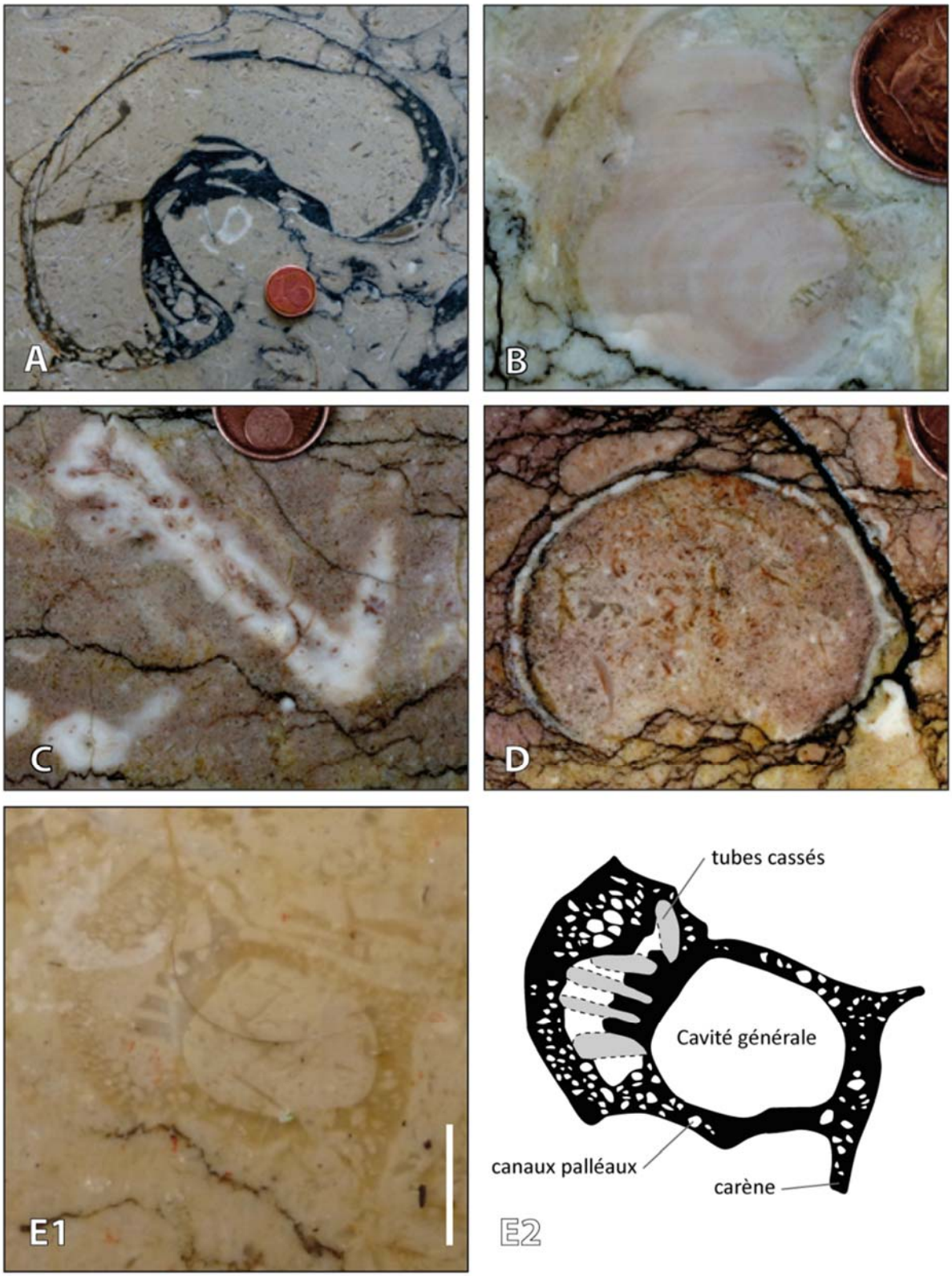


Figure 5. 6. Taxons rencontrés à la Bastide d'Orves et à la Barre des Aiguilles. A – *Caprinula boissyi* Orbigny 1840 (Caprinulidae, rudiste) ; B – Chaetetidae (Porifera), C – Cyclostomata (Bryozoa), D – Cidaroida (Echinoidea) ; E1 – Section transversale de valve droite de rudiste à canaux portant des tubes cassés. Echelle : 1 cm ; E2 – Schéma interprétatif de E1. Les carènes sont situées sur le côté postérieur (à droite). Le côté antérieur épaissi montre des tubes, cavités de taille plus importante que les canaux palléaux résultant de l'anastomose de carènes. Ces caractéristiques ne sont rencontrées que chez les Antillocaprinidae. Crédits photographiques : Marc Floquet et Aurélien Blénet. Echelle: 1 cm.

ou sur les différences entre une valeur et la suivante si une autocorrélation au sein d'une même variable est détectée. Les statististiques sont réalisées grâce aux logiciels PAST (Hammer *et al.* 2001) et R (R Development Core Team 2013).

4. Résultats

4.1. Définition des communautés à rudistes

Les communautés reconnaissables macroscopiquement dans la Barre de Fontblanche sont caractérisées par les occurrences de onze taxons, dont dix de rudistes (figs. 5.5 et 5.6) : *Sauvagesia sharpei* Bayle 1857 (Radiolitidae), *Durania arnaudi* Choffat 1891 (Radiolitidae), *Apricardia* sp. (Requienidae), *Ichthyosarcolites bicarinatus* Gemmellaro 1865 (Ichthyosarcolitidae ; identifiés comme *Ichthyosarcolites triangularis* selon Philip 1978, cependant, les spécimens complets montrent toujours plus d'une carène comme en Figure 5.5-D; voir chapitre 3 pour la révision du genre), *Caprinula boissyi* Orbigny 1840 (Caprinulidae), *Caprotina* sp. (Caprotinidae), *Chondrodonta joannae* Choffat 1886 (Chondrodontidae), Chaetetidae (Porifera), Scleractinia (Anthozoa), Cidaroida (Echinoidea). A la Barre des Aiguilles, dans la séquence S1, on retrouve également une unique occurrence d'un rudiste à canaux qui semble montrer des caractères spécifiques des taxons du clade 11 (fig. 5.6-E; Titanosarcolitinae sensu Mitchell 2013b; clade 24, voir chapitre 4). Cette sous-famille est jusqu'à présent restreinte au Domaine Caribéen où son apparition est datée du Santonien. Aucune coquille de ce taxon n'a pu être collectée sur le terrain mais si son étude confirmait une telle attribution taxonomique, celle-ci aurait des conséquences importantes sur la compréhension de l'histoire biogéographique du groupe. Le genre *Caprina* n'est identifié dans aucune des deux coupes sinon de façon très occasionnelle, au contraire des *Caprinula*. Les Monopleuridae sont très rares (présence relictuelle d'après Philip 1978). Les deux taxons *Sauvagesia sharpei* et *Durania arnaudi* (var. *intermedia* d'après Philip *et al.* 1989) sont facilement identifiables (au niveau de la sous-famille) grâce à leur couche calcitique celluleuse polygonale. Cependant, elles se rencontrent le plus souvent sous forme de débris et ne peuvent donc pas être identifiés au rang spécifique. Dans ce chapitre, pour les comptages, référence ne sera faite qu'à leur sous-famille, les Sauvagesiinae. *Sphaerulites foliaceus* a également été répertorié dans le

Cénomanien moyen de la Formation de la Marcouline (Philip 1980). Les fréquences de taxons comptabilisés sur les 40 points de comptages macroscopiques de la Bastide d'Orves (fig. 5.7) et de la Barre des Aiguilles (fig. 5.8) sont classés avec la CAH présentée en Figure 5.9. Quatre clusters sont retenus, illustrés en Figure 5.10 par PMNM, représentant les distances entre tous les points de comptage. Ces quatre clusters correspondent à quatre communautés biotiques homogènes en termes de fréquences de taxons :

1. La communauté à *Chondrodonta* et *Ichthyosarcolites* (fig. 5.10-A), définie par 13 points de comptages, est la plus hétérogène en termes de fréquences de taxons. Le taxon caractéristique de cet assemblage est *Chondrodonta joannae*. Cependant, sa fréquence est très variable, pouvant varier de 11,4% (point B9) à 95,8% (point B2). Les points d'abondance maximale de *C. joannae* correspondent à des événements d'accumulation de coquilles sans doute dus à des épisodes de tempêtes, les coquilles se trouvant souvent toutes orientées horizontalement, alors que le mode de vie de ces *Chondrodonta* était avec la coquille plantée dans le sable de manière oblique (Ayoub-Hannaa et Fürsich 2011). Lorsque les *C. joannae* sont présents en association, c'est généralement avec les Sauvagesiinae (maximum : B15, 51,4%), ainsi qu'avec *Ichthyosarcolites bicarinatus* (maximum : B5, 32,1%) et *Caprotina sp.* (maximum : B9, 23,7%). *Caprinula boissyi* et Chaetetidae sont présents occasionnellement.

2. La communauté à *Apricardia* et Sauvagesiinae (fig. 5.10-B), reconnue sur 22 points de comptage, est la plus courante. Cette communauté est beaucoup plus homogène, avec les deux taxons *Apricardia* et Sauvagesiinae largement dominants sur l'ensemble des points (minimum : A4, 72,7%; maximum : B19 et B20, 100%). Cette communauté présente deux pôles, bien illustrés par la PMNM (fig. 5.9 et 10) : un pôle à dominance d'*Apricardia* (B16 – *Apricardia* : 90,4%, Sauvagesiinae : 1,2%) et un pôle à dominance de Sauvagesiinae (A3 – Sauvagesiinae : 87%, *Apricardia* : 0%). Ces deux pôles ne peuvent cependant pas être séparés en deux

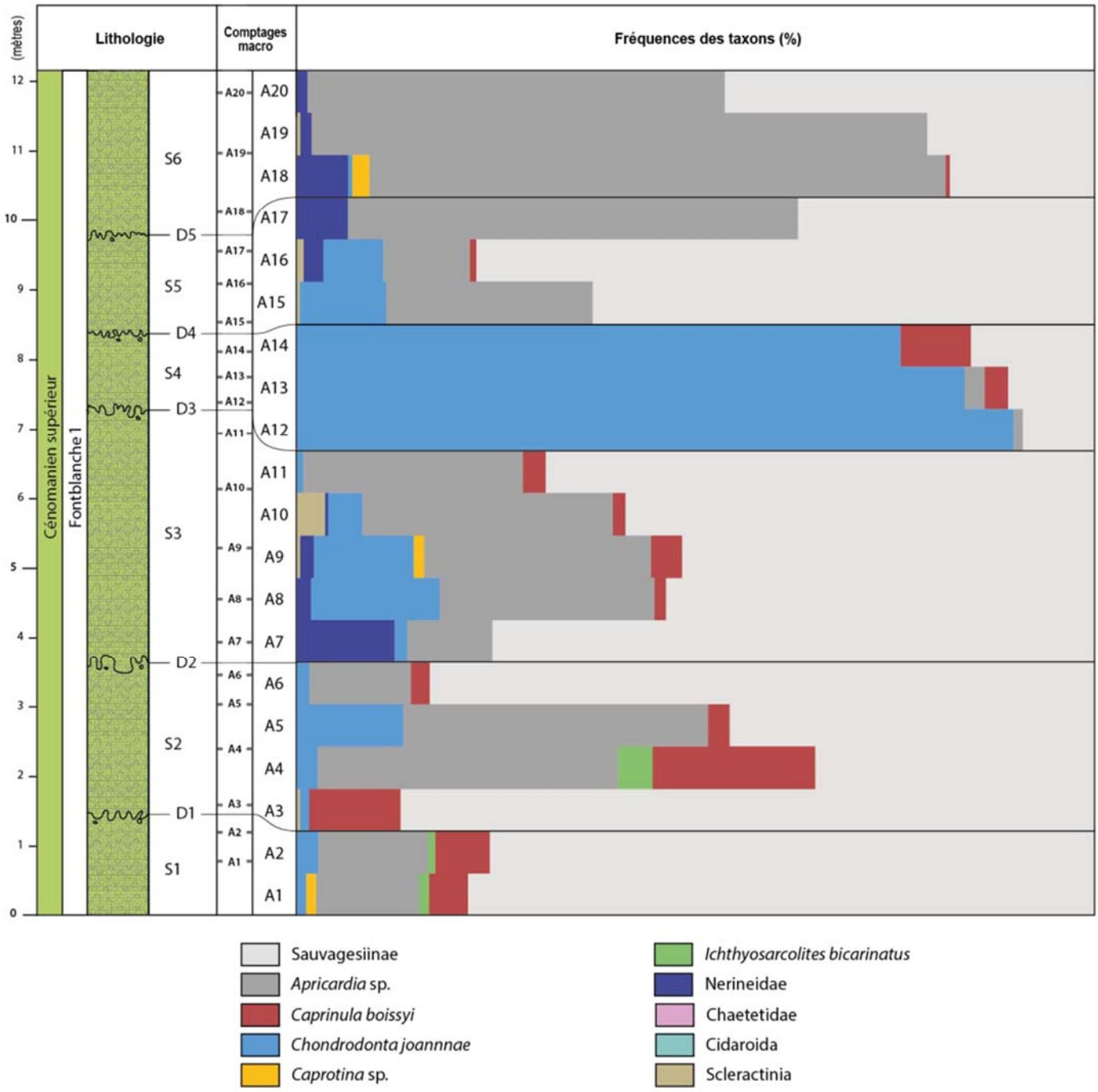


Figure 5. 7. Pourcentages d'occurrence des taxons à la Bastide d'Orves. Les fréquences des taxons sont reportées pour chaque point de comptage macroscopique. Gris clair : Sauvagesiinae ; marron : *Caprinula* ; vert : *Ichthyosarcolites* ; gris foncé : *Apricardia* ; jaune : *Caprotina* ; bleu clair : *Chondrodonta* ; bleu marine : *Nerinea* ; beige : *Scleractinia*.

communautés distinctes, car aucune limite claire entre ces deux pôles ne peut être assignée. Les *C. joannae* sont présents de manière ponctuelle. *Caprinula* est

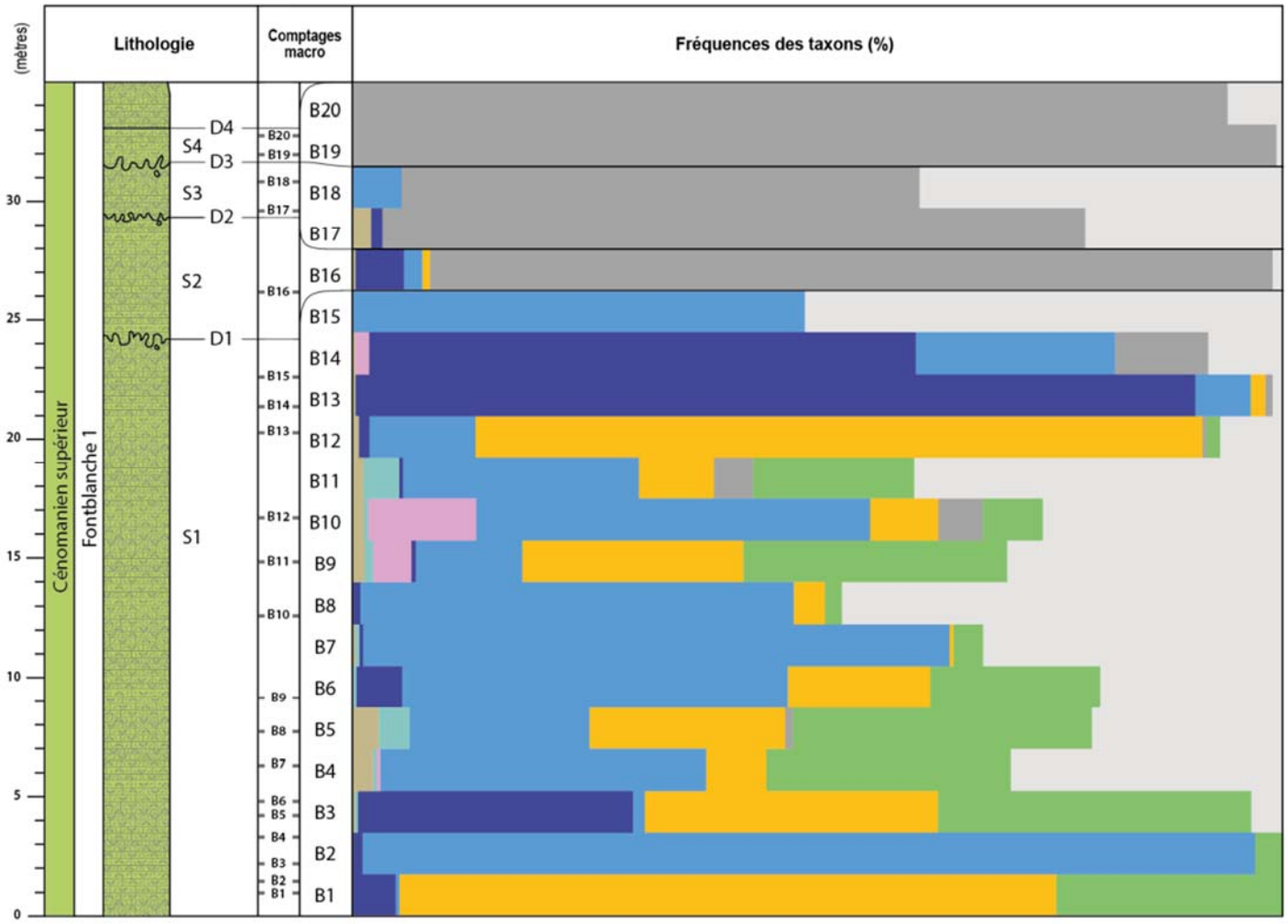


Figure 5. 8. Pourcentages d'occurrence des taxons à la Barre des Aiguilles. Les fréquences des taxons sont reportées pour chaque point de comptage macroscopique. Gris clair : Sauvagesiinae ; vert : *Ichthyosarcolites* ; gris foncé : *Apricardia* ; jaune : *Caprotina* ; bleu clair : *Chondrodonta* ; bleu marine : *Nerinea* ; beige : Scleractinia ; rose : Chaetetidae ; cyan : Cidaroida.

également présent de manière ponctuelle, mais uniquement dans les points ou les Sauvagesiinae sont dominants par rapport à *Apricardia*.

3. La communauté à *Caprotina* (fig. 5.10-C) n'est définie que par trois points. *Caprotina* est le taxon dominant (minimum : B3, 31,4%; maximum : B12, 78%). On trouve également *Ichthyosarcolites* dans deux des trois points (minimum : B12, 1.5%; maximum : B3, 33.6%) et *Nerinea* dans le point B3 (29,5%), de sorte que cette communauté apparaît hétérogène dans sa composition faunistique.

4. La communauté à *Nerinea* (fig. 5.10-D). *Nerinea* figure dans deux points de comptage de manière très dominante par rapport à tous les autres points (B14 : 58,7%, B13 : 90,1%). Cette forte abondance démarque ces deux points de tous les autres, de sorte qu'elle permet de distinguer cette quatrième communauté, qui n'apparaît qu'une seule fois, dans la Barre des Aiguilles. *C. joannae* peut y être associé de manière relativement abondante (B14 : 21,4%).

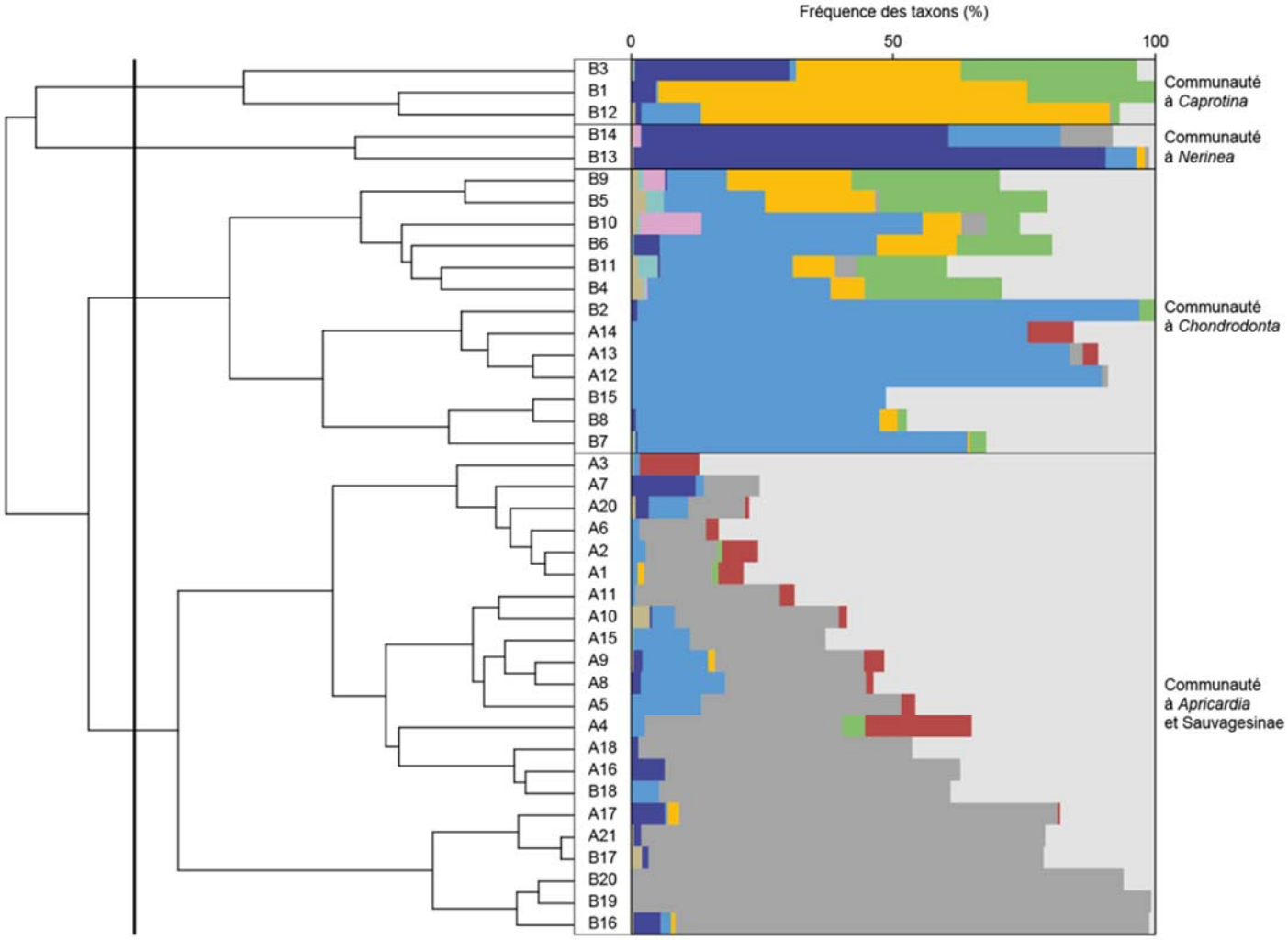


Figure 5. 9. Classification Ascendante Hiérarchique sur les fréquences des comptages de taxons, et les fréquences correspondant aux points de comptage. Quatre communautés sont discriminées : une communauté à *Caprotina* dominants, une communauté à *Nerinea* dominants, une communauté à *Chondrodonta* dominants accompagnées de *Ichthyosarcolites*, *Caprotina* et Sauvagesiinae, et une communauté à *Apricardia* et Sauvagesiinae dominants avec une grande variabilité entre les proportions des deux taxons dominants. Gris clair : Sauvagesiinae ; marron : *Caprinula* ; vert : *Ichthyosarcolites* ; gris foncé : *Apricardia* ; jaune : *Caprotina* ; bleu clair : *Chondrodonta* ; bleu marine : *Nerinea* ; beige : Scleractinia ; rose : Chaetetidae ; cyan : Cidaroida.

4.2. Taux de bioclastes et richesse spécifique

L'utilisation du coefficient de corrélation de Pearson ne révèle pas de corrélation significative entre le taux de bioclastes et la richesse spécifique au niveau de l'ensemble des deux coupes de la Bastide d'Orves et de la Barre des Aiguilles. Cependant, l'étude plus approfondie de ces coupes permet de mettre en évidence une importante corrélation entre le taux de bioclastes et la richesse spécifique des points b1 à b29 de la Barre des Aiguilles (coefficient de corrélation : 0,49; valeur-p : 0,02). Au sein de la Barre des Aiguilles, il n'y a pas de corrélation significative sur le reste des points de b30 à b42 (coefficient de corrélation : -0,51; valeur-p : 0,13). Une corrélation positive entre bioclastes et richesse spécifique existe dans l'ensemble de la coupe de la Bastide d'Orves (coefficient de corrélation : 0,36; valeur-p : 0,04). Finalement, l'utilisation d'une ANOVA permet de mettre en évidence que les communautés à rudistes définies précédemment n'expliquent qu'une très faible part de la variation du taux de bioclastes (moyenne des écarts au carré : 76,05 pour la part expliquée par les communautés, 249,39 pour la part non-expliquée par les communautés; fig. 5.11).

4.3. Stratigraphie séquentielle et successions biotiques

La succession le long des deux coupes des communautés précédemment définies est illustrée sur les figures 5.3 et 5.4. Ces communautés sont renfermées dans la plupart des cas dans des calcaires à textures floatstones où se mêlent des coquilles entières, ou cassées ou partiellement dissoutes et recristallisées en sparite sinon à vides de dissolution emplies de sédiments apportés postérieurement, le tout dans une matrice bioclastique à microbioclastique. Une forte bio-corrosion des coquilles pendant et juste après la mort des organismes des communautés liée à l'action d'organismes microperforants a entraîné (i) une fine fragmentation des coquilles sous forme de microbioclastes et (ii) à l'extrême, une micritisation, l'ensemble de ces produits constituant la matrice entre les coquilles. L'aragonite des coquilles est partiellement à totalement dissoute. Ce phénomène débute dès la mort de l'organisme (Sanders 1999; Brachert et Dullo 2000). La partie calcitique en revanche est généralement bien préservée, résultant en la préservation différentielle des coquilles composées à majorité de calcite (Sauvagesiinae) par rapport aux

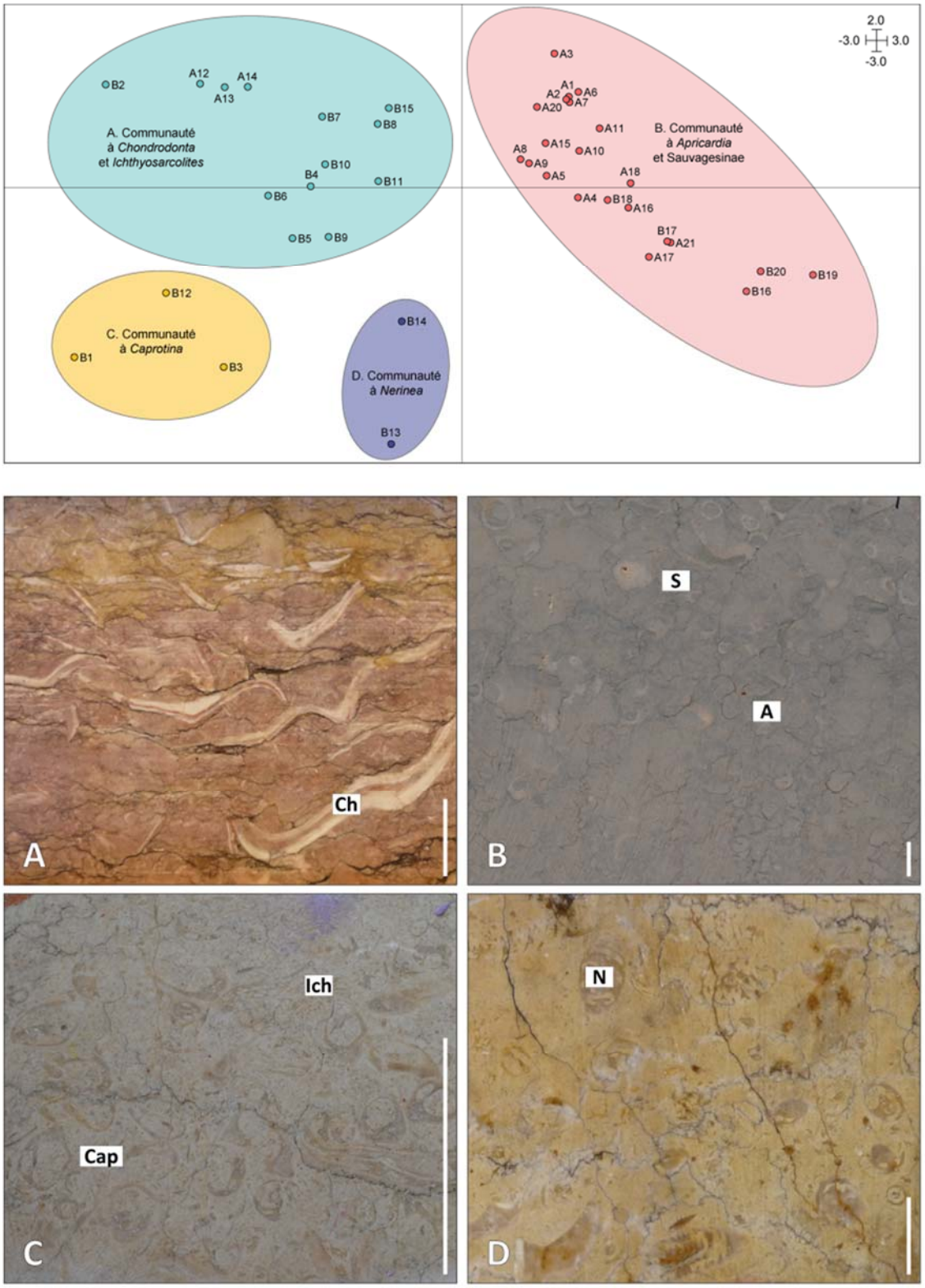


Figure 5. 10. En haut, représentation des communautés biotiques par Positionnement Multidimensionnel Non-Métrique permettant de visualiser les similarités entre les points de comptage des communautés proposées dans un espace en deux dimensions (stress : 0,129). En bas, illustration des quatre communautés biotiques retenues. A – communauté à *Chondrodonta* (Ch) dominants et *Ichthyosarcolites* ; B – communauté à Sauvagesiinae (S) et *Apricardia* (A) dominants ; C – communauté à *Caprotina* (Ca) dominants accompagnés de *Ichthyosarcolites* (I); D – communauté à *Nerinea* (N) dominants. Crédits photographiques (photos A, D) : Marc Floquet et Aurélien Blénet. Echelle: 5 cm.

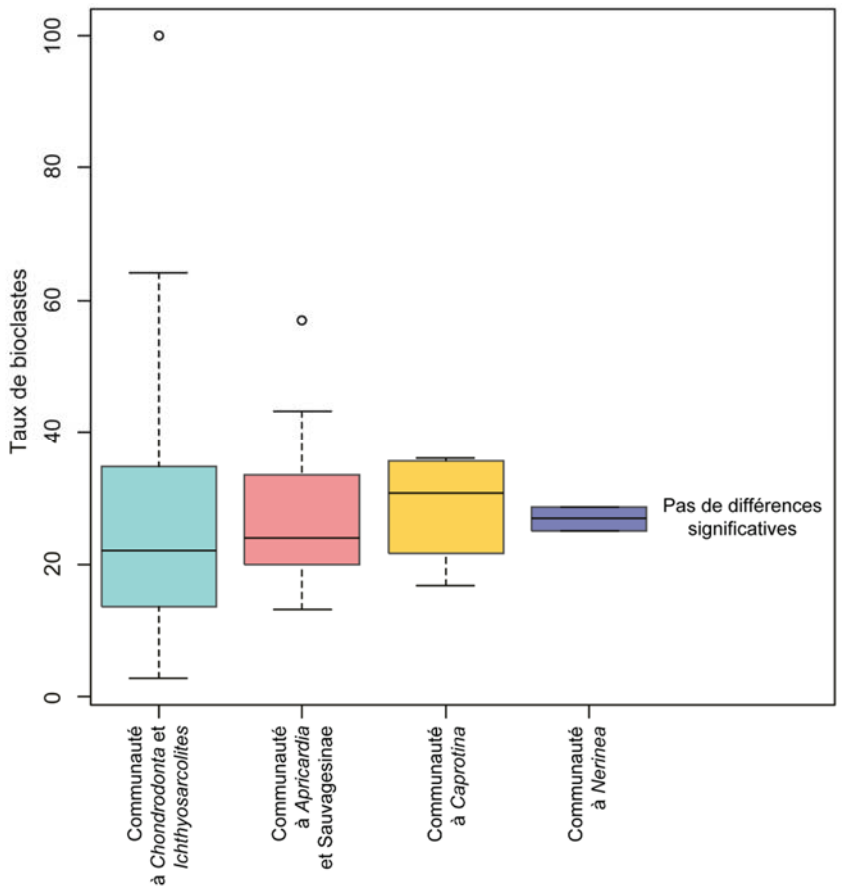


Figure 5. 11. Diagramme de variation du taux de bioclastes en fonction de la communauté macroscopique renseignée. Les quatre communautés ne sont quasiment pas affectées par le taux de bioclastes et partagent une médiane très similaire.

coquilles composées à majorité d'aragonite (*Caprinula boissyi*, *Ichthyosarcolites bicarinatus*, *Caprotina* sp.) au sein des communautés à rudistes. En lames minces, seule 1,7% de la fraction bioclastique sont identifiables, le reste des bioclastes demeurant indéterminés y compris les microbioclastes et a fortiori la micrite dérivant de l'altération ultime des bioclastes et reconnue comme telle. Concernant la part abiotique de la roche, seules l'unité stratigraphique Fb2 comporte des terrigènes sous forme de grains de quartz (environ 2%) et de glauconie.

Les assemblages micropaléontologiques (figs. 5.12 et 13) sont constitués en grande partie des bioclastes issus des rudistes, chondrodontes, gastéropodes, échinodermes, bryozoaires, éponges, algues rouges et algues vertes, plus des pithonelles, présents dans l'ensemble des séquences (fig. 5.14). Les foraminifères benthiques sont abondants dans

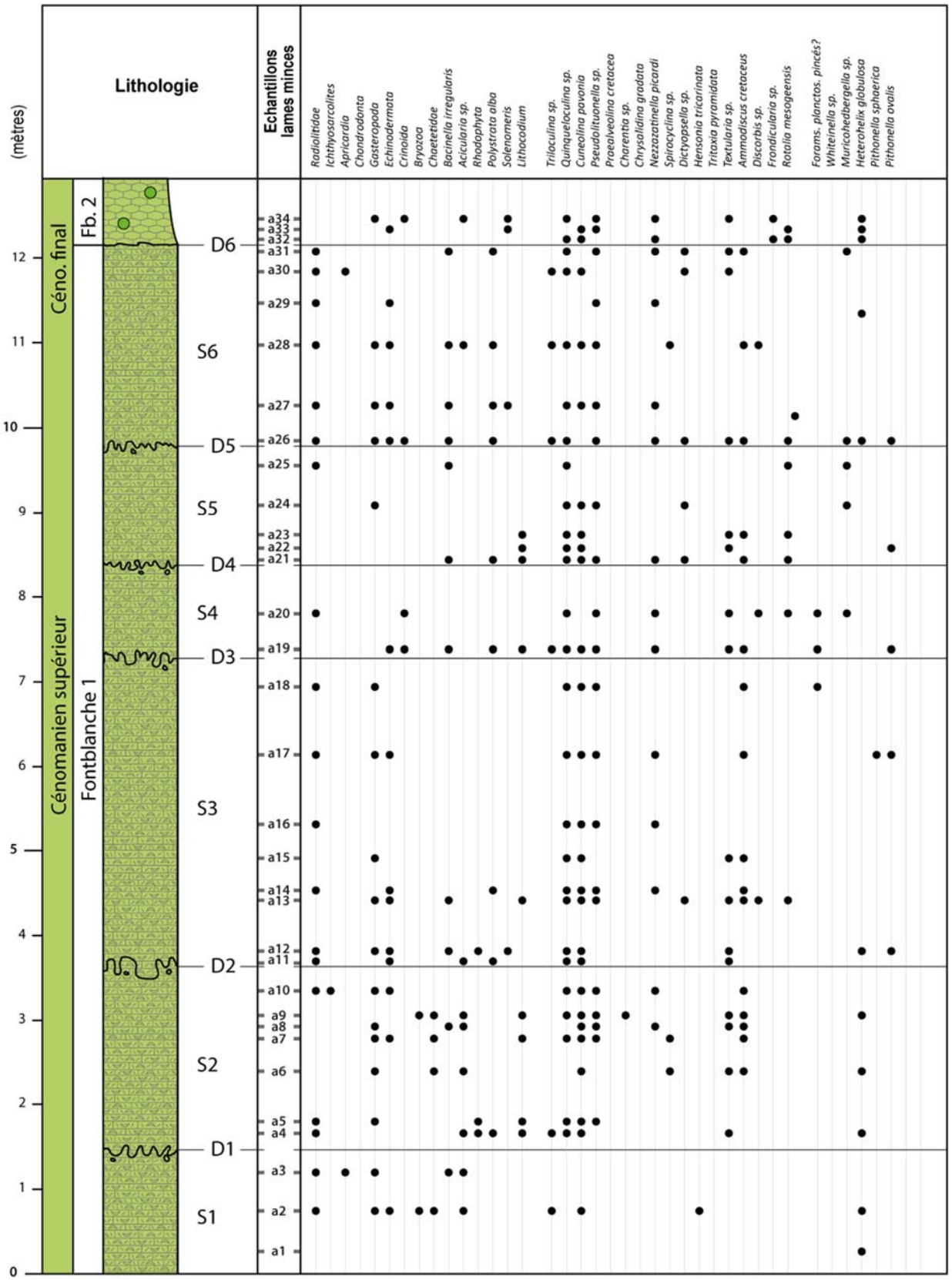


Figure 5. 12. Occurrences de taxons reconnus en lames minces de calcaires de la coupe de la Bastide d'Orves. La diversité semble constante dans toutes les séquences S1 à S6.

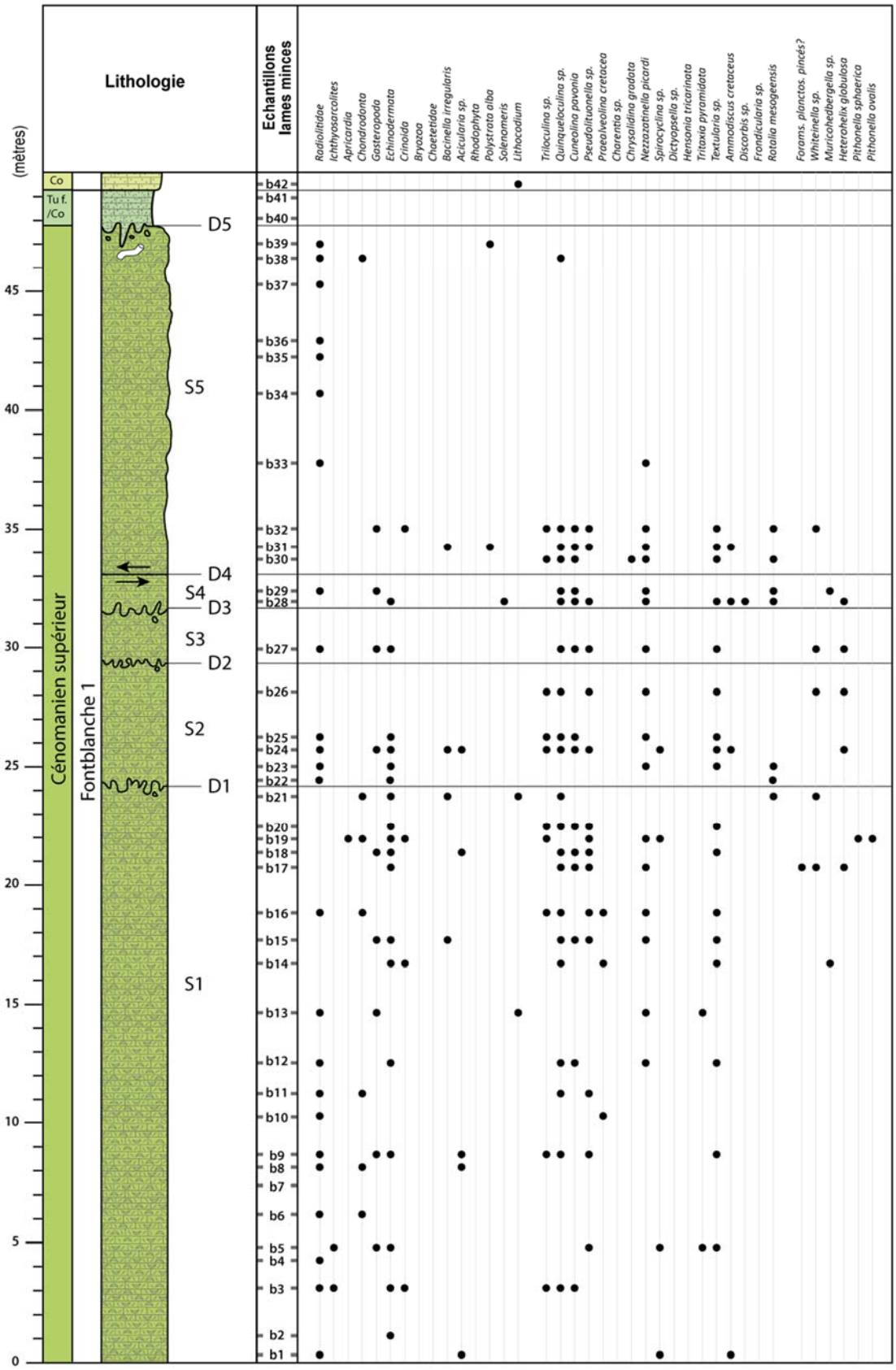


Figure 5. 13. Occurrences de taxons reconnus en lames minces de calcaires de la coupe de la Barre des Aiguilles. La biocénose est similaire à celle reconnue à partir de la coupe de la Bastide d'Orves. Toutefois, à partir de b33, la biodiversité s'effondre et quasiment aucun autre taxon de rudistes que celui des Radiolitidae voire de foraminifère benthique, ne sont reconnaissables.

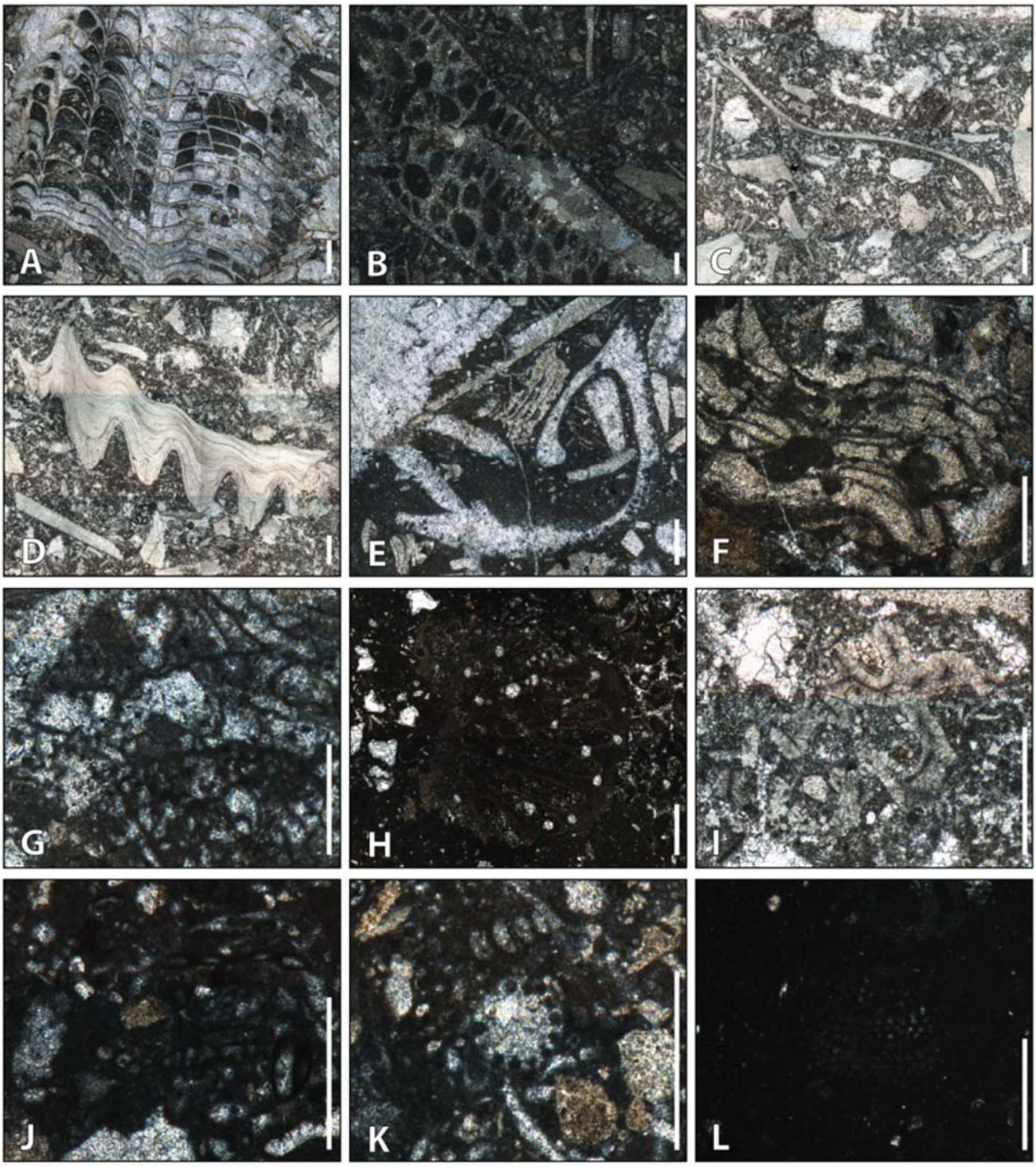


Figure 5. 14. Taxons rencontrés en lames minces de calcaires des coupes de la Bastide d'Orves et de la Barre des Aiguilles. A – Radiolitidae, B – *Ichthyosarcolites bicarinatus* Gemmellaro 1865 (Ichthyosarcolitidae, rudiste), C – *Apricardia* sp. (Requienidae, rudiste), D – *Chondrodonta joannae* Choffat 1886 (Chondrodontidae), E – Gasteropoda et Radiolitidae, F – *Polystrata alba* (Rhodophyta), G – *Bacinella irregularis* (Cyanobacteria), H – Bryozoa, I – Chaetetidae (Demosponga), J – *Lithocodium* sp. (Cyanobacteria), K – *Acicularia* sp. (Ulvophyceae)), L – *Solenomeris* sp.? (foraminifère benthique). Echelle: 500 µm.

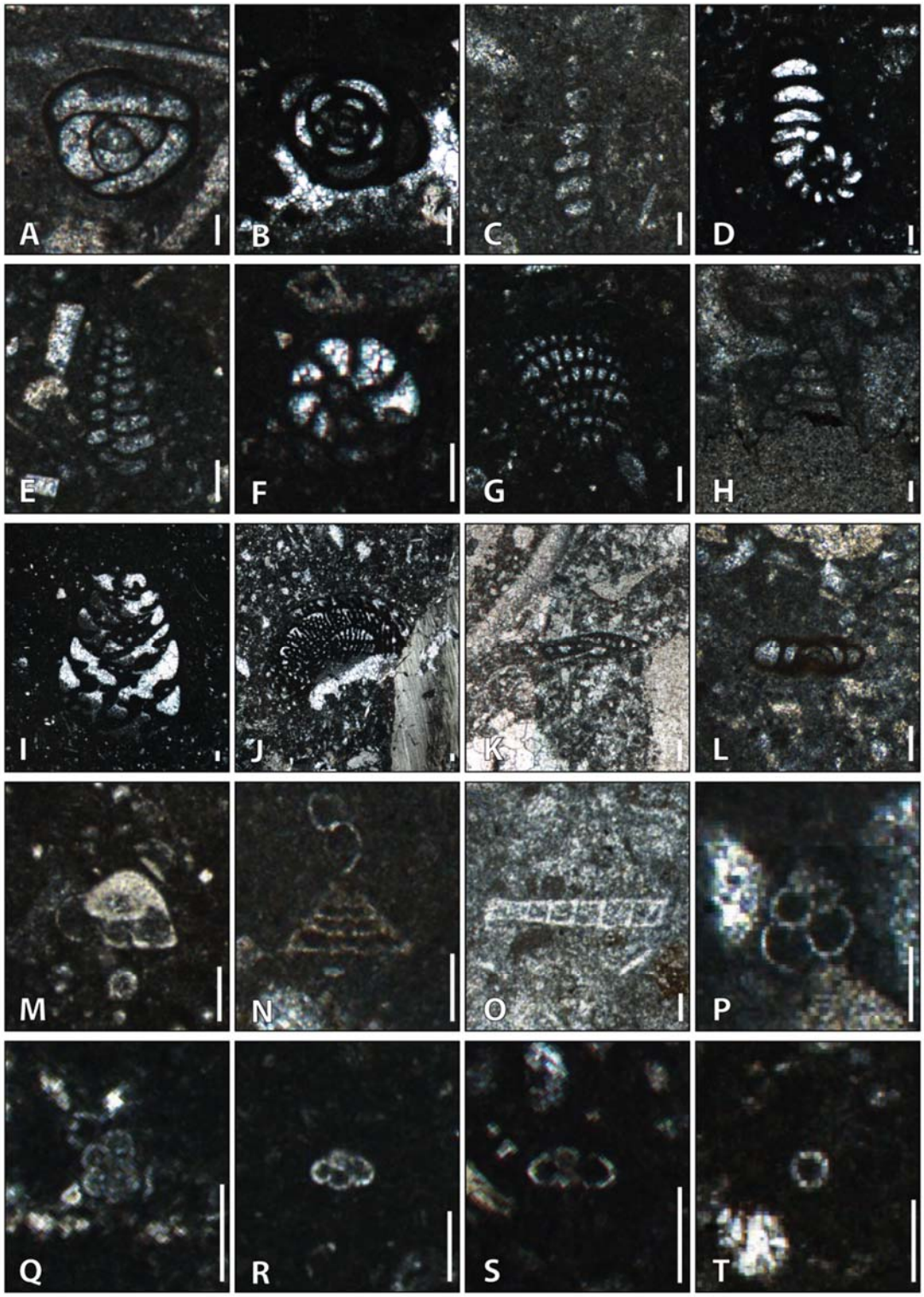


Figure 5. 15. Foraminifères (A à S) et kyste de dinoflagellé (T) rencontrés en lames minces de calcaires des coupes de la Bastide d'Orves et de la Barre des Aiguilles. A – *Triloculina* sp., B – *Quinqueloculina* sp., C – *Pseudolituonella* sp., D – *Spirocyclina* sp., E – *Textularia* sp., F – *Nezzazatinella picardi*, G – *Cuneolina pavonia*, H – *Charentia* sp., I – *Chrysalidina gradata*, J – *Praealveolina gr. cretacea*, K – *Dictyopsella* sp., L – *Planispira* sp., M – *Rotalia mesogeensis*, N – *Tritaxia pyramidata*, O – *Frondicularia sp.*, P – *Heterohelix globulosa*, Q-R – *Muricohedbergella sp.*, S – *Whiteinella sp.* ?, T – *Pithonella sphaerica*. Echelle : 100µm.

l'ensemble des séquences : miliolidés, *Cuneolina pavonia*, *Pseudolituonella sp.*, *Praealveolina cretacea*, *Charentia sp.*, *Chrysalidina gradata*, *Nezzazatinella picardi*, *Spirocyclina sp.*, *Dictyopsella sp.*, *Hensonia tricarinata*, *Tritaxia pyramidata*, *Textularia sp.*, *Ammodiscus cretaceus*, *Discorbis sp.*, *Frondicularia sp.*, *Rotalia mesogeensis* (fig. 5.15). Les foraminifères planctoniques sont en revanche très rares, mais se retrouvent dans tous les échantillons de la Bastide d'Orves, et entre b14 et b32 à la Barre des aiguilles (fig. 5.13). On trouve les taxons suivants, toujours isolés : *Whiteinella* sp*., Muricohedbergella sp.* et *Heterohelix globulosa*, ainsi que des foramnifières planctoniques à morphologie « pincée » non identifiables.

Les floatstones de la Barre des Aiguilles et de la Bastide d'Orves sont entrecoupés de discontinuités irrégulières parallèles à la stratification correspondant à des remaniements et épisodes de karstification dans lesquels on retrouve racines et cailloux noirs (Blénet, 2010; Ackouala, 2012). De très nombreux stylolithes parallèles à la stratification (stylolithes d'enfouissement et compaction) sont présents sur l'ensemble des séries des deux coupes – plus spécialement dans celle de la Bastide d'Orves. Des fractures – ainsi que des vides de dissolution de coquilles – sont remplies par des sédiments microbioclastiques, quartzo-glauconieux, laminaires géotropes (silts verdâtres, souvent oxydés en rouille-orangé) issus de FB2 au-dessus de la discontinuité sommitale de FB1 comme parfaitement visible dans la coupe de la Bastide d'Orves (D6 sur fig. 5.3). En conséquence, ces fractures peuvent être datées du Cénomanien final (Floquet *et al.*, 2013).

La série de la Bastide d'Orves présente six séquences (S1 à S6), chaque séquence se terminant par une discontinuité (D1 à D6). Le haut de la série de la coupe de la Bastide d'Orves est fait d'un calcaire noduleux, considéré comme représentant ici l'unité stratigraphique Fb2, caractérisé par la disparition complète des communautés à rudistes. La limite Cénomanien-Turonien serait incluse dans cette unité. La Barre des Aiguilles présente cinq séquences (S1-5 sur fig. 5.4) limitées par 5 discontinuités (D1-5 sur fig. 5.4). Des changements de communautés biotiques au sein des séquences lithologiques S1-6 permettent de définir des sous-séquences (par exemple S2a à S2c sur fig. 5.3 ou S1a à S1g sur fig. 5.4). Ainsi, neuf sous séquences sont reconnues dans la coupe de la Bastide d'Orves et quinze dans la coupe de la Barre des Aiguilles.

Les descriptions détaillées de chaque sous-séquence, combinant contenu faunique et texture, sont données dans le paragraphe suivant. Y sont aussi reportés le contenu biosédimentaire des lames minces, les taxons rencontrés (figs. 5.12 et 13), ainsi que le rapport bioclaste-microbioclaste/micrite du liant.

4.3.1. Bastide d'Orves

4.3.1.1. Séquence 1

La séquence S1 est un floatstone avec une seule communauté à Sauvagesiinae et *Apricardia* (S1a; fig. 5.16). Les Sauvagesiinae dominent (moyenne : 77,2%), accompagnés de quelques *Apricardia* (moyenne : 13,3%). Les coquilles ne sont pas préservées en position de vie mais ne semblent pas avoir été transportées. Le liant est constitué de micrite et de nombreux bioclastes très hétérométriques. La séquence se termine par la discontinuité D1 ou occurrent des coquilles noircies par adsorption de matière organique résultant d'une imprégnation en milieu anoxique due à des algues marines décomposées au niveau de l'étage médiolittoral (Bernier et Strasser, 1988).

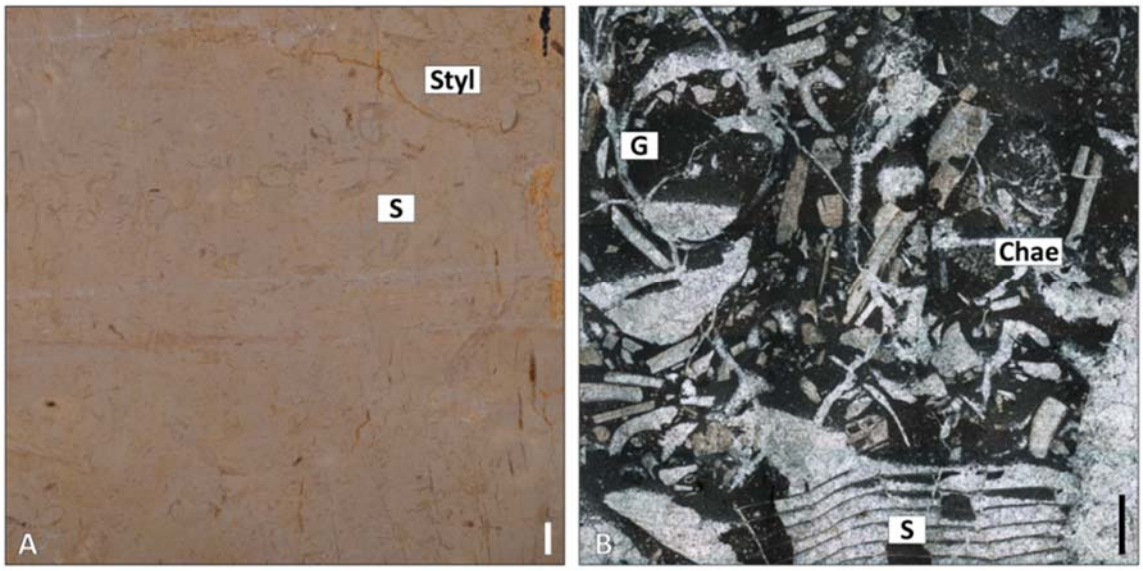


Figure 5. 16. Faciès de S1 de la coupe de la Bastide d'Orves. A – Surface de sciage avec Sauvagesiinae (S), stylolithes (Styl.). B – Liant en lame mince de a2, composé de bioclastes hétérométriques dans de la micrite microbioclastique. G : coquille de gastéropode avec un remplissage géotrope; S : test celluleux d'un Radiolitidae ; C : Chaetetidae. Echelles : 10 cm (A), 1 mm (B).

4.3.1.2. Séquence 2

La séquence S2 commence par des floatstones à Sauvagesiinae et Apricardia (fig. 5.17). Les Sauvagesiinae dominent d'abord très largement et aucun *Apricardia* n'est présent (S2a). Des coquilles de *Caprinula* accompagnent les Sauvagesiinae (11,4%). L'occurrence des *Caprinula* augmente (point suivant à 20,35%) conjointement avec les *Apricardia* très rapidement nombreuses (37,6%) au détriment des Sauvagesiinae (35,01%; communauté S2b). Les proportions de Sauvagesiinae par rapport aux *Apricardia* restent ensuite similaires tandis que les *Caprinula* diminuent fortement (2,7%) et que les *Chondrodonta* augmentent (13,4%). Au dernier point de S2 (communauté S2c), la population d'*Apricardia* diminue fortement (12,7%) et laisse place à une population très largement dominée par les Sauvagesiinae comme en S1 (83,3%). *Ichthyosarcolites* y est rare. Le liant est constitué de nombreux bioclastes parfois jointifs. Le pourcentage de bioclastes par rapport au liant est de plus en plus faible de bas en haut de S2 : les bioclastes sont jointifs de a4 à a7, puis deviennent plus rares et plus espacés. En a11, les bioclastes provenant de bivalves sont tous orientés parallèlement à la stratification. Cette séquence se termine avec la discontinuité D2 marquée par des traces d'émersion (passages de racines).

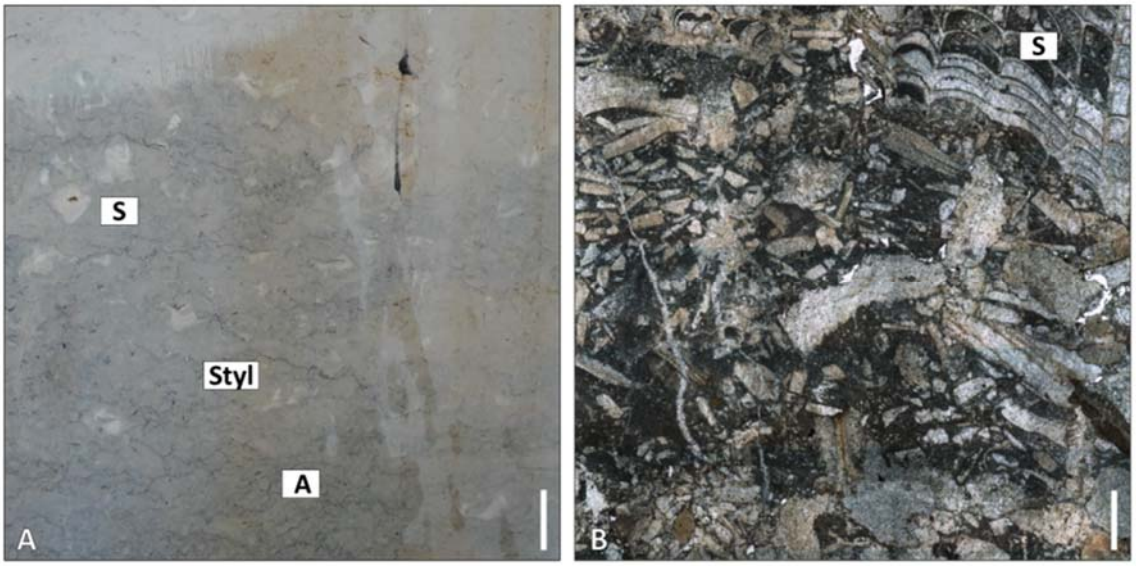


Figure 5. 17. Faciès de S2 de la coupe de la Bastide d'Orves. A – Surface de sciage avec Sauvagesiinae (S) majoritaires accompagnées d'*Apricardia* (A), stylolithes (Styl.). B – Liant en lame mince du floatsone a4, composé de bioclastes hétérométriques dans de la micrite microbioclastique. S : test celluleux d'un Sauvagesiinae. Echelles : 10 cm (A), 1 mm (B).

4.3.1.3. Séquence 3

La séquence S3 livre la même communauté à Sauvagesiinae et *Apricardia* que celles de S1 et S2 (fig. 5.18). Les Sauvagesiinae sont toujours largement dominants; cependant, on peut remarquer un accroissement temporaire de l'abondance des *Apricardia*. Les Sauvagesiinae forment au début de S3 (S3a) des populations nombreuses (compage macroscopique A7 sur fig. 5.3 : 75,5%), puis diminuent progressivement jusqu'à A9 (51,7%; S3b) et remontent ensuite progressivement jusqu'à la fin de S3 (68,83%; S3c). Les *Apricardia* sont au contraire peu nombreux (A7 : 10,7%) au début de S3, atteignent progressivement leur acmé en son milieu (A10 : 31,4%) où leur pourcentage se stabilise jusqu'à la fin (A11 : 27,5%). Des Nerineidae sont présents en tout début de S3 (A7 : 12,2%), puis très rares par la suite. Les *Chondrodonta* débutent dans S3 peu nombreux, ont une acmé en A8 (16,8%) puis diminuent progressivement. La texture floatstone possède un liant à bioclastes épars. De rares grains de glauconies se trouvent au début de S3.

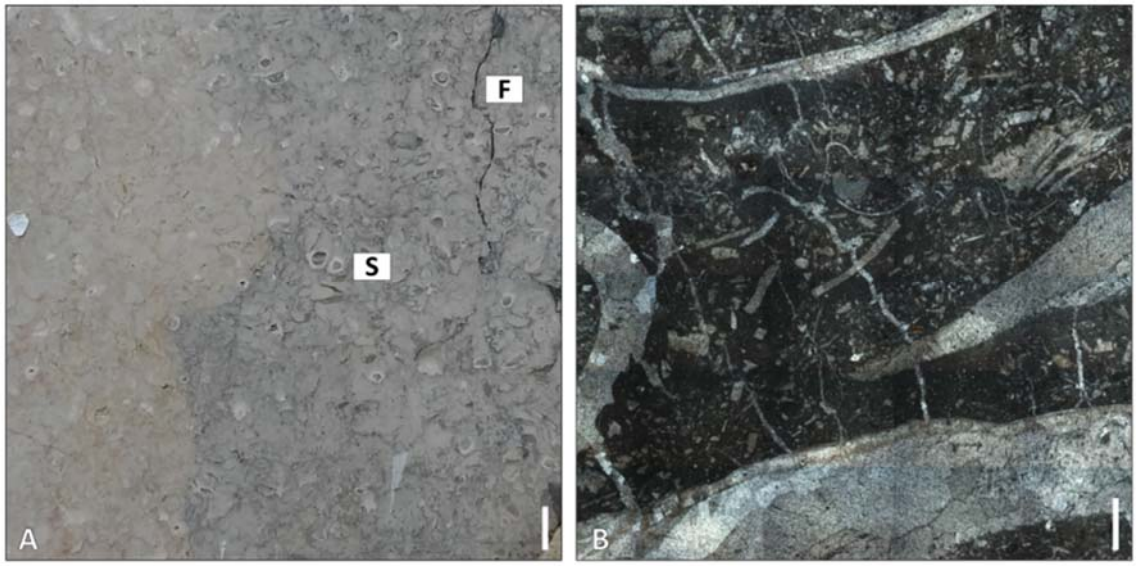


Figure 5. 18. Faciès de S3 de la coupe de la Bastide d'Orves. A – Surface de sciage avec Sauvagesiinae (S) majoritaires, remplissage laminé et géotrope des coquilles et de paléofractures (F). B – Liant en lame mince de a15, fait d'une biomicrite à bioclastes hétérométriques fragmentés. Echelles : 10 cm (A), 1 mm (B).

4.3.1.4. Séquence 4

Cette séquence est marquée par la seule quasi-absence de Sauvagesiinae et *Apricardia* dans la coupe de la Bastide d'Orves. Ce floatstone (fig. 5.19) est ici très

largement dominé par les *Chondrodonta* (moyenne : 83%; S4a). Les *Chondrodonta* sont, dans tout S4, disposés parallèlement à la stratification. Ils sont accompagnés de Sauvagesiinae (moyenne : 11,8%) et de *Caprinula* en fin de séquence (A14 : 8,8%). Le liant du floatsone est un wackestone/mudstone. De nombreux péloïdes apparaissent ponctuellement.

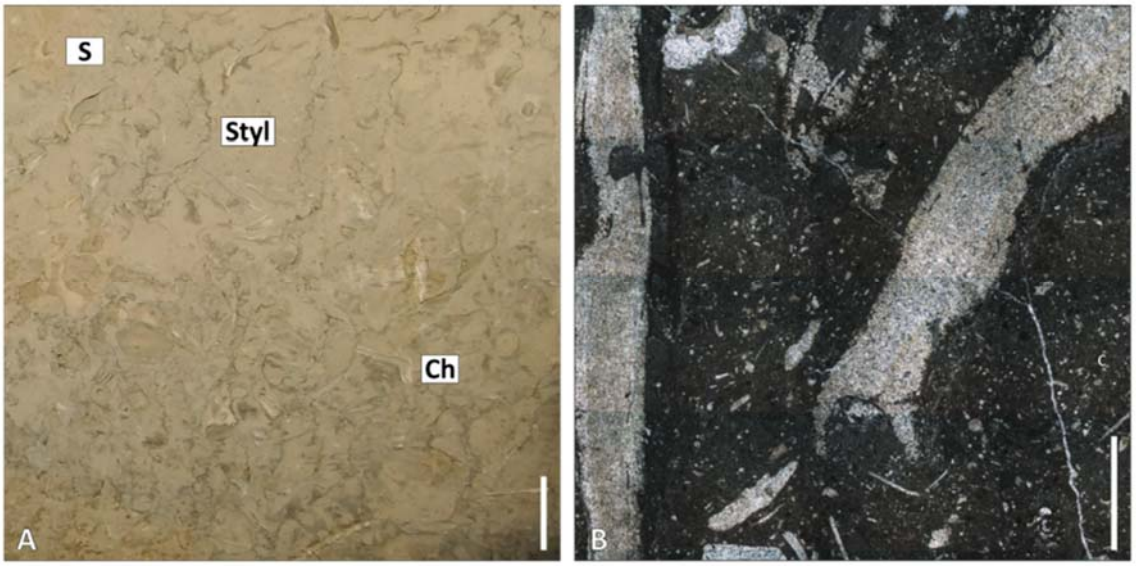


Figure 5. 19. Faciès de S4 de la coupe de la Bastide d'Orves. A – Surface de sciage avec *Chondrodonta* (C) et Sauvagesiinae (S), stylolithes (Styl.). B – Liant en lame mince de a19 composé d'une biomicrite à bioclastes hétérométriques (des microbioclastes au bioclastes de tailles centimétriques bioperforés et biocorrodés (C)). Echelles : 10 cm (A), 1 mm (B).

4.3.1.5. Séquence 5

La communauté à Sauvagesiinae et *Apricardia* revient dans S5 avec large dominance de Sauvagesiinae dont le pourcentage, de bas en haut de S5, augmente légèrement (A15 : 62,9%; A16 : 77,5%; de S5a à S5b) puis diminue drastiquement (A17 : 37,2%; de S5b à S5c). Le pourcentage des *Apricardia* évolue inversement (A15 : 25,8%; A16 : 10,8%; A17 : 56,4%; fig. 5.20). Cette évolution est inverse de celles de S2 et S3. Des *Chondrodonta* sont présents au début (A15 : 10,9%) auxquels succèdent des gastéropodes (A17 : 6,4%). Le liant est un wackestone à wackestone/mudstone (b23, b24, b25), à nombreux péloïdes.

4.3.1.6. Séquence 6

La séquence 6, toujours de texture floatstone, livre la dernière communauté à Sauvagesiinae et *Apricardia* (fig. 5.21). Elle se divise en deux sous-séquences : S6a très largement dominée par les *Apricardia* (moyenne : 74,6%) accompagnés de Sauvagesiinae (19,6%) puis S6b à quantités équilibrées d'*Apricardia* et de Sauvagesiinae (A20 : respectivement 52,2% et 46,4%). Seul le premier comptage macroscopique (A18) enregistre quelques gastéropodes (6,4%) et de rares *Caprotina* (2,1%). La richesse spécifique s'effondre de S6a (richesse spécifique moyenne : 13,7) à S6b (7,3). Le liant est un wackestone voire un wackestone/packstone en a26. La discontinuité S6 présente remaniements et karstification.

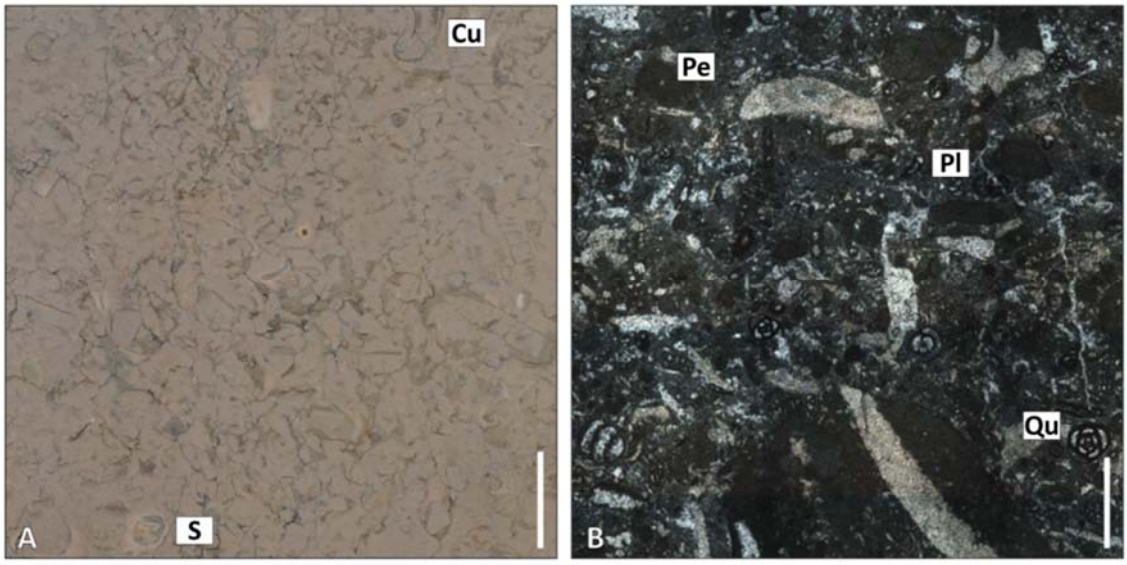


Figure 5. 20. Faciès de S5 de la coupe de la Bastide d'Orves. A – Surface de sciage avec Sauvagesiinae (S) dominantes accompagnées de *Caprinula* (Cu). B – Liant en lame mince de a23 composé d'une biomicroparite à bioclastes hétérométriques bioperforés, biocorrodés, micritisés (Pe : peloïde) et avec foraminifères benthiques (Pl : *Planispira* sp., Qu : *Quinqueloculina* sp.). Echelles : 10 cm (A), 1 mm (B).

4.3.1.7. Unité Fontblanche 2

L'unité sommitale de la coupe de la Bastide d'Orves, considérée comme représentant ici la formation dite de Fontblanche 2, d'âge probablement Cénomanien final à Turonien basal, est faite de calcaires noduleux ou disparaissent totalement les communautés à rudistes (fig. 5.22). La roche est un packstone fin, bioclastique, à plages de

Figure 5. 22. Faciès de F6 de la coupe de la Bastide d'Orves. A – Surface de dissolution du front de carrière montrant la succession du bas en haut de S6 de floatstones avec dominance des *Apricardia* (A ; S6a) vers des floatstones à Sauvagesiinae dominantes (S ; S6b). B – Liant en lame mince de a27 composé de biomicrite à bioclastes hétérométriques (microbioclastes fragmentés à bioclastes de tailles centimétriques cassés et biocorrodés (G : gastéropode ; fragments ambrés de coquilles d'*Apricardia*). Echelles : 10 cm (A), 1 mm (B).

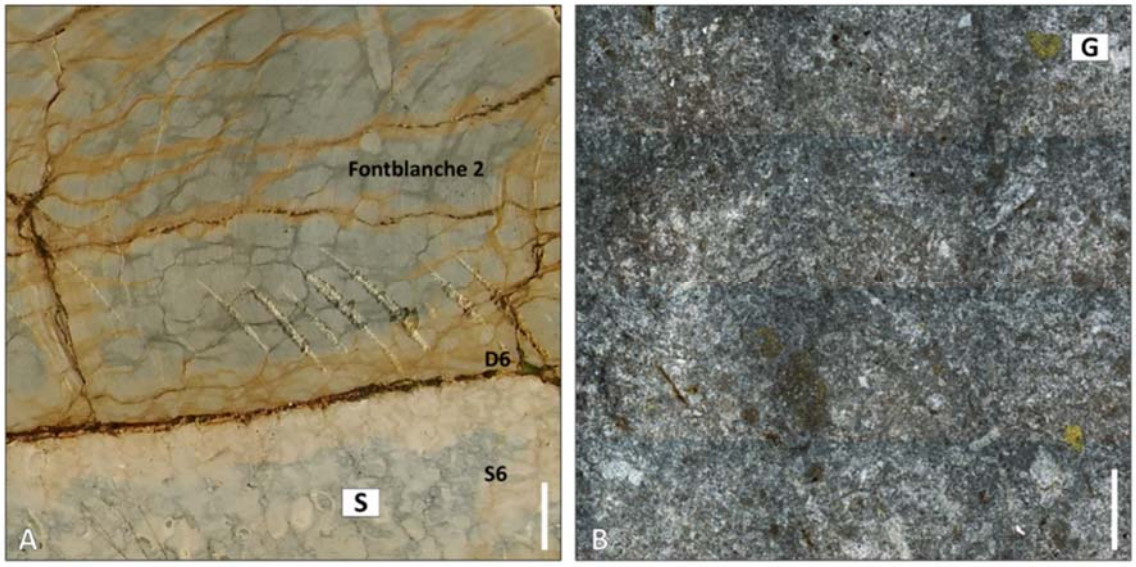


Figure 5. 22. A – Discontinuité de troncature D6 au toit de S6 de la coupe de la Bastide d'Orves (floatstone à Sauvagesiinae (S)) et marquant le sommet de l'unité stratigraphique Fontblanche 1 (Cénomanien supérieur). Calcaire noduleux de texture packstone bioclastique fin et recristallisé (biomicrosparite à fins débris dont bryozoaires, plus Pithonelles et *Heterohelix*…, grains de glauconie (G) et de quartz) et attribué à l'unité stratigraphique Fontblanche 2 (Cénomanien final à Turonien basal). Echelles : 10 cm (A), 1 mm (B).

sparite de recristallisation, à grains de glauconie et de quartz, à péloïdes issus de micritisation de bioclastes, à pyrite framboïde et débris végétaux charbonneux.

4.3.2. Barre des Aiguilles

4.3.2.1. Séquence 1

La séquence 1, de texture dominante floatstone, à liant micritique à nombreux bioclastes, est très riche en taxons dont les successions nettes ont permis de définir sept sous-séquences (S1a à S1g; figs. 5.4 et 5.23). S1a est faite d'un mince lit continu sur 50 m de *Caprotina* (B1 : 70,5%) conservées en position de vie et en bouquets. Les *Caprotina* sont accompagnées de quelques *Ichthyosarcolites* (B1 : 24,4%) et Nerineidae (B1 : 4,7%). S1b est fait d'un banc épais de 30 cm très largement dominé par des *Chondrodonta* orientés parallèlement à la stratification (B2 : 95,8%). S1c, épaisse de 2 mètres, comprend une communauté composée de Nerineidae (B3 : 29,5%), *Caprotina* (B3 : 31,4%) et *Ichthyosarcolites* (B3 : 33,6%) répartis équitablement. Les Sauvagesiinae apparaissent en très faible nombre (B3 : 3,5%) avec très rares *Apricardia* (par ailleurs rares dans tout S1 : maximum de 10% dans B14). Le spécimen Antillocaprinidae montré en figure 5.6-E a été trouvé dans S1c. La communauté à rudistes qui constitue S1d, épaisse de 13 mètres, est composée en majorité de *Chondrodonta* (moyenne : 35,5%) et *Apricardia* (moyenne : 30,6%), accompagnées d'*Ichthyosarcolites* (moyenne : 16,7%) et *Caprotina* (moyenne : 10,7%) et de rares chaetetidés, scléractiniaires et échinides. L'ensemble des trois premières communautés de S1a, S1b et S1c est accompagné par des foraminifères benthiques diversifiés (fig. 5.13). S1e est constituée en majorité par de très nombreuses *Caprotina* (B12 : 78%) accompagnées par d'abondantes *Praealveolina cretacea* tandis que les Sauvagesiinae et *Chondrodonta* sont peu représentées (B12 : respectivement 6,8% et 11,4%). Les *Ichthyosarcolites* sont encore présents mais rares (B12 : 1,52%; à partir de là, ils disparaissent définitivement dans cette coupe de la Barre des Aiguilles). Le début de S1f montre une communauté comprenant un peu plus de Sauvagesiinae dont la quantité moyenne passe de 1,2% à 8,1% et de *Chondrodonta* dont la quantité moyenne passe de 5,9% à 21,4%. Les *Caprotina* disparaissent quasiment au profit des Nerineidae ultra-dominantes (B13 : 90,1%; B14 : 58,6%), à coquilles toutes orientées dans la même

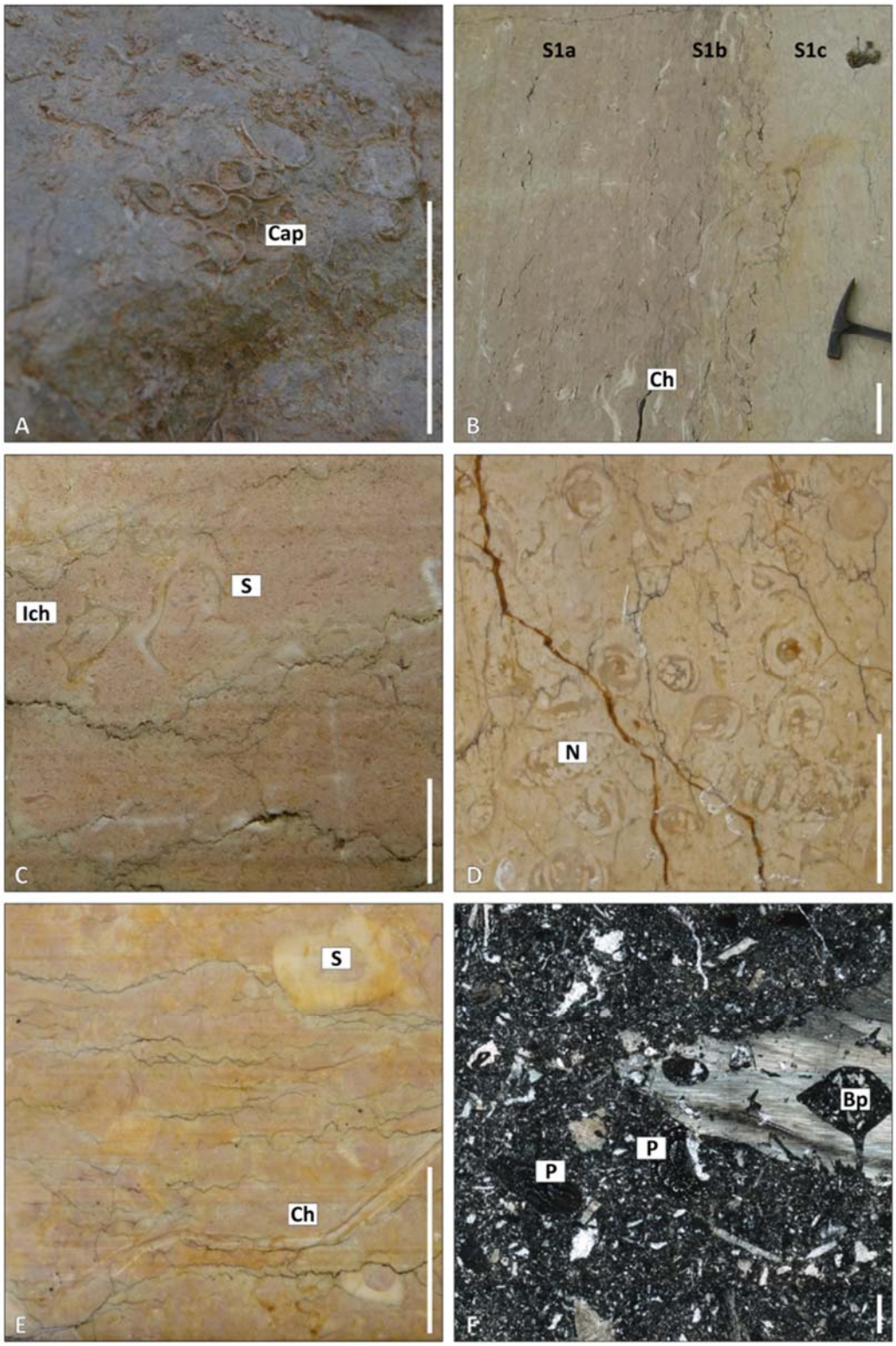


Figure 5. 23. Communautés de la séquence 1 dans la coupe de la Barre des Aiguilles. A – Bouquet de *Caprotina* conservé en position de vie (communauté de S1a). B – Succession des communautés de S1a (*Caprotina* et *Ichthyosarcolites*), S1b (*Chondrodonta* probablement accumulées sous l'action de tempête), S1c (*Caprotina*, *Ichthyosarcolites* et *Nerinea*). C – Communauté S1d à *Ichthyosarcolites* (Ich), *Caprotina* et Sauvagesiinae (S). D – Accumulation de *Nerinea* (N) (probablement sous l'action de tempête) = communauté de S1f. E – Association à Sauvagesiinae (S) et *Chondrodonta* (Ch). F – Liant microbioclastique et micritique du floatstone b14, avec bioclastes hétérométriques bioperforés (Bp) et biocorrodés, à foraminifères benthiques (P : *Praealveolina gr. cretacea*). Echelles : 10 cm (A, B, E), 1 mm (F), 1 cm (C).

direction. S'ajoutent de rares Chaetetidae. S1g comprend pour moitié des *Chondrodonta* (48,6%) et des Sauvagesiinae (51,4%). Le haut de S1 livre d'abondantes *Pithonella sphaerica* et *Pithonella ovalis* et de rares grains de glauconie. La biocorrosion y apparaît amplifiée. S1 se termine par un lit très irrégulier avec des lithoclastes, dont des cailloux noirs indicateurs d'un remaniement sédimentaire.

4.3.2.2. Séquence 2

La séquence S2, de texture floatstone (figs. 5.4 et 5.24) débute par S2a, épaisse de 1 mètre, caractérisée par une communauté à *Chondrodonta* et Sauvagesiinae. Suit S2b, épaisse de 4 mètres, caractérisée par un assemblage à *Apricardia* très largement dominantes (B16 : 90,4%), dans un liant micritique parfois sparitisé et à bioclastes hétérométriques parfois presque entièrement micritisés mais dont le contour est encore visible. Existent de rares Nerineidae (B16 : 5,2%), *Chondrodonta* (B16 : 2%) et *Caprotina* (B16 : 0,8%). Les *Caprotina* disparaissent ensuite totalement du reste de la coupe de la Barre des Aiguilles. Les foraminifères benthiques sont nombreux tandis que les foraminifères planctoniques sont rares (fig. 5.13). Une discontinuité irrégulière avec remaniements termine S2.

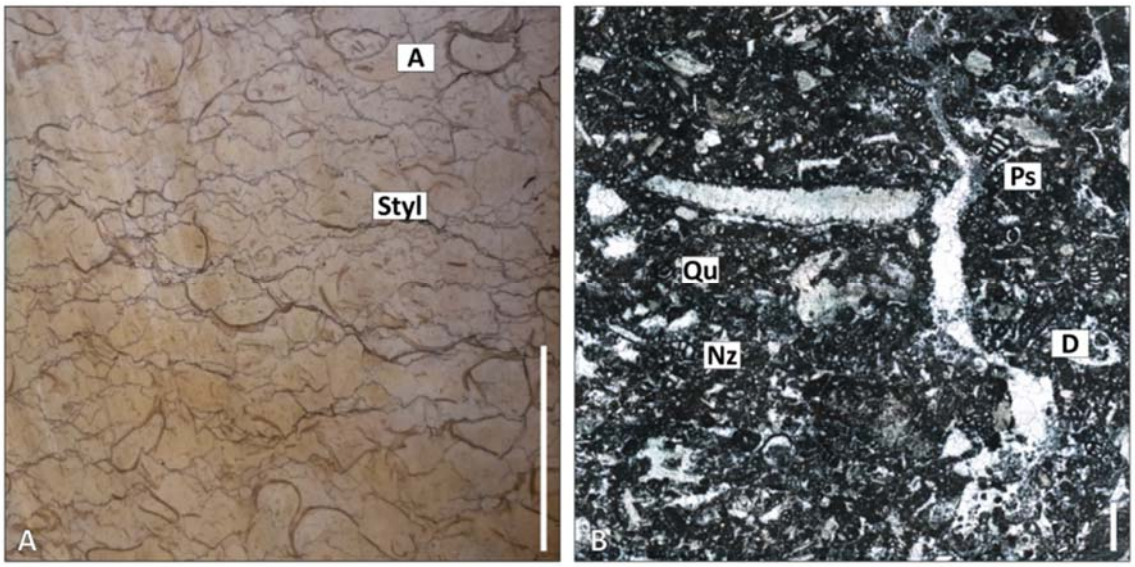


Figure 5. 24. Faciès de S2 de la coupe de la Barre des Aiguilles. A – Surface de sciage à *Apricardia* (A) majoritaire, nombreux stylolithes (Styl.). B – Liant en lame mince de b24, de biomicite à biomicrosparite à foraminifères benthiques (Ps : *Pseudolituonella* sp., D : *Dicyclina* sp., Nz : *Nezzazatinella picardi*, Qu : *Quinqueloculina* sp.), à bioclastes hétérométriques bioperforés et biocorrodés. Echelles : 10 cm (A), 1 mm (B).

4.3.2.3. Séquence 3

La séquence S3 est composée d'une unique communauté S3a où les *Apricardia*, dans une texture floatstone, sont dominants (moyenne : 65,5%; figs. 5.4 et 5.25), où les Sauvagesiinae apparaissant secondaires (moyenne : 30,2%) et les *Chondrodonta* rares (B18 : 5,3%). Des foraminifères benthiques et planctoniques sont présents (fig. 5.13). La richesse spécifique y est importante (11 taxons reconnus), le taux de bioclastes faible (15%). S3 se termine par une discontinuité comprenant des remaniements et un horizon irrégulier à cailloux noirs.

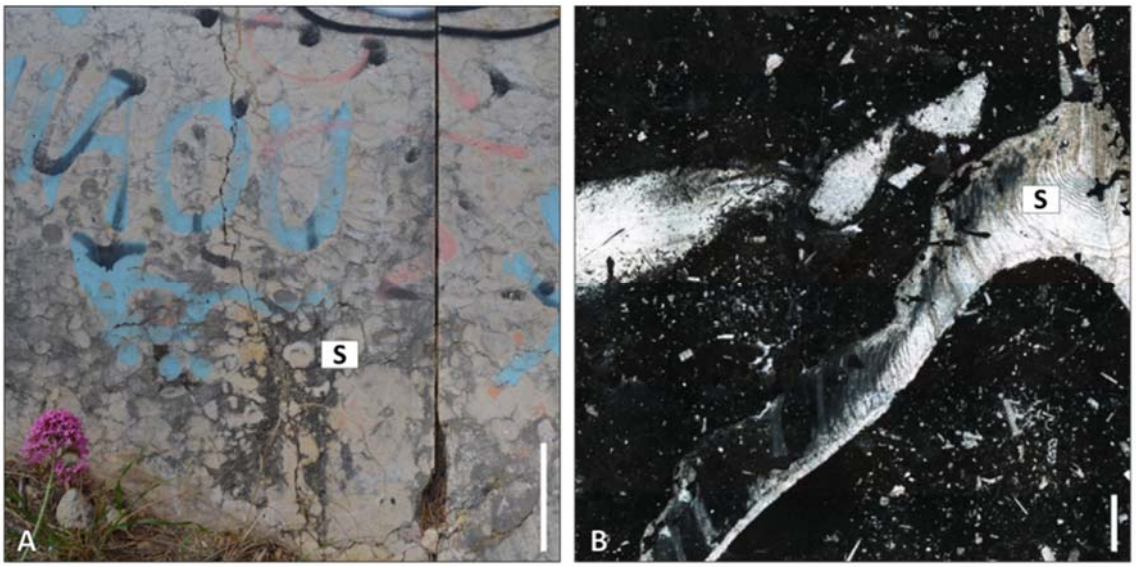


Figure 5. 25. Faciès de S3 de la coupe de la Barre des Aiguilles. A – Surface de sciage à Sauvagesiinae (S) et *Apricardia*. B – Liant en lame mince de b27, fait de bioclastes hétérométriques (microbioclastes et débris de Radiolitidae de taille centimétrique) bioperforés et biocorrodés (C). Echelles : 10 cm (A), 1 mm (B).

4.3.2.4. Séquence 4

S4 est caractérisée par un peuplement quasiment monospécifique à Apricardia (moyenne : 96,6%; figs. 5.4 et 5.26; communauté unique S4a) accompagnées de quelques rares Sauvagesiinae (moyenne : 3,4%) dans une texture floatstone dont la matrice est mudstone à wackestone faiblement bioclastique. La richesse spécifique y est élevée (11 taxons reconnus), avec foraminifères benthiques et planctoniques. S4 se termine par une discontinuité comprenant un lit très irrégulier à cailloux noirs mais dont l'observation est rendue difficile par la présence d'une faille décrochante sénestre.

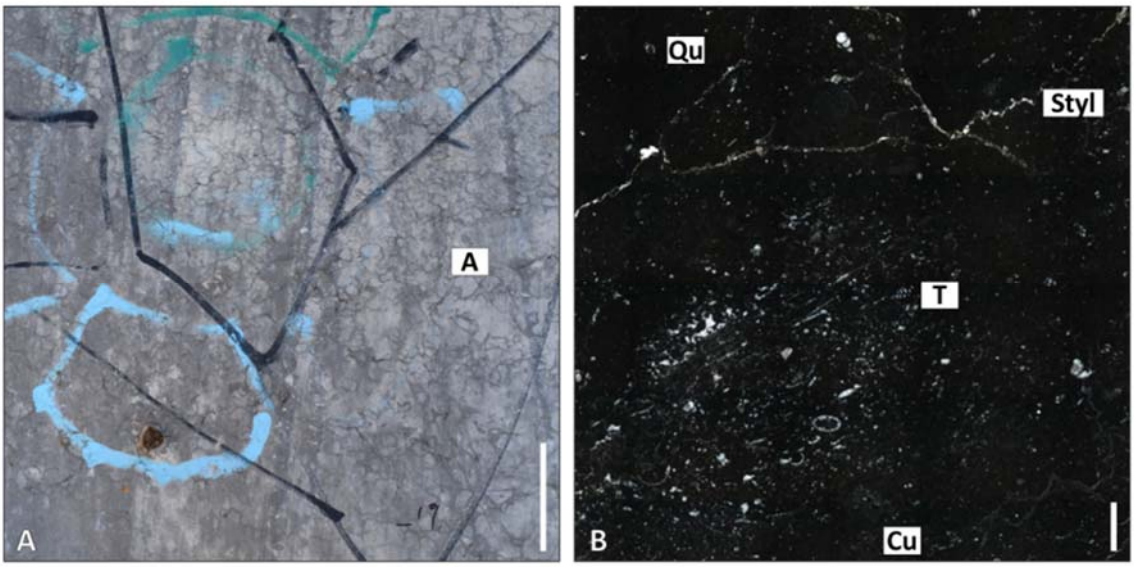


Figure 5. 26. Faciès de S4 de la coupe de la Barre des Aiguilles. A – Surface de sciage montrant une population très dense quasiment uniquement constituée d'*Apricardia* (A). B – Liant en lame mince de b28, de micrite dense à microbioclastes et foraminifères benthiques (Qu : *Quinqueloculina* sp., T : *Textularia* sp., Cu : *Cuneolina pavonia*), stylolithes (Styl). Echelles : 10 cm (A), 1 mm (B).

4.3.2.5. Séquence 5

La suite de la coupe de la Barre des Aiguilles correspond à la partie haute de l'ancienne carrière qui n'a pas été exploitée par sciage au câble (sauf tout au sommet aujourd'hui difficilement accessible) de sorte qu'il n'y a plus de surfaces lisses permettant un comptage précis. Cette partie, épaisse de 15 mètres (fig. 5.4), constitue S5 dans laquelle 4 successions biotiques ont été reconnues et qui définissent les 4 sous-séquences S5a à d (fig. 5.4). S5a, épais de moins d'1 mètre, comprend de nombreux *Chondrodonta* (fig. 5.27). De S5b à S5d, une communauté à *Apricardia* et Sauvagesiinae se développe, avec une très large dominance de Sauvagesiinae. Contrairement à la série de la Bastide d'Orves et aux séquences sous-jacentes dans la Barre des aiguilles, la richesse spécifique est anticorrélée au taux de bioclastes. Dans les 2 à 3 premiers mètres, la richesse spécifique est importante avec de nombreux foraminifères benthiques (fig. 5.13) et quelques foraminifères planctoniques; le taux de bioclastes dans le liant est très faible. A partir de b33, la richesse spécifique s'effondre et le taux de bioclastes grimpe. Les Sauvagesiinae sont ultra-dominants, très souvent sous forme de coquilles intactes dans une matrice incluant de nombreux bioclastes de Sauvagesiinae. Aucun autre taxon n'est précisément

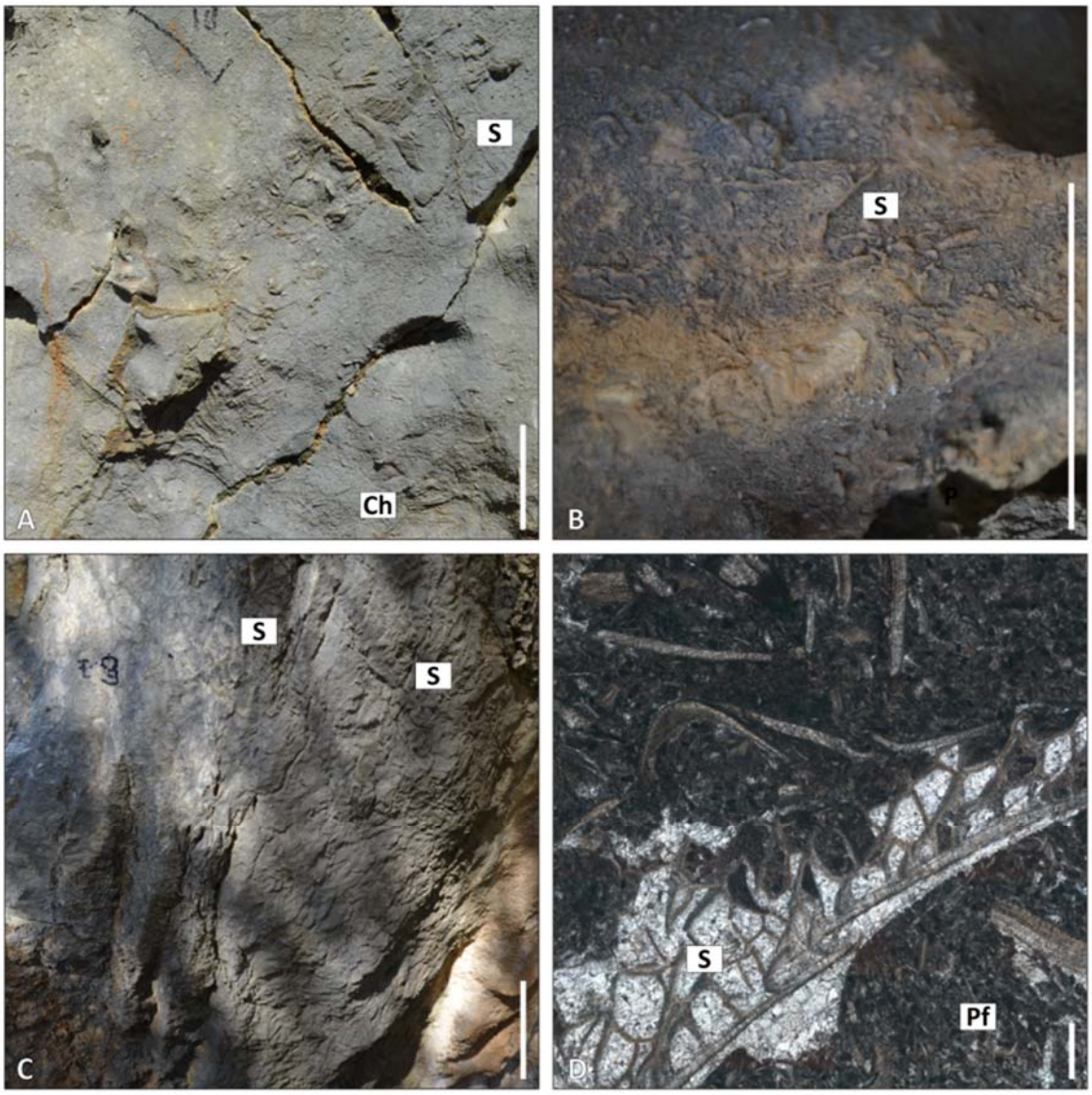


Figure 5. 27. Différents faciès de S5 de la coupe de la Barre des Aiguilles. A – Communauté S5c à Sauvagesiinae (S) et *Chondrodonta* (Ch). B – Accumulation de valves gauches operculaires de Sauvagesiinae (communauté de S5a). C – Biostrome à Sauvagesinae (S) en position de vie (communauté de S5b). D – Lame mince de b35 ; floatstone à gros débris de Sauvagesiinae (S) dans un liant essentiellement fait de bioclastes (de rudistes) hétérométriques, fragmentés et peu ou non biocorrodés, avec nombreux péloïdes de micritisation (Pf), et très peu de micrite. Echelles : 10 cm (A à C), 1 mm (B).

identifiable – quelques rares foraminifères benthiques. Des péloïdes sont abondants. Parfois, les Sauvagesiinae ont été préservés en position de vie (figs. 5.28) ce qui crée une texture bafflestone. Le liant ou matrice des rudistes entiers ou en gros débris est essentiellement finement bioclastique (taille de 20 µm en moyenne), les débris étant issus de Sauvagesiinae. Des valves gauches operculaires de Sauvagesiinae sont parfois

Figure 5. 28. A – Bioherme à Radiolitidae (probables Sauvagesiinae) conservés en position de vie. Les planchers de la couche calcitique des valves droites récipientes (cernées d'un trait) sont bien visibles. B – Tracés des individus de A (D : cavité générale de la valve droite). La flèche indique le haut. Echelle : 1 cm.

retrouvées ensemble, isolées de leurs valves droite. Les traces de biocorrosion sont rares. Quelques grains de quartz apparaissent au sommet de S5, infiltrés à partir de la formation surincombante suivant des bioturbations partant d'une discontinuité de troncature sommitale (D5 sur figs 4 et 12) et provenant de la séquence S6 et remaniés.

4.4. Couverture en discordance

Une formation de calcaires très quartzeux (moyenne : 26,7%), pauvres en bioclastes reconnaissables (foraminifères benthiques micritisés et écrasés, algues rouges) recouvre en discordance cartographique (Hennuy, 2003, pl. 8) la série de calcaires purs d'âge Cénomanien supérieur de la Barre des Aiguilles (fig. 5.29). Cette formation, nommée A3

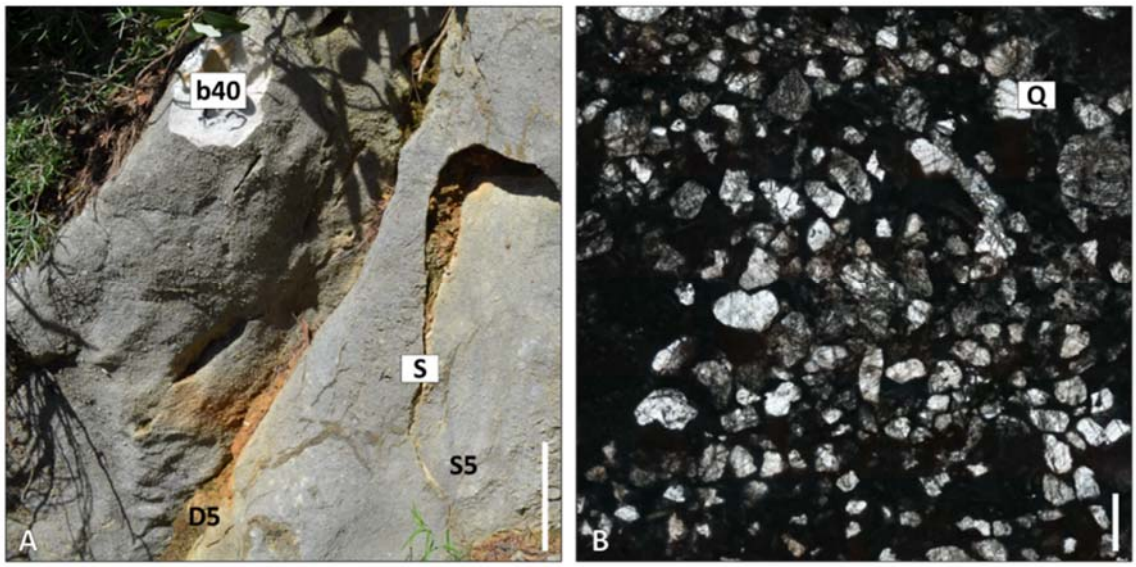


Figure 5. 29. A – Limite, marquée par la discontinuité D5, entre S5 (sommet de l'unité Fontblanche 1) du haut de la coupe de la Barre des Aiguilles, composée majoritairement de Sauvagesiinae (S) et les formations surincombantes de calcaire quartzeux dépourvu de rudistes. B – Lame mince de b40 pris dans la formation supra-Fb1 montrant une micrite très riche en quartz (Q, 30 à 40 %, de 300 µm à 1 mm de taille), faciès caractéristique de l'unité A4 au sens d'Hennuy (2003) d'âge Turonien final et en discordance cartographique sur la série cénomanienne de la Barre des Aiguilles (Hennuy 2003, planche 8). Echelles : 10 cm (A), 1 mm (B).

et datée du Turonien final par Hennuy (*ibidem*) est surmontée par une mégabrèche à éléments de calcaires à Vaccinites d'âge Coniacien inférieur (RSC4 d'Hennuy, *ibidem*).

5. Discussion

5.1. Paléoenvironnements et systèmes biosédimentaires associés

5.1.1. Unité Fontblanche 1

Le paléoenvironnement général reconnu à partir de l'étude des calcaires constituant la formation Fontblanche 1 dans les coupes de la Bastide d'Orves et de la Barre des Aiguilles est celui d'une mer peu profonde sous climat tropical, comme en attestent la faune et la flore reportée, et notamment les associations de rudistes, algues vertes (*Acicularia* sp.) et rouges (Philip 1970, 1972; Sanders 2001), foraminifères benthiques. *Bacinella*

irregularis se rencontre dans les eaux peu profondes en milieu marin ouvert, du subtidal à l'intertidal (Rameil *et al.* 2010).

La genèse des séquences biosédimentaires, particulièrement bien représentées dans la coupe de la Bastide d'Orves (S1 à S6) résulte de l'action combinée et répétée d'approfondissements puis de diminution de profondeur jusqu'à éventuelle émersion par accumulation biosédimentaire en milieux marins peu profonds de plate-forme carbonatée interne (classiquement décrites dans la littérature : *shallowing upward sequences*; Floquet, 1982). La production biosédimentaire due au développement des communautés à rudistes et organismes associés puis l'accumulation biosédimentaire faisant suite à la dégradation mécanique et biologique des coquilles entraînaient la diminution progressive de la tranche d'eau et son comblement jusqu'à émersion complète. Les émersions sont reconnaissables grâce aux discontinuités séparant les séquences (D1 à D6 dans la coupe de la Bastide d'Orves, D1 à D5 dans celle de la Barre des Aiguilles). Des évidences d'émersion sont livrées localement par certaines discontinuités : *fenestrae* d'étagement intertidal à supratidal, terriers ouverts et traces de passage de radicelles, accumulation en poches de cailloux noirs issus, remaniés, de milieux marécageux (marais maritimes ou milieux palustres; Strasser et Davaud 1983, Brouwer et Fitzpatrick 2002, Hennuy 2003) probablement situés à proximité. L'apparition d'une nouvelle séquence de dépôt correspondait à une remise en eaux marines et à la mise en place d'une tranche d'eau marine telle qu'elle permettait l'installation et le développement de telle ou telle communauté à rudistes et organismes commensaux.

Les très nombreux grains biocorrodés et entourés d'une gangue de micrite sont les témoins d'une forte activité biologique. Ceci est connu dans les mers tropicales peu profondes où pénètre la lumière, probablement quelques dizaines de mètres de profondeur au maximum (Bernard-Dumanois et Delance 1983). La biocorrosion due principalement à des bactéries se fait au cours de la vie des bivalves ainsi que juste après leur mort (Flügel, 2004). Les bioclastes provenant des rudistes sont également perclus d'*Entobia*, dus à des éponges endolithes, les Clionidae. L'importante corrélation entre le taux de bioclastes et la richesse spécifique des points b1 à b29 de la coupe de la Barre des Aiguilles et dans l'ensemble de la coupe de la Bastide d'Orves s'interprète comme une mesure du degré de biocorrosion (plus la corrosion est importante, plus l'identification des taxons au rang

spécifique est difficile et plus la richesse spécifique perçue diminue). La corrélation est uniquement testée dans les floatstones/bafflestones à rudistes ou on sait qu'occurrent les communautés. La matrice micritique (toujours avec microbioclastes) semble être le résultat ultime de la biocorrosion. Un continuum taphonomique existe donc entre les coquilles de rudistes entières, puis ensuite perforées, cassées et biocorrodées, générant des bioclastes de plus en plus petits et finalement une micrite liant les coquilles et les bioclastes restants. Le rapport bioclastes/micrite traduit donc l'intensité de la dégradation des coquilles, notamment par l'activité bactérienne : plus il y a de bioclastes et moins les biocorrosions/bioperforations sont intenses. Dit autrement, plus les bioclastes ont été micritisés et plus le pourcentage de micrite est élevé. Une biocorrosion importante entraîne donc une diminution de la biodiversité. Ceci est notoire pour les foraminifères, notamment benthiques, très souvent micitisés d'où leur faible richesse apparente. Cette diminution de biodiversité est donc artéfactuelle et traduit les conditions taphonomiques du milieu lorsque sont corrélés richesse spécifique et taux de bioclastes.

Dans la coupe de la Barre des Aiguilles, on ne retrouve plus cette corrélation significative entre richesse spécifique et taux de bioclastes sur les points b30 à b42. Là, la diminution de biodiversité traduit une réalité écologique et n'est pas conduite par des facteurs taphonomiques. Les Sauvagesiinae forment dans la séquence S5 des peuplements oligospécifiques.

5.1.2. Unité Fontblanche 2

Le changement radical de sédimentation au-dessus de la discontinuité sommitale de FB1 (tout spécialement au-dessus de D6 dans la coupe de la Bastide d'Orves) et clairement exprimé par les dépôts de la formation de Fontblanche 2, traduit l'apparition et l'installation d'un paléoenvironnement totalement différent du précédent. La biocénose (bryozoaires, échinodermes, lagénidés, foraminifères planctoniques) indique une nette ouverture vers les milieux marins infralittoraux associée à un approfondissement. En outre, l'abondance de *Pithonella* et d'*Heterohelix* serait indicatrice d'un haut contenu en nutriments et de conditions eutrophiques (Wendler *et al.* 2002). Ceci correspondrait à un bloom planctonique en relation avec l'événement anoxique global EAO2 qui advint à ce moment (Dias-Brito, 2000; Omaña *et al.*, 2014). Les nutriments, les nombreux débris végétaux

charbonneux, les fréquents quartz, ainsi que la genèse de glauconite, correspondent à des apports de matériel issu de milieux préalablement émergés, et remanié lors de la transgression marine associée aux ouverture et approfondissement. La conséquence finale est que le paléoenvironnement de plate-forme carbonatée superficielle à rudistes type Fontblanche 1 a disparu.

5.2. Ecotypes et communautés à rudistes

Les communautés à rudistes définies précédemment (communautés à *Nerineidae*, à *Caprotina*, à *Chondrodonta* et *Ichthyosarcolites*, à Sauvagesiinae et *Apricardia*) n'expliquent qu'une très faible part de la variation du taux de bioclastes. Ces communautés appartiennent donc toutes au même environnement sédimentaire; en revanche, on peut relier l'écologie des taxons de rudistes à un taux de production de sédiments (Skelton 1978, 1991; Gili et Skelton 2000). On référence trois écotypes de rudistes en fonction de l'orientation de leur plan de commissure (Skelton 1978, 1991) : les élévateurs sont, entre autres, adaptés à une sédimentation rapide, et peuvent développer des coquilles verticales rapidement; les organismes couchés (« *recumbent* ») sur le sédiment au contraire se développent dans des environnements à sédimentation lente, peu enfouissants, ou a substrat meuble; les organismes de type « *clinger* » formant un intermédiaire entre les types couchés et élévateurs, plantés de manière oblique dans un substrat meuble. Finalement, les organismes encroûtants nécessitent un substrat dur tout au long de leur développement. La production sédimentaire due au développement des communautés à rudistes et à la dégradation des coquilles exerce donc un rétrocontrôle sur ces mêmes populations à rudistes (Steuber et Löser 2000). Les Sauvagesiinae *Sauvagesia sharpei* et *Durania arnaudi* (Radiolitidae) sont des élévateurs. Ils sont associés à la plupart des fins de séries, avant un comblement. Ils sont également la plupart du temps accompagnés par *Apricardia* sp. (Requienidae) qui est en revanche une forme de type « *clinger* ». La communauté à *Apricardia* et Sauvagesiinae présente une grande amplitude d'occurrences entre ces deux taxons, et pourrait être reliée à ces deux écotypes différents, bien que proches. *Apricardia* se développe spécifiquement dans des environnements à taux de sédimentation plus faible que les Sauvagesiinae. *Caprinula boissyi* est également un taxon de type « *clinger* », mais se retrouve en très faibles proportions, comme taxon

accompagnant les taxons dominants *Chondrodonta,* Sauvagesiinae et *Apricardia*. La domination du substrat par *Apricardia* semble empêcher le développement des *Caprinula* (fig. 5.9). L'occurrence de Radiolitidae dans quasiment toutes les séquences des 2 coupes étudiées implique un succès écologique en relation avec l'explosion de la famille des Radiolitidae qui s'ensuit au Crétacé supérieur (Skelton 2003). Vu leurs écotypes, les rudistes couchés et « *clinger* » se développent préférentiellement au Crétacé inférieur sur des zones à faible taux de sédimentation. Avec l'avènement des formes élévatrices telles que les Radiolitidae et les Hippuritidae, les rudistes vont pouvoir coloniser des zones ou la sédimentation est nettement plus rapide, par une croissance rapide résultant en la formation d'importants biostromes.

Chondrodonta joannae est également un taxon de type « *clinger* », et accompagne effectivement les Sauvagesiinae tout comme *Apricardia*. On ne trouve en revanche jamais de populations dominées à la fois par *Chondrodonta* et *Apricardia*. Cependant, toutes les coquilles de *Chondrodonta* ne sont probablement pas préservées *in situ*. L'accumulation en grand nombre de *Chondrodonta* avec leurs 2 valves en connexion, tous orientés parallèlement à la stratification (Ayoub-Hannaa et Fürsich 2011), ainsi que l'occurrence de très nombreux Nerineidae orientés résulte probablement de l'action de tempêtes (tempestites). La communauté S1f de la coupe de la Barre des Aiguilles – l'unique communauté à Nerineidae – ainsi que les communautés à *Chondrodonta* S1b de la coupe de la Barre des Aiguilles et S4a de la coupe de la Bastide d'Orves sont considérées comme étant des tempestites, les coquilles apparaissant para-autochtones à allochtones. Les coquilles de *Chondrodonta* accompagnant les Sauvagesiinae et *Apricardia* des séquences S2 et S3 dans la coupe de la Bastide d'Orves sont probablement parautochtones (au sens de Kidwell et Holland 1991), ainsi que dans les communautés S1d et S1g dans la coupe de la Barre des Aiguilles où elles accompagnent les Sauvagesiinae, les *Ichthyosarcolites* et les *Caprotina*.

Les deux communautés à *Chondrodonta* et *Ichthyosarcolites* d'une part et à *Caprotina* d'autre part forment un écosystème très différent de celui des précédentes communautés. Les *Ichthyosarcolites* sont connus pour être des organismes pionniers de substrats meubles (Videt 2003, 2007), s'installant sur des vasières à faible taux de sédimentation, ce qui permet à ces organismes de type couché de ne pas être enfouis dans

le sédiment. Les *Ichthyosarcolites* sont des organismes de grande taille qui peuvent servir ensuite de substrat pour d'autres taxons. *Caprotina* est un taxon encroûtant qui forme une association caractéristique avec *Ichthyosarcolites*. Cette association est connue notamment dans le Bassin Aquitain jusqu'au Mans. Dans ce bassin, cette association est également accompagnée de Radiolitidae. Dans la coupe de la Barre des Aiguilles, les communautés à *Caprotina* et *Ichthyosarcolites* sont pionnières, et enclenchent la dynamique sédimentaire des plates-formes carbonatées. Si le milieu change avec en particulier une augmentation du taux de sédimentation telle que les formes encroûtantes et couchées ne peuvent plus s'établir, les formes élévatrices ou « *clinger* » telles que *Sauvagesia*, *Durania* et *Apricardia* les remplacent et se développent en occupant tout l'espace.

5.3. Succession temporelle des communautés à rudistes

5.3.1. Coupe de la Barre des Aiguilles

Dans S1, les *Ichthyosarcolites* associés aux *Caprotina* sont les taxons pionniers, les premiers couchés sur le sédiment meuble, les seconds encroûtant les coquilles des premiers (S1a, S1c, S1d). Les *Caprotina* apparaissent en minces lits dans cette séquence S1 (un lit de *Caprotina*, probablement une seule génération, s'étendant horizontalement mais pas verticalement), et forment aussi des bouquets ponctuels monospécifiques (S1a, S1c, S1e). Les *Chondrodonta* et les Sauvagesiinae s'installent à la suite de ces organismes pionniers et prennent progressivement de plus en plus d'ampleur (S1a à d). Des coquilles de *Chondrodonta* (S1b) ont probablement été rassemblées là par tempête – événement sans doute facilité par la morphologie de leur coquille. Les Sauvagesiinae en revanche restent en place. Les bouquets de *Caprotina* perdurent encore dans S1 avant de disparaître définitivement. L'accumulation suivante de Nerineidae pourrait être attribuée à l'action d'une tempête. Enfin, les Sauvagesiinae et *Chondrodonta* prennent définitivement le pas sur tous les autres taxons et ces élévateurs et « *clingers* » indiqueraient une accélération du taux de sédimentation. Cette séquence biosédimentaire et écologique S6 de la coupe de la Barre des Aiguilles (fig. 5.4) écologique se termine par émersion due au comblement de la tranche d'eau.

Les séquences S2, S3, S4 (Barre des Aiguilles) correspondent à des préservations de communautés majoritairement constituées d'*Apricardia* accompagnées de Sauvagesiinae, séparées par plusieurs épisodes d'émersion. Dans S4, la richesse en *Apricardia* en peuplements quasi monospécifiques qui envahissent le milieu infralittoral, l'abondance des foraminifères, traduisent une diminution de la tranche d'eau (zéro à quelques mètres; Masse et Fenerci-Masse, 2006).

Dans S5, aux *Apricardia* (S5a) succèdent les Sauvagesiinae (S5b). La richesse spécifique s'effondre alors et les Sauvagesiinae forment une communauté monospécifique. La faible biocorrosion/dégradation des coquilles de Sauvagesiinae sont probablement en relation avec une augmentation du taux de sédimentation.

5.3.2. Coupe de la Bastide d'Orves

L'ensemble des séquences S1 à S6 est là essentiellement composé de communautés à *Apricardia* et Sauvagesiinae, souvent avec *Caprinula* (contrairement aux associations dans la coupe de la Barre des Aiguilles) et entrecoupées d'un enrichissement local (S4) à *Chondrodonta*, probablement apportés par effet de tempête. Ces communautés peuvent se rapprocher de celles du « *Natih c-d type* » d'âge Cénomanien en Oman (Philip 1995) et elles sont indicatrices d'un fort taux de sédimentation. Les successions entre peuplements à dominance d'*Apricardia* et à dominance de Sauvagesiinae ne semblent pas répondre à un patron précis. Certaines séquences montrent cependant un renversement drastique et sans intermédiaire de l'influence du substrat sur la dominance d'un taxon ou de l'autre. Ces deux taxons correspondent probablement à un stade climacique dans un environnement infralittoral ouvert à substrat meuble, stade en équilibre dynamique pouvant conduire au remplacement rapide d'un taxon par l'autre.

Les 6 séquences biosédimentaires reconnues d'approfondissement – comblement – émersion sont là une excellente illustration des relations tectonique (subsidence locale) et production / accumulation biosédimentaire dans des conditions écologiques changeantes.

5.3.3. Les communautés à rudistes Cénomaniennes en méditerranée

Au Cénomanien, des communautés similaires à rudistes et organismes associés se retrouvent principalement sur le pourtour méditerranéen, avec des variations au niveau spécifique ou générique des taxons. Les communautés à rudistes au Cénomanien sont composées de 32 genres (Steuber *et al.* 2016). Les taxons majeurs sont des Caprotinidae (*Caprotina*, *Chaperia*), des Radiolitidae (*Eoradiolites*, *Durania*, *Sauvagesia*, *Distefanella*, *Praeradiolites*, *Radiolites*, *Sphaerulites*, *Neoradiolites*, *Polsakia*, *Paronaites*, *Maghrebites*), des rudistes à canaux (clade 15, voir chapitre 4 : *Neocaprina*, *Caprina*, *Caprinula*, *Sphaerucaprina*, *Orthoptychus*, *Schiosa*, *Oryxia*, *Ichthyosarcolites*) et des Requienidae (*Apricardia*, *Kugleria*, *Rutonia*). La présence de Monopleuridae (clade 2, voir chapitre 4) avec *Petalodontia*, *Araeopleura*, *Simacia*, ainsi que le genre *Cryptaulia* (famille indéterminée) peuvent également être reportés. Les communautés cénomaniennes sont très riches par rapport à celles du Crétacé inférieur en conséquence de radiations évolutives des Radiolitidae et des rudistes à canaux. Ces deux radiations entraînent un pullulement des communautés à rudistes sur les plates-formes carbonatées téthysiennes. De manière non-exhaustive, on peut reporter des communautés similaires en Charentes (Orbigny 1822; Bilotte et Philip 1985; Moreau 1993; Videt 2003) avec *Caprotina*, *Chaperia*, *Sphaerulites*, *Sauvagesia*, *Durania*, *Praeradiolites*, *Biradiolites*, *Ichthyosarcolites*, *Apricardia*, *Petalodontia*; en Istrie (Polšak 1967; Tišljar *et al.* 1998; Pleničar et Jurkovšek 1999; Pleničar *et al.* 1999; Korbar *et al.* 2001) où sont cités *Sauvagesia*, *Eoradiolites*, *Sphaerucaprina*, *Orthoptychus*, *Schiosa*, *Neocaprina*, *Caprina*, *Caprinula*, *Oryxia*, *Ichthyosarcolites*, *Toucasia*; en République Tchèque (El-Shazly *et al.* 2011) avec *Caprotina*, *Radiolites*, *Petalodontia*, *Araeopleura*, *Simacia*, *Caprina*, *Ichthyosarcolites*, *Cryptaulia*; en Jordanie (Özer et Ahmad 2016) avec *Sauvagesia*, *Durania*, *Caprinula*, *Neocaprina*, *Caprina*, *Apricardia*; en Egypte (El-Shazly *et al.* 2011) où sont mentionnés *Praeradiolites*, *Bournonia*, *Eoradiolites*, *Radiolites*, *Biradiolites*, *Agriopleura*, *Caprina*, *Ichthyosarcolites*; en Algérie (Chikhi-Aouimeur 1995, 2003) avec *Eoradiolites*, *Sauvagesia*, *Durania*, *Caprinula*, *Caprina*, *Sphaerucaprina*, *Orthoptychus*, *Neocaprina*, *Apricardia*; en Oman (Philip *et al.* 1995) avec *Sphaerulites*, *Praeradiolites*, *Eoradiolites*, *Sauvagesia*, *Durania*, *Caprina*, *Caprinula*, *Ichthyosarcolites*, *Apricardia*. Si les

Radiolitidae – et les Hippuritidae à partir du Turonien – continuent leur développement tout au long du Crétacé supérieur, la radiation des rudistes à canaux est stoppée à la limite Cénomanien/Turonien à partir de laquelle les rudistes à canaux (clade 20, chapitre 4) ne sont plus présents que ponctuellement sur les plates-formes téthysiennes (on n'y compte plus que les trois genres de Trechmanellidae : *Semailia*, *Dictyoptychus* et *Eodictyoptychus*), la majorité des formes à canaux : les Antillocaprinidae se trouvant dans la province caribéenne (Mitchell 2013b) ou des communautés similaires en termes de types écologiques peuvent s'observer, avec des Radiolitidae (plus Hippuritidae) élévateurs et « *clingers* », ainsi que des rudistes à canaux palléaux (plus Plagioptychidae) « *clingers* » et couchés (Mitchell 2002).

6. Conclusion du chapitre

La préservation et fossilisation des peuplements de rudistes dans les coupes de la Barre des Aiguilles et de la Bastide d'Orves dans des calcaires d'âge Cénomanien supérieur en Basse Provence permet de caractériser des communautés de manière précise, ainsi que d'enregistrer leurs successions. Ces communautés à rudistes vivaient toutes dans le même milieu, infralittoral, sur un substrat meuble. Les peuplements de rudistes correspondent à des communautés restées *in situ*. Les communautés définies sur des paramètres biotiques permettent de préciser leur comportement et la synécologie des populations face à divers facteurs abiotiques. Mises à part deux communautés non faites de rudistes mais de *Chondrodonta* et Nerineidae reconstituées à partir d'accumulations parautochtones à allochtones considérées comme étant des tempestites, les communautés à rudistes apparaissent avoir été préservées en place et non affectées outre mesure par les épisodes de tempêtes. Une des communautés à rudistes comporte le taxon pionnier *Ichthyosarcolites* qui se développe sur un substrat meuble avec un taux de sédimentation faible. Les *Ichthyosarcolites* peuvent être associés à d'autres taxons habitués aux substrats meubles tels que *Chondrodonta* et des Sauvagesiinae. Les *Caprotina* forment des bouquets monospécifiques à développement horizontalement limité et encroûtant des coquilles d'organismes pionniers. Lorsque les peuplements à rudistes prennent de l'ampleur et que le taux de sédimentation augmente, les Sauvagesiinae prennent le pas sur tous les autres

taxons. *Apricardia* et Sauvagesiinae forment une unique communauté représentant un stade climacique en équilibre dynamique avec domination alternative de l'un ou l'autre taxon.

Conclusion générale

A la base de notre rapport à la réalité, il y a la capacité que nous avons à reconnaître dans le chaos de nos perceptions des éléments communs. Nos mécanismes cognitifs sont fondés sur cette reconnaissance d'une relation entre perceptions. Les concepts à la base de la théorie cladistique sont exactement de cette nature. Le principal d'entre eux, le concept d'homologie (Saint-Hilaire 1818, 1830, Owen 1843, 1848), fonde notre relation à l'histoire des entités en évolution. Ces entités acquièrent des modifications au cours du temps, et les transmettent par hérédité à leurs descendants. La thèse défend l'analyse à trois éléments comme une méthode d'analyse cladistique cohérente et pertinente. La 3ia est née au sein d'un monde académique très largement dominé par la méthode de parcimonie, et s'est initialement présentée comme une manière d'augmenter la précision de cette dernière (Nelson et Platnick 1991). La suite de l'histoire de la 3ia s'est faite parallèlement au développement d'une nouvelle manière de voir la reconstruction phylogénétique. Aujourd'hui, avec ses développements les plus récents (Cao *et al.* 2007; Cao 2008; Lelièvre *et al.* 2008; Zaragüeta i Bagils 2011), la 3ia forme une méthode autonome et un ensemble cohérent plour laquelle la référence aux autres méthodes est inutile quand elle ne la dessert pas. Le premier chapitre a pour but d'expliciter la cohérence générale de la 3ia. De la manière dont elle se construit, il nécessite un renversement complet dans la manière d'appréhender certains des concepts élémentaires de la discipline phylogénétique. A cet effet, nous proposons un glossaire avec une redéfinition claire des termes usuels. A partir de cette partie, nous apportons des exemples de construction de caractère qui posent d'importants problèmes dans les autres méthodes cladistiques (Pleijel 1995). Nous proposons également à partir du concept de triplet une métrique permettant de décrire le contenu informatif des nœuds dans un cladogramme. Une partie permet également d'appréhender les problèmes algorithmiques qui se posent avec l'implémentation logicielle de la 3ia. Nous explorons les pistes qui s'offriront dans le futur afin d'optimiser les temps de calculs lors d'analyses de 3ia. Une meilleure connaissance des relations de combinaison

entre triplets a été entreprise. Les relations de combinaison entre triplets nous permettent également de jauger les différentes méthodes de consensus : nous proposons que le consensus strict ainsi que le RCC soient les seules structures à pouvoir être interprétées – dans le cadre d'une analyse donnant plusieurs arbres optimaux – comme l'unique cladogramme vrai et non-ambigu d'une analyse phylogénétique.

Dans le second chapitre, nous revenons sur les arguments théoriques et méthodologiques de la littérature qui ont émaillé les débats entre 3ia et parcimonie. Quand la 3ia permet la construction de caractères hiérarchiques, faisant écho à la nature relationnelle de l'homologie (Nelson 1994; Prin 2016), la parcimonie utilise la partition – dont les conséquences sont le pas de transformation et la matrice –, une structure intrinsèquement paraphylétique et donc, par définition, incompatible avec la théorie cladistique (dans son courant « taxique »; Williams et Ebach 2006; Zaragüeta i Bagils 2011; Prin 2012). Les deux études basées sur des simulations informatiques sont là pour mettre en perspective ces discussions méthodologiques et théoriques. Si la 3ia ainsi que la théorie cladistique refondée à partir des travaux de Nelson et Platnick (Nelson 1979, 1994; Platnick 1979; Brady 1982; Ebach *et al.* 2007, Zaragüeta i Bagils 2011, Prin 2012) semblent être arrivées à une maturité qui devrait permettre d'accéder à une révolution conceptuelle de la systématique phylogénétique et de la perception de l'évolution, semblent d'après le nombre de travaux s'en réclamant être un échec total (Nelson 2004). Cette contribution que nous proposons à la théorie cladistique et à la 3ia ne peut se perpétuer que par la mise en relief de sa cohérence, en particulier en paléontologie, vis-à-vis de ce que l'on peut d'ores et déjà appeler la révolution probabiliste (Maximum de Vraisemblance et Inférence Bayésienne). La révolution probabiliste n'apporte qu'un raffinement de la parcimonie et soufre des mêmes défauts théoriques et méthodologiques – la partition et le pas de transformation. Ce n'est que le succès ou l'échec de cette entreprise qui permettra ou non dans le futur à la 3ia de continuer son développement, par l'application concrète à des cas empiriques. C'est ce mouvement qui doit se prolonger et se propager après ce travail cladistique sur l'évolution des rudistes.

Dans la deuxième partie de cette thèse, nous avons proposé une mise en application de la cladistique et de l'analyse à trois éléments afin de proposer un nouveau regard sur l'histoire

phylogénétique des rudistes au cours du Crétacé et leur relation avec leur biotope. Dans un premier temps, nous avons entrepris la révision du genre *Ichthyosarcolites* à partir de la littérature, de matériel de collections ainsi que de fossiles collectés dans le Cénomanien Charentais (France) qui constitue le chapitre 3. A partir d'une réflexion sur l'homologie des carènes et les contraintes structurales sur la cavité générale de la coquille, seules quatre espèces sur les 19 que compte la littérature doivent être retenues. La révision du genre nous permet de mettre en évidence un nouveau genre et une nouvelle espèce basés sur un assemblage jusqu'alors inconnu de caractères. Cette nouvelle espèce, *Oryxia sulcata*, ainsi que le genre *Ichthyosarcolites*, sont des taxons clefs dont la connaissance nous permet une meilleure appréhension de la phylogénie. *O. sulcata* n'est répertoriée que dans le Cénomanien de Slovénie et probablement d'Istrie (Croatie). De nouvelles fouilles de ces localités pourraient délivrer de nouvelles informations sur ce genre dont il manque à notre connaissance la totalité de la valve gauche.

Le chapitre 4 traite de la reconstruction de la phylogénie des rudistes s'attachant sur leur valve droite au substrat – sous-ordre Hippuritidina (Skelton 2013a, 2013b) – en 3ia. Le travail d'anatomie comparée de 27 espèces de rudistes dans un cadre cladistique nous a permis de formaliser de nouvelles hypothèses d'homologies ainsi que de nouveaux caractères débarrassés des problèmes de formalisation et de dépendance. La mise en congruence de ces 40 caractères donne deux cladogrammes. La phylogénie obtenue permet de tester les états de caractères construits et de mettre en évidence que les myophores comme les canaux palléaux sont source d'homoplasie, à différents niveaux hiérarchiques cependant. La phylogénie que nous obtenons ainsi que les caractères qui ont servi à la construire sont discutés par rapport aux précédentes analyses (Chartrousse 1994; Skelton et Masse 1998; Steuber 1999; Stone et Telford 1999; Skelton et Smith 2000; Sano *et al.* 2012; Mitchell 2013b). L'IRN nous permet également de décrire avec précision le contenu informatif de la phylogénie obtenue. Nous discutons également des relations des 154 genres du sous ordre Hippuritidina (Steuber *et al.* 2016) avec les clades reconnus par l'analyse. La phylogénie est datée et permet d'établir une chronologie de l'évolution des rudistes. Cette chronologie nous permet de mettre en évidence deux radiations évolutives subséquentes aux deux grands évènements anoxiques du Crétacé. Nous discutons également de l'impact des fluctuations du niveau marin et de l'étendue des plates-formes carbonatées de l'époque.

Afin de préciser ce scénario évolutif, il sera nécessaire de construire de nouveaux cladogrammes, en particulier au sein des Radiolitidae et des Hippuritidae qui comptent pour une grande partie de la diversité taxonomique des rudistes. Une analyse de ces familles permettra de tester plus en profondeur l'hypothèse que nous proposons ici sur l'origine des Hippuritidae qui semblent s'être différenciés au sein des Radiolitidae par la forte torsion de la coquille au niveau des bandes siphonales, l'acquisition du système pores/canaux situé sur la valve gauche ou encore la densification de leur couche calcitique externe. La confirmation ou le rejet de cette hypothèse aura des effets majeurs sur les études sur la diversification des rudistes, et nous permettra de comprendre si les Radiolitidae et les Hippuritidae ont connu deux épisodes indépendants de diversification ou bien si, comme nous le soutenons, il s'agit d'une explosion radiative unique dont les Hippuritidae suivent la dynamique.

Le dernier chapitre reprend ce cadre phylogénétique et taxonomique afin de discuter d'évènements à échelle locale. L'étude de la Bastide d'Orves et de la Barre des Aiguilles (Var, France) nous a permis de mettre en évidence de manière quantitative la succession des communautés à rudistes au cours du Cénomanien supérieur à terminal dans le Bassin Sud-Provençal. Nous avons pu montrer d'une part une communauté à *Ichthyosarcolites* pionniers et à *Caprotina* encroûtant sur les précédents et formant des lits ponctuels envahissant tout l'espace. Ces populations vivent avec des Sauvagesiinae et des *Chondrodonta*. Un deuxième type de communauté est dominé alternativement par les Sauvagesiinae et les *Apricardia*, formant des peuplements paucispécifiques, et dont la localisation est plus proximale par rapport au littoral que les communautés à *Ichthyosarcolites*, *Caprotina* et *Chondrodonta*. Ces communautés, à leur stade climacique, participent à l'accélération de la dynamique sédimentaire sur la plate-forme carbonatée sur laquelle ils vivent.

Bibliographie

Ackouala, P. 2012. Diagenèse Des Calcaires À Rudistes de La Bastide d'Orves. Mémoire de master 1 (Spécialité : Géosystèmes), sous la direction de Floquet, M., Marseille, Université Aix-Marseille, 34 p.

Aguilar-Pérez, J. 2008. Rudistas Del Cretacio Inferior Y Medio, Barremiano-Cenomaniano, Noreste, Centro Y Oeste de México. Thèse de doctorat, Universitat Autònoma de Barcelona, 139 p.

Adams, E. N. 1972. Consensus Techniques and the Comparison of Taxonomic Trees. Systematic Biology, **21**, 390–397.

Adams, E. N. 1986. N-trees as nestings: Complexity, similarity, and consensus. Journal of Classification, **3**, 299–317.

Aho, A. V., Sagiv, Y., Szymanski, T. G. & Ullman, J. D. 1981. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions. SIAM Journal on Computing, **10**, 405–421.

Almeida, M. T. & Bisby, F. A. 1984. Simple method for establishing taxonomic characters from measurement data. Taxon, **33**, 405–409.

Alroy, J. 2001. A Multispecies Overkill Simulation of the End-Pleistocene Megafaunal Mass Extinction. Science, **292**, 1893–1896.

Archie, J. W. 1985. Methods for Coding Variable Morphological Features for Numerical Taxonomic Analysis. Systematic Biology, **34**, 326–345.

Archie, J. W. 1989. Homoplasy Excess Ratios: New Indices for Measuring Levels of Homoplasy in Phylogenetic Systematics and a Critique of the Consistency Index. Systematic Biology, **38**, 253–269.

Ayoub-Hannaa, W. & Fürsich, F. T. 2011. Functional morphology and taphonomy of Cenomanian (Cretaceous) oysters from the eastern Sinai Peninsula, Egypt. Palaeobiodiversity and Palaeoenvironments, **91**, 197–214.

Baer, K. E. von. 1828. Uber Entwicklungsgeschichte Der Thiere. Beobachtung Und Reflexion. Königsberg, Bornträger, 306 p.

Barbut, M. 1968. Partitions d'un ensemble fini: leur treillis (cosimplexe) et leur représentation géométrique. Mathématiques et sciences humaines, **22**, 5–22.

Bardin, J., Rouget, I., Yacobucci, M. & Cecca, F. 2013. Increasing the number of discrete character states for continuous characters generates well-resolved trees that do not reflect phylogeny. Integrative Zoology, **9**, 531–541.

Barthélémy, J.-P. & Guénoche, A. 1991. Trees and Proximity Representations. Chichester, John Wiley & Sons.

Baum, B. 1992. Combining Trees as a Way of Combining Data Sets for Phylogenetic Inference, and the Desirability of Combining Gene Trees. Taxon, **41**, 3–10.

Bayle, E. 1855. Observations sur la structure des coquilles des Hippurites, suivies de quelques remarques sur les *Radiolites*. Bulletin de la Société Géologique de France, **2**, 772–807.

Belon, P. 1555. L'histoire de La Nature Des Oyseaux, Avec Leurs Descriptions et Naïfs Portraicts Retirez Du Naturel. Guillaume Cauellat, Paris, 381 p.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the False Discovery Rate: A practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, **57**, 289–300.

Bernard-Dumanois, A. & Delance, J.-H. 1983. Microperforations par algues et champignons sur les coquilles des 'Marnes a *Ostrea Acuminata*' (Bajocien supérieur) de Bourgogne (France), relations avec le milieu et utilisation paleobathymétrique. Geobios, **16**, 419–429.

Bernier, P. & Strasser, A. 1988. Formation de galets noirs et coquilles noircies dans l'actuel (ile de Noirmoutier, France). Bulletin de la Société Géologique de France, **4**, 341–347.

Berry, V. & Gascuel, O. 2000. Inferring evolutionary trees with strong combinatorial evidence. Theoretical Computer Science, **240**, 271–298.

Berthou, P.-Y. 1973. Le Cénomanien de L'estrémadure Portugaise. Lisboa, Serviços geologicos de portugal, **23**, 169 p.

Bilotte, M. & Philip, J. 1985. Les faunes de rudistes du Crétacé supérieur Charentais du chantier de l'autoroute 'l'Aquitaine'. Cretaceous Research, **6**, 79–84.

Blénet, A. 2010. Relations Entre Les Écosystèmes À Rudistes et Le Système Sédimentaire Carbonaté de La Plate-Forme Provençale Au Cénomanien Supérieur. Mémoire de master 2 (Spécialité : Géosciences de l'environnement), sous la direction de Floquet, M. & Villier, L., Marseille, Université Paul Cézanne – Université de Provence, 38 p.

Bobkova, N. N. 1968. New Late Cretaceous rudists from central Asia. Trudy vsesojuznogo nauchno-issledovatel'skogo geologicheskogo Instituta (VSEGEI), **2**, 285–290.

Bouchet, P. & Rocroi, J.-P. 2010. Nomenclator of bivalve families; with a classification of bivalve families by Bieler, R., Carter, J. G. & Coan, E. V. Malacologia, **52**, 1–184.

Bourdon, E. 2006. L'avifaune Du Paléogène Des Phosphates Du Maroc et Du Togo : Diversité, Systématique et Apports À La Connaissance de La Diversification Des Oiseaux Modernes (Neornithes). Paris, Muséum national d'histoire naturelle. Thèse de doctorat, Muséum national d'Histoire Naturelle, 330 p.

Brachert, T. C. & Dullo, W. C. 2000. Shallow burial diagenesis of skeletal carbonates: selective loss of aragonite shell material (Miocene to Recent, Queensland Plateau and Queensland Trough, NE Australia). Implications for shallow cool-water carbonates. Sedimentary Geology, **136**, 169–187.

Brady, R. H. 1982. Theoretical issues and 'pattern cladistics'. Systematic Zoology, **31**, 286–291.

Brazeau, M. D. 2011. Problematic character coding methods in morphology and their effects. Biological Journal of the Linnean Society, **104**, 489–498.

Bremer, K. 1990. Combinable Component Consensus. Cladistics, **6**, 369–372.

Bronn, H. G. 1858. Untersuchungen Über Die Entwickelungs-Gesetze Der Organischen Welt Während Der Bildungs-Zeit Unserer Erd-Oberfläche. Stuttgart, E. Schweizerbart.

Brooks, D. R. 1990. Parsimony Analysis in Historical Biogeography and Coevolution: Methodological and Theoretical Update. Systematic Zoology, **39**, 14–30.

Brooks, D. R., Van Veller, M. G. & McLennan, D. A. 2001. How to do BPA, really. Journal of Biogeography, **28**, 345–358.

Brouwer, J. & Fitzpatrick, R. 2002. Relations between soil macro-morphology and current soil hydrology in a toposequence in SE Australia – Relations entre macromorphologie des sols et hydrologie actuelle dans une toposéquence en Australie du SE. Soil Research, **40**, 903–906.

Brower, A. V. Z. 2014. Transformational and taxic homology revisited. Cladistics, **31**, 197–201.

Brower, A. V. Z. & de Pinna, M. C. C. 2012. Homology and errors. Cladistics, **28**, 529–538.

Brower, A. V. Z. & de Pinna, M. C. C. 2014. About nothing. Cladistics, **30**, 330–336.

Broyles, S. S. & Moss, B. 1986. Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22-kDa subunits. Proceedings of the National Academy of Sciences, 83, 3141–3145.

Bryant, D. 2003. A classification of consensus methods for phylogenetics. DIMACS series in discrete mathematics and theoretical computer science, **61**, 163–184.

Buneman, P. 1971. The recovery of trees from measures of dissimilarity. In: Hodson, F. R., Kendall, D. G., & Tăutu, P. (eds) Mathematics in the Archaeological and Historical Sciences. Edinburgh, Edinburgh University Press, 387–395.

Caldwell, W. G. E. & Evans, J. K. 1963. A cretaceous rudist from Canada and a redescription of the holotype of *Ichthyosarcolites coraloidea* (Hall & Meek). Journal of Paleontology, **36**, 615–620.

Callaway, E. 2011. Ancient DNA reveals secrets of human history. Nature News, **476**, 136–137.

Camin, J. H. & Sokal, R. R. 1965. A method for deducing branching sequences in phylogeny. Evolution, **19**, 311–326.

Canoville, A. & Laurin, M. 2010. Evolution of humeral microanatomy and lifestyle in amniotes, and some comments on palaeobiological inferences. Biological Journal of the Linnean Society, **100**, 384–406.

Cao, N. 2008. Analyse à trois éléments et anatomie du bois des Fagales Engl. Thèse de doctorat, Muséum national d'histoire naturelle, 239 p.

Cao, N., Zaragüeta i Bagils, R. & Vignes-Lebbe, R. 2007. Hierarchical representation of hypotheses of homology. Geodiversitas, **29**, 5–15.

Cao, N., Bourdon, E., El Azawi, M. & Zaragüeta i Bagils, R. 2009. Three-item analysis and parsimony, intersection tree and strict consensus: a biogeographical example. Bulletins de la société géologique de France, **180**, 13–15.

Carannante, G., Ruberti, D. & Simone, L. 1993. Rudists and related sediments in Late Cretaceous open shelf settings. A case history from Matese area (central Apennines, Italy). Giornale di Geologia, **55**, 21–36.

Carannante, G., Graziano, R., Ruberti, D. & Simone, L. 1997. Upper Cretaceous temperate-type open shelves from northern (Sardinia) and southern (Apennines-Apulia) Mesozoic Tethyan margins. In: James, N. P. & Clarke, J. A. D. (eds) Cool-Water Carbonates. SEPM Special Publication, **56**, 309–325.

Carannante, G., Graziano, R., Pappone, G., Ruberti, D. & Simone, L. 1999. Depositional System and Response to Sea Level Oscillations of the Senonian Rudist-bearing Carbonate Shelves. Examples from Central Mediterranean Areas. Facies, **40**, 1–24.

Carannante, G., Ruberti, D. & Sirna, M. 2000. Upper Cretaceous ramp limestones from the Sorrento Peninsula (southern Apennines, Italy): micro-and macrofossil associations and their significance in the depositional sequences. Sedimentary geology, **132**, 89–123.

Carbone, F., Praturlon, A. & Sirna, G. 1971. The Cenomanian shelf edge facies of Rocca di Cave. Geologica romana, **10**, 131–197.

Carine, M. A. & Scotland, R. W. 1999. Taxic and transformational homology: different ways of seeing. Cladistics, **15**, 121–129.

Carter, J. G., Altaba, C. R., Anderson, L. R., Araujo, R., Biakov, A. S., Bogan, A. E., Campbell, D. C., Campbell, M., Chen, J., Cope, J. C. W., Delvene, G., Dijkstra, H. H., Fang, Z., Gardner, R. N., Gavrilova, V. A., Goncharova, I. A., Harries, P. J., Hartman, J. H., Hautmann, M., Hoeh, W. R., Hylleberg, J., Jiang, B., Johnston, P., Kirkendale, L., Kleemann, K., Koppka, J., Kříž, J., Machado, D., Malchus, N., Márquez-Aliaga, A., Masse, J.-P., McRoberts, C. A., Middelfart, P. U., Mitchell, S., Nevesskaja, L. A., Özer, S., Jr, P., J, Inga V. Polubotko, J. M. P., Yancey1, T., Polubotko, I. V., Pons, J. M., Popov, S., Sánchez, T., Sartori, A. F., Scott, R. W., Sey, I. I., Signorelli, J. H., Silantiev, V. V., Skelton, P. W., Steuber, T., Waterhouse, J. B., Wingard, G. L. & Yancey, T. 2011. A synoptical classification of the Bivalvia (Mollusca). Paleontological contributions, **4**, 1–47.

Cavalli-Sforza, L. L. & Edwards, A. W. F. 1967. Phylogenetic analysis. Models and estimation procedures. American Journal of Human Genetics, **19**, 233–257.

Cestari, R., Pons, J. M. & Sirna, G. 1998. Undescribed *Ichthyosarcolites* from Sicily, belonging to Gemmellaro's collection. Geobios, **31**, 69–73.

Chappill, J. A. 1989. Quantitative Characters in Phylogenetic Analysis. Cladistics, **5**, 217–234.

Chartrousse, A. 1994. Les Caprinidae (Rudistes) Du Crétacé Inférieur. Thèse de doctorat, Université de Provence, 281 p.

Chartrousse, A. 1998. The myocardinal organization of coalcomaninid rudists revisited. Geobios, **31**, 75–85.

Chikhi-Aouimeur, F. 1995. *Caprinula* aff. *cedrorum* (Blanckenhorn, 1890) from the upper cenomanian of western Algeria. Revista Mexicana de Ciencias Geológicas, **12**, 179–184.

Chikhi-Aouimeur, F. 2003. Stratigraphic and Geographic Distribution of Rudists in Algeria: A State of the Art. In: Gili, E., El Hédi Negra, M. & Skelton, P. W. (eds) North African Cretaceous Carbonate Platform Systems, **28**, 161–170.

Chikhi-Aouimeur, F. 2010. L'Algérie À Travers Son Patrimoine Paléontologique - Les Rudistes. Sarl Baosem, 270 p.

Colless, D. H. 1985. On 'Character' and Related Terms. Systematic Zoology, **34**, 229–233.

Colonius, H. & Schulze, H. H. 1981. Tree structures for proximity data. British Journal of Mathematical and Statistical Psychology, **34**, 167–180.

Corvez, A. 2012. L'origine de La Mégaphylle Chez Les Monilophytes. Thèse de doctorat, Muséum national d'Histoire naturelle, 276 p.

Cox, L. R. 1933. The evolutionary history of the Rudists. Proceedings of the Geologists' Association, **44**, 379–388.

Critchlow, D. E., Pearl, D. K. & Qian, C. 1996. The triples distance for rooted bifurcating phylogenetic trees. Systematic Biology, **45**, 323–334.

Croizat, L. 1958. Panbiogeography or An Introductory Synthesis of Zoogeography, Phytogeography, and Geology; with Notes on Evolution, Systematics, Ecology, Anthropology, Etc. Caracas, publié à compte d'auteur, volume I: 1018 p., volume IIa: 771 p., volume IIb 1731 p.

Damestoy, G. 1970. Essai de classification phylogénétique des caprinidés (Lamellibranches). Bulletin du Museum d'Histoire Naturelle, **42**, 1003–1008.

Darlu, P. & Tassy, P. 1993. La Reconstruction Phylogénétique. Concepts et Méthodes. Paris, Masson, 244 p.

Day, W. H. E. & Sankoff, D. 1986. Computational complexity of inferring phylogenies by compatibility. Systematic Biology. **35**, 224–229.

Dechaseaux, C. 1969. Superfamily Hippuritacea Gray, 1848. In: Moore, R. C. (ed) Treatise on Invertebrate Paleontology. Part N. Mollusca 6. Bivalvia. The Geological Society of America, Inc. and The University of Kansas, 749–817.

Decraene, L. R. & Smets, E. F. 1994. Merosity in flowers: definition, origin, and taxonomic significance. Plant Systematics and Evolution, **191**, 83–104.

Deshayes, G. P. 1825. Dictionnaire classique d'Histoire naturelle. Ichthyosarcolite. **8**, 499–501.

Desmarest, A. G. 1812. Mémoire sur deux genres de coquilles fossiles cloisonnées et à siphon. Bulletin de la Société des Sciences physiques, de médecine et d'Agriculture d'Orléans, **5**, 308–324.

Desmarest, M. A. G. 1817. Mémoire sur deux genres de coquilles fossiles cloisonnées et à siphon. Journal de Physique, de Chimie et d'Histoire naturelle, 42–51.

DesMoulins, C. 1826. Essai sur les Sphérulites qui existent dans les collections de MM. F. Jouannet. Bulletin d'Histoire naturelle de la Société Linnéenne de Bordeaux, **1**, 1–156.

Deshayes, G. P. 1825. Quelques observations sur les genres *Hippurite* et *Radiolite*. Annales des sciences naturelles, **5**, 205–211.

Dhondt, A. V. & Dieni, I. 1992. Non rudistid bivalves from Cretaceous rudist formations. Geologica Romana, **28**, 211–218.

Dhondt, A. V. & Dieni, I. 1993. Non-rudistid bivalves from Late Cretaceous rudist limestones of NE Italy (Col dei Schiosi and Lago di S. Croce areas). Memorie de Sicenze Geologiche, **45**, 165–241.

Dias-Brito, D. 2000. Global stratigraphy, palaeobiogeography and palaeoecology of Albian–Maastrichtian pithonellid calcispheres: impact on Tethys configuration. Cretaceous Research, **21**, 315–349.

Douvillé, H. 1886. Essai sur la morphologie des rudistes. Bulletin de la Société Géologique de France, **3**, 389–404.

Douvillé, H. 1887. Sur quelques formes peu connues de la famille des chamidés. Bulletin de la Société Géologique de France, **3**, 756–802.

Douvillé, H. 1888. Etudes sur les Caprines. Bulletin de la Société Géologique de France, 16, 699–730.

Douvillé, H. 1889. Sur quelques rudistes du terrain crétacé inférieur des Pyrénées. Bulletin de la Société géologique de France, **3**, 627–635.

Douvillé, H. 1896. Observations sur la charnière des lamellibranches hétérodontes. Bulletin de la Société géologique de France, **24**, 26–28.

Douvillé, H. 1935. Les rudistes et leur évolution. Bulletins de la Société Géologique de France, **5**, 319–358.

Ducrotay de Blainville, H. M. 1827. Manuel de Malacologie et de Conchyologie. Levrault, Paris, 647 p.

Dunham, R. J. 1962. Classification of carbonate rocks according to depositional textures. In: Ham, W. E. (ed.) Classification of Carbonate Rocks. Memoirs of the American Association of Petroleum Geologists, **1**, 108–121.

Ebach, M. C., Morrone, J. J. & Williams, D. M. 2007. A new cladistics of cladists. Biology & Philosophy, **23**, 153–156.

El-Shazly, S., Košťák, M., Kloučková, B., Saber, S. G., Felieh Salama, Y., Mazuch, M. & Žák, K. 2011. Carbon and oxygen stable isotopes of selected Cenomanian and Turonian rudists from Egypt and Czech Republic, and a note on changes in rudist diversity. Bulletin of Geosciences, **86**, 209–226.

Embry, A. F. I. & Klovan, J. E. 1971. A late Devonian reef tract on northeaster banks island. Bulletin of Canadian Petroleum Geology, **19**, 730–781.

Estabrook, G. F., Johnson, J. C. S. & McMorris, F. R. 1976a. A mathematical foundation for the analysis of cladistic character compatibility. Mathematical Biosciences, **29**, 181–187.

Estabrook, G. F., Johnson, J. C. S. & McMorris, F. R. 1976b. An algebraic analysis of cladistic characters. Discrete mathematics, **16**, 141–147.

Estabrook, G. F., McMorris, F. R. & Meacham, C. A. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. Systematic Zoology, **34**, 193–200.

Faith, D. P. 1992. Systematics and Conservation: On Predicting the Feature Diversity of Subsets of Taxa. Cladistics, **8**, 361–373.

Farris, J. S. 1970. Methods for Computing Wagner Trees. Systematic Zoology, **19**, 83–92.

Farris, J. S. 1977. Phylogenetic Analysis Under Dollo's Law. Systematic Zoology, **26**, 77–88.

Farris, J. S. 1982. Outgroups and parsimony. Systematic Zoology, **31**, 328–334.

Farris, J. S. 1983. The logical basis of phylogenetic analysis. Advances in cladistics, **2**, 7–36.

Farris, J. S. 1989. The Retention Index and the Rescaled Consistency Index. Cladistics, **5**, 417–419.

Farris, J. S. 1990. Phenetics in Camouflage. Cladistics, **6**, 91–100.

Farris, J. S. 1997. Who really is a statistician? Paper presented at the Sixteenth Meeting of the Willi Hennig Society. George Washington University, Washington, DC.

Farris, J. S. 2010. Systematic foundering. Cladistics, **26**, 1–15.

Farris, J. S. 2012. 3ta Sleeps with the fishes: Book review. Cladistics, **28**, 422–436.

Farris, J. S. 2014. 'Taxic homology' is neither. Cladistics, **30**, 113–115.

Farris, J. S. & Kluge, A. G. 1998. A/The Brief History of Three-Taxon Analysis. Cladistics, **14**, 349–362.

Farris, J. S., Kluge, A. G. & Eckardt, M. J. 1970. A Numerical Approach to Phylogenetic Systematics. Systematic Biology, **19**, 172–189.

Farris, J. S., Källersjö, M., Albert, V. A., Allard, M., Anderberg, A., Bowditch, B., Bult, C., Carpenter, J. M., Crowe, T. M. & Laet, J. 1995. Explanation. Cladistics, **11**, 211–218.

Farris, J. S., Kluge, A. G. & De Laet, J. 2001. Taxic Revisions. Cladistics, **17**, 79–103.

Faure-Brac, M., Rineau, V. & Zaragüeta i Bagils, R. Issues of Representation: A Hierarchical Point of View. In prep.

Felsenstein, J. 1978. The Number of Evolutionary Trees. Systematic Zoology, **27**, 27–33.

Felsenstein, J. 1985. Phylogenies and the comparative method. The American naturalist, **125**, 1–15.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annual review of genetics**, 22**, 521–565.

Felsenstein, J. 2004. Inferring Phylogenies. Sunderland, Massachusetts, Sinauer Associates, 580 p.

Fenerci-Masse, M. 2006. Les Communautés À Rudistes Du Crétacé Inférieur de La Marge Ouest Européenne de La Téthys. Thèse de doctorat, Université de Provence, 436 p.

Fenerci-Masse, M., Masse, J.-P. & Chazottes, V. 2004. Quantitative analysis of rudist assemblages: a key for palaeocommunity reconstructions. The Late Barremian record from SE France. Palaeogeography, Palaeoclimatology, Palaeoecology, **206**, 133–147.

Fenerci-Masse, M., Masse, J. P. & Pernarcic, E. 2005. Quantitative stratigraphy of rudist limestones and its bearing on spatial organisation of rudist communities: the Late Barremian, Urgonian, sequences of Provence (S.E. France). Palaeogeography, Palaeoclimatology, Palaeoecology, **215**, 265–284.

Fenerci-Masse, M., Skelton, P. & Masse, J.-P. 2011. The rudist bivalve genus *Gorjanovicia* (Radiolitidae, Hippuritoidea) a revision of species based on quantitative analysis of morphological characters. *Palaeontology*, **54**, 1–23.

Filkorn, H. F. 2002. A new species of *Mexicaprina* (caprinidae, coalcomaninae) and review of the age and paleobiogeography of the genus. Journal of Paleontology, **76**, 672–691.

Fischer, P. 1887. Manuel de Conchyliologie et de Paléontologie Conchyliologique Ou Histoire Naturelle Des Mollusques Vivants et Fossiles. Paris, Savy, 1369 p.

Fitch, W. M. 1970. Distinguishing homologous from analogous proteins. Systematic Zoology, **19**, 99–113.

Fitch, W. M. 1971. Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. Systematic Biology, **20**, 406–416.

Fitch, W. M. 2000. Homology: a personal view on some of the problems. Trends in genetics, **16**, 227–231.

Floquet, M. 1982. Un modèle de Comblement de Plate-forme Interne Carbonatée: Santonien Supérieur-Campanien des Chaines Ibérique Nord. Cretaceous research, **3**, 69–81.

Floquet, M. & Hennuy, J. 2001. Anatomy of a carbonate resedimentation within the latest Turonian–earliest Coniacian South Provençal Basin. Géologie Méditerranéenne, **28**, 67–71.

Floquet, M. & Hennuy, J. 2003. Evolutionary gravity flow deposits in the middle Turonian-Early Coniacian southern Provence basin (SE France). In: Locat, J. & Mienert, J. (eds) Origins and depositional processes. Submarine Mass Movements and Their Consequences. Dordrecht, Boston, London, Kluwer Academic Books, 417–424.

Floquet, M., Gari, J., Hennuy, J., Léonide, P. & Philip, J. 2005. Sédimentations gravitaires carbonatées et silicoclastiques dans un bassin en transtension, séries d'âge Cénomanien à Coniacien moyen du Bassin Sud-Provencal. In: Floquet, M., Gari, J., Hennuy, J., Léonide, P. & Philip, J. (eds) Livret - Guide d'Excursion Géologique – 10ème Congrès Français de Sédimentologie. Paris, Publication de l'Association des Sédimentologistes Français, **52**, 80 p.

Floquet, M., Philip, J., Léonide, P. & Gari, J. 2006. Sédimentation et Géodynamique du Bassin Sud-Provençal au Crétacé Supérieur. In: Floquet, Borgomano, J. & Cecca, F. (eds) Journées Jean Philip – Histoire et dynamique des plate-formes carbonatées et de leurs biotas durant le Phanérozoïque. Marseille, Université de Provence – Aix-Marseille I, 69 p.

Floquet, M., Léonide, P., Villier, L., Ackouala, A.-P. & Blenet, A. 2013. Expressions biosédimentaire, diagénétique, géochimique et tectonique de l'événement anoxique cénomano-turonien (OAE 2) par les plate-formes carbonatées Sud-Provençales. Présentation au 14ème Congrès de l'Association des Sédimentologistes, Paris.

Flügel, E. 1982. Evolution of Triassic reefs: current concepts and problems. Facies, **6**, 297–327.

Flügel, E. 2004. Microfacies Analysis of Carbonate Rocks. Analysis, Interpretation and Application. 2nd Edition. Springer, Berlin, 984 p.

Forey, P. L. & Kitching, I. J. 2000. Experiments in coding multistate characters. In: Scotland, R. & Pennington, R. T. (eds) Homology and Systematics. Coding Characters for Phylogenetic Analysis. London and New York, Systematics Association Special Volume, 54–80.

Frege, G. 1971. Écrits Logiques et Philosophiques. Éditions du Seuil, Paris, 234 p.

Garland, T. J., Dickerman, A. W., Janis, C. M. & Jones, J. A. 1993. Phylogenetic Analysis of Covariance by Computer Simulation. Systematic Biology, **42**, 265–292.

Gemmellaro, G. G. 1865. Caprinellidi della zona superiore della Ciaga dei dintorni di Palermo. Atti dell'Accademia Gioenia di Scienze naturali di Catania. Serie seconda, **20**, 198–238.

Gili, E. & Skelton, P. W. 2000. Factors regulating the development of elevator rudist congregations. Geological Society, London, Special Publications, **178**, 109–116.

Gili, E., Masse, J.-P. & Skelton, P. W. 1995. Rudists as gregarious sediment-dwellers, not reef-builders, on Cretaceous carbonate platforms. Palaeogeography, Palaeoclimatology, Palaeoecology, **118**, 245–267.

Goldberg, E. E. & Igić, B. 2008. On Phylogenetic Test of Irreversible Evolution. Evolution, **62**, 2727–2741.

Götz, S. 2003. Larval settlement and ontogenetic development of *Hippuritella vasseuri* (Douvillé) (Hippuritoidea, Bivalvia). 56, 123–131.

Gradstein, F. M., Ogg, J. G., Schmitz, M. & Ogg, G. 2012. The Geologic Time Scale. 1st edition. Amsterdam, Elsevier, 1176 p.

Grand, A., Corvez, A., Duque Velez, L. M. & Laurin, M. 2013. Phylogenetic inference using discrete characters: performance of ordered and unordered parsimony and of three-item statements. Biological Journal of the Linnean Society, **110**, 914–930.

Gray, J. E. 1854. A revision of the arrangement of the families of bivalve shells (Conchifera). Annals and Magazine of Natural History, series 2, **13**, 408–418.

Gray, G. S. & Fitch, W. M. 1983. Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resistance gene from *Staphylococcus aureus*. Molecular Biology and Evolution, **1**, 57–66.

Guettard, J-E. 1786. Quatorzième mémoire, sur les Pierres figurées, pour servir à l'histoire des Préjugés en Minéralogie, et à l'intelligence de plusieurs endroits de l'Histoire Naturelle de Pline. In: Guettard, J-E. (ed) Nouvelle Collection de Mémoires Sur Différentes Parties Intéressantes Des Sciences et Arts. Tome Second, Faisant Le Ve. de La Collection. Paris, Lamy, 503–614.

Hairabian, A. 2007. Reconstitution Paléoenvironnementale de La Bordure Sud Du Bassin Sud-Provencal (Secteur Du Val d'Aren, Var) Au Cénomanien Supérieur. Mémoire de master 2 (Spécialité : Géosciences de l'environnement), sous la direction de Floquet, M., Villier, L. & Léonide. P, Marseille, Université Paul Cézanne – Université de Provence, 31 p.

Hall, J. & Meek, F. B. 1855. Descriptions of New Species of Fossils, from the Cretaceous Formations of Nebraska, with Observations upon *Baculites ovatus* and *B. compressus*, and the Progressive Development of the Septa in *Baculites*, *Ammonites*, and *Scaphites*. Memoirs of the American Academy of Arts and Sciences, **5**, 379–411.

Hammer, Ø., Harper, D. A. T. & Ryan, P. D. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica, **4**, 9 p.

Hammer, Ø., Harper, D. A. T. 2008. Paleontological Data Analysis. Wiley Blackwell, 368 p.

Harvey, A. W. 1992. Three-taxon statements: more precisely, an abuse of parsimony? Cladistics, **8**, 345–354.

Hauser, D. L. & Presch, W. 1991. The effect of ordered characters on phylogenetic reconstruction. Cladistics, **7**, 243–265.

Hennig, E. 1916. Die Fauna der deutsch-ostafrikanischen Urgonfazies. Zeitschrift der deutschen geologischen Gesellschaft, **68**, 441–476.

Hennig, W. 1950. Grundzüge einer Theorie der Phylogenetischen Systematik. Deutscher Zentralverlag, Berlin, 396 p.

Hennig, W. 1965. Phylogenetic systematics. Annual Review of Entomology, **10**, 97–116.

Hennig, W. 1966. Phylogenetic Systematics. Urbana, Chicago, London, University of Illinois Press, 284 p.

Hennuy, J. 2003. Sédimentation Carbonatée et Silicoclastique Sous Contrôle Tectonique, Le Bassin Sud- Provençal et Sa Plate-Forme Carbonatée Au Turonien Moyen Au Coniacien Moyen. Evolutions Séquentielle, Diagénétique, Paleogéographique. Thèse de doctorat, Université de Provence, 252 p.

Hennuy, J. & Floquet, M. 2001. Evolution de La Plate-Forme Carbonatée Sud-Provençale Au Turonien Moyen pro-Parte - Coniacien Inférieur. Relations Avec Les Resédimentations Carbonatées Dans Le Bassin Sud-Provençal. Rapport de finalisation de contrat de recherche établi entre l'Université de Provence et la société Elf-Aquitaine, réf. EP/T/IRD N°00-465, inédit, 89 p.

Herrmannsen, A.N. 1846-1852. *Indicis generum malacozoorum primordia. Nomina subgenerum, generum, familiarum, tribuum, ordinum, classium: adjectis auctoribus, temporibus, locis systematicis atque literariis, etymus, synonymis. Praetermittuntur Cirripedia, Tunicata et Rhizopoda*. T. Fischeri, Cassel, 1-2, 1-232 (1846), 233-637 (1847-1), 1-352 (1847-2), 353-492 (1848), 493-717 (1849), supplément (1852).

Jardine, N. 1969. A logical basis for biological classification. Systematic Zoology, **18**, 37–52.

Kannan, S., Warnow, T. & Yooseph, S. 1998. Computing the local consensus of trees. SIAM Journal on Computing, **27**, 1695–1724.

Kearney, M. 2002. Fragmentary taxa, missing data, and ambiguity: mistaken assumptions and conclusions. Systematic biology, **51**, 369–381.

Kennedy, W. J. & Taylor, J. D. 1968. Aragonite in rudists. Proceedings of the Geological Society of London, **1645**, 326–331.

Kidwell, S. M. & Holland, S. M. 1991. Field description of coarse bioclastic fabrics. Palaios, **6**, 426–434.

Kidwell, S. M. & Holland, S. M. 2002. The Quality of the Fossil Record: Implications for Evolutionary Analyses. Annual Review of Ecology and Systematics, **33**, 561–588.

Kitching, I. J., Forey, P. L., Humphries, C. J. & Williams, D. M. 1998. Cladistics: The Theory and Practice of Parsimony Analysis. Second Edition. Oxford University Press, Oxford, 252 p.

Klappa, C. F. 1978. Biolithogenesis of *Microcodium*: elucidation. Sedimentology, **25**, 489–522.

Kluge, A. 1994. Moving targets and shell games. Cladistics, **10**, 403–413.

Kluge, A. G. & Farris, J. S. 1969. Quantitative Phyletics and the Evolution of Anurans. Systematic Zoology, **18**, 1–32.

Kohlsdorf, T., Lynch, V. J., Rodrigues, M. T., Brandley, M. C. & Wagner, G. P. 2010. Data and Data Interpretation in the Study of Limb Evolution: a Reply to Galis *et al*. on the Reevolution of Digits in the Lizard Genus *Bachia*. Evolution, **64**, 2477–2485.

Korbar, T., Fucek, L., Husinec, A., Vlahovic, I., Ostric, N., Maticek, D. & Jelaska, V. 2001. Cenomanian Carbonate Facies and Rudists along Shallow Intraplatform Basin Margin - the Island of Cres (Adriatic Sea, Croatia). Facies, **45**, 39–58.

Laet de, J. & Smets, E. 1998. On the three-taxon approach to parsimony analysis. Cladistics, **14**, 363–381.

Lamarck, J.-B. 1819. Histoire Naturelle Des Animaux sans Vertèbres. Volume 6. Paris, Verdière, 232 p.

Lamarck, J.-B. 1835. Histoire Naturelle Des Animaux sans Vertèbres. Volume 6. Deuxième édition. Paris, Baillière, 600 p.

Lapointe, F. J., Wilkinson, M. & Bryant, D. 2003. Matrix representations with parsimony or with distances: two sides of the same coin? Systematic Biology, **52**, 865–868.

Lapointe, F.-J. & Cucumel, G. 1997. The average consensus procedure: combination of weighted trees containing identical or overlapping sets of taxa. Systematic Biology, **46**, 306–312.

Laurin, M. 2004. The Evolution of Body Size, Cope's Rule and the Origin of Amniotes. Systematic Biology, **53**, 594–622.

Laurin, M. 2010. Assessment of the Relative Merits of a Few Methods to Detect Evolutionary Trends. Systematic Biology, **59**, 689–704.

Laurin, M. & Germain, D. 2011. Developmental Characters in Phylogenetic Inference and Their Absolute Timing Information. Systematic Biology, **60**, 630–644.

Laurin, M., Gussekloo, S. W. S., Marjanović, D., Legendre, L. & Cubo, J. 2012. Testing gradual and speciational models of evolution in extant taxa: the example of ratites: Testing models of evolution. Journal of Evolutionary Biology, **25**, 293–303.

Lebbe, J. 1991. Représentation Des Concepts En Biologie et Médecine. Introduction À L'analyse Des Connaissances et À L'identification Assistée Par Ordinateur. Thèse de doctorat, Université Pierre et Marie Curie, 281 p.

Leitner, T., Escanilla, D., Franzén, C., Uhlén, M. & Albert, J. 1996. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proceedings of the National Academy of Sciences, **93**, 10864–10869.

Lelièvre, H., Zaragüeta i Bagils, R. Z. & Rouget, I. 2008. Temporal information, fossil record and phylogeny. Comptes Rendus Palevol, **7**, 27–36.

Linnaeus Von, C. 1758. *Systema Naturæ per Regna Tria Naturæ, Secundum Classes, Ordines, Genera, Species, Cum Characteribus, Differentiis, Synonymis, Locis*. 10e edition. Laurentii Salvii.

Liu, J. 2016. *Yuanansuchus maopingchangensis* sp. nov., the second capitosauroid temnospondyl from the Middle Triassic Badong Formation of Yuanan, Hubei, China. PeerJ, **4,** e1903.

Lu, J., Zhu, M., Ahlberg, P. E., Qiao, T., Zhu, Y., Zhao, W. & Jia, L. 2016. A Devonian predatory fish provides insights into the early evolution of modern sarcopterygians. Science advances, **2**, e1600154.

Mac Gillavry, H. J. 1937. Geology of the province of Camagüey, Cuba, with revisional studies in rudist paleontology (mainly based upon collections from Cuba). Geographische en Geologische Mededeelingen, Physiographisch-Geologische series Reeks, **14**, 168 p.

Maddison, W. 1989. Reconstructing character evolution on polytomous cladograms. Cladistics, **5**, 365–377.

Maddison, W. P. 1991. Squared-Change Parsimony Reconstructions of Ancestral States for Continuous-Valued Characters on a Phylogenetic Tree. Systematic Biology, **40**, 304–315.

Maddison, W. P. 1993. Missing data versus missing characters in phylogenetic analysis. Systematic Biology, **42**, 576–581.

Maddison, W. & Maddison, D. 2001. Mesquite: a modular system for evolutionary analysis. Version 2.6. http://mesquiteproject.org.

Maddison, W. P., Donoghue, M. J. & Maddison, D. R. 1984. Outgroup Analysis and Parsimony. Systematic Biology, **33**, 83–103.

Malchus, N. 1995. The meaning of 'inversion' in Chamids and rudists (Bivalvia) reviewed and an unbiassed theoretical approach to late jurassic-early cretaceous rudist phylogeny. Revista Mexicana de Ciencias Geológicas, **12**, 211–223.

Margush, T. & McMorris, F. R. 1981. Consensus n-trees. Bulletin of Mathematical Biology, **43**, 239–244.

Marjanović, D. & Laurin, M. 2008. A reevaluation of the evidence supporting an unorthodox hypothesis on the origin of extant amphibians. Contributions to Zoology, **77**, 149–199.

Maslin, T. P. 1952. Morphological criteria of phyletic relationships. Systematic Zoology, **1**, 49–70.

Masse, J.-P. & Fenerci-Masse, M. 2006. Carbonate production by rudist bivalves. The record of Late Barremian requieniid communities from Provence (SE France). Palaeogeography, Palaeoclimatology, Palaeoecology, **234**, 239–257.

Masse, J.-P. & Fenerci-Masse, M. 2010. *Mathesia Mainelli* (Hippuritoidea, Monopleuridae) from the late Aptian-Albian of the Mediterranean region, a revision. Turkish Journal of Earth Sciences, **19**, 543–556.

Masse, J.-P. & Fenerci-Masse, M. 2014. Evolution of the rudist bivalve *Agriopleura* Kühn (Radiolitidae, Hippuritida) from the Mediterranean region. Palaeontology, **58**, 71–100.

Masse, J. & Montaggioni, L. 2001. Growth history of shallow-water carbonates: control of accommodation on ecological and depositional processes. International Journal of Earth Sciences, **90**, 452–469.

Masse, J.-P. & Philip, J. 1976. Paléogéographie et tectonique du Crétacé moyen en Provence: révision du concept d'isthme durancien. Revue de Géographie physique et de Géologie dynamique, **18**, 49–66.

Masse, J.-P., Philip, J. & Toomey, D. F. 1981. Cretaceous coral-rudist buildups of France. In: Toomey., D. F. (ed) European Fossil Reef Models. SEPM Special Publication, **30**, 399–426.

Masse, J.-P., Skelton, P. & Slišković, T. 1984. *Glossomyophorus costatus* nov. gen. nov. sp., rudiste (Caprotinidae) nouveau de l'aptien du domaine Méditerranéen central et oriental. Geobios, 723–732.

Masse, J., Fenerci-Masse, M. & Pernarcic, E. 2003. Palaeobathymetric reconstruction of peritidal carbonates. Late Barremian, Urgonian, sequences of Provence (SE France). Palaeogeography, Palaeoclimatology, Palaeoecology, **200**, 65–81.

Masse, J.-P., Fenerci-Masse, M., Arias, C. & Vilas, L. 2015. Description of a new species of Offneria (Hippuritida, Caprinidae) from the lower Aptian of southeast Spain. Stratigraphic, evolutionary, palaeobiogeographic and palaeoenvironmental implications. Cretaceous Research, **53**, 153–166.

Masse, J.-P., Tüysüz, O., Fenerci-Masse, M., Özer, S. & Sari, B. 2009. Stratigraphic organisation, spatial distribution, palaeoenvironmental reconstruction, and demise of Lower Cretaceous (Barremian-lower Aptian) carbonate platforms of the Western Pontides (Black Sea region, Turkey). *Cretaceous Research*, **30**, 1170–1180.

Mayr, E. 1986. La systématique évolutionniste et les quatre étapes du processus de classification. In: Tassy, P. (ed.) L'ordre et La Diversité Du Vivant. Paris, Fayard, Fondation Diderot, 143–160.

McLachlan, G. J. 2004. Discriminant Analysis and Statistical Pattern Recognition. Wiley, New York, Wiley series in probability and mathematical statistics, 526 p.

McMorris, F. R. & Powers, R. C. 2003. The arrovian program from weak orders to hierarchical and tree-like relations. In: Janowitz, M. F., Lapointe, F. J., McMorris, F. R., Mirkin, B. & Roberts, F. S. (eds) Bioconsensus. Providence, DIMACS series in discrete mathematics and theoretical computer sciences, **61**, 37–45.

Merle, D. 2005. The spiral cords of the Muricidae (Gastropoda, Neogastropoda): importance of ontogenetic and topological correspondences for delineating structural homologies. Lethaia, **38**, 367–379.

Mickevich, M. F. 1982. Transformation Series Analysis. Systematic Biology, **31**, 461–478.

Mickevich, M. F. & Johnson, M. S. 1976. Congruence Between Morphological and Allozyme Data in Evolutionary Inference and Character Evolution. Systematic Biology, **25**, 260–270.

Mickevich, M. F. & Platnick, N. I. 1989. On the information content of classifications. Cladistics, **5**, 33–47.

Miller, A. I. & Foote, M. 1996. Calibrating the Ordovicien Radiation of Marine Life: Implications for Phanerozoic Diversity Trends. Paleobiology, **22**, 304–309.

Mitchell, S. F. 2002. Palaeocology of corals and rudists in mixed volcaniclastic-carbonate small-scale rhythms (Upper Cretaceous, Jamaica). Palaeogeography, Palaeoclimatology, Palaeoecology, **2927**, 1–24.

Mitchell, S. F. 2010. Revision of the antillocaprinid rudist bivalve *Antillosarcolites* Chubb with a description of a new genus: Implications for evolution of Antillocaprinidae. Geobios, **43**, 511–517.

Mitchell, S. F. 2013a. A new rudist bivalve *Curtocaprina clabaughikinsorum* gen. et sp. nov. from the Middle Albian of Texas and its bearing on the origin of the Ichthyosarcolitidae Douvillé. Caribbean Journal of Earth Science, **45**, 85–89.

Mitchell, S. F. 2013b. Revision of the Antillocaprinidae Mac Gillavry (Hippuritida, Bivalvia) and their position within the Caprinoidea d'Orbigny. Geobios, **46**, 423–446.

Mitchell, S. F. & Gunter, G. C. 2006. New tube-bearing Antillocaprinid rudist bivalves from the Maastrichtian of Jamaica. Palaeontology, **49**, 35–57.

Möbius, K. A. 1877. Die Auster Und Die Austernwirthschaft. Verlag von Wiegandt. Berlin, Hemple & Parey, 126 p.

Moreau, P. 1993. La Transgression Cénomanienne Sur La Marge Septentrionale Du Bassin de L'Aquitaine (Charentes). Flanc Nord Du Synclinal de Saintes et de l'Angoumois. Modalité D'une Invasion Marine. Thèse de doctorat, Université de Poitiers, 505 p.

Müller, J. & Reisz, R. R. 2006. The Phylogeny of Early Eureptiles: Comparing Parsimony and Bayesian Approaches in the Investigation of a Basal Fossil Clade. Systematic Biology, **55**, 503–511.

Munier-Chalmas, H. 1882. Etudes critiques sur les Rudistes. Bulletins de la société géologique de France, **3**, 482–494.

Nelson, G. 1970. Outline of a theory of comparative biology. Systematic Zoology, **19**, 373–384.

Nelson, G. 1972. Phylogenetic relationship and classification. Systematic Zoology, **21**, 227–231.

Nelson, G. 1973. Classification as an expression of phylogenetic relationships. Systematic Zoology, **22**, 344–359.

Nelson, G. 1978. Ontogeny, phylogeny, paleontology and the biogenetic law. Systematic Zoology, **27**, 324–345.

Nelson, G. 1979. Cladistic Analysis and Synthesis: Principles and Definitions, with a Historical Note on Adanson's Familles Des Plantes (1763–1764). Systematic Biology, **28**, 1–21.

Nelson, G. 1994. Homology and systematics. In: Hall, B. K. (ed.) The Hierarchical Basis of Comparative Biology. San Diego, Academic Press, 101–149.

Nelson, G. 1996. Nullius in Verba. American Museum of Natural History, New York, and School of Botany, University of Melbourne, publié à compte d'auteur, 24 p.

Nelson, G. 2004. Cladistics: its arrested development. Systematics Association Special Volume, **67**, 127–148.

Nelson, G. & Ladiges, P. Y. 1991a. Standard assumptions for biogeographic analysis. Australian Systematic Botany, **4**, 41–58.

Nelson, G. & Ladiges, P. Y. 1991b. Three-area statements: standard assumptions for biogeographic analysis. Systematic Biology, **40**, 470–485.

Nelson, G. & Ladiges, P. Y. 1992. Information content and fractional weight of three-item statements. Systematic biology, **41**, 490–494.

Nelson, G. & Ladiges, P. Y. 1994. Three-item consensus empirical test of fractional weighting. In: R. W., Scotland, D. J., Siebert, & D. M., Williams (eds) Models in phylogeny reconstruction. Clarendon Press, Oxford, 193–209.

Nelson, G. J. & Ladiges, P. Y. 1996. Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. American Museum novitates, **3167**, 58.

Nelson, G. & Platnick, N. 1981. Systematics and Biogeography: Cladistics and Vicariance. Columbia University Press, 567 p.

Nelson, G. & Platnick, N. I. 1991. Three-Taxon Statement: a more precise use of parsimony? Cladistics, **7**, 351–366.

Nelson, G., Williams, D. M. & Ebach, M. C. 2003. A question of conflict: Three-item and standard parsimony compared. Systematics and Biodiversity, **1**, 145–149.

Newell, N. D. 1965. Classification of the Bivalvia. American Museum novitates, **2206**, 1–25.

Newman, M. E. J. & Sibani, P. 1999. Extinction, diversity and survivorship of taxa in the fossil record. Proceedings of the Royal Society of London. Series B: Biological Sciences, **266,** 1593–1599.

Nopcsa, F. B. 1923. Reversible and Irreversible Evolution; a Study based on Reptiles. By Dr. Francis, Baron Nopcsa. Journal of Zoology, **93**, 483–1097.

Omaña, L., Doncel, R. L., Torres, J. R. & Alencaster, G. 2012. Biostratigraphy and paleoenvironment of the Cenomanian/Turonian boundary interval based on foraminifera from W Valles-San Luis Potosí Platform, Mexico. Micropaleontology, **58**, 457–485.

Omaña, L., Torres, J. R., López Doncel, R., Alencáster, G. & López Caballero, I. 2014. A pithonellid bloom in the Cenomanian-Turonian boundary interval from Cerritos in the western Valles–San Luis Potosí platform, Mexico: Paleoenvironmental significance. Revista Mexicana de Ciencias Geológicas, **31**, 28–44.

Orbigny de, A. 1822. Notice sur quelques espèces nouvelles de mollusques fossiles, du Département de la Charente-Inférieure. Mémoires du Muséum d'Histoire naturelle, **8**, 98–110.

Orbigny de, A. 1840. Note sur le genre Caprine. Revue zoologique. Par la Société Cuvierienne, **1839**, 168–170.

Orbigny de, A. 1842a. Quelques considérations géologiques sur les rudistes. Bulletin de la Société Géologique de France, **13**, 148–163.

Orbigny de, A. 1842b. Quelques considérations zoologiques et géologiques sur les Rudistes. Annales des sciences naturelles, **2**, 173–192.

Orbigny de, A. 1850. Paléontologie française. Description zoologique et géologique de tous les animaux mollusques et rayonnés fossiles de France. Terrains Crétacés. Brachiopoda. Texte et atlas. 408 p.

Owen, R. 1843. Lectures on Comparative Anatomy and Physiology of the Invertebrate Animals, Delivered at the Royal College of Surgeons in 1843. Longman, Brown, Green & Longmans. London p.

Owen, R. 1848. On the Archetype and Homologies of the Vertebrate Skeleton. London, Voorst, 203 p.

Özer, S. & Ahmad, F. 2016. *Caprinula* and *Sauvagesia* rudist faunas (Bivalvia) from the Cenomanian of NW Jordan. Stratigraphy and taxonomy. Cretaceous Research, **58**, 141–159.

Özer, S., Sözbilir, H., Özkar, İ., Toker, V. & Sari, B. 2001. Stratigraphy of Upper Cretaceous–Palaeogene sequences in the southern and eastern Menderes Massif (western Turkey). International Journal of Earth Sciences, **89**, 852–866.

Page, R. D. M. 1989. Comments on Component-Compatibility in Historical Biogeography. Cladistics, **5**, 167–182.

Palissy, B. 1563. Recepte Véritable Par Laquelle Tous Les Hommes de La France Pourront Apprendre À Multiplier et Augmenter Leurs Thrésors. De l'Imprimerie de Bathélémy Berton, la Rochelle.

Paquier, V. 1903. Les rudistes urgoniens. II . Série inverse. Mémoires de la société géologique de France, Paléontologie, **29**, 47–102.

Parona, C. F. 1921. Fauna del neocretacico della Tripolitania. Memorie per servire alla Descrizione della Carta geologica d'Italia, **8**, 1–21.

Patel, D. D., & Pickup, D. J. 1989. The second-largest subunit of the poxvirus RNA polymerase is similar to the corresponding subunits of procaryotic and eucaryotic RNA polymerases. Journal of virology, **63**, 1076–1086.

Patterson, C. 1982. Morphological characters and homology. In: Joysey, K. A. & Friday, E. (eds) Problems of Phylogenetic Reconstruction. New York, Academic Press, 21–74.

Patterson, C. & Johnson, G. D. 1995. The intermuscular bones and ligaments of teleostean fishes. Smithsonian Contributions to Zoology (USA), **559**, 1–85.

Philip, J. 1967. Les zones de Rudistes du Cénomanien Provençal. Bulletin de la Société Géologique de France, **7**, 497–503.

Philip, J. 1970. Les Formations Calcaires À Rudistes Du Crétacé Supérieur Provençal et Rhodanien. Thèse de doctorat, Université de Provence, 2011, 438 p.

Philip, J. 1972. Paléoécologie des formations a rudistes du crétacé supérieur-l'exemple du sud-est de la France. Palaeogeography, Palaeoclimatology, Palaeoecology, **12**, 205–222.

Philip, J. 1978. Stratigraphie et paléoécologie des formations à rudistes du Cénomanien: l'exemple de la Provence. Géologie méditerranéenne, **5**, 155–168.

Philip, J. 1980. Crétacé supérieur de Provence. Geobios mémoire spécial, **4**, 99–109.

Philip, J. 1981. Les rudistes du Crétacé moyen de la province Méditerranéenne occidentale. Evolution, paléoécologie, paléobiogéographie. *Cretaceous Research*, **2**, 395–403.

Philip, J. 1993. Late Cretaceous Carbonate-Siliciclastic Platforms of Provence, Southeastern France. In: Simo, J. A. T., Scott, R. W. & Masse, J.-P. (eds) Cretaceous Carbonate Platforms. Memoirs of the American Association of Petroleum Geologists, 375–385.

Philip, J. 1998. Biostratigraphie et paléobiogéographie des rudistes: évolution des concepts et progrès récents. Bulletins de la société géologique de France, **169**, 689–709.

Philip, J., Airaud, C. & Tronchetti, G. 1989. Événements paléogéographiques en provence (SE France) au passage Cénomanien-Turonien Modifications biosédimentaires-Causes géodynamiques. Geobios, **22**, 107–117.

Philip, J., Borgomano, J. & Al-Maskiry, S. 1995. Cenomanian-Early Turonian carbonate platform of Northern Oman: stratigraphy and palaeo-environments. Palaeogeography, Palaeoclimatology, Palaeoecology, **119**, 77–92.

Phillips, C. & Warnow, T. J. 1996. The asymmetric median tree — a new model for building consensus trees. Discrete Applied Mathematics **71**, 311–335.

Picot de Lapeirousse, P. I. 1781. Description de Plusieurs Nouvelles Espèces d'Orthocératites et d'Ostracites. Wolfgang Walther, Erlang, 45 p.

Pimentel, R. A. & Riggins, R. 1987. The Nature of Cladistic Data. Cladistics, **3**, 201–209.

Pinna de, M. C. 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics, **7**, 367–394.

Pinna de, M. C., Bockmann, F. A. & Zaragüeta i Bagils, R. 2016. Unrooted trees discovered independently in philology and phylogenetics: a remarkable case of methodological convergence. Systematics and Biodiversity, **14**, 1–10.

Platnick, N. I. 1979. Philosophy and the transformation of cladistics. Systematic Zoology, **28**, 537–546.

Platnick, N. I. 2013. Less on homology. Cladistics, **29**, 10–12.

Platnick, N. I., Humphries, C. J., Nelson, G. & Williams, D. M. 1996. Is Farris Optimization perfect?: three-taxon statements and multiple branching. Cladistics, **12**, 243–252.

Pleijel, F. 1995. On character coding for phylogeny reconstruction. Cladistics, **11**, 309–315.

Pleijel, F. & Rouse, G. W. 2000. Least-inclusive taxonomic unit: a new taxonomic concept for biology. Proceedings of the Royal Society of London. Series B: Biological Sciences, **267**, 627–630.

Pleničar, M. 1965. New discoveries of rudists in the region of Kocevski Rog, southeastern Slovenia. Geologija, **8**, 92–101.

Pleničar, M. & Dozet, S. 1994. Contribution to the Knowledge of Upper Cretaceous Beds in Kočevje and Gorski Kotar Area (NW Dinarides). Geologija, **36**, 183–194.

Pleničar, M. & Jurvošek, B. 1999. Rudists from the Cenomanian bioherms of Hrušica and Nanos, Slovenia. Geologija, **42**, 69–116.

Pleničar, M., Jurkovšek, B. & Kolar-Jurkovšek, T. 1999. Stop 1: Cenomanian-Turonian bioherm on Hrusica. In: Höfling, R., & Steuber, T. (eds) Fifth International Congress on Rudists - Abstracts and field trip guides. Erlangen, Erlanger geologische Abhandlungen Sb3, 118–121.

Počta, F. 1887. Vorläufiger Bericht über die Rudisten der böhmischen Kreideformation. Vestnik královské ceské spolecnosti Nauk, Trída mathematicko-prírodovedecká, **1886**, 194–208.

Počta, F. 1889. Ueber Rudisten, eine ausgestorbene Familie der Lamellibranchiaten, aus der böhmischen Kreideformation. Rozpravy královské Ceské spolecnosti Nauk, **7**, 78 p.

Poe, S. 2006. Test of Von Baer's Law of the Conservation of Early Development. Evolution, **60**, 2239–2245.

Poe, S. & Wake, M. H. 2004. Quantitative tests of general models for the evolution of development. The American naturalist, **164**, 415–422.

Pogue, M. G. & Mickevich, M. F. 1990. Character definitions and character state delineation: the bête noire of phylogenetic inference. Cladistics, **6**, 319–361.

Polšak, A. 1967. Kredna makrofaun a južne Istre - Macrofaune crétacée de l'Istrie méridionale (Yougoslavie). Palaeontologica jugoslavica, **8**, 219 p.

Pons, J. M. & Vicens, E. 2008. The structure of the outer shell layer in radiolitid rudists, a morphoconstructional approach. Lethaia, **41**, 219–234.

Pons, J. M., Vicens, E., Oviedo, A., Aguilar, J., García-Barrera, P. & Alencáster, G. 2013. The Rudist Fauna of the Cárdenas Formation, Maastrichtian, San Luis Potosí State, Mexico. Journal of Paleontology, **87**, 726–754.

Popper, K. R. 1998. The World of Parmenides. Essays on the Presocratic Enlightenment. Petersen, A., Routledge, London and New York, 328 p.

Prin, S. 2012. Structure Mathématique Des Hypothèses Cladistiques et Conséquences Pour La Phylogénie et L'évolution. Avec Une Perspective Sur L'analyse Cladistique. Thèse de doctorat, Muséum national d'Histoire naturelle, 305 p.

Prin, S. 2016. The relational view of phylogenetic hypotheses and what it tells us on the phylogeny/classification relation problem. In: Williams, D., Schmitt, M. & Wheeler, Q. (eds) The Futur of Phylogenetic Systematics: The Legacy of Willi Hennig. The Systematics Association, Special Volume Series, 431–468.

Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J. & Green, R. E. 2010. Computational challenges in the analysis of ancient DNA. Genome biology, **11**, R47.

Pyron, R. A. 2011. Divergence Time Estimation Using Fossils as Terminal Taxa and the Origins of Lissamphibia. Systematic Biology, **60**, 466–481.

Quesne Le, W. J. 1969. A method of selection of characters in numerical taxonomy. Systematic Zoology, **18**, 201–205.

Quesne Le, W. J. 1972. Further studies based on the uniquely derived character concept. Systematic Biology, **21**, 281–288.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ragan, M. 1992. Phylogenetic Inference Based on Matrix Representation of Trees. Molecular Phylogenetics and Evolution, **1**, 53–58.

Rameil, N., Immenhauser, A., Warrlich, G., HillgäRtner, H. & Droste, H. J. 2010. Morphological patterns of Aptian *Lithocodium-Bacinella* geobodies: relation to environment and scale: *Lithocodium-Bacinella*, Aptian of Oman. Sedimentology, **57**, 883–911.

Rasband, W. 1997. ImageJ. U. S. National Institutes of Health, Bethesda, Maryland, USA. https://imagej.nih.gov/ij/.

Redondo, C. 1986. Etude Des Sédiments Détritiques Du Crétacé Supérieur Marin de La Provence Occidentale et Recherche Des Zones D'apport. Sédimentologie ; Pétrographie ; Minéralogie. Thèse de doctorat, Université de Provence, 2011, 474 p.

Remane, A. 1952. Die Besiedlung des Sandbodens im Meere und die Bedeutung der Lebensformtypen für die Ökologie. Zool Anz, **16 (Suppl)**, 327–359.

Rineau, V., Grand, A., Zaragüeta i Bagils, R. & Laurin, M. 2015. Experimental systematics: sensitivity of cladistic methods to polarization and character ordering schemes. Contributions to Zoology, **84**, 129–148.

Robinson, D. F. & Foulds, L. R. 1981. Comparison of phylogenetic trees. Mathematical Biosciences, **53**, 131–147.

Roduit, N. 2008. JMicroVision Version 1.2. 7: Image Analysis Toolbox for Measuring and Quantifying Components of High-Definition Images.

Rohlf, J. 1982. Consensus indices for comparing classifications. Mathematical Biosciences, **59**, 131–144.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L. & Rasnitsyn, A. P. 2012. A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera. Systematic Biology, **61**, 973–999.

Rosen, D. R. 1979. Fishes from the uplands and intermontane basins of Guatemala: revisionary studies and comparative geography. Bulletin American Museum of Natural History, **162**, 269–375.

Ross, D. & Skelton, P. 1993. Rudist formations of the Cretaceous: a palaeoecological, sedimentological and stratigraphical review. In: Wright P. (ed) Sedimentology Review, **1**, 73–91.

Ruberti, D. 1997. Facies analysis of an Upper Cretaceous high-energy rudist-dominated carbonate ramp (Matese Mountains, central-southern Italy): subtidal and peritidal cycles. Sedimentary Geology, **113**, 81–110.

Saint-Hilaire, G. 1818. Philosophie anatomique des organes respiratoires: sous le rapport de la détermination et de l'identité de leurs pièces osseuses: avec figures des 116 nouvelles préparations d'anatomie. Mequignon-Marvis, Paris.

Saint-Hilaire, G. 1830. Principes de philosophie zoologique, discutés en mars 1830, au sein de l'Académie Royale des Sciences. Pichon et Didier, Paris.

Sanders, D. 1999. Shell disintegration and taphonomic loss in rudist biostromes. Lethaia, **32**, 101–112.

Sanders, D. 2001. Syndepositional dissolution of calcium carbonate in neritic carbonate environments: geological recognition, processes, potential significance. African Earth Sciences, **36**, 99–134.

Sankoff, D. & Rousseau, P. 1975. Locating the vertices of a Steiner tree in an arbitrary metric space. Mathematical Programming, **9**, 240–246.

Sano, S.-I., Skelton, P. W., Watarai, M., Iba, Y., Kondo, Y. & Sato, Y. 2012. First record of an Early Barremian caprinid rudist from Japan - implications for the palaeobiogeography of the Caprinidae (Bivalvia). *Palaeontology*, **55**, 843–851.

Sano, S.-I., Iba, Y., Skelton, P. W., Masse, J.-P., Aguilar, Y. M. & Kase, T. 2014. The evolution of canaliculate rudists in the light of a new canaliculate polyconitid rudist from the Albian of the Central Pacific. Palaeontology, **57,** 951–962.

Sansom, R. S., Gabbott, S. E. & Purnell, M. A. 2011. Decay of vertebrate characters in hagfish and lamprey (Cyclostomata) and the implications for the vertebrate fossil record. Proceedings of the Royal Society B: Biological Sciences, **278**, 1150–1157.

Sarı, B. & Özer, S. 2009. Upper Cretaceous rudist biostratigraphy of the Bey Dağları Carbonate Platform, Western Taurides, SW Turkey. Geobios, **42**, 359–380.

Sauvages de, P. A. 1746. Mémoire contenant des observations de Lithologie, pour servir à l'Histoire Naturelle du Languedoc, et à la théorie de la Terre. 1re partie. Paris, Mémoire de l'académie royale des sciences, 713–758.

Schneider, S. & Skupien, P. 2015. Palaeoecology of *Valletia antiqua* Favre in Joukowsky & Favre, 1913 (Bivalvia, Hippuritida, Diceratidae); with comments on the taxonomy and distribution of the genus *Valletia*. Bulletin of Geosciences, **90**, 613–632.

Schopf, J. W., Kudryavtsev, A. B., Czaja, A. D. & Tripathi, A. B. 2007. Evidence of Archean life: Stromatolites and microfossils. Precambrian Research, **158**, 141–155.

Schumann, D. 2000. Paleoecology of Late Cretaceous rudists settlements in Central Oman. In: Alsharhan, A. S. & Scott, R. W. (eds). Middle East Models of Jurassic/Cretaceous Carbonate Systems, SEPM Special Publication, **69**, 143–153.

Scotland, R. W. 2000. Taxic homology and three-taxon statement analysis. Systematic biology, **49**, 480–500.

Scott, R. W. 1995. Cretaceous rudists of Guatemala. Revista Mexicana de Ciencias Geológicas, **12**, 294–306.

Seilacher, A. 1984. Constructional morphology of bivalves: evolutionary pathways in primary versus secondary soft-bottom dwellers. Palaeontology, **27**, 207–237.

Seilacher, A. 1990. Aberrations in bivalve evolution related to photo-and chemosymbiosis. Historical Biology, **3**, 289–311.

Sharkey, M. J. & Leathers, J. W. 2001. Majority Does Not Rule: The Trouble with Majority-Rule Consensus Trees. Cladistics, **17**, 282–284.

Siebert, D. J. & Williams, D. M. 1998. Recycled. Cladistics, **14**, 339–347.

Simon, C. 1983. A New Coding Procedure for Morphometric Data with an Example from Periodical Cicada Wing Veins. In: Felsenstein, J. (ed.) Numerical Taxonomy. Springer Berlin, NATO ASI Series, **1**, 378–382.

Sirna, G. 1982. Quelques rudistes cénomaniens du Monte Pellegrino (Palermo, Sicile). Geologia romana, **21**, 79–87.

Skelton, P. W. 1976. Functional morphology of the Hippuritidae. *Lethaia*, **9**, 83–100.

Skelton, P. W. 1978. The Evolution of Functional Design in Rudists (Hippuritacea) and Its Taxonomic Implications. Philosophical Transactions of the Royal Society B: Biological Sciences, **284**, 305–318.

Skelton, P. W. 1979. Preserved ligament in a radiolitid rudist bivalve and its implications of mantle marginal feeding in the group. *Paleobiology*, **5**, 90–106.

Skelton, P. W. 1991. Morphogenetic versus environmental cues for adaptive radiations. In: Schmidt-Kittler, N. & Vogel, K. (eds) Constructional Morphology and Evolution. Berlin, Heidelberg, 375–388.

Skelton, P. W. 2003. Rudist evolution and extinction - A north african perspective. In: Gili, E., El Hédi Negra, M. & Skelton, P. W. (eds) North African Cretaceous Carbonate Platform Systems. NATO Science Series, **IV**, 215–227.

Skelton, P. W. 2004. *Oedomyophorus shaybahensis*, A New Genus And Species Of Caprinid (?) Rudist From The Lower Aptian Shu'aiba Formation Of Eastern Saudi Arabia. Courier Forschungsinstitut Senckenberg, **247**, 35–47.

Skelton, P. W. 2013a. Rudist classification for the revised Bivalvia volumes of the 'Treatise on Invertebrate Paleontology'. Caribbean Journal of Earth Science, **45**, 9–33.

Skelton, P. W. 2013b. Rudist classification: nomenclatural correction of 'Suborder Radiolitidina Skelton, 2013' to 'Suborder Hippuritidina Newell, 1965'. Caribbean journal of Earth Science, **45**, 34–34.

Skelton, P. W. 2014. Why myophores are fundamental in rudist systematics. In: Pons, J. M, & Vicens, E (eds) Tenth International Congress on Rudist Bivalves, Bellaterra, June 22-27, Scientific Program and Abstracts. Barcelona, Universitat Autònoma de Barcelona, 31.

Skelton, P. W. & Gili, E. 2012. Rudists and carbonate platforms in the Aptian: a case study on biotic interactions with ocean chemistry and climate: Aptian rudists and carbonate platforms. Sedimentology, **59**, 81–117.

Skelton, P. W. & Masse, J.-P. 1998. Revision of the Lower Cretaceous rudist genera *Pachytraga* Paquier and *Retha* Cox (Bivalvia: Hippuritacea), and the origins of the Caprinidae. Geobios, **31**, 331–370.

Skelton, P. W. & Smith, A. B. 2000. A preliminary phylogeny for rudist bivalves: sifting clades from grades. Geological Society, London, Special Publications, **177**, 97–127.

Skelton, P. W., Masse, J.-P. 2000. Synoptic guide to Lower Cretaceous rudist bivalves of Arabia. In: Scott, R. W. & Alsharhan, A. S. (eds) Middle East Models of Jurassic/Cretaceous Carbonate Systems. SEPM Special Publication, **69**, 89–99.

Slišković, T. 1966. Zwei neue arten der gattung *Ichthyosarcolites* aus der oberkreide (ablagerungen) der südherzegowina. Conseil des Académies des Sciences et des Arts de la RSF de Yougoslavie, Bulletin scientifique, **A**, 177–178.

Slowinski, J. B. 1993. 'Unordered' versus 'ordered' characters. Systematic Biology, **42**, 155–165.

Smith, A. B. 1994. Systematics and the fossil record: documenting evolutionary patterns. Oxford, Blackwell scientific publications, 223 p.

Smith, A. M. & Nelson, C. S. 2003. Effects of early sea-floor processes on the taphonomy of temperate shelf skeletal carbonate deposits. Earth-Science Reviews, **63**, 1–31.

Sokal, R. R. & Rohlf, F. J. 1981. Taxonomic Congruence in the Leptopodomorpha Re-Examined. Systematic Zoology, **30**, 309–325.

Soua, M. & Tribovillard, N. 2007. Modèle de sédimentation au passage Cénomanien/Turonien pour la formation Bahloul en Tunisie. Comptes Rendus Geoscience, **339**, 692–701.

Spencer, M., Davidson, E. A., Barbrook, A. C. & Howe, C. J. 2004. Phylogenetics of artificial manuscripts. Journal of Theoretical Biology, **227**, 503–511.

Stephenson, L. W. 1938. A new upper cretaceous rudistid from the Kemp clay of Texas. Shorter contributions to general geology, **193**, 1–15.

Steuber, T. 1999. Cretaceous Rudists of Boeotia, Central Greece. Special papers in paleontology, **61**, 229 p.

Steuber, T. & Bachmann, M. 2002. Upper Aptian–Albian rudist bivalves from northern Sinai, Egypt. Palaeontology, **45**, 725–749.

Steuber, T. & Löser, H. 2000. Species richness and abundance patterns of Tethyan Cretaceous rudist bivalves (Mollusca: Hippuritacea) in the central-eastern Mediterranean and Middle East, analysed from a palaeontological database. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **162**, 75–104.

Steuber, T., Scott, R. W., Mitchell, S. F. & Skelton, P. W. 2016. Part N, revised, volume 1, chapter 26c: stratigraphy and diversity dynamics of Jurassic–Cretaceous Hippuritida (rudist bivalves).

Stevens, P. F. 1991. Character states, morphological variation, and phylogenetic analysis: a review. Systematic Botany, **16**, 553–583.

Stone, J. R. & Telford, M. 1999. Using critical path method to analyse the radiation of rudist bivalves. *Palaeontology*, **42**, 231–242.

Strasser, A. & Davaud, E. 1983. Black pebbles of the Purbeckian (Swiss and French Jura): lithology, geochemistry and origin. Eclogae geol. Helv, **76**, 551–580.

Strong, E. E. & Lipscomb, D. 1999. Character coding and inapplicable data. Cladistics, **15**, 363–371.

Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (*and Other Methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.

Swinburne, N. H. M. 1990. The Extinction of the Rudist Bivalves. Thèse de doctorat, Open University, 175 p.

Tassy, P. 2005. Fait et théorie : quelles connaissances de base pour la cladistique structurale ? In: Deleporte, P. & Lecointre, G. (eds) Biosystema 24 : Philosophie de La Systématique. Paris, Société Française de Systématique, 63–74.

Tavaré, S. 1986. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Lectures on Mathematics in the Life Sciences, **17**, 57–86.

Thiele, K. 1993. The Holy Grail of the Perfect Character: The Cladistic Treatment of Morphometric Data. Cladistics, **9**, 275–304.

Thorley, J. L. & Page, R. D. M. 2000. RadCon: phylogenetic tree comparison and consensus. Bioinformatics, **16**, 486–487.

Thorley, J. L., Wilkinson, M. & Charleston, M. 1998. The Information Content of Consensus Trees. In: Rizzi, P. A., Vichi, P. M. & Bock, P. D. H-H. (eds) Advances in Data Science and Classification. Springer, Berlin, Heidelberg, 91–98.

Tintant, H. 1977. Le polymorphisme intraspécifique en paléontologie (exemples pris chez les Ammonites). Haliotis, **6**, 49–69.

Tišljar, J., Vlahović, I., Velić, I., Matičec, D. & Robson, J. 1998. Carbonate facies evolution from the Late Albian to Middle Cenomanian in southern Istria (Croatia): influence of synsedimentary tectonics and extensive organic carbonate production. Facies, **38**, 137–151.

Tomasovych, A. & Schlogl, J. 2008. Analyzing Variations in Cephalopod Abundances in Shell Concentrations: The Combined Effects of Production and Density-Dependent Cementation Rates. PALAIOS, **23**, 648–666.

Trechmann, C. T. 1924. The Cretaceous limestones of Jamaica and their Mollusca. The Geological Magazine, **61**, 396–410.

Tuomey, M. 1854. Description of Some New Fossils, from the Cretaceous Rocks of the Southern States. Proceedings of the Academy of Natural Sciences of Philadelphia, **7**, 167–172.

Udovicic, F., Mcfadden, G. I. & Ladiges, P. Y. 1995. Phylogeny of Eucalyptus and Angophora Based on 5S rDNA Spacer Sequence Data. Molecular Phylogenetics and Evolution, **4**, 247–256.

Uhlenbeck, G. E. & Ornstein, L. S. 1930. On the Theory of the Brownian Motion. Physical Review, **36**, 823–841.

Videt, B. 2003. Dynamique Des Paléoenvironnements À Huîtres Du Crétacé Supérieur Nord-Aquitain (SO France) et Du Mio-Pliocène Andalou (SE Espagne): Biodiversité, Analyse Séquentielle, Biogéochimie. Thèse de doctorat, Université Rennes 1, 263 p.

Videt, B. 2007. Typologie des assemblages d'ostréidés (Bivalvia) et interprétations paléoenvironnementales. Comptes Rendus Palevol, **6**, 47–58.

Vignes, R. 1991. Caractérisation Automatique de Groupes Biologiques. Thèse de doctorat, Université Pierre et Marie Curie, 268 p.

Villier, L. 2009. Linking evolution of fossil communities with evolution of carbonate sedimentary systems. PALAIOS, **24**, 551–552.

Vinther, J., Stein, M., Longrich, N. R. & Harper, D. A. T. 2014. A suspension-feeding anomalocarid from the Early Cambrian. Nature, **507**, 496–499.

Vogel, K. 1975. Endosymbiotic algae in rudists? Palaeogeography, Palaeoclimatology, Palaeoecology, **17**, 327–332.

Walsh, D., Carswell, R. F. & Weymann, R. J. 1979. 0957+ 561 A, B: twin quasistellar objects or gravitational lens? Nature, **279**, 381–384.

Warnow, T., Ringe, D. & Taylor, A. 1995. Reconstructing the evolutionnary history of natural languages. Institute for Research in Cognitive Science, University of Pennsylvania, technical report, **129**, 18 p.

Wendler, J., Gräfe, K.-U. & Willems, H. 2002. Palaeoecology of calcareous dinoflagellate cysts in the mid-Cenomanian Boreal Realm: implications for the reconstruction of palaeoceanography of the NW European shelf sea. Cretaceous Research, **23**, 213–229.

Wheeler, Q. D. & Platnick, N. I. 2000. The phylogenetic species concept (sensu Wheeler and Platnick). In: Wheeler., Q. & Meier, R. (eds) Species Concepts and Phylogenetic Theory: A Debate. New York, Columbia University Press, 55–69.

Whiting, M., Bradler, S. & Maxwell, T. 2003. Loss and recovery of wings in stick insects. Nature, **421**, 264–267.

Wiens, J. J. 2001. Character analysis in morphological phylogenetics: problems and solutions. Systematic Biology, **50**, 689–699.

Wiley, E. O. 1976. The phylogeny and biogeography of fossil and recent gars (Actinopterygii: Lepisosteidae). Miscellaneous publications of the Museum of Natural History, University of Kansas, **64**, 1–111.

Wiley, E. O. 1986. Methods in vicariance biogeography. In: Hovenkamp, P. (ed.) Systematics and Evolution: A Matter of Diversity. Utrecht, 283–306.

Wiley, E. O. 1988a. Parsimony Analysis and Vicariance Biogeography. Systematic Zoology, **37**, 271–290.

Wiley, E. O. 1988b. Vicariance biogeography. Annual Review of Ecology and Systematics, **19**, 513–542.

Wiley, E. O., Siegel-Causey, D., Brooks, D. R. & Funk, V. A. 1991. The Compleat Cladist: A Primer of Phylogenetic Procedures. Lawrence, Kansas, University of Kansas, Museum of Natural History, special publication, 168 p.

Wilkinson, M. 1992. Ordered versus unordered characters. Cladistics, **8**, 375–385.

Wilkinson, M. 1994a. Common Cladistic Information and its Consensus Representation: Reduced Adams and Reduced Cladistic Consensus Trees and Profiles. Systematic Biology, **43,** 343–368.

Wilkinson, M. 1994b. Three-taxon statements: when is a parsimony analysis also a clique analysis? Cladistics, **10**, 221–223.

Wilkinson, M. 1995. More on Reduced Consensus Methods. Systematic Biology, **44,** 435–439.

Wilkinson, M. 2001. REDCON 3.0: Software and Documentation. Department of Zoology. The Natural History Museum, London p.

Wilkinson, M. & Benton, M. J. 1995. Missing data and rhynchosaur phylogeny. Historical Biology, **10**, 137–150.

Wilkinson, M., Cotton, J. A., Lapointe, F.-J. & Pisani, D. 2007. Properties of Supertree Methods in the Consensus Setting. Systematic Biology, **56**, 330–337.

Willerslev, E. & Cooper, A. 2005. Ancient DNA. Proceedings of the Royal Society B: Biological Sciences, **272**, 3–16.

Williams, D. M. 1996. Characters and cladograms. Taxon, 275–283.

Williams, D. M. & Ebach, M. C. 2005. Drowning by Numbers: Rereading Nelson's 'Nullius in Verba'. The Botanical Review, **71**, 415–430.

Williams, D. M. & Ebach, M. C. 2006. The data matrix. Geodiversitas, **28**, 409–420.

Wilson, E. O. 1965. A consistency test for phylogenies based on contemporaneous species. Systematic Zoology, **14**, 214–220.

Woodward, S. P. 1855. On the structure and affinities of the Hippuritidae. Proceedings of the geological society, **11**, 40–61.

Wright, A. M. & Hillis, D. M. 2014. Bayesian Analysis Using a Simple Likelihood Model Outperforms Parsimony for Estimation of Phylogeny from Discrete Morphological Data. PLoS One, **9**, e109210.

Yonge, C. M. 1967. Form, habit and evolution in the Chamidae (Bivalvia) with reference to conditions in the rudists (Hippuritacea). Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, **252**, 49–105.

Zakhera, M. S. 2010. Distribution and abundance of rudist bivalves in the Cretaceous platform sequences in Egypt: time and space. Turkish Journal of Earth Sciences, **19**, 745–755.

Zaragüeta i Bagils, R. 2002. Test Morphologique et Moléculaire Des Hypothèses de Phylogénie Des Clupeomorpha (Actinopterygii, Teleostei). Muséum national d'histoire naturelle, 534 p.

Zaragüeta i Bagils, R. 2011. Ancêtres. In: Malécot, V., Léger, N. & Tassy, P. (eds) Biosystema 28 : L'arbre Du Vivant Existe-Il? Paris, Société Française de Systématique, 41–62.

Zaragüeta i Bagils, R. & Bourdon, E. 2005. *Pithecanthropus erectus*. Science et Avenir, Hors-série, 142, 58–63.

Zaragüeta i Bagils, R. & Bourdon, E. 2007. Three-item analysis: Hierarchical representation and treatment of missing and inapplicable data. Comptes Rendus Palevol, **6**, 527–534.

Zaragüeta i Bagils, R. & Lelièvre, H. 2001. Âges et noms versus caractères et phylogénie: le problème de l'ajustement entre la stratigraphie et la phylogénie. In: Tassy, P. & De Ricqlès, A. (eds) Biosystema 19 : Systématique et Paléontologie. Paris, Société Française de Systématique, 169–185.

Zaragüeta i Bagils, R. & Pécaud, S. 2016. Why should cladograms be dichotomous? In: Williams, D., Schmitt, M. & Wheeler, Q. (eds) The Future of Phylogenetic Systematics: The Legacy of Willi Hennig. Cambridge, Systematics Association Special Volume Series, 230–257.

Zaragüeta i Bagils, R., Lelièvre, H. & Tassy, P. 2004. Temporal paralogy, cladograms, and the quality of the fossil record. Geodiversitas, **26**, 381–389.

Zaragüeta i Bagils, R., Ung, V., Grand, A., Vignes-Lebbe, R., Cao, N. & Ducasse, J. 2012. LisBeth: New cladistics for phylogenetics and biogeography. Comptes Rendus Palevol, **11**, 563–566.

Zhang, C., Stadler, T., Klopfstein, S., Heath, T. A. & Ronquist, F. 2016. Total-Evidence Dating under the Fossilized Birth–Death Process. Systematic Biology, **65**, 228–249.

Zwicky, F. 1937. Nebulae as Gravitational lenses. Physical Review, **51**, 290.

Lexique

Partie 1 : Analyse cladistique

Analyse à trois aires – (anglais : « *Three-area analysis* ») Analyse de 3ia appliquée à la discipline biogéographique. Les objets sont des Unités Biogéographiques (UB) et le but est de reconstruire des Aires Hypothétiques (TH).

Analyse à trois éléments – (anglais : « *Three-item analysis* ») Méthode d'analyse cladistique en quatre étapes : (1) proposition de relations de parenté entre parties d'objets (relation hiérarchique), (2) standardisation des parties d'objets en objets, (3) décomposition des relations de parenté en triplets et assignation d'une pondération fractionnaire à chaque triplet, (4) choix du cladogramme contenant le poids maximal de triplets.

Analyse à trois taxons – (anglais : « *Three-taxon analysis* ») Analyse de 3ia appliquée à la discipline phylogénétique. Les objets sont des Unités Taxonomiques (UT) et le but est de reconstruire des Taxons Hypothétiques (TH).

Analyse de Parcimonie de Brooks – (anglais : « *Brooks Parsimony Analysis* ») Analyse de parcimonie appliquée à la discipline biogéographique. Les objets sont des Unités Taxonomiques (UT) et le but est de reconstruire des Taxons Hypothétiques (TH).

Aire Hypothétique – Dans une analyse biogéographique donnée, toute aire résultant de l'analyse. Clade d'Unités biogéographiques (UB).

Aire biogéographique – Synonyme d'aire d'endémisme. Se rapporte aux AH et aux UB sans distinction.

Analyse en sous-arbres libres de paralogie – (anglais : « *Free-paralogy subtree analysis* »). Décomposition d'un caractère comportant un ou plusieurs nœuds paralogues en un ensemble de sous-arbres dépourvus de paralogues.

Apomorphie – Homologue informatif caractérisant un clade. Se rapporte aux autapomorphies (une Unité Taxonomique Minimale) et aux synapomorphies (plusieurs Unités Taxonomiques).

Aréagramme – Cladogramme d'unités biogéographiques (UB).

Autapomorphie – (étymologie : « *autó»* propre, « *apo »* dérivé de, « *morphos* » forme). Homologue porté par une unique Unité Taxonomique Minimale et permettant de caractériser celle-ci.

Biogéographie cladistique – Science de l'étude de l'histoire des aires biogéographiques.

Branche – Élément d'un arbre hiérarchique spécifiant une relation d'inclusion entre deux nœuds.

Caractère – (étymologie : « *kharakter* » signe, empreinte). Hiérarchie d'états de caractères. Un caractère phylogénétique définit des relations de parenté entre Unités Taxonomiques. Un caractère biogéographique définit des relations de parenté entre Unités Biogéographiques.

Clade – Groupe monophylétique. Dans l'ontologie cladistique, seul type de groupe réel.

Cladogramme – Hiérarchie de clades.

Classe – Synonyme de nœud. Regroupement d'un ou de plusieurs objets. Une classe peut être la racine, un singleton ou une classe informative.

Consensus – Arbre(s) dont l'objectif est de synthétiser l'information provenant de plusieurs cladogrammes équi-optimaux d'une même analyse.

État de caractère – Nœud interne d'un caractère. Hypothèse de clade d'Unités Taxonomiques déduite d'un homologue après standardisation.

Extra-groupe – Unité Taxonomique dont on sait *a priori* de l'analyse cladistique qu'elle porte une partie se branchant directement sur l'homologue ancestral. Une Unité Taxonomique extra-groupe permet de polariser une homologie.

Feuille – Synonyme de nœud externe. Nœud terminal d'un cladogramme, toujours relié à un unique nœud interne. Spécifie à la fois un objet (Unité Taxonomique) et son singleton.

Groupe – Synonyme de nœud interne. Ensemble d'au minimum deux objets.

Groupe-frère – Nœud le plus proche parent d'un autre nœud.

Hiérarchie cladistique – Classification spécifiant des relations de parenté entre objets. Composée d'une racine, de singletons et d'au moins une classe informative. S'applique indistinctement à : homologie, caractère, cladogramme, triplet.

Holomorphe – (étymologie : « *hólos»* entier, « *morphé* » forme) Organisme à un instant donné, portant des attributs morphologiques, anatomiques, comportementaux, biochimiques, etc.

Homologie – (étymologie : « *homós* » semblable, pareil, « *lógos* » étude) Hiérarchie cladistique spécifiant des relations de degré de parenté entre parties d'Unités Taxonomiques (UT). Une homologie est une hiérarchie d'homologues.

Homologue – Nœud interne d'une homologie. Hypothèse de regroupement de parties d'Unités Taxonomiques (UT).

Homologue ancestral – Racine d'une homologie.

Homologue dérivé – Synonyme d'homologue informatif. Tout homologue qui n'est pas la racine.

Homonomie – Relation de parenté entre parties d'une même Unité Taxonomique.

Homoplasie – (étymologie : « *homós* » semblable, pareil, « *plásis* », façonner, modeler) Etat de caractère révélé paraphylétique *a posteriori* d'une analyse cladistique. Caractérise les états rejetés du cladogramme par la mise en congruence.

Horogramme – Cladogramme de temps. Hiérarchie de PC. Définit le temps cladistique, ou chaque période se différencie en la période suivante. Un horogramme est toujours une orthologie.

Indice de Concordance Hiérarchique – (anglais : « *Hierarchical fit Index* ») Mesure l'adéquation entre un cladogramme et un horogramme.

Informatif – S'applique aux homologues et aux états. Se dit d'un groupe d'objets qui n'est ni la racine ni un singleton. Permet de postuler une relation de degré de parenté.

Indice de Rétention – (anglais : « *Retention Index* ») Mesure la proportion de triplets d'un caractère retenus dans le cladogramme.

Indice de Rétention Inter-Arbres – (anglais : « *Inter-Tree Retention Index* ») Mesure la proportion de triplets d'une hiérarchie contenue dans une autre hiérarchie.

Indice de Rétention Nodal – Mesure la proportion de triplets de l'ensemble des caractères contenue dans un nœud informatif d'un cladogramme.

Méthode d'analyse cladistique – Méthode dont le but est la construction d'un cladogramme d'objets en proposant des relations de parenté entre les parties de ces objets.

Monophylétique – (étymologie : « *mónos* », unique, « *phŷlon* » groupe, tribu, race) S'applique à un groupe, défini comme complet en termes de parenté. Les membres du groupe sont tous plus proches parents entre eux que par rapport aux non-membres.

Nœud – Synonyme de classe.

Nœud externe –Synonyme de singleton.

Nœud interne –Nœud incluant au minimum deux nœuds. Dans une analyse donnée, synonyme de groupe.

Objet (cladistique) – Dans une analyse cladistique donnée, tout élément postulé *a priori* de l'analyse dont on cherche les relations de parenté par rapport à d'autres objets. S'applique aux Unités Taxonomiques.

Orthologie –Synonyme d'anagenèse, lignée anagénétique. Hiérarchie cladistique dépourvue de nœuds paralogues.

Orthologue – (étymologie : « *orthós* » droit, « *lógos* » étude). Nœud interne incluant uniquement un nœud interne.

Paralogie temporelle – Phénomène de duplication d'une hiérarchie temporelle à un paralogue d'un cladogramme. A partir du nœud paralogue se diversifient des lignées indépendantes possédant chacun une hiérarchie temporelle qui lui est propre.

Paralogue – (étymologie « *pará* » à côté de, « *lógos* » étude) Nœud interne incluant directement au minimum deux nœuds internes. Clade duquel partent plusieurs clades indépendants les uns des autres. Instance de diversification.

Paraphylétique – (étymologie : « *pará* », à côté de, « *phŷlon* » groupe, tribu, race) S'applique à un groupe, défini comme incomplet en termes de parenté. Il manque des Unités Taxonomiques à un groupe paraphylétique pour être monophylétique. Groupe non-monophylétique. Groupe artificiel.

Parcimonie – (anglais : « *Parsimony* », « *Maximum of Parsimony* ») Méthode d'analyse cladistique en quatre étapes : (1) proposition de relations de parenté entre parties d'objets au sein d'une partition, (2) standardisation des parties d'objets en objets, (3) choix de l'arbre non-enraciné contenant le minimum de distorsion des partitions (minimum de pas de transformation), (4) polarisation de l'arbre non-enraciné en hiérarchie.

Partition – Classification spécifiant des relations de parenté entre objets. Composée de classes mutuellement exclusives. Une partition représente un caractère en parcimonie. La polarisation permet de proposer un sens à la partition qui devient une partition ordonnée.

Pas de transformation – Passage d'un état à un autre en parcimonie. Résulte du plaquage du caractère-partition sur une hiérarchie.

Période Cladistique – Clade d'un horogramme. Classe de temps monophylétique définie par un début mais pas par une fin.

Phylogénie – (étymologie : « *phŷlon* » groupe, tribu, race, « *géneia* » qui engendre) Cladogramme d'Unités Taxonomiques.

Polarisation – Action de spécifier l'emboîtement des homologues afin de former des homologies.

Polyphylétique – (étymologie : « *pará* », à côté de, « *phŷlon* » groupe, tribu, race) S'applique à un groupe, défini comme incomplet en termes de parenté. Les taxons d'un groupe polyphylétique ne peuvent être rassemblés en un groupe monophylétique; ils se répartissent dans plusieurs groupes monophylétiques (au moins deux). Groupe non-monophylétique. Groupe artificiel.

Pondération fractionnaire – (anglais : « *Fractionnal weighting* ») Poids associé à un triplet permettant de le rendre indépendant par suppression de redondances.

Pseudo-caractère – Résultat de la standardisation d'une homologie ou certaines Unités Taxonomiques se trouvent *in fine* répétées par homonomie. La suppression des Unités Taxonomiques répétées transforme le pseudo-caractère en caractère.

Pseudo-état – Classe d'un pseudo-caractère.

Racine – Étant donné une hiérarchie, classe incluant tous les objets de la hiérarchie. Classe la plus inclusive, non-informative.

Sémaphoronte – (étymologie : « *sêma* » signe, « *phore* » porteur, « *ontos* » ce qui est) Ensemble d'holomorphes d'un même organisme. Organisme à travers tous les changements au cours de son cycle de vie.

Singleton – Classe minimale incluant un unique objet. Classe nécessaire de la hiérarchie cladistique, non-informative.

Standardisation – Étape de l'analyse cladistique ou les parties d'objets sont remplacées par les objets eux-mêmes. Passage de l'homologie au caractère.

Synapomorphie – (étymologie « *sún* » commun, « *apo* » dérivé de, « *morphos* » forme) Homologue informatif caractérisant un clade.

Symplésiomorphie – Homologue racine d'une homologie.

Systématique – Discipline de la classification des taxons.

Taxon – (étymologie : « *taxis* » ordre). Groupe d'organismes fondé sur au moins une propriété. S'applique indistinctement aux unités taxonomiques (UT) et aux taxons hypothétiques (TH).

Taxon Hypothétique – (anglais : « *Hypothetical Taxa* ») Dans une analyse phylogénétique donnée, tout taxon résultant de l'analyse. Clade assemblant des Unités Taxinomiques (UT).

Théorie cladistique – Théorie des clades et des cladogrammes. Définit une méthode permettant de construire des cladogrammes, la méthode cladistique.

Taxon à Large Répartition – (anglais : « *Multiple Area Single Taxa* ») Dans une analyse biogéographique donnée, quand une UT est portée par plusieurs UB. Engendre un problème lors de la standardisation.

Triplet – Relation de parenté entre trois objets définissant que deux sont plus apparentés entre eux, par rapport à un troisième. Hiérarchie cladistique minimale. Unité d'information cladistique. Structure en laquelle sont décomposés les caractères dans l'analyse en 3iA.

Triplet primaire – Triplet généré par la décomposition des caractères.

Triplet n-aire – Triplet généré par la combinaison d'autres triplets.

Unité Biogéographique – Dans une analyse biogéographique donnée, toute aire biogéographique postulé a priori de l'analyse et considéré comme objet de l'analyse.

Unité Taxonomique – Dans une analyse phylogénétique donnée, tout taxon postulé a priori de l'analyse et considéré comme objet de l'analyse.

Unité Taxonomique Minimale – Tout taxon dans lequel il est impossible de définir un sous-taxon à l'intérieur. Relatif à un état de connaissance donné.

Part 2 : Rudist morphoanatomy

Accessory cavity – Descriptive term applying to various shell cavities, irrelative to homologies. Includes ectomyophoral cavities (cavities lying on the outer side of a myophore) and endomyophoral cavities (cavities lying on the inner side of a myophore). Not employed in this work except for discussion of previous works.

Anterior myophore – Attachment structure for the animal muscle, that can be horizontal (shelf), tilted or vertical (protruding or not into the opposite valve). Recognizable by its anterior position and sometimes by muscle scars. There is an anterior myophore in each valve. In Hippuritidae, the anterior myophore is stretched and divided in two daughter myophores.

Anterior tooth – Protrusion of shell wall and inserted into a cavity of the opposite valve. Located in the anterior side shell, it is generally as large as the posterior tooth, or larger.

Anterior tooth socket – Cavity where the anterior tooth is inserted. The socket lies anteriorly to the central tooth. It can be filled with radial blades. Contact with the posterior tooth socket are possible in case of elongation on the ventral side of the shell.

Blade – Aragonitic shell thickening protruding out of the commissural plane and faces a cavity wall of the other valve. Generally carrying a myophore. When thick, the blade can have a tooth-like morphology.

Calcitic canals – Radial canals in the outer shell layer of the left valve, communicating with the outside by pores and communicating with the inside by opening on the shell margin.

Central tooth – Protrusion of shell wall between anterior tooth socket and posterior tooth socket. Can be highly regressed, as in Radiolitidae. Can surround the anterior tooth socket, as in Diceratidae.

Central tooth socket – Cavity in which the central tooth fits. The cavity lies between anterior tooth, posterior tooth and general cavity. It can be filled with radial blades, and be in contact with the posterior accessory cavity on the ventral side of the shell.

Commissural plane – Plane between the left valve and the right valve, passing through the margins of both shells in transversal view. Protrusion is a term applied to any structure passing through the commissural plane, such as teeth or myophoral blades.

Costulations – Close little outgrowths of the outer calcitic shell layer all around the shell surface.

Dentition – Group of all teeth and sockets of a shell. In most rudists, dentition includes the anterior, posterior (left valve), and central teeth (right valve) and their associated sockets in their corresponding valves. In *Diceras*, a tooth in the left valve is missing and a tooth in the right valve is surnumerary. Some supplementary teeth are found in *Monopleura aliena* or in some Antillocaprinidae.

Ectomyophoral cavity – See accessory cavity.

Endomyophoral cavity – See accessory cavity.

Flange(s) – Longitudinal outgrowths of the inner shell layer that can reach important sizes. Highly variable in number and development.

General cavity – General cavity that occupied the mass of the organism. It is usually the largest cavity of the shell.

Inner layer – Aragonitic internal layer of the shell.

Left valve – Left valve, free valve or superior valve, that can be developed to operculae. Can be coiled or not.

Ligamentary cavity – Longitudinal cavity where the ligament is inserted.

Ligamentary groove – Small inflexion of the shell allowing insertion of the ligament, and oriented toward the exterior of the shell.

Longitudinal section – Section of a rudist fossil perpendicular to the commissural plane.

Margins – Flat zone all around the shell where the left valve and the right valve are in contact.

Myophores – Components of the myophoral apparatus. Composed of four myophores (six in Hippuritidae): an anterior and a posterior, on each valve. The myophores are attachment structures for the muscle, with a specific microstructure.

Myocardinal apparatus – Ensemble of dentition and myophores of a shell.

Outer layer – Calcitic external layer of the shell.

Pallial canals of the LV – Canals resulting of the posterior accessory cavity and/or the central tooth socket filled by radial walls.

Pallial canals of the RV – Canals resulting of the anterior tooth socket and/or the posterior tooth socket filled by radial walls.

Pores – Numerous small holes of various morphologies allowing the calcitic canals of the left valve to communicate with the shell outside.

Posterior accessory cavity – Cavity developed between the posterior myophore of the left valve and the posterior side of the shell. It can be filled with radial blades, and can be in contact with the central tooth socket on the ventral side of the shell.

Posterior myophore – Attachment structure for the muscle of the animal, that can be horizontal (shelf), tilted or vertical (protruding or not). Recognizable by its posterior position and sometimes by its muscle scar. There is a posterior myophore in each valve.

Posterior myophoral cavity – Free cavity in *Ichthyosarcolites* and *Oryxia* that surrounds the posterior myophore of the right valve in *Titanosarcolites* and *Trechmanellidae*.

Posterior tooth – Protrusion of shell wall on the posterior side present in all rudists attached with their left valve except *Diceras*.

Posterior tooth socket – Cavity where the posterior tooth is inserted. The cavity lies posteriorly to the central tooth in the so-called inverse rudists. It can be filled with radial blades, and it can be in contact with the anterior tooth socket on the ventral side of the shell.

Pillar – Two posterior radial folds in Hippuritidae that correspond to radial bands in other taxa. The most dorsal fold is the P1, the most ventral fold is the P2.

Radial bands – Two radial zones situated on the outer calcitic shell layer, extended from the umbo to the commissure of either valves or only the RV. They are easily recognizable by a different morphology than the rest of the shell, modifying the ornamentation.

Right valve – Right valve, attaching valve or inferior valve. It is generally the most developed valve. Can be coiled or not.

Tabulae – Skeletal wall partitioning the animal cavities parallel to the commissural plane. The designation of tabulae is followed by the indication of the implied cavity.

Transverse section – Section of a rudist fossil parallel to the commissural plane.

Tubes – Tubular cavities running along the shell and probably produced by anastomosis of flanges. Independent structures from the pallial canals.

Umbo – Apex and oldest part of the shell valves. It is usually coiled.

Wall – The term wall is used to refer to a delimitation between shell cavities, as the CTS/G wall, separating G from the CTS. All rudist cavities can be partitioned with walls. When a cavity is filled with numerous walls, the resulting cavities are generally described as pallial canals (with the reference of the cavity of origin).

Abbréviations et symboles - Abbreviations and symbols

Partie 1: Analyse cladistique

$\{a, b\}$ – ensemble d'objets

(a, b) – classe d'objets

α – holomorphe

α_1 – sémaphoronte, partie de l'holomorphe α

a – feuille, unité taxonomique (UT)

a_x – feuille, unité taxonomique (UT) en tant que porteur de l'état dérivé X

a_p – feuille, partie de l'unité taxonomique UT a

a_h – homologue autapomorphique

A – clade, hypothèse de clade, taxon hypothétique (TH), état de caractère

A_h – homologue synapomorphique

2/3 – pondération d'un triplet

$C(X)$ – clade C caractérisé par un état de caractère X

UB_1 – feuille, unité biogéographique (UB)

Ω – clade, hypothèse de clade, aire hypothétique (AH), état de caractère biogéographique

$(a, b) \rightarrow (c, d)$ – relation d'ordre entre deux classes d'une partition (en parcimonie : polarisation; résulte de l'enracinement)

3aa – (tiré de l'anglais pour « *Three-area analysis* ») Analyse à trois aires.

3ia – (tiré de l'anglais pour « *Three-item analysis* ») Analyse à trois éléments.

3ta – (tiré de l'anglais pour « *Three-taxon analysis* ») Analyse à trois taxons.

APB – (anglais : BPA, « *Brooks Parsimony Analysis* ») Analyse de Parcimonie de Brooks.

AH – Aire Hypothétique.

ICH – (anglais : HiFi, « *Hierarchical fit Index* ») Indice de Concordance Hiérarchique.

IR – (anglais : RI, « *Retention index* ») Indice de Rétention.

IRIA – (anglais : ITRI, « *Inter-Tree Retention Index* ») Indice de Rétention Inter-Arbres.

IRN – (anglais : NRI, « *Nodal Retention Index* ») Indice de Rétention Nodal.

PC – Période Cladistique.

PF – (anglais : FW, « *Fractionnal Weighting* ») Pondération fractionnaire.

TH – (anglais : HT, « *Hypothetical Taxa* ») Taxon Hypothétique.

TLR – (anglais : MAST, « *Multiple Area Single Taxa* ») Taxon à Large Répartition.

UB – Unité Biogéographique.

UT – (anglais : TU, « *Taxonomic Unit* ») Unité Taxonomique.

UTM – (anglais : LITU, « *Least Inclusive Taxonomic Unit* ») Unité Taxonomique Minimale.

Part 2 : Rudist morphoanatomy

AM – Anterior myophore

AT – Anterior tooth

ATS – Anterior tooth socket

CT – Central tooth

CTS – Central tooth socket

D – General cavity of the right valve

G – General cavity of the left valve

GC – General cavity

LV – Left valve

PAC – Posterior accessory cavity

PM – Posterior myophore

PMC – Posterior myophoral cavity

PT – Posterior tooth

PTS – Posterior tooth socket

P1 – posterodorsal pillar

P2 – posteroventral pillar

RV – Right valve

Wall – shell material separating two cavities. For example, the CTS/G wall separates the CTS from G.

Annexes

Annexe 1

Shell references and Landmark coordinates of the general cavity of 23 *Ichthyosarcolites* shells using PhyloNimbus.

Specimen number	Reference	Number of flanges	Valve	Localisation
1		6	LV	Guerrero (Mexico)
2		6	RV	Guerrero (Mexico)
3	UNAM 6970	5	LV	Guerrero (Mexico)
4		4	LV	Guerrero (Mexico)
5	UNAM 6979	4	RV	Guerrero (Mexico)
6	UNAM 6972	4	RV	Guerrero (Mexico)
7	UNAM 6978	4	RV	Guerrero (Mexico)
8		4	RV	Guerrero (Mexico)
9		4	LV	Tamaulipas (Mexico)
10		4	RV	Tamaulipas (Mexico)
11		1	RV	Istria (Croatia)
12		1	RV	Istria (Croatia)
13		1	RV	Istria (Croatia)
14		1	RV	Istria (Croatia)
15		1	RV	Istria (Croatia)
16		1	RV	Istria (Croatia)
17	fig. 88-3 p.96 in Chikhi-Amouimeur, 2010; MPUP 085	3	RV	Jebel-Seloum (Tunisia)
18	ICH1, Masse collection	0	RV	La Bédoule (France)
19	fig. 2-1 p. 71 in Cestari *et al*., 1998; num 796	0	RV	Sicilia (Italia)
20	fig. 2-2 p. 71 in Cestari *et al*., 1998; num 796	0	RV	Sicilia (Italia)
21	fig. 87-4 p.96 in Chikhi-Amouimeur, 2010	2	RV	Aqaba (Jordania)
22		1	RV	île d'Oléron (France)
23		1	RV	île d'Oléron (France)

Specimen	S1X	S1Y	S2X	S2Y	S3X	S3Y	S4X	S4Y	S5X	S5Y	S6X	S6Y	S7X	S7Y	S8X	S8Y	S9X	S9Y	S10X	S10Y
1	1105.81	750.49	1176.46	798.18	1235.77	861.79	1281.95	937.12	1227.83	1016.02	1134.34	1031.79	1042.49	1016.56	997.82	925.93	987.10	831.84	1068.16	761.79
2	976.72	472.85	878.45	376.17	747.56	438.77	630.94	512.38	555.85	630.33	656.97	754.90	777.37	824.73	921.60	823.31	1007.99	676.68	1002.07	529.88
3	1307.51	435.01	1424.53	511.39	1507.51	626.74	1511.87	759.15	1448.61	891.80	1287.33	915.31	1140.98	880.87	1125.94	693.09	1226.66	560.58	1274.33	433.78
4	1249.74	1303.34	1161.27	1325.13	1130.49	1414.56	1071.55	1486.85	1127.19	1572.71	1226.40	1586.32	1322.15	1571.35	1353.07	1467.12	1354.69	1369.57	1266.92	1307.64
5	1094.74	946.41	1271.19	964.49	1375.80	1103.91	1419.84	1277.01	1367.60	1436.97	1193.59	1481.45	1020.45	1434.95	951.17	1260.76	902.00	1101.87	1017.31	955.25
6	1029.98	1393.27	1186.33	1445.26	1239.92	1603.06	1230.80	1776.70	1176.53	1939.24	998.80	1963.14	836.76	1891.18	748.18	1732.39	874.74	1542.90	973.02	1404.66
7	1403.13	875.02	1507.80	974.97	1549.98	1111.75	1561.65	1251.90	1424.35	1309.41	1281.15	1332.53	1181.76	1216.00	1146.50	1073.53	1307.27	974.64	1332.67	851.54
8	555.00	385.42	497.75	282.91	366.08	239.18	228.17	254.57	110.92	329.89	126.60	483.84	165.05	621.41	345.81	578.48	446.28	462.52	555.00	400.83
9	609.96	492.10	539.63	522.96	487.46	588.97	499.40	670.39	558.42	726.87	628.20	774.77	702.87	738.88	728.69	651.81	687.59	563.10	625.32	500.63
10	1666.26	1112.27	1731.29	951.46	1690.91	786.86	1578.51	656.87	1412.53	639.79	1264.85	725.19	1312.65	891.44	1384.40	1049.63	1484.82	1192.19	1647.05	1127.63
11	933.33	314.67	1033.66	455.14	1013.68	625.54	877.83	720.91	732.84	802.17	556.96	735.06	490.21	555.98	577.95	387.70	730.69	290.16	915.56	309.33
12	626.67	204.44	438.60	312.51	422.04	525.05	581.70	678.17	740.01	818.58	959.03	780.87	1122.60	626.45	1073.93	398.20	899.70	239.39	666.67	206.67
13	697.16	197.28	866.16	252.26	839.96	427.57	743.71	579.59	604.72	695.51	450.37	685.78	293.76	597.30	390.29	403.88	505.38	260.92	682.73	195.97
14	338.81	100.75	427.15	125.96	423.55	215.65	376.51	297.99	304.01	360.31	223.11	372.53	133.63	338.69	168.28	229.22	227.66	153.04	314.44	112.13
15	808.89	679.11	930.21	561.98	876.79	398.15	714.86	338.70	550.82	278.42	392.69	398.23	275.57	531.40	454.19	636.37	570.12	776.34	771.56	695.11
16	521.28	720.88	449.89	639.94	335.69	638.56	271.97	732.78	200.35	816.86	287.61	928.77	370.91	1003.61	486.52	974.14	533.62	849.87	523.59	730.14
17	569.09	347.15	701.80	321.17	830.84	280.71	962.26	261.68	977.10	425.17	920.42	548.16	799.81	580.25	665.00	565.54	553.60	487.97	570.05	348.12
18	400.08	229.47	542.62	295.10	589.68	442.50	519.97	579.38	492.78	732.72	306.54	758.83	197.79	610.30	182.83	442.18	185.07	282.44	380.85	230.67
19	303.64	119.61	389.55	149.32	426.99	230.49	399.22	315.09	359.15	398.29	252.52	395.09	175.06	334.37	159.98	238.69	180.76	144.93	296.48	121.99
20	360.21	134.23	498.28	192.93	570.77	321.80	512.27	451.75	407.30	554.52	262.44	596.76	123.12	507.72	120.67	328.36	165.53	179.97	345.73	139.06
21	904.88	261.59	804.40	236.55	695.53	258.60	600.63	308.40	555.92	394.91	609.95	493.40	724.35	510.52	830.09	462.12	925.06	400.88	931.87	280.20
22	1702.40	567.91	1167.70	912.62	981.18	1551.85	991.54	2221.34	1265.74	2836.48	1923.20	2844.43	2182.07	2210.17	2701.92	1735.17	2330.18	980.17	1756.11	576.17
23	3047.98	1030.24	2980.00	1183.82	3072.49	1325.16	3230.44	1294.98	3372.91	1202.69	3476.73	1067.62	3506.15	903.01	3330.04	806.28	3149.08	865.33	3064.46	1006.43

Annexe 2

Lisbeth .3ia file of the phylogenetic analysis.

```
;
Dimensions
  ntax=27 nchar=40
;
Charlist
  [1]  LV-Teeth-1
  [2]  LV-Teeth-2
  [3]  LV-Teeth-3
  [4]  LV-Myophores-1
  [5]  LV-Myophores-2
  [6]  LV-Myophores-3
  [7]  LV-Myophores-4
  [8]  LV-Vertical_blades-1
  [9]  LV-Vertical_blades-2
  [10] LV-Vertical_blades-3
  [11] LV-Canals-1
  [12] LV-Canals-2
  [13] LV-Canals-3
  [14] LV-Canals-4
  [15] LV-Canals-5
  [16] LV-Canals-6
  [17] LV-Horizontal_blades-1
  [18] LV-Horizontal_blades-2
  [19] LV-Horizontal_blades-3
  [20] LV-Pore_system-1
  [21] LV-Pore_system-2
  [22] LV-Coiling
  [23] RV-Teeth-1
  [24] RV-Teeth-2
  [25] RV-Teeth-3
  [26] RV-Myophores-1
  [27] RV-Myophores-2
  [28] RV-Vertical_blades-1
  [29] RV-Vertical_blades-2
  [30] RV-Canals-2
  [31] RV-Canals-3
  [32] RV-Canals-4
  [33] RV-Horizontal_blades-1
  [34] RV-Horizontal_blades-2
  [35] LV+RV-pillars
  [36] LV+RV-outer_layer
  [37] LV+RV-ligament-1
  [38] LV+RV-ligament-2
  [39] LV+RV-ornamentation-1
  [40] LV+RV-ornamentation-2
;
```

```
Taxa
  00 = Diceras_arietinum
  01 = Valletia_tombecki
  02 = Monopleura_michaillensis
  03 = Glossomyophorus_costatus
  04 = Radiolites_radiosus
  05 = Agriopleura_marticensis
  06 = Hippurites_radiosus
  07 = Barrettia_monilifera
  08 = Mexicaprina_minuta
  09 = Amphitriscoelus_waringi
  10 = Pachytraga_paradoxa
  11 = Offneria_rhodanica
  12 = Dictyoptichus_morgani
  13 = Semailia_smithi
  14 = Titanosarcolites_giganteus
  15 = Clinocaprina_lowentsami
  16 = Ichthyosarcolites_alatus
  17 = Curtocaprina_clabaughikinsorum
  18 = Plagioptychus_arnaudi
  19 = Coralliochama_orcutti
  20 = Polyconites_verneuili
  21 = Tepeyacia_corrugata
  22 = Caprotina_striata
  23 = Chaperia_costata
  24 = Caprinula_boissyi
  25 = Sellaea_americana
  26 = Oryxia_sulcata

;
Characters

  [1]  (00,(01,02,03,08,09,10,11,17,20,21,22,23,(04,05,06,07,14,16,18,19,24,25),(12,13,15)))
  [2]  (00,(01,02,03,04,05,06,07,08,09,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25))
  [3]  (00,01,02,03,06,07,08,09,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,(04,05))
  [4]  (00,(01,02,(11,18,19,24,(03,10,22,23,25,(20,21),(04,05,08,09,12,14,15,16,17,(06,07,13))))))
  [5]  (00,01,02,03,05,08,09,10,11,15,16,17,18,19,20,21,22,23,24,25,(04,06,07,12,13,14))
  [6]  (00,(01,02,03,20,22,23,(18,19),(10,11,17,25,(04,05,08,12,13,14,15,16,21,(06,07)))))
  [7]  (00,01,02,03,05,06,07,08,09,10,11,12,15,16,17,18,19,20,21,22,23,24,25,(04,14))
  [8]  (00,(01,02,22,23,25,(04,05,06,07,08,09,10,11,12,13,14,15,16,17,18,19,20,21,24)))
  [9]  (00,01,02,03,05,08,09,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,(04,06))
  [10]  (00,01,02,08,09,10,11,18,19,20,21,22,23,24,25,(03,04,05,06,07,12,13,14,15,16,17))
  [11]  (00,01,02,04,05,06,07,(03,09,10,20,21,22,23,25,18,19,08,11,12,13,14,15,16,17,24))
  [12]  (00,01,02,04,05,06,07,20,03,09,10,21,22,23,25,18,19,(08,11,12,13,14,15,16,17,24))
  [13]  (00,01,02,03,04,05,06,07,20,21,23,25,22,(09,10,12,(08,13,14,15,16,17,24),(11,18,19)))
  [14]  (00,01,02,03,04,05,06,07,09,10,11,12,14,15,18,19,20,21,22,23,25,(08,(13,16,17,24)))
  [15]  (00,01,02,03,04,05,06,07,08,10,11,12,13,14,15,16,17,18,19,20,21,22,23,25,(09,24))
  [16]  (00,01,02,03,04,05,06,07,08,09,10,11,12,16,17,18,20,21,22,23,24,25,(13,14,15,19))
  [17]  (00,01,02,03,04,05,06,07,08,09,10,11,12,13,15,17,18,19,20,21,22,23,24,25,(14,16))
  [18]  (00,01,02,03,04,05,06,07,08,09,10,12,13,15,17,18,19,20,21,22,23,24,25,(11,14,16))
  [19]  (00,01,02,03,04,05,06,08,09,10,12,13,15,17,18,19,20,21,22,23,(11,14,24))
```

[20] (00,01,02,03,04,05,08,09,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,(06,07))
[21] (00,01,02,03,04,05,08,09,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,(06,07))
[22] (00,01,02,03,08,09,10,11,12,13,14,15,16,17,18,19,20,22,23,24,25,(04,05,06,07))
[23] (00,(01,02,03,04,05,06,07,08,09,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26))
[24] (00,01,02,03,09,10,18,20,22,23,25,(04,05,06,07,08,11,12,13,14,15,16,17,19,21,24))
[25] (00,01,02,10,20,(04,05,06,07,08,11,12,13,14,15,16,17,18,19,21,22,23,24,25))
[26] (00,(04,05,17,20,21),(01,02,18,19,(03,22,23,25,(10,11,24)),(06,07,08,09,15,(12,13,14,16,26))))
[27] (00,04,05,06,07,10,12,16,17,21,22,23,26,(01,02,03,11,15,18,19,20,25,(08,14)))
[28] (00,01,02,03,04,05,06,07,18,19,20,21,22,23,(08,09,10,11,12,13,14,15,16,17,24,25,26))
[29] (00,04,(01,02,03,05,06,07,18,19,20,21,26,(08,09,10,11,12,13,14,15,16,17,22,23,24,25)))
[30] (00,01,02,03,04,05,06,07,09,10,18,19,20,21,22,23,25,(08,11,12,13,14,15,16,17,24,26))
[31] (00,01,02,03,04,06,07,18,19,20,21,22,23,25,(05,09,10,11,12,17,(08,13,14,15,16,24,26)))
[32] (00,01,02,03,04,05,06,07,08,09,10,11,12,13,15,19,18,20,21,22,23,24,25,(14,17,16,26))
[33] (00,01,02,03,05,08,09,10,11,13,17,18,19,20,21,22,23,24,25,(04,06,07,12,14,15,16,26))
[34] (00,01,02,03,04,05,06,07,08,09,10,13,15,17,18,19,20,21,22,23,24,25,26,(11,12,14,16))
[35] (00,01,02,03,04,05,08,09,10,11,12,13,15,16,17,18,19,20,22,23,24,25,26,(06,07,21))
[36] (00,01,02,03,07,08,09,10,11,13,14,15,16,17,18,19,22,23,24,25,26,(04,05,06,12,20,21))
[37] (00,01,(02,03,04,05,06,07,08,09,10,11,18,19,20,21,22,23,24,25,26,(12,13,15,(14,16,17))))
[38] (00,01,02,03,05,07,08,09,10,11,12,13,15,16,17,18,19,20,21,22,23,24,25,(04,06,26))
[39] (00,01,15,16,17,26,(02,03,04,05,06,07,08,09,10,11,12,13,14,18,19,20,21,22,23,25))
[40] (00,01,02,03,04,05,06,07,08,09,10,11,12,15,18,19,20,21,22,23,24,25,(13,14,16,17,26))

Table des illustrations

Chapitre 1

Figure 1.1. Représentation hiérarchique de la différenciation d'un être humain au cours du temps. Le vieillard se différencie à partir de l'adulte qui se différencie à partir de l'enfant. Il s'ensuit qu'un vieillard est un adulte car ce qui constitue l'adulte existe encore chez le vieillard, sous une forme ou une autre, et que l'adulte est un enfant. 16

Figure 1.2. Représentation sur une même classification hiérarchique d'un groupe monophylétique et d'un groupe paraphylétique (en grisé). Le groupe de gauche est monophylétique car il est complet en termes de parenté : *a* et *b* partagent un degré de parenté exclusif. Le groupe de droite est paraphylétique car il est incomplet en termes de parenté : *b* et *c* partagent un degré de parenté non-exclusive, *a* possédant le même degré de parenté mais étant exclu du groupe.....19

Figure 1.3. Deux classifications hiérarchiques construites sur le modèle de la transformation (à gauche) et de la différenciation (à droite). La hiérarchie de gauche montre la transformation de e en d et c, puis la transformation de *d* en *a* et *b*. Tous les taxons *a*, *b*, *c*, *d* et *e* sont mutuellement exclusifs, un organisme ne peut appartenir qu'à un seul de ces taxons. La hiérarchie de droite montre la différenciation de {*a*, *b*, *c*} en {*a*, *b*} et {*c*}, et la différenciation de {*a*, *b*} en {*a*} et {*b*}. Deux taxons sont soit mutuellement exclusifs ({*a*, *b*} et {*c*}), soit liés par une relation d'inclusion ({*a*} est inclus dans {*a*, *b*, *c*}). {*d*} et {*e*} sont des TH paraphylétiques, {*a*, *b*} et {*a*, *b*, *c*} sont des TH monophylétiques.....21

Figure 1.4. Deux représentations de la même information. En haut, un diagramme d'Euler, en bas, un arbre enraciné. Les deux représentations montrent une hiérarchie de type différenciation, avec trois objets à classer *a*, *b* et *c*, et une classification composée des deux classes {*a*, *b*, *c*} et {*a*, *b*}. La classification montre une relation de parenté entre a et b par rapport à c. Elle peut se représenter de manière parenthésée (*c,* (*a, b*)). On notera par la suite ce type de classe par une majuscule. 23

Figure 1.5 En haut. Processus général d'analyse/synthèse. Le problème général complexe est décomposé en problèmes partiels plus simples. Chaque problème partiel est résolu par une solution partielle argumentée (args.). La synthèse de toutes les solutions partielles donne une solution générale. En bas. Processus d'analyse/synthèse dans le cadre de la méthode cladistique. Le problème général est de construire un cladogramme sur un ensemble d'objets, ici {*a*, *b*, *c*, *d*}. Ce problème général est décomposé en problèmes partiels consistant à rechercher les relations de parenté entre les parties des objets de l'ensemble (au moins trois). Chaque solution donne une relation de degré de parenté ou deux parties sont plus apparentées entre elles que par rapport à une troisième. La combinaison de toutes ces solutions partielles donne la solution générale : le cladogramme. 24

Figure 1.6. Schéma général de la méthode d'analyse cladistique en 3ia. Les homologies sont des structures établissant des relations entre parties d'UT. Les parties d'UT sont ensuite remplacées par les UT elles-mêmes lors de l'étape de standardisation pour former des caractères. Les caractères composés de relations entre UT sont décomposés en triplets, les relations minimales entre UT. Le cladogramme est la solution générale au problème des relations de parenté entre UT. Il est construit comme l'ensemble maximisant le nombre de triplets congruents. 30

Figure 1.7. Composition d'une UTM. α_1 est un sémaphoronte. α_1, α_2 et α_3 sont les différents stades de vie de l'holomorphe α. Aucune relation de degré de parenté ne peut être établie entre α, β, γ, δ et ε car il n'y a aucun homologue que certains posséderaient et pas d'autres. *a* est une UTM qui rassemble α, β, γ, δ et ε par la présence de l'homologue X_h. α, β, γ, δ et ε sont donc plus proches parents entre eux que de tout autre holomorphe. Une UTM comme toute autre classe peut se définir en intension (X_h) ou en extension {α, β, γ, δ, ε}. 32

Figure 1.8. Etapes de construction d'une homologie à deux homologues. (1) On procède à la comparaison de quatre parties *p* de quatre UT *a*, *b*, *c* et *d*. Lors de la comparaison, on remarque des ressemblances entre a_p et b_p, entre b_p et c_p et entre a_p et c_p (flèches noires). (2) Ces ressemblances sont généralisées en un homologue Y_h incluant a_p, b_p et c_p. Les étapes (1) et (2) se répètent jusqu'à ce qu'on ne puisse plus générer d'homologue. (3) On ajoute un homologue racine X_h qui inclut toutes les parties et un

singleton pour chacune des parties. L'ensemble des singletons de parties a_p, b_p, c_p et d_p et les homologues X_h et Y_h forment une homologie. 37

Figure 1.9. Quatre homologies différentes à partir du même ensemble a_p, b_p, c_p et d_p. L'homologie 1 présente un homologue U_h (a_p, b_p). L'homologie 2 présente également un homologue V_h (a_p, b_p). d_p est manquant et ne peut pas être observé. L'homologie 3 présente un homologue X_h (a_p, b_p) inclus dans un homologue W_h (a_p, b_p, c_d). X_h et Y_h incluent moins de deux nœuds internes, ils sont orthologues. L'homologie 4 présente un homologue Y_h (a_p, b_p) et un homologue Z_h (c_p, d_p) qui sont disjoints. Ici, la racine est paralogue : elle contient deux nœuds internes, Y_h et Z_h, qui sont indépendants. 38

Figure 1.10. Illustration des nœuds orthologues et paralogues. (1) Orthologie, tous les nœuds internes sont en relation d'inclusion les uns par rapport aux autres. Aucune diversification ne peut être postulée. (2) Hiérarchie présentant un nœud paralogue *X* d'où partent deux lignées d'orthologues *G* et *M* indépendantes l'une de l'autre. 40

Figure 1.11. Illustration tirée de Belon, 1555 : *Histoire de la nature des oyseaux, avec leurs descriptions et naïfs portraicts retirez du naturel* (p. 40-41). L'affinité entre l'humain et l'oiseau est mise en évidence par l'utilisation du même symbole pour une partie de l'un et de l'autre. Le squelette de l'oiseau est modifié pour mettre en évidence l'identité des parties avec l'humain. 42

Figure 1.12. Illustration du problème de polarisation. La comparaison permet d'identifier deux ensembles {a_p, b_p} et {c_p, d_p}. Il y a trois possibilités d'homologies. La première possibilité (gauche) est de considérer que X_h {a_p, b_p} est le seul homologue informatif, et que des arguments ont été produits impliquant l'extension de la similarité de c_p, d_p à a_p et b_p. La seconde possibilité (centre) est de considérer cette fois que c'est Y_h {c_p, d_p} qui est l'homologue dérivé, a_p et b_p faisant partie d'un homologue ancestral que possèdent finalement également c_p et d_p. La troisième solution est de construire deux homologues dérivés X_h {a_p, b_p} et Y_h {c_p, d_p} à partir de l'homologue racine Z_h. 45

Figure 1.13. Standardisation d'homologies en caractères. En haut, correspondance entre parties d'UT a_p, b_p, c_p, d_p et UT *a*, *b*, *c*, *d*. En bas, les homologies sont composées de parties d'UT. Les parties sont remplacées par leur UT correspondante et forment les caractères. (1) Exemple de standardisation parfaite ou chaque partie correspond

exclusivement à une UT. La standardisation ici ne fait perdre aucune information. (2) Exemple de standardisation non informative, ou homonomie. Toutes les parties correspondent à la même UT. Le caractère résultant de la standardisation n'apporte aucune information sur les relations de parenté entre UT. 47

Figure 1.14. Standardisation d'homologies en caractères. En haut, correspondance entre parties d'UT (a_p, b_p, c_p, d_p) et UT (a, b, c,). En bas, l'homologie standardisée en caractère. L'UT b est polymorphique. Elle porte à la fois l'état X et Y. X étant inclut dans Y, toute feuille incluse dans X est également incluse dans Y. Dans ce cas, on peut supprimer la feuille b incluse dans Y, redondante avec l'autre instance de b. On obtient le caractère (d, (a, b)) (en bas). 49

Figure 1.15. Analyse en sous-arbres libres de paralogie sur l'arbre 1. A partir du nœud paralogue X partent deux orthologies. L'analyse produit donc deux sous-arbres dépourvus de nœuds paralogues. Chaque sous-arbre contient également les feuilles appartenant à l'autre orthologie qui n'a pas été conservée. S'il y a du polymorphisme, celui-ci peut être résolu facilement à cause de la structure orthologique des sous-arbres. 51

Figure 1.16. Analyse en sous-arbres libres de paralogie sur l'arbre 1. Trois sous-arbres 2a, 2b et 2c sont générés. 2a et b sont générés à partir des orthologies branchées au paralogue X. 2c est généré à partir de l'arbre source ou tous les nœuds internes inclus dans le paralogue X (i à r) sont supprimés. Puisque tous les nœuds internes du nœud X sont supprimés, le nœud X devient orthologue. 52

Figure 1.17. Illustration de Fitch 1970 et traitement par l'analyse en sous-arbres libres de paralogie. (1) L'arbre en noir et rouge correspond à l'homologie sur les parties a_M, b_α, c_α, b_β et c_β, l'arbre en gris correspond aux relations entre les UT a, b et c. Cette représentation met en évidence la diversification des parties (paralogue) sans diversification du tout. (2) Représentation en 3ia : (2a) homologie, (2b) caractère, (2c) sous-caractères libres de paralogie, (2d) suppression des UT redondants. Les deux premiers arbres sont non-informatifs (seulement deux UT). Le résultat (a, (b, c))) correspond au schéma (1). 54

Figure 1.18. Représentation de différents types de relations entre deux triplets. La dépendance est un cas particulier de combinabilité qui est un cas particulier de

compatibilité. Des exemples montrent la combinaison de deux triplets (sauf pour les relations simplement compatibles). Les quatre types de combinaison sont détaillés : la combinaison *in-in* et les trois combinaisons qui impliquent l'émergence de nouveaux triplets dépendants (dépendance), combinaison *in-ex*, *paralogue*, *orthologue*. Les points noirs signalent les feuilles communes au couple de triplets 57

Figure 1.19. Quatre types de combinaison de couples de triplets en fonction de la position de leurs feuilles communes (notée avec un point noir). La disposition de haut en bas est : triplet primaire *x* + triplet primaire *y* = arbre *Z* = triplet secondaire *1* + triplet secondaire *2*. La combinaison *n*ts *in-in* ne produit aucun triplet. La combinaison *n*ts *in-ex* produit un triplet supplémentaire. La combinaison *in-in* et *in-ex* produisent un arbre *Z* avec un seul nœud interne. La combinaison paralogue produit un arbre *Z* dont la racine est paralogue. La combinaison orthologue produit un arbre *Z* qui est une orthologie. Les combinaisons paralogues et orthologues génèrent deux triplets supplémentaires. 58

Figure 1.20. Trois triplets primaires (cercles simples) décomposés à partir du caractère (*a*, (*b*, *c*, *d*)) et les triplets secondaires déductibles (cercles doubles) de la combinaison de couples dépendants. Chaque triplet secondaire est relié par un trait noir à ses triplets parents. Les flèches grisent montrent la répétition du même triplet, entraînant la redondance. Il suffit de deux triplets primaires parmi les trois pour reconstruire le caractère initial. 60

Figure 1.21. Calcul des PF pour le caractère (*a*, (*b*, (*c*, *d*))) en deux étapes. (1- redondance intra-état, noir) Chaque état de caractère (*X* et *Y*) est décomposé en triplets (3 et 2). La PF des triplets de l'état *X* vaut 2/3, celle des triplets de *Y* vaut 1. (2- redondance inter-états, violet) Le triplet (*a*, (*c*, *d*)) est présent dans les deux états. Sa suppression de *x* permet de lever l'ambigüité de la PF de 2/3 des triplets *X* (« *choisir deux parmi trois* »).62

Figure 1.22. Calcul des PF pour le caractère multi-états (*a*, (*b*, (*c*, (*d*, *e*)))) en deux étapes. (1- redondance intra état, noir) Chaque état de caractère (*X*, *Y* et *Z*) est décomposé en triplets (6, 6 et 3). La PF des triplets de *X* vaut 1/2, celle des triplets de *y* vaut 2/3 et celle de *y* 1. (2- redondance inter-états, violet) On vérifie la redondance par couples d'états, des états les plus inclusifs aux moins inclusifs. On supprime les triplets redondants de *X* par

rapport à *Y*, puis ceux de *Y* par rapport à *Z*. Le caractère étant une orthologie, toute ambiguïté est supprimée (toutes les PF=1). 64

Figure 1.23. Illustration du concept d'information cladistique. Les caractères sont représentés par leurs nœuds internes, les chiffres correspondent aux feuilles qui s'insèrent à cet endroit. Le caractère le plus à gauche est l'entité informationnelle minimale : son poids est de 1. Les deux caractères du milieu ont un poids de 998 : il suffit de se tromper sur l'observation de 2 feuilles pour que le caractère soit faux. Le caractère de droite est le plus équilibré. Il faudrait se tromper sur l'observation de 500 feuilles pour que le caractère soit faux. Son poids de 249 500 souligne le degré de confiance que l'on peut lui accorder. 65

Figure 1.24. Illustration de l'incompatibilité secondaire entre triplets. Trois triplets primaires (cercles simples) sont compatibles deux à deux. Chaque couple est lié par de la dépendance et génèrent des triplets secondaires (cercles doubles). Certains triplets secondaires sont incompatibles (flèches rouges) avec les triplets primaires, rendant impossible la combinaison des trois triplets. 69

Figure 1.25. Exemple de construction des cladogrammes optimaux par la méthode du bas vers le haut pour l'ensemble de caractères {(*a*, *b*, *c*, (*d*, *e*)), (*a*, (*d*, *e*)), (*a*, (*c*, *d*)), (*e*, (*a*, *c*))}. (1) Liste des triplets primaires issus des caractères et leur PF associée. (2) déduction des triplets secondaires. (3) Liste des combinaisons de triplets incompatibles (si un triplet secondaire est présent dans une combinaison, il est remplacé par l'ensemble de ses triplets parents, entre crochets dans l'étape 2). (4) Arbre de décision de l'ensemble de triplets de poids minimal à supprimer pour faire disparaître toutes les relations d'incompatibilité. Pour une branche, on ne compte qu'une fois un même triplet. (5) deux solutions apparaissent : supprimer (*a*, (*c*, *d*)), et supprimer (*e*, (*a*, *c*)). Les deux listes de triplets optimales conduisent à 16 cladogrammes optimaux (6). 71

Figure 1.26. Graphe montrant le nombre d'arbres, de composantes et de triplets possibles (ordonnée, logarithmique) par rapport au nombre de taxons de l'analyse (abscisse, linéaire). 74

Figure 1.27. Quatre morphologies possibles de la caractéristique *X*. Les taxons a et b ne portent pas la caractéristique *X* et ne sont donc concernés par aucune de ses variantes.

c et *d* possèdent la morphologie 1 (*X* rond et noir); e et f possèdent la morphologie 2 (*X* rond et strié); *g* et *h* possèdent la morphologie 3 (*X* carré et noir); et *i* et *j* possèdent la morphologie 4 (*X* carré et strié). 75

Figure 1.28. Propositions de Pleijel, réinterprétées en caractères hiérarchiques. La proposition 1 présente cinq caractères binaires (interprétation du codage D de Pleijel 1995 en hiérarchies) : le premier rassemble les UT par la présence de la caractéristique *X* (1.a); le second par la forme ronde (1.b); le troisième par la forme carrée (1.c); le quatrième par la pigmentation noire (1.d); et le dernier par la pigmentation striée (1.e). Cette proposition vaut 86 triplets pondérés. La proposition 2 présente trois caractères distincts (interprétation du codage A de Pleijel 1995 en hiérarchies) : le premier rassemble par la présence de la caractéristique *X* (2.a); le second rassemble selon la forme et réunit les UT sous deux états dérivés possibles, « *X rond* » et « *X carré* » (2.b); et le troisième rassemble selon les variations de pigmentation de *X*, distinguant les états dérivés « *X noir* » et « *X strié* » (2.c). Cette proposition vaut 62 triplets pondérés. Le *X* désigne l'apparition de la caractéristique *X* et les différents symboles renvoient aux formes et pigmentations associés..... 77

Figure 1.29. Nouvelles hypothèses hiérarchiques présentées ici. La proposition 1 correspond à l'hypothèse rassemblant les différentes formes et pigmentations sous un nœud *X*; la proposition 2 rassemble ces formes et pigmentations sous deux nœuds distincts « *forme de X* » et « *pigmentation de X* »; la proposition 3 cumule ces deux approches, rassemblant les nœuds « forme de X » et « *pigmentation de X* » sous le nœud *X*..... 78

Figure 1.30. Résultat de l'analyse en sous arbres libre de paralogie. L'ensemble des trois propositions de la figure 3 donnent toutes le même résultat et valent toutes la même quantité d'information, soit 100 triplets pondérés. On note l'incertitude des noms apparaissant à certains nœuds : cela vient du fait que les nœuds collapsent. Même s'ils sont identiques en termes d'extension, ce n'est pas vrai pour leur intension, qui est ambigüe. 80

Figure 1.31. Calcul de l'indice de rétention du caractère (*a*, *b*, *c*, (*d*, *e*)) par rapport au cladogramme (*a*, (*b*, (*c*, (*d*, *e*)))). On retire du caractère trois triplets pondérés à 1. Parmi ces trois triplets, un est rejeté par le cladogramme. L'indice de rétention de 2/3 indique la proportion de l'information du caractère participant au cladogramme..... 86

Figure 1.32. Exemples de répartition des valeurs de NRI sur un cladogramme pour un caractère donné. (1) Valeurs de NRI d'un arbre pectiné pour un caractère (*a*, (*b*, *c*, *d*, *e, f, g, h, i, j, k, l, m, n, o*)). L'état dérivé vaut 35 triplets pondérés qui se retrouvent tous dans le clade {*b*, *c*, *d*, *e, f, g, h, i, j, k, l, m, n, o*}. Cependant, certains triplets participent également à des nœuds moins inclusifs. (2) Valeurs de NRI d'un arbre pectiné pour un caractère (*a*, *b*, *c*, *d*, *e, f, g, h, i, j, k, l, m,* (*n, o*)). (3) Valeurs de NRI d'un arbre pectiné pour un caractère (*a*, *b*, *c*, *d*, *e, f, i, j, k, l, m, n, o,* (*g, h*)). (4) Valeurs de NRI d'un arbre avec un paralogue pour un caractère (*a*, *b*, *c*, *d*, *e, k, l, m, n, o, (f, g, h, i, j)*). Le nœud paralogue fonctionne comme un point de diversification à partir duquel deux lignées deviennent indépendantes l'une de l'autre. (5) Valeurs de NRI d'un arbre avec un paralogue pour un caractère rejeté (*a*, *b*, *c*, *d*, *e, f, g, h, k, l, m, (i, j, n, o*). 89

Figure 1.33. Correspondance entre un homologue X_h, hypothèse de clade sur des parties d'UT, et l'état de caractère correspondant (*C*) après standardisation, hypothèse de clade sur des UT (flèche grise). Assignation de l'état *X* au clade *C* (flèche noire). L'état *X* est la synapomorphie du clade *C*, que l'on note *C*(*X*). Le clade *C* regroupe les taxons *c*, *d* et *e* par la propriété « *avoir* X_h ». 91

Figure 1.34. Caractérisation d'un état *X* à un clade du cladogramme (*a*, (*b*, (*c*, *d*))). A l'étape caractérisation des clades, on reprend le cladogramme sur lequel on n'affiche que les UT qui portent l'état *X*. Le cladogramme est composé de trois clades, il n'y a donc que trois possibilités. (1) L'état *X* correspond parfaitement au clade *C*. (2) L'état *X* est associé au clade *B*, *d* ayant potentiellement subi une réversion. (3) L'état *X* est associé à la racine *A*, *c* et *d* ayant potentiellement subi une réversion. Un état peut toujours être assigné à un clade si le cladogramme est une orthologie. 93

Figure 1.35. Test d'un état dérivé *X* sur un cladogramme (1) contenant un paralogue et illustration du nœud paralogue comme instance de test : un paralogue rejetant un état est noté en rouge, un paralogue acceptant un état en vert, et un paralogue non impliqué dans le test (les instances de *X* dans un seul nœud interne inclus dans le paralogue) en blanc. Les tests 2, 3, 4, 5 et 6 montrent le cladogramme et uniquement les UT qui sont également contenues dans l'état *X*. *X* est accepté partout et caractérise un clade sauf en (4) ou l'état est rejeté car apparaissant dans deux lignées indépendantes. 96

Figure 1.36. Schéma du test d'un état sur un nœud paralogue *Z* contenant les nœuds internes *V*, *W*, *X* et *Y*. Le clade caractérisé par *X* est indiqué en bleu. (1) Si *X* et *Y* sont caractérisés par l'état, alors *Z* aussi. Si seul *X* l'est, *Z* ne contient pas l'état dérivé. Si aucun nœud n'est caractérisé par l'état, *Z* non plus. L'état est rejeté si on trouve des instances dans différentes lignées d'un paralogue qui n'est pas lui-même caractérisé par l'état. (4) et (5) pour que le nœud paralogue soit caractérisé par l'état, toutes les lignées indépendantes doivent être caractérisées par l'état. 97

Figure 1.37. Test d'un état dérivé *X* sur un cladogramme contenant un paralogue lui-même inclus dans une orthologie. Un paralogue rejetant un état est noté en rouge, un paralogue acceptant un état en vert, et un paralogue non impliqué dans le test (les instances de *X* dans un seul nœud interne inclus dans le paralogue) en blanc. Seuls les UT inclus dans l'état dérivé sont affichés. (1) et (2) l'état *X* est accepté. (3) l'état est rejeté car apparaissant dans deux lignées indépendantes. L'état est rejeté car il ne passe pas le test du nœud paralogue rouge). 98

Figure 1.38. Représentations du temps. (1) Echelle des temps géologiques. Mathématiquement, une échelle est une partition. Une période commence lorsque l'autre s'arrête. (2) Hiérarchie temporelle représentée sous la forme d'un diagramme d'Euler. Une période commence à un moment donné mais ne s'arrête jamais. Une période se différencie en une autre. La hiérarchie temporelle est une classification : les classes sont des périodes, ou unités chronologiques (UC) et les feuilles sont des instances particulières de ces unités. (3) Même hiérarchie représentée sous forme d'arbre. La hiérarchie temporelle est une classification d'instants temporels (feuilles). Le temps hiérarchique est un temps cladistique. Les nœuds internes sont colorés selon la convention de l'échelle des temps géologique. 101

Figure 1.39. Datation d'un cladogramme-orthologie. (1) Cladogramme. (2) Hiérarchie temporelle composée de périodes (UC – nœuds internes) et de dates particulières (instances d'UC – feuilles). (3) Cladogrammes datés. On place une UC au clade le plus inclusif ne contenant que des instances de cette UC. Dans le cas (a), hiérarchie temporelle et cladogramme sont parfaitement congruents. Dans le cas (b), il y a de l'incongruence. Les nœuds *B*, *D* et *F* ne sont pas caractérisables. L'UC *Maastrichtien+* ne peut être placée sur un clade (un nœud interne). Dans le cas (c), l'incongruence est très forte : seul le nœud *A*

est défini, par la période Permien+ bien que ne clade daté au Permien+ contienne des UT beaucoup plus récentes. 102

Figure 1.40. Datation d'un cladogramme contenant un paralogue *X*. (1) Cladogramme. (2) Décomposition du cladogramme en trois sous-arbres libres de paralogie et datation (en Ma). On place une UC au clade le plus inclusif ne contenant que des instances de cette UC. (3) La réunion des sous-arbres met en évidence un phénomène de paralogie temporelle : à partir du nœud paralogue, la hiérarchie temporelle se scinde en deux chronologies indépendantes l'une de l'autre. 104

Figure 1.41. Diverging objectives of consensus tree methods. A consensus tree in agreement with the soft consensus aim must represent only that information shared by all the optimal trees. Its efficiency is calculated using the Consensus Retention Index (CRI). A consensus tree following the hard aim must summarize graphically all the information conveyed by a set of optimal trees. Its efficiency is calculated using the Consensus Efficiency (CE) index. 110

Figure 1.42. Diverging interpretations of the informative content of a hierarchical tree and the relations between phylogenetic statements. Soft interpretation of polytomies – the optimal tree *AB(CD)* shows a soft polytomy that expresses uncertain resolution. Under the soft interpretation, phylogenetic information consists of the positive statements *A(CD)* and *B(CD)* and their two orthogonal disqualifiers not*D(AC)* and not*C(AD)*, and not*C(BD)* and not*D(BC)*, respectively. Hard interpretation of polytomies – the optimal tree *AB(CD)* shows a hard polytomy that expresses a multiple speciation event. Its deduced statements include all disqualifiers (e.g. not*A(BC)*) and thus forbid any possible resolution of the polytomy. Because these disqualifiers are not linked to any positive information, the orthogonality rule is broken. 112

Figure 1.43. Two optimal cladograms (Tree 1 and Tree 2) and their differential treatment by applying the hard consensus aim (left – the consensus must represent the set of optimal trees) or the soft consensus aim (right – the consensus represents the congruent information between all optimal trees) on triads and their disqualifiers. The set of optimal trees is composed of the two trees *A(B(CD))* and *B(A(CD)*. Optimal trees are decomposed into triads and orthogonal disqualifiers. Hard aim – during the decomposition step, the

orthogonality rule is broken (black crosses) in order to maximize consensus efficiency. 2/4 of triads and 6/8 of disqualifiers common to the set of optimal trees are retained (grey area). The recombination of these statements produces *AB(CD)* and not*CD(AB)* in order to forbid the tree *(AB)(CD)* that is not one of the two optimal trees. Trees permitted by the consensus are identical to the optimal trees, which leads to a maximal Consensus Efficiency index. Soft aim – during the decomposition step, the orthogonality is conserved (circles surround each mother triad and its two child disqualifiers). 2/4 of triads and 4/8 of disqualifiers common to the set of optimal trees are retained (grey area). The recombination of these statements produces *AB(CD)*, which is the only unambiguous true information. The Consensus Retention index is maximal since all the common information is displayed by the consensus. 114

Figure 1.44. Two optimal cladograms *A(B(CD))* and *B(C(AD))* and their differential treatment by applying the hard or the soft consensus aims on triads and their disqualifiers. Hard aim – during the decomposition step, the orthogonality rule is broken (black crosses) in order to maximize consensus efficiency. 1/4 of triads and 5/8 of disqualifiers common to the set of optimal trees are retained (grey area). The recombination of these statements produces *B(CD)* and not*D(C(AB))*. The Consensus Efficiency index cannot be maximal because there is another kind of information that cannot be displayed on a hierarchical tree: the mutually exclusive statements [*A*(*BC*) and *A*(*BD*) and *A*(*CD*)] OR [*B*(*AC*) and *B*(*AD*) and *C*(*AD*)]. Breaking the orthogonality rule implies loss of information on the origin of disqualifiers. Thus, several triads are present in the set of optimal tree but can never be gathered into the same optimal tree. Soft aim – during the decomposition step, the orthogonality is conserved. 1/4 of triads and 2/8 of disqualifiers common to the set of optimal trees are retained (grey area). The recombination of these statements produces *B(CD)*, which is the only unambiguous true information. The Consensus Retention index is maximal since all the common information is displayed by the consensus. 116

Figure 1.45. Illustration of the concepts of compatibility and combinability using triads. Compatibility – Left, examples of non-compatible triads. Triads within each couple cannot be true at the same time, because they necessarily contradict each other. Right, compatible triads that can be true at the same time. Triads are compatible if they are not orthogonal. Two triads are orthogonal if they have the same set of OTUs but different

relationships between these OTUs. Combinability – Left, triads within each couple are non-combinable because they have less than two common OTUs. The strict consensus of all possible trees that contain two non-combinable triads is always non-informative: the triads cannot be combined into one single non-ambiguous hierarchy. Ten possible trees contain the triads of the first couple *A*(*BC*) - *A*(*DE*), e.g. ((*BC*)(*A*(*DE*))) and (*A(B*(*D*(*CE*)))). Right, triads within each couple are combinable because they share two OTUs. The strict consensus of all possible trees that contain two combinable triads always includes an informative internal node. 126

Figure 1.46. Schéma général d'une analyse en 3ia. Les homologies sont standardisées en caractères. Les caractères sont décomposés en triplets (à chaque triplet est associé une PF). Le cladogramme répond à la question des relations de parenté au sein de {*a*, *b*, *c*, *d*, *e*}. 132

Chapitre 2

Figure 2.1. Schéma général de la méthode d'analyse cladistique en parcimonie. (i) Les caractères sont des partitions non ordonnées (langage mathématique : non-ordonnée; langage des systématiciens : non polarisée), rassemblées en une matrice. (ii) Les partitions sont ensuite optimisées sur un arbre non-enraciné de manière à diminuer au maximum la distorsion des partitions – minimisation du nombre de classes créées. Ces structures ne sont pas mathématiquement équivalentes et n'ont pas le même contenu sémantique. (iii) Les partitions sont ordonnées (i. e. polarisées), le plus souvent grâce à l'un des taxons désigné comme extra-groupe. (iv) La polarisation des caractères entraîne la polarisation de l'arbre qui devient hiérarchique et sur lequel est possible de déduire si un groupe est monophylétique ou non. Le cladogramme optimal est celui qui implique le minimum de pas de transformation. Les caractères sont tous paraphylétiques par leur état plésiomorphe – l'état ancestral d'un caractère-partition ne contenant jamais tous ses descendants. 139

Figure 2.2. Exemple d'optimisation en parcimonie d'une partition sur un arbre. (1) Partition ordonnée (*a*, *b*, *c*, *d*) → (*e*, *f*, *g*) optimisée sur les arbres (2) et (3) (en haut, cladogramme; en bas, partition résultant de l'optimisation. Sur l'arbre (2), la partition n'est pas distordue : l'état dérivé est synapomorphique du clade (*e*, *f*, *g*), et l'état ancestral (*a*, *b*,

c, *d*) est paraphylétique – et donc artificiel. Sur l'arbre (3), la partition est distordue en trois classes au lieu de deux (distorsion de type « réversion »). Une nouvelle hypothèse de synapomorphie (*c*, *d*) est générée à partir de l'état ancestral paraphylétique. L'état dérivé (*e*, *f*, *g*) devient paraphylétique – et donc artificiel. 143

Figure 2.3. Exemple de polymorphisme en parcimonie. Une partition impose des classes d'états mutuellement exclusives. Or, un taxon polymorphique possède les deux états à la fois, et génère une pseudo-partition. La pseudo-partition ne peut pas être utilisée en l'état : elle doit être interprétée en partition. Deux interprétations sont possibles, conduisant à la partition 1 ou à la partition 2. Le choix de la bonne partition devrait se faire au moment même de la construction du caractère, selon des critères intrinsèques au caractère, par exemple sa structure. Aucun algorithme de parcimonie n'arrive à gérer le problème du polymorphisme, et le contourne en utilisant un critère extrinsèque totalement indépendant du problème : le choix de la partition se fait selon le critère de minimisation de la distorsion de l'ensemble des caractères. 145

Figure 2.4. Exemple d'optimisation ambigüe en parcimonie d'une partition (*a*, *d*) → (*b*, *c*) sur un arbre : il y a incertitude sur quels clades sont soutenus par le caractère-partition. (2) Optimisation de type *transformation accélérée* (« *ACCTRAN* »). Le clade (*b*, *c*, *d*) est soutenu par l'état dérivé. (3) Optimisation de type *transformation retardée* (« *DELTRAN* »). L'état dérivé ne groupe pas de taxons au sein d'un clade et est non-informatif. Savoir si l'état dérivé (*b*, *c*) permet de former un groupe monophylétique est indécidable. 147

Figure 2.5. Problème énoncé par Nelson en 1996 et sa résolution en 3ia et en parcimonie. A partir de l'énoncé cladistique selon lequel au sein de l'ensemble {*a*, *b*, *c*, *d*}, on peut proposer des hypothèses de regroupement suivantes : {*b*, *c*} (carré jaune), {*b*, *d*} (carré vert), {*a*, *d*} (carré violet). En dessous, codage en matrice (parcimonie) et en hiérarchies (3ia), et consensus strict. Sous la méthode de parcimonie, aucune information non-ambigüe ne se dégage des hypothèses proposées. En 3ia, un clade (*b*, *c*, *d*) est retrouvé de manière non-ambigüe. 148

Figure 2.6. Suite du problème énoncé par Nelson en 1996 et sa résolution en 3ia et en parcimonie. Les hypothèses de l'énoncé peuvent être multipliées à l'infini : le taxon *a*

ne porte jamais d'état ancestral, et les taxons *b*, *c* et *d* peuvent potentiellement partager deux à deux une infinité d'états dérivés. En dessous, consensus strict relatif à la parcimonie et à la 3ia. Sous la méthode de parcimonie, aucune information non-ambigüe ne se dégage des hypothèses proposées. En 3ia, un clade (*b*, *c*, *d*) est retrouvé de manière non-ambigüe. 150

Figure 2.7. Illustration de la sensibilité de l'indice de rétention (IR) en parcimonie au nombre de taxons. L'indice de rétention calcule l'information cladistique du caractère-partition (1) retenu sur un cladogramme (2). (a) Exemple d'un caractère optimisé sur un cladogramme dont l'IR vaut 0,5. Un clade (*c*, *d*, *e*, *f*, *g*) est soutenu. De l'information est cependant perdue à cause de la distorsion du caractère. (b) En rajoutant un taxon *x* porteur de l'état ancestral à la racine, l'IR du caractère augmente et passe à 0,66. Cette augmentation est artificielle puisque l'information retenue ne change pas : le caractère est distordu de la même manière qu'en (a) avec deux pas de transformation, et soutient le même clade (*c*, *d*, *e*, *f*, *g*). 152

Figure 2.8. Standardisation en biogéographie. En haut, correspondance entre UT (*a*, *b*, *c*, *d*, *e*, *f*, *g*) et UB (UB_1, UB_2, UB_3, UB_4). On remplace sur un cladogramme de taxons les UT par les UB, ce qui forme un caractère biogéographique. Le caractère biogéographique pose des hypothèses de clades d'UB. (1) Standardisation complète. (2) Standardisation non informative due à une répétition de terminaux ne permettant pas la génération d'un caractère bioghéographique. On constate l'identité totale de la procédure entre phylogénie et biogéographie. 156

Figure 2.9. Correspondance entre étapes de 3ta et étapes de 3aa. L'homologie présente des hypothèses de parenté entre parties d'UT. L'homologie est standardisée en caractère ou les parties d'UT sont remplacées par les UT elles-mêmes. L'ensemble des caractères est décomposé en triplets. La recombinaison de l'ensemble de congruence maximale permet de générer la phylogénie, cladogramme de taxons. (3aa) Le cladogramme présente des hypothèses de parenté entre UT. Le cladogramme est standardisée en caractère biogéographique ou les UT sont remplacées par les UB. L'ensemble des caractères biogéographiques est décomposé en triplets. La recombinaison de l'ensemble de congruence maximale permet de générer l'aréagramme, cladogramme d'aires..... 158

Figure 2.10. Illustration de la méthode de parcimonie en analyse biogéographique (Analyse de Parcimonie de Brooks). (1) La source des hypothèses sur les aires biogéographiques provient de la connaissance des relations de parenté entre taxons. (2) Le caractère biogéographique, intrinsèquement hiérarchique du fait de son origine, est transformé en caractère-partition. (3) Les partitions sont optimisées sur un arbre. Polarisé, l'arbre devient un aréagramme. L'étape (2) modifie le contenu sémantique des hypothèses biogéographiques. 159

Figure 2.11. Three examples of character representation: unordered in parsimony (A; non-additive) and ordered in parsimony (B; additive) form the class of parsimony characters, associated with their stepmatrix (Sankoff and Rousseau, 1975) necessary in a matrix representation, and the other classes of character trees, hierarchies treated in 3ta (C). A symbolizes unordered character states, were all transformations between states have the same cost. B represents ordered states as an unrooted tree and the equivalent stepmatrix, and C represents a character tree, were nodes are homology hypotheses under the hierarchy. The differences in their representations (graph, unrooted tree, hierarchy) traduce both their content and the way in which corresponding methods treat them (how they deal with incongruence they can generate; for example, in parsimony graphs A and B, a step can correspond to transformation from 1 to 2 or from 2 to 1, which make no sense in the hierarchical representation C, symbolizing a 3ta treatment). Letters and numbers (0-5, Z-W) represent named nodes as homologs. 163

Figure 2.12. Fractional weighting (FW) and dependence among triplets derived from the same character. A. A character relating five terminal taxa is decomposed into its elementary relationships, expressed as a list of triplets. The decomposition results into a list of six triplets. However, only two of the three triplets [*e.g.* (A (C D)) and (A (C E)) that relate three terminals (C D, E) to other, more distant ones (*e.g.* A] are necessary to build the initial hierarchical character, as shown in the set diagram. FW assigns an information content/weight of 2/3 to each triplet of the two groups. B. A multi-state character is analyzed into components and then into triplets. 3ta takes into account the fact that each component is not independent. Thus, the triplets derived from different components of the same character may be redundant. Consider the triplets (A (D E)): it appears in the three components. However, it represents a single node in the original character. The triplets

derived from the third component have a FW=1, *i.e.* they are non-redundant. The same triplets in the first and second components are thus redundant and can be removed. The FW of the remaining triplets of the component is then recalculated from FW=2/3 to FW=1 in the second component, and FW<1 in the first. Etc… Note that the matrix treatment of the triplets is immaterial with their status of dependent or independent, contrary to characters in parsimony analysis.....166

Figure 2.13. An example of character history according to cladistics. A is a modified phylogeny from Whiting *et al*. (2003) with transformations (obtained using parsimony under acctran optimization) including reversals (denoted by a '-' sign) and acquisitions or reacquisitions ('+') from a binary character absence/presence (the 3ta character tree is represented as a hierarchy). The degree of gains and losses are denoted by numbers (1 for primary, 2 for secondary, etc.) in the circles placed at the corresponding nodes. The degree of the resulting state (presence or absence, appearing in parentheses below and to the left of the symbols for gains or losses) is denoted by the asterisks; primary gains or losses have none; secondary gains or losses have one; ternary ones have two, etc. B shows the wingless Phasmatodea *Leptynia hispanica* (Bolivar, 1878), *Phyllium bioculatum* (Gray, 1832) with wings and *Pseudophasma acanthonotum* (Redtenbacher, 1906) with wings and *Pseudophasma acanthonotum* (Redtenbacher, 1906) with convergently re-acquired wings (copyrights, respectively: Fritz Geller-Grimm and Felix Grimm - CC BY-SA 3.0, Drägus - GFDL and Drägus – PD-self). C represents with a character tree the history reconstructed a posteriori from the analysis (with the mapping of the character on the optimal tree) of the character 'wing' of phasms. Each node represent a parsimony apomorphy, contrary to the 3ta for which the only apomorphy is the node 1 (as only secondary homology because the only primary homology postulated was the homology between all wings).....168

Figure 2.14. An example of optimal cladogram obtained both in parsimony and 3ta with effect of polarization error with an outgroup criterion. The true ancestral state is 0, and 1 is a true synapomorphy of D, E and F, but the outgroup criterion leads to an erroneous ancestral condition (1). Parsimony may be less sensitive to errors in polarization than 3ta, which cannot support a clade with what appears to be a reversal from 0 to 1.....171

Figure 2.15. Trees used for our simulations: A, pectinate; B, symmetric; C, equiprobable; D, equiprobable with branch length set to 1; E-G, equiprobable with steady

increase of internal/external branch length ratio. Each color represent a specific outgroup branch length expressed as a proportion of total tree depth (A-C: blue, 1; green, 1/2; red, 1/4; D: dark blue, 1; light blue, 2/3; green, 1/2; yellow, 1/3; orange, 1/10; red, 0). A–C represent trees with a neontological ingroup; trees D–G represent paleontological trees (with diachronous tips). 172

Figure 2.16. Histograms showing resolving power and artefactual resolution obtained by ITRI mean percentages sorted by outgroup branch length (blue, 1/4; green, 1/2; red, 1), by method (ordered parsimony, 3ta and unordered parsimony), and by topology (pectinate, equiprobable and symmetric; see S8 in http://www.ctoz.nl/vol84/nr02/a03). Each ITRI mean represents 100 matrices of 100 characters and 21 taxa (outgroup included). 177

Figure 2.17. Impact of outgroup branch length expressed as a proportion of total tree depth (0, 1/10, 1/3, 1/2, 2/3, 1) on resolving power and artefactual resolution. This was obtained through ITRI mean percentages from the tree shown in fig. 2.15D (also see S8 in http://www.ctoz.nl/vol84/nr02/a03). 182

Figure 2.18. Character evolution under Brownian motion. In A, the unrooted tree represents the relationships between states of a character simulated under Brownian motion and discretized into 10 states. B shows the simulation of characters on a known phylogeny under Brownian motion. The left graphic represents fluctuations of a continuous character, with cladogenesis and terminal taxa shown as red nodes. After discretization through states 0, 1 and 2, the known evolutionary model and a topology (known or inferred) allow inferring character history. Here, the example treats characters under the parsimony method, with acctran optimization. 187

Figure 2.19. Ultrametric tree used in our simulations. The randomly generated phylogeny comprises an ingroup of 21 OTU and a single outgroup OTU. 196

Figure 2.20. Two scenarios of errors in character ordering (in a character with 10 states) with five stages of degradation. Scenario 1 represents the case with slight permutation errors involving neighbors. Scenario 2 represents the worst case possible, with states displaced as far as possible from its neighbors. Coloured states represent permuted states and arrows represent permutations. 197

Figure 2.21. Impact of number of states in ordered, unordered MP and 3ta on the efficiency of the methods. The difference in efficiency between methods was assessed by iterative recoding of the same matrices with two, four, six and eight states. Orange: ordered MP; red: 3ta; black: unordered MP.....201

Figure 2.22. Impact of ordering errors shown in figure 2 on MP efficiency. The difference in efficiency between methods was assessed by iterative recoding of the same matrices with the various ordering schemes with more and more permutations. Scenario 1 shows MP's behavior with slight errors; scenario 2 shows the worst possible errors in the ordering scheme. Orange: ordered MP; red: 3ta. Unordered MP is constant because the ordering scheme is not coded, and is represented as the black dotted line.....202

Figure 2.23. 1 – Impact of reversals on phylogenetic reconstruction methods. Left: the evolutionary model used is an irreversible model, generating only convergences, showing better performance of MP than 3ta (p-value: 0.02153). Right: a standard Brownian Motion model is used, generating both convergences and reversals; 3ta performs better than MP (p-value: 0.0004). 2 – Impact of randomly generated errors on phylogenetic reconstruction methods. Left: standard Brownian motion model (no significant difference between methods; p-value: 0.4402). Right: same model with random errors added; 3ta performs better than MP (p-value: <0.0001). Orange : ordered MP; red: 3ta; black: unordered MP.....204

Figure 2.24. Example of differential behaviors between MP and 3ta. The example shows two phylogenies, A and B, and a character *X*. Character *X* is represented in a matrix in MP, and through a hierarchical character (*a*, *b*, *d*, *e*, (*c*,*f*), with (*c*,*f*) as derived character state. In MP, the optimization of both trees always leads to two steps. Regarding only character *X*, they are both suboptimal. In 3ta, the optimal tree is chosen as the structure retaining the maximal number of hypotheses (i. e. triplets resulting from the character decomposition). Tree A is completely incompatible with character *X*, as no triplet is retained. Regarding only character *X*, tree B is a better phylogeny as it retains one triplet from the initial character. In 3ta, triplets tend to 'attract' nodes that independently acquire a derived state, whereas this phenomenon does not occur in MP.....210

Chapitre 3

Figure 3. 1. A – Earliest illustration of *I. triangularis* as "Cucroïdes" by Guettard (1786), who thought the fossil to be circular ("*pierres circulaires*"). B – Original drawings of the internal molds of *I. triangularis* (Desmarest 1817). C – Specimen of *I. triangularis* MNHN.F.R08003 from Île d'Oléron illustrated in d'Orbigny's Paléontologie Française (1847); this specimen is designated here as neotype of *Ichthyosarcolites triangularis* Desmarest 1812. 224

Figure 3. 2. Types of *Ichthyosarcolites* species. All scale bars = 1cm. A – Holotype of *Ichthyosarcolites rotundus* Polšak, 1967 (LGPZ.778) showing a cylindrical shell. Picture from original publication. Cenomanian, Fontanela bay (Istria, Croatia). B – Holotype of *Ichthyosarcolites monocarinatus* Sliskovic, 1966 (MB.3470) showing a small flange and round canals. Cenomanian, Dinaric Alps. Picture provided by Alceo Tarlao. C – Syntype of *Ichthyosarcolites bicarinatus* (Gemmellaro collection, MGUP.021.2-110) showing two flanges and a twisted body along the umbo-aperture axis. Cenomanian, Piemonte (Italy). Picture provided by Tony Costagliola. D – Holotype of *Ichthyosarcolites polycarinatus* Sliskovic, 1966 (MB.3482) showing seven elongated flanges and small capillary canals. Upper Cenomanian, Dinaric Alps. Picture provided by Alceo Tarlao. E – Holotype of *Ichthyosarcolites alatus* Filkhorn, 2002 (UNAM.6979) showing five flanges and a presumed juvenile attached to its ventral side. Left valve. The myocardinal apparatus and its associated cavities are visible. Lower Cenomanian, Guerrero (Mexico). Picture provided by Javier Aguilar-Pérez. F – Lectotype of *Ichthyosarcolites doublieri* d'Orbigny, 1847 (MNHN.F.R07971), lacking tabulae. Internal mold showing the highly developed right valve and the short, left valve. Upper Cenomanian, Martigues (Bouches-du-Rhône, France). 226

Figure 3. 3. Shells internal features of *Ichthyosarcolites*. All scale bars = 1cm. A – *I. rotundus* without any flange (AMU.ICH1) from La Bédoule (Provence, France). Arrow: posterior accessory cavity. B – Picture of *I. bicarinatus* from Chiki-Aouimeur (2010) showing two elongated flanges and big round canals (Aqaba, Jordania). C – Sliced right valve of *I. triangularis* with three flanges, the two sockets and the posterior myophoral cavity (AMU.HBES 85-a). D – Specimen of *I. bicarinatus* from Cañón de la Servilleta

(Tamaulipas, Mexico) showing four flanges and thin tabulae. Picture provided by Javier Aguilar-Pérez. E – Mexican right valve with four flanges *I. bicarinatus* (from Filkhorn, 2002). The myocardinal apparatus from the left valve is visible in connection (UNAM.6970). Picture provided by Javier Aguilar-Pérez. F – Sliced specimen of *I. coraloidea* (GMUS.Mp1) showing general cavity filled of canals, a thick shell and three cavities invaded by capillary canals. Picture provided by Brian Pratt.....227

Figure 3. 4. Photographs of *Ichthyosarcolites* shells illustrating tabulae and coiling patterns. All scale bars = 1cm. A – Longitudinal slice of *I. triangularis* from Île d'Oléron showing thick tabulae in the general cavity and the sockets. Arrow: tabulae. B – Internal molds of *I. triangularis* from Île d'Oléron (Charentes, France) showing the partial or complete dissolution of the thin tabulae in the general cavity and the sockets. Arrow: tabulae. C – Specimen of *I. triangularis* from Île d'Oléron with intersecting tabulae. D – Crushed internal mold of *I. triangularis* from Île Madame, showing tabulae and one socket (d'Orbigny collection, MNHN.6517–B). E – Almost complete coiled right valve internal mold from Provence (France). F – Specimen of *I. triangularis* from Île d'Oléron with uncoiled internal mold. 230

Figure 3. 5. A- Method for measurement of the angular positions of flanges on rudist shells. A circle approximates the general cavity. White circles correspond to the projection of the flange stem on the circle. Angles are measured clockwise with 0° corresponds to the presumed position of the central tooth or central tooth socket. Scale bar = 1cm. Specimen UNAM.6978, picture provided by Javier Aguilar-Pérez. B- Plot of flanges angle measurement by cos and sin coordinates. C- Hierarchical classification of flanges positions. The two main clusters represent the dichotomy between *I. monocarinatus* ventral flanges and *I. triangularis* dorsal flanges. Each point corresponds to a flange position, and the color corresponds to the number of flanges the shell from which the flange is from bear.....231

Figure 3. 6. Discriminant analysis plot of 23 general cavities of *Ichthyosarcolites* shells landmarks. Four clusters are designated from the number of flanges: *I. rotundus* (no flange), *I. monocarinatus* (one ventral flange), *I. triangularis* (one dorsal flange) and *I. bicarinatus*. Analyzed data are presented as supplemental Online Material. 234

Figure 3. 7. Geographical map showing the fossil localities of Cenomanian age bearing *Ichthyosarcolites*. Yellow dots represent *I. rotundus*, green dots *I. triangularis*, brown dots *I. monocarinatus*, orange dots *I. bicarinatus*, red dots *I. polycarinatus* and blue stars *Oryxia sulcata* nov. gen. nov. sp. Black dots represent *Ichthyosarcolites* occurrences from which the identification to species level cannot be verified. The paleogeographic map was generated using the ODSN Plate Tectonic Reconstruction Service (http://www.odsn.de/odsn/services/paleomap/paleomap.html). 241

Figure 3. 8. Holotype of *Oryxia sulcata* nov. gen. nov. sp. (right valve; BJ.1901-HR2) sliced in three samples. A – Transversal view of a slice. B – Interpretative scheme of A with main morphoanatomical features and homologs. AM: anterior myophore; PM: posterior myophore; ATS: anterior tooth socket; PTS: posterior tooth socket; CT: central tooth; D: general cavity of the right valve; AAC: anterior accessory cavity; LC: ligamentary cavity; LB: ligamentary bulge. C – Dorsal view of the three samples of the right valve. D – transversal view of the paratype BJ.1914-HR2. All pictures provided by Bogdan Jurkovšek. 259

Chapitre 4

Figure 4. 1. Evolutionary scheme transcribing the view of Douvillé (1935). Four 'rameaux' (branches) evolve from a *Pterocardium* origin. Each branch is supported by morphological features. The 'Rameau des *Diceras'* is much detailed because it's the starting point to two other lineages, the *Hippurites* branch and the *Radiolites* branch. ..266

Figure 4. 2. Evolutionary scheme according to Skelton's 2003 tree, replacing all of the rudist families in an evolutionary context. Uncertainties remain in the three Trechmanellidae, Antillocaprinidae and Ichthyosarcolitidae families. The Diceratidae family is partitioned in two branches by their attachment valve. LVA: Left Valve Attachment; RVA: Right Valve Attachment; PC: platform crisis.....270

Figure 4. 3. Strict consensus tree redrawn from Skelton and Smith's (2000). The corresponding families are replaced for each genus. One exclamation (!) mark means paraphyly and two means polyphyly (!!). 272

Figure 4. 4. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Hippurites radiosus*, interior of left valve; Senonian, Charentes, France (MNHN.F.A58757). B – *Hippurites radiosus*, interior of right valve; Senonian, Charentes, France (MNHN.F.A58757). C – *Monopleura michaillensis*, interior of left valve; Barremian, Brouzet, Gard, France (UPMC.1272). D – *Monopleura michaillensis*, interior of right valve; Barremian, Brouzet, Gard, France (UPMC.1272). E – *Plagioptychus arnaudi*, interior of left valve; France (UPMC.1273). F – *Plagioptychus arnaudi*, interior of right valve; France (UPMC.1273). G – *Valletia germani*, interior of left valve; lower Cretaceous, Le Fourney, Savoie, France (MNHN.F.J05798). H – *Valletia germani*, interior of right valve, lower Cretaceous, Le Fourney, Savoie, France (MNHN.F.J05798). Scale bars, 10 mm. 276

Figure 4. 5. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Radiolites radiosus*, interior of left valve; Turonian, Charentes, France (MNHN.F.J09021). B – *Radiolites radiosus*, interior of right valve; Turonian, Charentes, France (MNHN.F.B17070). C – *Mexicaprina minuta*, interior of left valve; Albian-Cenomanian, Querétaro, Mexico (JPMA.14097-5). D – *Mexicaprina minuta*, interior of right valve Albian-Cenomanian, Querétaro, Mexico (JPMA.14097-1). E – *Caprinula boissyi*, interior of left valve; Turonian, Corbières, Aude, France (Coll. D'Orbigny, MNHN.6924). F – *Caprinula boissyi*, interior of right valve; Turonian, Corbières, Aude, France (Coll. D'Orbigny, MNHN.6924). G – *Caprotina striata*, interior of left valve; Cenomanian, Le Mans, Sarthe, France (Collection Guéranger, MV 2003.2.5147). H – *Caprotina striata*, interior of right valve; Cenomanian, Le Mans, Sarthe, France (Collection Guéranger, MV 2003.2.5147). Scale bars, 10 mm. 277

Figure 4. 6. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Ichthyosarcolites bicarinatus*, interior of left valve; Cenomanian, Guerrero, Mexico (IGM.6970). B – *Ichthyosarcolites bicarinatus*, interior of right valve; Cenomanian, Guerrero, Mexico (IGM.6972); pictures provided by Javier Aguilar-Pérez (A and B). C – *Dictyoptychus morgani*, interior of left valve; Maastrichtian, Bakhtiari country, Iran (NHM.L-58427). D – *Dictyoptychus morgani*, interior of right valve; Maastrichtian, Bakhtiari country, Iran (NHM.L-58424); pictures provided by Peter W. Skelton (C and D). E – *Pachytraga paradoxa*, interior of left valve; Aptian, Rustrel, Vaucluse, France (JPMA.

17738-1). F – *Pachytraga paradoxa*, interior of right valve; Aptian, Rustrel, Vaucluse, France (JPMA.17738-2). G – *Titanosarcolites giganteus*, interior of left valve; Maastrichtian, Ocozocuautla Formation, Chiapas, Mexico (UT.47533). H – *Titanosarcolites giganteus*, interior of right valve; Maastrichtian, Ocozocuautla Formation, Chiapas, Mexico (UT.47533). Scale bars, 10 mm. 278

Figure 4. 7. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Diceras arietinum*, interior of left valve, Coulonges, Yonne, France (MNHN.without collection number). B – *Diceras arietinum*, interior of right valve; Oxfordian, Saint-Mihiel, Meuse, France (MNHN.F.A52439). C – *Glossomyophorus costatus*, interior of left valve; Aptian, Roub al Khali, Saudi Arabia (MNHN.F.R06779). D – *Glossomyophorus costatus*, interior of right valve, Aptian, Roub al Khali, Saudi Arabia (MNHN.F.A27280). E – *Amphitriscoelus waringi*, interior of left valve. F – *Amphitriscoelus waringi*, interior of right valve; pictures provided by Jean-Pierre Masse (E and F). G – *Offneria rhodanica*, interior of left valve; Urgonian, Saint-Montand, Ardèche, France (UCBL.R275). H – *Offneria rhodanica*, interior of right valve, Urgonian, Saint-Montand, Ardèche, France (UCBL.R781); pictures provided by Emmanuel Robert (G and H). Scale bars, 10 mm. 279

Figure 4. 8. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Clinocaprina lowentsami*, interior of left valve; Campanian, Stapleton Saint James, Jamaica (USNM.245241). B – *Clinocaprina lowentsami*, interior of right valve; Campanian, Stapleton Saint James, Jamaica (USNM.245241). C – *Curtocaprina clabaughikinsorum*, interior of left valve; Albian, Kerr County, Texas, USA (UT.10937). D – *Curtocaprina clabaughikinsorum*, interior of right valve; Albian, Kerr County, Texas, USA (UT.10922). E – *Coralliochama orcutti*, interior of left valve; late Campanian/early Maastrichtian, Punta Banda, Baja California, Mexico (USNM.186662); picture provided by Jean-Pierre Masse. F – *Coralliochama orcutti*, interior of right valve; late Campanian/early Maastrichtian, Punta Banda, Baja California, Mexico (USNM.186665). G – *Oryxia sulcata*, interior of right valve; Cenomanian of Hrušica, Slovenia (BJ.1901-HR2); picture provided by Bogdan Jurkovšek. Scale bars, 10 mm. .280

Figure 4. 9. Photographies of specimens from our taxonomic sampling showing LV (left) and RV (right). A – *Chaperia costata*, interior of left valve; Cenomanian, île d'Aix,

Charente-Maritime, France (MNHN.F.R07950). B – *Chaperia costata*, interior of right valve; Cenomanian, île d'Aix, Charente-Maritime, France (MNHN.F.R07942). C – '*Sellaea' sp.*, interior of left valve; Mc Lennan County, Texas, USA (WSA3058). D – '*Sellaea' sp.*, interior of right valve; Mc Lennan County, Texas, USA (WSA3058). E – *Agriopleura marticensis*, interior of right valve; Martigues, Bouches-du-Rhône, France (fig. 15-G in Masse and Fenerci-Masse 2014; JPMA.15172-2). F – *Tepeyacia corrugata*, interior of right valve; Turonian, Puebla, Mexico (CAS.66657.01). Scale bars, 10 mm.281

Figure 4. 10. Drawings of specimens from our taxonomic sampling showing orientation of various homologous features on left valve. A – *Monopleura michaillensis* (UPMC.1272). B – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013; Aptian, València, Spain). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Mexicaprina minuta* (JPMA.14097-5). E – *Radiolites radiosus* (MNHN.F.J09137). F – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm. 282

Figure 4. 11. Drawings of specimens from our taxonomic sampling showing orientation of various homologous features on left valve. A – *Diceras arietinum*. B – *Valletia germani* (MNHN.F.J05798). C – *Caprotina striata* (MV 2003.2.5147). D – *Chaperia costata* (MNHN.F.R07950). E – *Glossomyophorus costatus* (MNHN.F.R06779). Scale bars, 10 mm. 283

Figure 4. 12. Drawings of specimens from our taxonomic sampling showing orientation of various homologous features on right valve. A – *Monopleura michaillensis* (UPMC.1272). B – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013; Aptian, València, Spain). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Mexicaprina minuta* (JPMA.14097-1). E – *Radiolites radiosus* (MNHN.F.J09021). F – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm. 284

Figure 4. 13. Drawings of specimens from our taxonomic sampling showing orientation of various homologous features on right valve. A – *Diceras arietinum* (MNHN.F.A52439). B – *Valletia germani* (MNHN.F.J05798). C – *Chaperia costata* (MNHN.F.R07942). D – *Caprotina striata* (MV 2003.2.5147). E – *Offneria rhodanica* (UCBL.R781). Scale bars, 10 mm. 285

Figure 4. 14. Interpreted cross section into left valves with colors corresponding to homologous features. A – *Valletia germani* (MNHN.F.J05798). B – *Monopleura michaillensis* (UPMC.1272). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). E – *Caprotina striata* (MV 2003.2.5147). F – *Pachytraga paradoxa* (JPMA.17738-1). G – *Caprinula boissyi* (MNHN.6924). H – *Mexicaprina minuta* (JPMA.14097-5). I – *Radiolites radiosus* (MNHN.F.J09021). J – *Hippurites radiosus* (MNHN.F.A58757). K – *Dictyoptychus morgani* (NHM.L-58427). L – *Titanosarcolites giganteus* (UT.47533). M – *Ichthyosarcolites bicarinatus* (IGM.6970). The entire shell of *Titanosarcolites giganteus* is invaded by very small capillary canals (purple), including the teeth (purple hatches). All rudists possess an outer calcitic layer, which is not is usually not visible on fossils when very thin. Scale bars, 10 mm. 292

Figure 4. 15. Interpreted cross section into left valves with colors corresponding to homologous features. A – *Diceras arietinum* (MNHN). B – *Glossomyophorus costatus* (MNHN.F.R06779). C – *Amphitriscoelus waringi*. D – *Offneria rhodanica* (UCBL.R275). E – *Semailia smithi* (BM.LL41931). F – *Clinocaprina lowentsami* (USNM.245242). G – *Curtocaprina clabaughikinsorum* (UT.10937). H – *Coralliochama orcutti* (USNM.186662). I – *Tepeyacia corrugata* (CAS.66657.01). J – *Chaperia costata* (MNHN.F.R07950). K – '*Sellaea' sp* (WSA3058). All rudists possess an outer calcitic layer, which is not is usually not visible on fossils when very thin. Scale bars, 10 mm. 293

Figure 4. 16. Interpreted cross section into right valves with colors corresponding to homologous features. A – *Valletia germani* (MNHN.F.J05798). B – *Monopleura michaillensis* (UPMC.1272). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). E – *Caprotina striata* (MV 2003.2.5147). F – *Pachytraga paradoxa* (JPMA.17738-2). G – *Caprinula boissyi* (MNHN.6924). H – *Mexicaprina minuta* (JPMA.14097-1). I – *Radiolites radiosus* (MNHN.F.J09021). J – *Hippurites radiosus* (MNHN.F.A58757). K – *Dictyoptychus morgani* (NHM.L-58424). L – *Titanosarcolites giganteus* (UT.47533). M – *Ichthyosarcolites bicarinatus* (IGM.6972). The entire shell of *Titanosarcolites giganteus* is invaded by very small capillary canals (purple), including the teeth (purple hatches). All rudists possess an outer calcitic layer, which is not is usually not visible on fossils when very thin. Scale bars, 10 mm. 294

Figure 4. 17. Interpreted cross section into right valves with colors corresponding to homologous features. A – *Diceras arietinum* (MNHN.F.A52439). B – *Glossomyophorus costatus* (MNHN.F.A27280). C – *Agriopleura marticensis* (JPMA.15172-2). D – *Barrettia monilifera* (fig. 6B in Mitchell 2010; UWIGM.RUD.2006.223). E – *Amphitriscoelus waringi*. F – *Offneria rhodanica* (UCBL.R781). G – *Semailia smithi* (BM.LL41931). H – *Clinocaprina lowentsami* (USNM.245242). I – *Curtocaprina clabaughikinsorum* (UT.10922). J – *Coralliochama orcutti* (USNM.186665). K – *Tepeyacia corrugata* (CAS.66657.01). L – *Chaperia costata* (MNHN.F.R07942). M – '*Sellaea' sp* (WSA3058). N – *Oryxia sulcata* (BJ.1901-HR2). The entire shell of *Semailia smithi and Oryxia sulcata* is invaded by very small capillary canals (purple), including the teeth (purple hatches). All rudists possess an outer calcitic layer, which is not is usually not visible on fossils when very thin. Scale bars, 10 mm. 295

Figure 4. 18. Principal sections (red lines; transversal, longitudinal, tangential) on rudist fossils with various shell shapes. Modified from Dechaseaux (1969). Arrows: commissural plane. 296

Figure 4. 19. Confronting hypotheses about the CTS and the 'ichthyosarcolitid toothlet'. Left: our hypothesis, illustrating the two states of the character 10. Right: Mitchell (2013) hypothesis about the 'ichthyosarcolitid toothlet', illustrating its character 13. The dotted arrow show two manifestations of the states 1 and 2 that are indistinguishable. Scale bars, 10 mm. Middle: photographies illustrating the two diverging character states. A – *Plagioptychus arnaudi* (UPMC.1273), B – *Clinocaprina lowentsami* (USNM.245242), C – *Curtocaprina clabaughikinsorum* (UT.10937), D – *Ichthyosarcolites bicarinatus* (IGM.6970); picture provided by Javier Aguilar-Pérez. 300

Figure 4. 20. Myophoral arrangements encountered in our taxonomic sampling in transversal views. Positions and orientations of the anterior and posterior muscles are displayed in blue. Corresponding character states are given in table 4.3. 302

Figure 4. 21. Hypotheses about the origins of pallial canals. Nine character states are implied in their evolution: four on LV (#6.4, 11.1, 12.1, 13.1) and five on RV (#28.1, 29.1, 29.2, 30.1, 31.1). Left valve: digging of a PAC (#11.1), anterior expansion of the CTS (#6.4), cavities invaded with blades (#13.1), ventral connection between PAC and CTS

(#12.1). Right valve: digging of a PTS (#29.1), posterior expansion of the PTS (#29.2), anterior expansion of the ATS (#28.1), cavities invaded with blades (#31.1), ventral connection between PTS and ATS (#30.1). Left valves: A – *Monopleura michaillensis* (UPMC.1272), B – *Caprotina striata* (MV 2003.2.5147), C – '*Sellaea' sp*. (WSA3058), D – *Amphitriscoelus waringi* (picture provided by Jean-Pierre Masse), E – *Offneria rhodanica* (UCBL.R275; picture provided by Emmanuel Robert). Right valves: F – *Plagioptychus arnaudi* (UPMC.1273), G – *Caprotina striata* (MV 2003.2.5147), H –'*Sellaea*' *sp*. (WSA3058), I – *Amphitriscoelus waringi* (picture provided by Jean-Pierre Masse), J – *Caprinula boissyi* (MNHN.6924). Scale bars, 10 mm. 307

Figure 4. 22. Illustration of the CTS anterior development. A – '*Sellaea' sp* (UT.48167), sliced left valve. B – *Retha sp*. (WSA3831.4), left valve. Scale bars, 10 mm. 308

Figure 4. 23. Illustration of confronting hypotheses on CTS homology. A – *Valletia germani* (MNHN.F.J05798); B – *Plagioptychus arnaudi* (UPMC 1273); C – *Chaperia costata* (MNHN.F.R07950). CTS development hypothesis supported herein: the CTS in *Valletia germani* is primitively stranded between the AT and the PM (myophoral shelf): in *Plagioptychus arnaudi*, the posterior myophoral plate becomes oblique, causing the expansion of the CTS; in *Chaperia costata*, the expanded CTS is divided by a thin blade. Myophoral cavity hypothesis: the CTS in *Valletia germani* is primitively stranded between the AT and the PM (myophoral shelf); in *Chaperia costata*, a myophoral cavity fills the position of the posterior myophoral shelf – myophoral cavity and CTS are separated by a blade; in *Plagioptychus arnaudi*, the CTS/myophoral cavity blade disappear. Scale bars, 10 mm. 309

Figure 4. 24. Strict consensus tree of the two optimal trees obtained in 3ta, with corresponding character states. Squares represent the nodes of the tree, and numbers refer to clade labels. Circles represent character states that characterize clades. At the right of the tree, panels correspond to families and superfamilies according to Skelton (2013). Families – white: paraphyletic group, grey: monophyletic group, colored: polyphyletic group. Superfamilies – white: Radiolitoidea, grey: Caprinoidea. 318

Figure 4. 25. Description of the phylogeny and clades by their *Nodal Retention Index* (NRI) displaying the distribution of cladistic information amongst the tree. A triplet is generally counted in the NRI of several ortholog clades (ortholog clades being dependent). A – Absolute NRI values; B – proportions of NRI by kinds of characters; C – proportions of NRI between canals and myophores; D – proportions of NRI by valve..320

Figure 4. 26. Dated phylogeny highlighting two radiations of clades. Clades are characterized by their age of origin. 321

Figure 4. 27. Probable position of pallial canals character states on the phylogeny. A – Left valve pallial canals. B – Right valve pallial canals. 324

Figure 4. 28. Stratigraphic extension of 154 rudist genera of the Hippuritidina suborder, from Oxfordian to Maastrichtian, modified from Steuber et *al*. (2016). Genera are grouped according to the clade to which they probably belong (or clades for the probably polyphyletic genus *Offneria* and *Caprina*). 326

Figure 4. 29. Detail of the structuration of the aragonitic layer in A – *Coralliochama orcutti* (USNM.186665) and B – *Oryxia sulcata* (BJ.1901-HR2; picture provided by Bogdan Jurkovšek) right valves. *Coralliochama* 'pseudocanals' are not homologous to *Oryxia*'s pallial canals. Scale bars, 10 mm. 332

Figure 4. 30. Sketch illustrating *Hippurites* and *Radiolites* kinship (RV). The siphonal bands of *Radiolites* develop into pillars in *Hippurites* (P1 and P2). These infoldings are the source of numerous structural deformations in Hippuritid shells: (i) the ligament moves posteriorly, (ii) the vertical posterior muscle becomes stranded between the pillar P1 and the cardinal apparatus, and (iii) the anterior muscle stretch itself and become divided into two parts. A – *Radiolites radiosus* (MNHN.F.J09021). B – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm. 335

Figure 4. 31. Number of rudist genera from Oxfordian to Maastrichtian compared to the spatial extent of carbonate plateforms of the New world (North and South America) and the old World (Europe, Africa, and Arabia; adapted from Skelton 2003 and Steuber et *al*. 2016). The arrows represent the two Hauterivian and Albian rudist radiations. 344

Figure 4. 32. Example of hierarchical character with associated captions. Scale bars, 10 mm. 348

Figure 4. 33. Hierarchical character #1. A – *Diceras arietinum* (MNHN). B – *Mexicaprina minuta* (JPMA.14097-5). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Clinocaprina lowentsami* (USNM.245242). E – *Clinocaprina lowentsami* (USNM.245241). Scale bars, 10 mm. 352

Figure 4. 34. Hierarchical character #2. A – *Diceras arietinum* (MNHN). B – *Valletia germani* (MNHN.F.J05798). Scale bars, 10 mm. 354

Figure 4. 35. Hierarchical character #3. A – *Valletia germani* (MNHN.F.J05798). B – *Agriopleura marticensis*. C – *Radiolites radiosus* (MNHN.F.J09021; picture provided by Jean-Pierre Masse). Scale bars, 10 mm. 356

Figure 4. 36. Hierarchical character #4, its character states intensions, and its treatment with paralogy-free subtree analysis. 360

Figure 4. 37. Hierarchical character #4.1. A – *Diceras arietinum* (MNHN). B – *Valletia germani* (MNHN.F.J05798). C – *Plagioptychus arnaudi* (UPMC.1273). D – *Caprotina costata* (MV 2003.2.5147). Scale bars, 10 mm. 362

Figure 4. 38. Hierarchical character #4.2. A – *Caprotina costata* (MV 2003.2.5147). B – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). Scale bars, 10 mm. ... 364

Figure 4. 39. Hierarchical character #4.3. A – *Caprotina costata* (MV 2003.2.5147). B – *Mexicaprina minuta* (JPMA.14097-5). C – *Radiolites radiosus* (MNHN.F.J09021). D – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm. 366

Figure 4. 40. Hierarchical character #5. A – *Mexicaprina minuta* (JPMA.14097-5). B – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm. 368

Figure 4. 41. Hierarchical character #6, its character states intensions, and its treatment with paralogy-free subtree analysis. 372

Figure 4. 42. Hierarchical character #6.1. A – *Diceras arietinum* (MNHN). B – *Caprotina costata* (MV 2003.2.5147). Scale bars, 10 mm. 374

Figure 4. 43. Hierarchical character #6.2. A – *Caprotina costata* (MV 2003.2.5147). B – *Coralliochama orcutti* (USNM.186662). C – *Plagioptychus arnaudi* (UPMC.1273) Scale bars, 10 mm. 376

Figure 4. 44. Hierarchical character #6.3. A – *Caprotina costata* (MV 2003.2.5147). B – '*Sellaea' sp.* (WSA3058). C – *Mexicaprina minuta* (JPMA.14097-5). D – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm. 378

Figure 4. 45. Hierarchical character #7. A – *Chaperia costata* (MNHN.F.R07950). B – *Radiolites radiosus* (MNHN.F.J09021). Scale bars, 10 mm.....380

Figure 4. 46. Hierarchical character #8. A – *Diceras arietinum* (MNHN). B – *Monopleura michaillensis* (UPMC.1272). C – *Plagioptychus arnaudi* (UPMC.1273). Scale bars, 10 mm.....382

Figure 4. 47. Hierarchical character #9. A – *Monopleura michaillensis* (UPMC.1272). B – *Hippurites radiosus* (MNHN.F.A58757). Scale bars, 10 mm. 384

Figure 4. 48. Hierarchical character #10. A – *Plagioptychus arnaudi* (UPMC.1273). B – *Titanosarcolites giganteus* (UT.47533). C – *Curtocaprina clabaughikinsorum* (UT.10937). Scale bars, 10 mm. 386

Figure 4. 49. Hierarchical character #11. A – *Monopleura michaillensis* (UPMC.1272). B – *Caprotina costata* (MV 2003.2.5147). Scale bars, 10 mm. 388

Figure 4. 50. Hierarchical character #12. A – *Amphitriscoelus waringi*. B – *Offneria rhodanica* (UCBL.R275). Scale bars, 10 mm.....390

Figure 4. 51. Hierarchical character #13. A – *'Sellaea' sp.* (WSA3058). B – *Amphitriscoelus waringi*. C – *Titanosarcolites giganteus* (UT.47533). D – *Schiosa ramosa* (R485, Université Claude Bernard Lyon 1; Coalcoman, Mexico). E – *Coralliochama orcutti* (USNM.186662). Scale bars, 10 mm. 392

Figure 4. 52. Hierarchical character #14. A – *Offneria rhodanica* (JPMA.16494). B – *Kimbleia capacis* (PUAB-81521). C – *Mexicaprina minuta* (JPMA.14097-5). D – *Caprinula boissyi* (MNHN.6924). E – *Caprinuloidea sp.* (UT.10922-2). Scale bars, 10 mm. 394

Figure 4. 53. Hierarchical character #15. A – *Monopleura michaillensis* (UPMC.1272). B – *Caprinula boissyi* (MNHN.6924). Scale bars, 10 mm. 396

Figure 4. 54. Hierarchical character #16. A – *Mexicaprina minuta* (JPMA.14097-5). B – *Coralliochama orcutti* (USNM.186662). Scale bars, 10 mm. 398

Figure 4. 55. Hierarchical character #17. A – *Amphitriscoelus waringi*. B – *Titanosarcolites giganteus* (UT.47533). Scale bars, 10 mm. 400

Figure 4. 56. Hierarchical character #18. A – *Caprinula boissyi* (MNHN.6924). B – *Offneria rhodanica* (JPMA.16494). Scale bars, 10 mm. 402

Figure 4. 57. Hierarchical character #19. A – *Amphitriscoelus waringi*. B – *Offneria rhodanica* (UCBL.R275). Scale bars, 10 mm.....404

Figure 4. 58. Hierarchical character #20. A – *Radiolites radiosus* (MNHN.F.J09021; umbonal view). B – *Hippurites radiosus* (MNHN.F.A58757; umbonal view). Scale bars, 10 mm. 406

Figure 4. 59. Hierarchical character #21. A – *Radiolites radiosus* (MNHN.F.J09021; umbonal view). B – *Hippurites radiosus* (MNHN.F.A58757; umbonal view). C – *Barrettia monilifera* (fig. 3D in Mitchell 2010; UWIGM.2010.01.0003). Scale bars, 10 mm.....408

Figure 4. 60. Hierarchical character #22. A – *Glossomyophorus costatus* (MNHN.F.R06779). B – *Hippurites radiosus* (MNHN.F.A58757). C – *Radiolites radiosus* (MNHN.F.J09021). Scale bars, 10 mm. 410

Figure 4. 61. Hierarchical character #23. A – *Diceras arietinum* (MNHN.F.A52439). B – *Monopleura michaillensis* (UPMC.1272). Scale bars, 10 mm. 412

Figure 4. 62. Hierarchical character #24. A – *Monopleura michaillensis* (UPMC.1272). B – *Mexicaprina minuta* (JPMA.14097-1). Scale bars, 10 mm.....414

Figure 4. 63. Hierarchical character #25. A – *Monopleura michaillensis* (UPMC.1272). B – *Plagioptychus arnaudi* (UPMC without number). Scale bars, 10 mm. 416

Figure 4. 64. Hierarchical character #26, its character states intensions, and its treatment with paralogy-free subtree analysis. 420

Figure 4. 65. Hierarchical character #26.1. A – *Diceras arietinum* (MNHN.F.A52439). B – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). Scale bars, 10 mm.....422

Figure 4. 66. Hierarchical character #26.2. A – *Diceras arietinum* (MNHN.F.A52439). B – *Plagioptychus arnaudi* (UPMC.1273). Scale bars, 10 mm.....424

Figure 4. 67. Hierarchical character #26.3. A – *Plagioptychus arnaudi* (UPMC.1273). B – *Caprotina costata* (MV 2003.2.5147). C – *Offneria rhodanica* (UCBL.R781). Scale bars, 10 mm. 426

Figure 4. 68. Hierarchical character #26.4. A – *Plagioptychus arnaudi* (UPMC.1273). B – *Amphitriscoelus waringi*. C – *Titanosarcolites giganteus* (UT.47533). Scale bars, 10 mm. 428

Figure 4. 69. Hierarchical character #27. A – *Chaperia costata* (MNHN.F.R07942). B – *Plagioptychus arnaudi* (UPMC.1273). C – *Mexicaprina minuta* (JPMA.14097-1). Scale bars, 10 mm. 430

Figure 4. 70. Hierarchical character #28. A – *Radiolites radiosus* (MNHN.F.J09021). B – '*Sellaea' sp.* (WSA3058). Scale bars, 10 mm. 432

Figure 4. 71. Hierarchical character #29. A – *Diceras arietinum* (MNHN.F.A52439). B – *Radiolites radiosus* (MNHN.F.J09021). C – *Chaperia costata* (MNHN.F.R07942). Scale bars, 10 mm. 434

Figure 4. 72. Hierarchical character #30. A – *Amphitriscoelus waringi*. B – *Caprinula boissyi* (MNHN.6924). Scale bars, 10 mm.....436

Figure 4. 73. Hierarchical character #31. A – *Caprotina costata* (MV 2003.2.5147). B – *Amphitriscoelus waringi*. C – *Ichthyosarcolites bicarinatus* (Upper Cenomanian, D559a road, south exit of La Bédoule, Var, France; picture provided by Marc Floquet). Scale bars, 10 mm. 438

Figure 4. 74. Hierarchical character #32. A – *Diceras arietinum* (MNHN.F.A52439). B – *Titanosarcolites giganteus* (UT.47533). C – *Curtocaprina clabaughikinsorum* (UT.10922). Scale bars, 10 mm. 440

Figure 4. 75. Hierarchical character #33. A – '*Sellaea*' *sp.* (UT.48167). B – *Hippurites sp.* (UPMC without number). Scale bars, 10 mm. 442

Figure 4. 76. Hierarchical character #34. A – *Caprinula boissyi* (MNHN.6924). B – *Caprinuloidea sp.* (JPMA.without number). Scale bars, 10 mm. 444

Figure 4. 77. Hierarchical character #35. A – *Polyconites verneuili* (Ch02 in Pascual-Cebrian et *al*. 2013). B – *Glossomyophorus costatus* (MNHN.F.R06779). C – *Hippurites radiosus* (MNHN.F.A58757). D – *Barrettia monilifera* (UWIGM.RUD.2006.223). E – *Tepeyacia corrugata* (CAS.66657.01). Scale bars, 10 mm. 446

Figure 4. 78. Hierarchical character #36. A – *Glossomyophorus costatus* (MNHN.F.R06779). B – *'Sellaea' sp.* (WSA3058). C – *Radiolites radiosus* (MNHN.F.J09021). D – *Hippurites radiosus* (MNHN.F.A58757). E – *Dictyoptychus morgani* (NHM.L-58427; picture provided by Peter W. Skelton). Scale bars, 10 mm. ..448

Figure 4. 79. Hierarchical character #37. A – *Valletia germani* (MNHN.F.J05798). B – *Glossomyophorus costatus* (MNHN.F.A27280). C – *Caprinula boissyi* (MNHN.6924). D – *Clinocaprina lowentsami* (USNM.245241). E – *Curtocaprina clabaughikinsorum* (UT.10937). Scale bars, 10 mm. 450

Figure 4. 80. Hierarchical character #38. A – *Glossomyophorus costatus* (MNHN.F.A27280). B – *Radiolites radiosus* (MNHN.F.J09021). Scale bars, 10 mm...452

Figure 4. 81. Hierarchical character #39. A – *Diceras arietinum* (MNHN.F.A52439). B – *Curtocaprina clabaughikinsorum* (UT.10922). C – *Chaperia costata* (MNHN.F.R07942). D – *Tepeyacia corrugata* (CAS.66657.01). Scale bars, 10 mm. 454

Figure 4. 82. Hierarchical character #40. A – *Glossomyophorus costatus* (MNHN.F.R06779). B – *Ichthyosarcolites bicarinatus* (MGUP.021.2-110; picture provided by Carolina d'Arpa and Tony Costaglia). C – *Curtocaprina clabaughikinsorum* (UT.10937). D – *Oryxia sulcata* (BJ.1901-HR2). Scale bars, 10 mm.....456

Chapitre 5

Figure 5. 1. Carte paléogéographique de l'Europe de l'Ouest au Cénomanien supérieur - Turonien, modifiée à partir de Floquet et Hennuy 2001. Marron : terre émergée ; jaune : côtier ; bleu clair : plate-forme carbonatée ; vert : plate-forme externe ; bleu foncé : marin profond.....463

Figure 5. 2. Carte géologique de la région du Bassin Sud-Provençal (actuel synclinal du Beausset), Sud-Est de la France. La ville de La Ciotat est indiquée en point de repère, ainsi que les deux localités étudiées de la Barre des Aiguilles et de la Bastide d'Orves. Echelle : 2 km. Modifié à partir des cartes géologiques à 1.50000 du BRGM Aubagne-Marseille et La Ciotat.....464

Figure 5. 3. Coupe de la Bastide d'Orves avec les unités stratigraphiques Fontblanche 1 (Cénomanien supérieur) et Fontblanche 2 (Cénomanien terminal/Turonien inférieur). Les séquences S1 à S6 et les discontinuités D1 à D6 (matériel et méthodes) sont reportées ainsi que les macrocommunautés (1 à 4), la texture, la richesse spécifique, le taux de bioclastes et les valeurs de dominance pour les macrocommunautés.....466

Figure 5. 4. Coupe de la Barre des Aiguilles avec les unités stratigraphiques Fontblanche 1 (Cénomanien supérieur) et Fontblanche 2 (Fb2 ; Cénomanien terminal/Turonien inférieur). Les séquences S1 à S5 et les discontinuités D1 à D5 sont reportées ainsi que les macrocommunautés (1 à 4), la texture, la richesse spécifique, le taux de bioclastes et les valeurs de dominance pour les macrocommunautés.....467

Figure 5. 5. Taxons rencontrés à la Bastide d'Orves et à la Barre des Aiguilles. A – *Sauvagesia sharpei* Bayle 1857 (Radiolitidae, rudiste); B – *Apricardia* sp. (Requienidae, rudiste), C – *Caprotina* sp. (Caprotinidae, rudiste); D – *Ichthyosarcolites bicarinatus* Gemmellaro 1865 (Ichthyosarcolitidae, rudiste); E – *Chondrodonta joannae* Choffat 1886 (Chondrodontidae); F – Nerinea sp. (Nerineidae). Crédit photographique (photo F) : Marc Floquet et Aurélien Blénet. Echelle: 1 cm.....468

Figure 5. 6. Taxons rencontrés à la Bastide d'Orves et à la Barre des Aiguilles. A – *Caprinula boissyi* Orbigny 1840 (Caprinulidae, rudiste) ; B – Chaetetidae (Porifera), C – Cyclostomata (Bryozoa), D – Cidaroida (Echinoidea) ; E1 – Section transversale de valve droite de rudiste à canaux portant des tubes cassés. Echelle : 1 cm ; E2 – Schéma interprétatif de E1. Les carènes sont situées sur le côté postérieur (à droite). Le côté antérieur épaissi montre des tubes, cavités de taille plus importante que les canaux palléaux résultant de l'anastomose de carènes. Ces caractéristiques ne sont rencontrées que chez les Antillocaprinidae. Crédits photographiques : Marc Floquet et Aurélien Blénet. Echelle: 1 cm.....469

Figure 5. 7. Pourcentages d'occurrence des taxons à la Bastide d'Orves. Les fréquences des taxons sont reportées pour chaque point de comptage macroscopique. Gris clair : Sauvagesiinae ; marron : *Caprinula* ; vert : *Ichthyosarcolites* ; gris foncé : *Apricardia* ; jaune : *Caprotina* ; bleu clair : *Chondrodonta* ; bleu marine : *Nerinea* ; beige : *Scleractinia*. 472

Figure 5. 8. Pourcentages d'occurrence des taxons à la Barre des Aiguilles. Les fréquences des taxons sont reportées pour chaque point de comptage macroscopique. Gris clair : Sauvagesiinae ; vert : *Ichthyosarcolites* ; gris foncé : *Apricardia* ; jaune : *Caprotina* ; bleu clair : *Chondrodonta* ; bleu marine : *Nerinea* ; beige : Scleractinia ; rose : Chaetetidae ; cyan : Cidaroida. 473

Figure 5. 9. Classification Ascendante Hiérarchique sur les fréquences des comptages de taxons, et les fréquences correspondant aux points de comptage. Quatre communautés sont discriminées : une communauté à *Caprotina* dominants, une communauté à *Nerinea* dominants, une communauté à *Chondrodonta* dominants accompagnées de *Ichthyosarcolites*, *Caprotina* et Sauvagesiinae, et une communauté à *Apricardia* et Sauvagesiinae dominants avec une grande variabilité entre les proportions des deux taxons dominants. Gris clair : Sauvagesiinae ; marron : *Caprinula* ; vert : *Ichthyosarcolites* ; gris foncé : *Apricardia* ; jaune : *Caprotina* ; bleu clair : *Chondrodonta* ; bleu marine : *Nerinea* ; beige : Scleractinia ; rose : Chaetetidae ; cyan : Cidaroida. 474

Figure 5. 10. En haut, représentation des communautés biotiques par Positionnement Multidimensionnel Non-Métrique permettant de visualiser les similarités entre les points de comptage des communautés proposées dans un espace en deux dimensions (stress : 0,129). En bas, illustration des quatre communautés biotiques retenues. A – communauté à *Chondrodonta* (Ch) dominants et *Ichthyosarcolites* ; B – communauté à Sauvagesiinae (S) et *Apricardia* (A) dominants ; C – communauté à *Caprotina* (Ca) dominants accompagnés de *Ichthyosarcolites* (I); D – communauté à *Nerinea* (N) dominants. Crédits photographiques (photos A, D) : Marc Floquet et Aurélien Blénet. Echelle: 5 cm.....476

Figure 5. 11. Diagramme de variation du taux de bioclastes en fonction de la communauté macroscopique renseignée. Les quatre communautés ne sont quasiment pas affectées par le taux de bioclastes et partagent une médiane très similaire. 477

Figure 5. 12. Occurrences de taxons reconnus en lames minces de calcaires de la coupe de la Bastide d'Orves. La diversité semble constante dans toutes les séquences S1 à S6. 478

Figure 5. 13. Occurrences de taxons reconnus en lames minces de calcaires de la coupe de la Barre des Aiguilles. La biocénose est similaire à celle reconnue à partir de la coupe de la Bastide d'Orves. Toutefois, à partir de b33, la biodiversité s'effondre et quasiment aucun autre taxon de rudistes que celui des Radiolitidae voire de foraminifère benthique, ne sont reconnaissables. 479

Figure 5. 14. Taxons rencontrés en lames minces de calcaires des coupes de la Bastide d'Orves et de la Barre des Aiguilles. A – Radiolitidae, B – *Ichthyosarcolites bicarinatus* Gemmellaro 1865 (Ichthyosarcolitidae, rudiste), C – *Apricardia* sp. (Requienidae, rudiste), D – *Chondrodonta joannae* Choffat 1886 (Chondrodontidae), E – Gasteropoda et Radiolitidae, F – *Polystrata alba* (Rhodophyta), G – *Bacinella irregularis* (Cyanobacteria), H – Bryozoa, I – Chaetetidae (Demosponga), J – *Lithocodium* sp. (Cyanobacteria), K – *Acicularia* sp. (Ulvophyceae)), L – *Solenomeris* sp.? (foraminifère benthique). Echelle: 500 µm. 480

Figure 5. 15. Foraminifères (A à S) et kyste de dinoflagellé (T) rencontrés en lames minces de calcaires des coupes de la Bastide d'Orves et de la Barre des Aiguilles. A – *Triloculina* sp., B – *Quinqueloculina* sp., C – *Pseudolituonella* sp., D – *Spirocyclina* sp., E – *Textularia* sp., F – *Nezzazatinella picardi*, G – *Cuneolina pavonia*, H – *Charentia* sp., I – *Chrysalidina gradata*, J – *Praealveolina gr. cretacea*, K – *Dictyopsella* sp., L – *Planispira* sp., M – *Rotalia mesogeensis*, N – *Tritaxia pyramidata*, O – *Frondicularia sp.*, P – *Heterohelix globulosa*, Q-R – *Muricohedbergella sp.*, S – *Whiteinella sp.* ?, T – *Pithonella sphaerica*. Echelle : 100µm. 481

Figure 5. 16. Faciès de S1 de la coupe de la Bastide d'Orves. A – Surface de sciage avec Sauvagesiinae (S), stylolithes (Styl.). B – Liant en lame mince de a2, composé de bioclastes hétérométriques dans de la micrite microbioclastique. G : coquille de

gastéropode avec un remplissage géotrope; S : test celluleux d'un Radiolitidae ; C : Chaetetidae. Echelles : 10 cm (A), 1 mm (B). 483

Figure 5. 17. Faciès de S2 de la coupe de la Bastide d'Orves. A – Surface de sciage avec Sauvagesiinae (S) majoritaires accompagnées d'*Apricardia* (A), stylolithes (Styl.). B – Liant en lame mince du floatsone a4, composé de bioclastes hétérométriques dans de la micrite microbioclastique. S : test celluleux d'un Sauvagesiinae. Echelles : 10 cm (A), 1 mm (B). 484

Figure 5. 18. Faciès de S3 de la coupe de la Bastide d'Orves. A – Surface de sciage avec Sauvagesiinae (S) majoritaires, remplissage laminé et géotrope des coquilles et de paléofractures (F). B – Liant en lame mince de a15, fait d'une biomicrite à bioclastes hétérométriques fragmentés. Echelles : 10 cm (A), 1 mm (B). 485

Figure 5. 19. Faciès de S4 de la coupe de la Bastide d'Orves. A – Surface de sciage avec *Chondrodonta* (C) et Sauvagesiinae (S), stylolithes (Styl.). B – Liant en lame mince de a19 composé d'une biomicrite à bioclastes hétérométriques (des microbioclastes au bioclastes de tailles centimétriques bioperforés et biocorrodés (C)). Echelles : 10 cm (A), 1 mm (B). 486

Figure 5. 20. Faciès de S5 de la coupe de la Bastide d'Orves. A – Surface de sciage avec Sauvagesiinae (S) dominantes accompagnées de *Caprinula* (Cu). B – Liant en lame mince de a23 composé d'une biomicroparite à bioclastes hétérométriques bioperforés, biocorrodés, micritisés (Pe : peloïde) et avec foraminifères benthiques (Pl : *Planispira* sp., Qu : *Quinqueloculina* sp.). Echelles : 10 cm (A), 1 mm (B). 487

Figure 5. 22. Faciès de F6 de la coupe de la Bastide d'Orves. A – Surface de dissolution du front de carrière montrant la succession du bas en haut de S6 de floatstones avec dominance des *Apricardia* (A ; S6a) vers des floatstones à Sauvagesiinae dominantes (S ; S6b). B – Liant en lame mince de a27 composé de biomicrite à bioclastes hétérométriques (microbioclastes fragmentés à bioclastes de tailles centimétriques cassés et biocorrodés (G : gastéropode ; fragments ambrés de coquilles d'*Apricardia*). Echelles : 10 cm (A), 1 mm (B).....488

Figure 5. 22. A – Discontinuité de troncature D6 au toit de S6 de la coupe de la Bastide d'Orves (floatstone à Sauvagesiinae (S)) et marquant le sommet de l'unité

stratigraphique Fontblanche 1 (Cénomanien supérieur). Calcaire noduleux de texture packstone bioclastique fin et recristallisé (biomicrosparite à fins débris dont bryozoaires, plus Pithonelles et *Heterohelix*…, grains de glauconie (G) et de quartz) et attribué à l'unité stratigraphique Fontblanche 2 (Cénomanien final à Turonien basal). Echelles : 10 cm (A), 1 mm (B). 488

Figure 5. 23. Communautés de la séquence 1 dans la coupe de la Barre des Aiguilles. A – Bouquet de *Caprotina* conservé en position de vie (communauté de S1a). B – Succession des communautés de S1a (*Caprotina* et *Ichthyosarcolites*), S1b (*Chondrodonta* probablement accumulées sous l'action de tempête), S1c (*Caprotina*, *Ichthyosarcolites* et *Nerinea*). C – Communauté S1d à *Ichthyosarcolites* (Ich), *Caprotina* et Sauvagesiinae (S). D – Accumulation de *Nerinea* (N) (probablement sous l'action de tempête) = communauté de S1f. E – Association à Sauvagesiinae (S) et *Chondrodonta* (Ch). F – Liant microbioclastique et micritique du floatstone b14, avec bioclastes hétérométriques bioperforés (Bp) et biocorrodés, à foraminifères benthiques (P : *Praealveolina gr. cretacea*). Echelles : 10 cm (A, B, E), 1 mm (F), 1 cm (C). 490

Figure 5. 24. Faciès de S2 de la coupe de la Barre des Aiguilles. A – Surface de sciage à *Apricardia* (A) majoritaire, nombreux stylolithes (Styl.). B – Liant en lame mince de b24, de biomicite à biomicrosparite à foraminifères benthiques (Ps : *Pseudolituonella* sp., D : *Dicyclina* sp., Nz : *Nezzazatinella picardi*, Qu : *Quinqueloculina* sp.), à bioclastes hétérométriques bioperforés et biocorrodés. Echelles : 10 cm (A), 1 mm (B). 491

Figure 5. 25. Faciès de S3 de la coupe de la Barre des Aiguilles. A – Surface de sciage à Sauvagesiinae (S) et *Apricardia*. B – Liant en lame mince de b27, fait de bioclastes hétérométriques (microbioclastes et débris de Radiolitidae de taille centimétrique) bioperforés et biocorrodés (C). Echelles : 10 cm (A), 1 mm (B)..... 492

Figure 5. 26. Faciès de S4 de la coupe de la Barre des Aiguilles. A – Surface de sciage montrant une population très dense quasiment uniquement constituée d'*Apricardia* (A). B – Liant en lame mince de b28, de micrite dense à microbioclastes et foraminifères benthiques (Qu : *Quinqueloculina* sp., T : *Textularia* sp., Cu : *Cuneolina pavonia*), stylolithes (Styl). Echelles : 10 cm (A), 1 mm (B). 493

Figure 5. 27. Différents faciès de S5 de la coupe de la Barre des Aiguilles. A – Communauté S5c à Sauvagesiinae (S) et *Chondrodonta* (Ch). B – Accumulation de valves gauches operculaires de Sauvagesiinae (communauté de S5a). C – Biostrome à Sauvagesinae (S) en position de vie (communauté de S5b). D – Lame mince de b35 ; floatstone à gros débris de Sauvagesiinae (S) dans un liant essentiellement fait de bioclastes (de rudistes) hétérométriques, fragmentés et peu ou non biocorrodés, avec nombreux péloïdes de micritisation (Pf), et très peu de micrite. Echelles : 10 cm (A à C), 1 mm (B). 494

Figure 5. 28. A – Bioherme à Radiolitidae (probables Sauvagesiinae) conservés en position de vie. Les planchers de la couche calcitique des valves droites récipientes (cernées d'un trait) sont bien visibles. B – Tracés des individus de A (D : cavité générale de la valve droite). La flèche indique le haut. Echelle : 1 cm. 495

Figure 5. 29. A – Limite, marquée par la discontinuité D5, entre S5 (sommet de l'unité Fontblanche 1) du haut de la coupe de la Barre des Aiguilles, composée majoritairement de Sauvagesiinae (S) et les formations surincombantes de calcaire quartzeux dépourvu de rudistes. B – Lame mince de b40 pris dans la formation supra-Fb1 montrant une micrite très riche en quartz (Q, 30 à 40 %, de 300 µm à 1 mm de taille), faciès caractéristique de l'unité A4 au sens d'Hennuy (2003) d'âge Turonien final et en discordance cartographique sur la série cénomanienne de la Barre des Aiguilles (Hennuy 2003, planche 8). Echelles : 10 cm (A), 1 mm (B). 496

Table des tableaux

Chapitre 1

Table 2. 1. . Statistical significance of differences in resolving power and artefactual resolution (calculated through mean ITRI) between methods, outgroup branch lengths, and topology of the reference tree. The probabilities were calculated using Wilcoxon signed-rank tests for the comparisons between methods and with a Mann-Whitney test for the outgroup branch length and the shape of the reference tree. Non-significant results with α=0.05 are shown in blue, and non-significant results after correction for multiple tests using the false discovery rate (Benjamini and Hochberg, 1995) are shown in brown. Individual values can be found in S7. 180

Table 2. 2. Impact of outgroup branch length (values of 0, 1/10, 1/3, 1/2, 2/3, 1, expressed as a proportion of total tree depth, were tested) on resolving power and artefactual resolution. Probabilities of the null hypothesis (no impact) were tested through linear regressions of ITRI means on outgroup branch length. These results were obtained from the tree shown in figure 2.15D. Resolving power, slope: -7.2 (ordered), -10.1 (unordered), -13.5 (3ta); intercept (1000 indicates complete power or error): 991.9 (ordered), 675.3 (unordered), 989 (3ta). Artefactual resolution, slope: 4.6 (ordered), 18.3 (unordered), 19.8 (3ta); intercept: 28.9 (ordered), 160 (unordered), 30.8 (3ta). Individual values can be found in S7. 183

Table 2. 3. Impact of the number of states on MP and 3ta efficiency. Probabilities of the null hypothesis (no impact of number of states on efficiency) were tested through linear regressions. 201

Table 2. 4. Impact of errors in the ordering scheme on MP and 3ta efficiency, based on characters with 10 states. Probabilities of the null hypothesis (no impact) were tested through linear regressions. Two kinds of errors were tested: minor (scenario 1) and major (scenario 2) errors. See figure 2 for explanations of both scenarios. Unordered parsimony for these data yields an efficiency of 69.0%. 203

Table 2. 6. Impact of reversals in the evolutionary model on MP and 3ta. An irreversible model (matrices modeled without reversals and binary characters) is analysed with both methods, and compared to Brownian motion with binary characters. A Wilcoxon test is used to test the significance of the difference of the results obtained with ordered MP and 3ta. Given that the characters used for this simulation are binary, there is no freedom in ordering scheme and hence, no difference between ordered and unordered MP. 205

Table 2. 5. Impact of homoplasy by random errors on phylogenetic reconstruction on ordered MP, unordered MP and 3ta. Random errors are generated on matrices generated with Brownian motion on four states and compared to analyses of the same matrices without random error. A Wilcoxon test is used to test the significance between ordered MP and 3ta. 205

Table 2. 8. Empirical test of cladistic analysis by unordered MP and 3ta of HIV virion. The analysis is made with two alternative outgroups, a close HIV and a much more remote SIV. Results present the number of trees for each analysis, the number of steps for MP (the concept being meaningless in 3ta), the retention index (RI) and the efficiency composed of TR and FR. 206

Table 2. 7. Empirical test of cladistic analysis by MP and 3ta of copied manuscripts. Results present the number of trees for each analysis, the number of steps for MP, the retention index (RI) of the characters on the obtained tree, and the efficiency composed of TR and FR. 206

Chapitre 3

Table 3. 1. Comparison of the known morphological features and geographic distribution of all 19 nominal *Ichthyosarcolites* species, plus the herein recognized taxon *Oryxia sulcata* nov. gen. nov. sp. “Unknown” means that the information was unavailable from the literature or the surveyed specimens. 232

Chapitre 4

Table 4. 1 List of specimens studied for the phylogenetic analysis. For each specimen is specified the family, species, reference, status, publication and type of material when available. 275

Table 4. 2. Glossary of terms used for description of rudist morphology in the present study, their description and their link to other terms that can be found in the literature (Douvillé, 1935; Skelton & Smith 2000; Skelton 2013; Chartrousse, 1998)...289

Table 4. 3. Overview of the general myophoral organization (PM-LV, AM-LV, PM-RV, AM-RV) of rudists from our sampling. For each species are displayed the myophoral characters. The last column is the morphotype number of the species. There is a number for each combination of myophoral states that are displayed in Figure 4.20. 303

Table 4. 4. Matrix of 3ta characters with their hierarchical coding and their assignation to OTUs of this analysis.....315

Table 4. 5. Overview of the results of the character testing procedure using 3ta. List of hierarchical characters with their localization (LV, RV or both), their number in this paper, their kind, their number of informative states (i. e. excluding the root), their retention index, and the character states numbers that pass the congruence test and become synapomorphies on the cladogram.....319

Table 4. 6. Table of families and superfamilies following Skelton 2013 and their phylogenetic status (monophyletic, paraphyletic, polyphyletic, or not tested) under Skelton and Smith phylogenetic test, under Skelton 2013 presumptions on the rudist taxonomy, and under our phylogenetic analysis. Empty lines imply monophyly.....323

Table des matières

Remerciements vii

Sommaire x

Introduction générale 1

PARTIE 1. ANALYSE CLADISTIQUE 7

Chapitre 1. Cladistique et analyse à trois éléments 9

1. Introduction 11

2. Théorie et méthode cladistique 14

2.1. Théorie cladistique 14
2.2. Méthode cladistique 24

3. Reconstruction phylogénétique 30

3.1. Hypothèses 31
3.1.1. Unité taxonomique 31
3.1.1.1. Constituants 31
3.1.1.2. Structure 33
3.1.2. Homologie 34
3.1.2.1. Structure 35
3.1.2.2. Arguments 41
3.1.3. Caractère 46
3.1.3.1. Standardisation 46
3.1.3.2. Polymorphisme 48
3.2. Cladogramme 55
3.2.1.1. Propriétés des triplets 56
3.2.1.2. Décomposition 59
3.2.1.2.1. Caractères binaires 59
3.2.1.2.2. Caractères à plus de deux états 61
3.2.1.2.3. Le triplet comme unité informationnelle 63
3.2.1.3. Recomposition 67
3.2.1.3.1. Méthode du cladogramme vers les triplets 67
3.2.1.3.2. Méthode des triplets vers le cladogramme 68
3.2.1.3.3. Méthode mixte 73
3.3. Caractères hiérarchiques et 3ia : application au problème de Pleijel (1995) 75
3.4. Logique cladistique et la méthode de 3ia 81
3.5. Interprétation 84
3.5.1. Indice de rétention 85
3.5.2. Test des états de caractères 91
3.5.3. Datation 99

4. Consensus trees: interpretation, informativity and efficiency 107

4.1. Abstract 107
4.2. Introduction 108
4.3. Phylogenetic Information 108

4.4. Aim of consensus trees ... 110
4.4.1. Treatment of phylogenetic information ... 110
4.4.2. Hard aim of consensus trees ... 113
4.4.3. Soft aim of consensus trees ... 117
4.5. Informativeness of consensus trees ... 118
4.6. Comparison of consensus methods ... 119
4.6.1. Majority rule Methods ... 121
4.6.2. Strict retention Methods ... 124
4.6.2.1. Strict consensus ... 124
4.6.2.2. Adams consensus and triad consensus ... 124
4.6.2.3. Reduced cladistic consensus ... 128
4.7. Conclusion ... 130

5. Conclusion du chapitre ... 131

Chapitre 2. Analyse à trois éléments et parcimonie : éléments de comparaison ... 135

1. Introduction ... 137

2. Méthode de parcimonie et 3ia : comparaison théorique et méthodologique ... 138

2.1. Cohérence théorique ... 138
2.1.1. Pas de transformation versus caractères hiérarchiques ... 138
2.1.2. « *Optimisation* » : quelques problèmes découlant de l'incohérence partition/hiérarchie 141
2.1.2.1. Principe d'analyse logique et l'hypothèse de réversion en parcimonie ... 141
2.1.2.2. Données manquantes, données non-applicables et polymorphisme ... 144
2.1.2.3. Ambiguïtés d'optimisation ... 146
2.2. Adéquation empirique ... 147
2.2.1. Nullius in verba ... 147
2.2.2. Indice de rétention et contenu informatif ... 149
2.3. Étendue du pouvoir explicatif des méthodes cladistiques ... 154
2.3.1. Analyse à trois aires ... 155
2.4. Analyse de parcimonie de Brooks ... 159

3. Experimental systematics: sensitivity of cladistic methods to polarization and character ordering schemes ... 161

3.1. Introduction ... 161
3.2. Material and methods ... 171
3.2.1. Simulated character sets ... 171
3.2.1.1. Reference topologies ... 171
3.2.1.2. Reference trees ... 173
3.2.1.3. Simulated matrices ... 173
3.2.1.4. Character coding ... 173
3.2.2. Tree searches ... 174
3.2.3. Tree comparisons ... 174
3.2.4. Testing the results ... 176
3.3. Results ... 176
3.4. Discussion ... 179
3.4.1. Brownian motion in cladistics ... 179
3.4.2. Reversals in phylogenetics ... 179
3.4.3. Uncertain polarization ... 185
3.4.4. Character states and ordering schemes ... 185
3.4.5. Tree shape and branch length ... 186
3.4.6. Implications on simulation-based studies and evolutionary models ... 187

3.5. Conclusion 189

4. Impact of errors on cladistic inference: simulation-based comparison between parsimony and three-taxon analysis 191

4.1. Introduction 191
4.2. Material and methods 195
4.2.1. Simulated character sets 195
4.2.1.1. Reference tree 195
4.2.1.2. Simulations and matrix coding 196
4.2.1.3. Tree searches 198
4.2.1.4. Tree comparisons 198
4.2.1.5. Test of the significance of the results 199
4.2.2. HIV character set 199
4.2.3. Copied manuscript character set 200
4.3. Results 200
4.3.1. Ordering and number of states 200
4.3.2. Ordering and errors 202
4.3.3. Irreversible characters 203
4.3.4. Uncorrelated errors 203
4.3.5. Empirical tests 204
4.4. Discussion 205
4.4.1. Effect of the number of states 205
4.4.2. Errors in ordering schemes 207
4.4.3. Effect of evolutionary reversals and convergences 208
4.4.4. Random errors, homoplasy and phylogenetic inference 211
4.4.5. Empirical studies 212
4.5. Conclusion 213

5. Conclusion du chapitre 215

PARTIE 2. LES RUDISTES: ANATOMIE COMPARÉE, PHYLOGÉNIE, SYSTÉMATIQUE, PALÉOÉCOLOGIE 219

Chapitre 3. Taxonomic revision of the genus *Ichthyosarcolites* Desmarest, 1812 (Bivalvia, Hippuritida): an example of high intraspecific variability 221

1. Introduction 223

2. Material and methods 225

2.1. Studied material 225
2.2. Morphological analysis 229

3. Morphology 235

3.1. Pallial canals 235
3.2. Tabulae 235
3.3. Shell coiling 236
3.4. Ligament 236
3.5. Flanges 237
3.6. General cavity 238

4. Systematic paleontology 239

4.1. Genus *Ichthyosarcolites* Desmarest, 1812 240
4.1.1. Diagnosis 240
4.1.2. Remarks 240

4.1.3. *Ichthyosarcolites triangularis* Desmarest, 1812 242
4.1.3.1. Type material 243
4.1.3.2. Diagnosis 243
4.1.3.3. Description 243
4.1.3.4. Remarks 244
4.1.3.5. Occurrences 245
4.1.4. *Ichthyosarcolites rotundus* Polšak, 1967 245
4.1.4.1. Type material 245
4.1.4.2. Diagnosis 246
4.1.4.3. Description 246
4.1.4.4. Remarks 247
4.1.4.5. Occurrences 247
4.1.5. *Ichthyosarcolites monocarinatus* Slišković, 1966 247
4.1.5.1. Type material 248
4.1.5.2. Diagnosis 248
4.1.5.3. Description 248
4.1.5.4. Remarks 249
4.1.5.5. Occurrences 249
4.1.6. *Ichthyosarcolites bicarinatus* Gemmellaro, 1865 249
4.1.6.1. Type material 251
4.1.6.2. Diagnosis 251
4.1.6.3. Description 251
4.1.6.4. Remarks 253
4.1.6.5. Occurrences 254
4.1.7. *Incertae sedis* species 254
4.2. Genus *Oryxia* gen. nov. 257
4.2.1. Derivation of the name 257
4.2.2. Diagnosis 257
4.2.3. Remarks 257
4.2.4. *Oryxia sulcata* sp. nov. 258
4.2.4.1. Derivation of the name 258
4.2.4.2. Type material 258
4.2.4.3. Diagnosis 258
4.2.4.4. Description 258
4.2.4.5. Remarks 260
4.2.4.6. Occurrences 260

5. Chapter conclusion 260

Chapitre 4. Homologies and the phylogeny of rudists (Bivalvia, Hippuritida): new insights from comparative anatomy 263

1. Introduction 265

2. Historical background 267

3. Material and methods 273

3.1. Taxonomic sampling 273
3.2. Homologies and characters 286
3.2.1. Approach 286
3.2.2. Rudists morphology 290
3.2.2.1. Teeth 298
3.2.2.2. Myophores 301
3.2.2.3. Cavities and pallial canals 305
3.2.2.3.1. Left valve 306

3.2.2.3.1.1. Canals ... 306
3.2.2.3.1.2. Central tooth socket ... 308
3.2.2.3.2. Right valve ... 310
3.2.2.3.3. Tabulae ... 311
3.2.2.4. Calcitic outer shell layer ... 311
3.2.2.5. Ligament ... 312
3.2.3. Characters ... 313
3.2.4. Parameters of the phylogenetic analyses ... 314

4. Results ... 317

5. Discussion ... 322

5.1. Cladogram ... 322
5.1.1. Clade 1 ... 325
5.1.2. Clade 2 ... 325
5.1.3. Clade 3 ... 328
5.1.4. Clade 4 ... 329
5.1.5. Clade 5 ... 330
5.1.6. Clade 6: Polyconitidae ... 330
5.1.7. Clade 7 ... 331
5.1.8. Clade 8: Plagioptychidae ... 331
5.1.9. Clade 9 ... 332
5.1.10. Clade 10 ... 333
5.1.11. Clade 11 ... 333
5.1.12. Clade 12 ... 334
5.1.13. Clade 13: Radiolitidae ... 336
5.1.14. Clade 14: Hippuritidae ... 337
5.1.15. Clade 15 ... 337
5.1.16. Clade 16 ... 338
5.1.17. Clade 17 ... 339
5.1.18. Clade 18 ... 339
5.1.19. Clade 19 ... 340
5.1.20. Clade 20 ... 341
5.1.21. Clade 21 ... 342
5.1.22. Clade 22 ... 342
5.1.23. Clade 23 ... 342
5.1.24. Clade 24 ... 343
5.2. Time calibration of the tree and historical trends in rudit diversity ... 343

6. Chapter conclusion ... 345

7. Character list ... 347

7.1. Character 1 ... 350
7.2. Character 2 ... 354
7.3. Character 3 ... 356
7.4. Character 4 ... 358
7.5. Character 5 ... 368
7.6. Character 6 ... 370
7.7. Character 7 ... 380
7.8. Character 8 ... 382
7.9. Character 9 ... 384
7.10. Character 10 ... 386
7.11. Character 11 ... 388
7.12. Character 12 ... 390

7.13. Character 13 392
7.14. Character 14 394
7.15. Character 15 396
7.16. Character 16 398
7.17. Character 17 400
7.18. Character 18 402
7.19. Character 19 404
7.20. Character 20 406
7.21. Character 21 408
7.22. Character 22 410
7.23. Character 23 412
7.24. Character 24 414
7.25. Character 25 416
7.26. Character 26 418
7.27. Character 27 430
7.28. Character 28 432
7.29. Character 29 434
7.30. Character 30 436
7.31. Character 31 438
7.32. Character 32 440
7.33. Character 33 442
7.34. Character 34 444
7.35. Character 35 446
7.36. Character 36 448
7.37. Character 37 450
7.38. Character 38 452
7.39. Character 39 454
7.40. Character 40 456

Chapitre 5. Paléoécologie, paléoenvironnements et événements biosédimentaires enregistrés dans la succession de communautés à rudistes du Bassin Sud-Provençal au Cénomanien supérieur 459

1. Introduction 461
2. Cadre géologique 462
3. Matériel et méthodes 464
4. Résultats 470

4.1. Définition des communautés à rudistes 470
4.2. Taux de bioclastes et richesse spécifique 475
4.3. Stratigraphie séquentielle et successions biotiques 475
4.3.1. Bastide d'Orves 483
4.3.1.1. Séquence 1 483
4.3.1.2. Séquence 2 484
4.3.1.3. Séquence 3 485
4.3.1.4. Séquence 4 485
4.3.1.5. Séquence 5 486
4.3.1.6. Séquence 6 487
4.3.1.7. Unité Fontblanche 2 487
4.3.2. Barre des Aiguilles 489
4.3.2.1. Séquence 1 489
4.3.2.2. Séquence 2 491
4.3.2.3. Séquence 3 492

4.3.2.4. Séquence 4 492
4.3.2.5. Séquence 5 493
4.4. Couverture en discordance 495
5. Discussion 496
5.1. Paléoenvironnements et systèmes biosédimentaires associés 496
5.1.1. Unité Fontblanche 1 496
5.1.2. Unité Fontblanche 2 498
5.2. Ecotypes et communautés à rudistes 499
5.3. Succession temporelle des communautés à rudistes 501
5.3.1. Coupe de la Barre des Aiguilles 501
5.3.2. Coupe de la Bastide d'Orves 502
5.3.3. Les communautés à rudistes Cénomaniennes en méditerranée 503
6. Conclusion du chapitre 504
Conclusion générale 507
Bibliographie 513
Lexique 553
Partie 1 : Analyse cladistique 553
Part 2 : Rudist morphoanatomy 559
Abbréviations et symboles - Abbreviations and symbols 563
Partie 1: Analyse cladistique 563
Part 2 : Rudist morphoanatomy 565
Annexes 567
Table des illustrations 573
Table des tableaux 613
Table des matières 617

Un nouveau regard cladistique sur l'anatomie comparée, la phylogénie, la systématique et la paléoécologie des rudistes (Bivalvia, Hippuritida)

Résumé :
Les rudistes bivalves forment un groupe monophylétique qui s'étend dans le registre fossile de 160 à 66 millions d'années. L'objectif de cette thèse est de porter un nouveau regard sur l'histoire évolutionnaire de ce groupe dans le cadre de la théorie cladistique. Dans une première partie sont posées les bases théoriques, méthodologiques et techniques de l'analyse à trois éléments. Le concept de triplet permet de proposer des arguments sur la pertinence des arbres consensus. La méthode d'analyse à trois éléments est ensuite comparée à la méthode de parcimonie grâce à des arguments théoriques et méthodologiques, ainsi que par des simulations basées sur des modèles d'évolution. Dans une seconde partie, nous appliquons la théorie et la méthode à la reconstruction de l'histoire des rudistes. Le genre *Ichthyosarcolites* est révisé, et des analyses statistiques permettent de tester la pertinence des hypothèses d'homologie basée sur la forme de la coquille. Les hypothèses d'homologie sont revues et de nouveaux caractères basés sur les myophores et les canaux palléaux sont formalisés pour l'ensemble des Hippuritida. La nouvelle phylogénie qui en découle est cohérente avec l'âge géologique et permet de raffiner l'histoire évolutive des rudistes au Crétacé. L'étude de couches Cénomaniennes du Bassin Sud-Provençal (Var, France) nous permettent de faire un lien entre l'histoire de la diversification des rudistes et leur paléoécologie.

Mots clés : Rudistes bivalves, Hippuritida, Cladistique, Analyse à trois éléments, Systématique Phylogénétique, Paléoécologie, Paléocommunautés.

A new, cladistic insight to the comparative anatomy, phylogeny, systematics and paleoecology of rudists (Bivalvia, Hippuritida)

Abstract :
The rudist bivalves form a monophyletic group that extends in the fossil record from 160 to 66 million years. The objective of this thesis is to propose new insights at the evolutionary history of this group within the framework of cladistic theory. In the first part, we lay the theoretical, methodological and technical foundations of the three-item analysis. The concept of triplet allows us to propose arguments on the relevance of consensus trees in cladistics. The three-item analysis method is then compared to the method of parsimony using theoretical and methodological arguments, and simulations based on evolutionary models. In a second part, we apply the theory and method to the reconstruction of the rudist history. The genus *Ichthyosarcolites* is reviewed, and statistical analyses are used to test the relevance of homology hypotheses based on shell shape. Hypotheses of homology and new characters based on myophores and pallial canals are tested on the Hippuritida. The new, resulting phylogeny is consistent with known geological occurrences of the group and further our understanding of rudist evolutionary history during the Cretaceous. The study of Cenomanian outcrops of the South Provence Basin (Var, France) allows us to make a link between the history of the diversification of rudists and their paleoecology.

Keywords : Rudist bivalves, Hippuritida, Cladistics, Three-item analysis, Phylogenetics Systematics, Paleoecology, Paleocommunities.