



HAL
open science

Les cténophores : de leur position dans l'arbre des métazoaires (approche phylogénomique) à leur diversité taxonomique (phylogénie moléculaire et anatomie comparée)

Paul Simion

► **To cite this version:**

Paul Simion. Les cténophores : de leur position dans l'arbre des métazoaires (approche phylogénomique) à leur diversité taxonomique (phylogénie moléculaire et anatomie comparée). Zoologie des invertébrés. Université Pierre et Marie Curie - Paris VI, 2014. Français. NNT : 2014PA066448 . tel-01165014

HAL Id: tel-01165014

<https://theses.hal.science/tel-01165014v1>

Submitted on 18 Jun 2015

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.



THESE DE DOCTORAT DE L'UNIVERSITE PIERRE ET MARIE CURIE – PARIS VI

École Doctorale 515 « Complexité du Vivant »

Présentée pour obtenir le grade de
Docteur de l'Université Pierre et Marie Curie par

Paul SIMION

Les cténophores :
de leur position dans l'arbre des métazoaires
(approche phylogénomique)
à leur diversité taxonomique
(phylogénie moléculaire et anatomie comparée)

Soutenance prévue le 27 novembre 2014

Devant le jury composé de :

Maximilian TELFORD, Professeur, University College London	Rapporteur
Didier CASANE, Professeur, Université Paris VII	Rapporteur
Evelyn HOULISTON, Directrice de Recherche CNRS	
Nicole BOURY-ESNAULT, Directrice de Recherche CNRS honoraire	
Frédéric DELSUC, Chargé de recherche CNRS	
Michaël MANUEL, Professeur, Université Paris VI	Directeur de thèse

Résumé

Les cténophores représentent l'un des quatre embranchements animaux extérieurs aux Bilateria. La majorité des espèces sont planctoniques et gélatineuses, et sont reconnaissables à leurs huit rangées de peignes dont le battement permet la nage. Leur systématique est encore de nos jours mal comprise. L'anatomie des cténophores offre peu de caractères aussi bien pour placer la lignée dans l'arbre des métazoaires, que pour établir les relations de parenté au sein de l'embranchement et délimiter les espèces, et jusqu'à présent les données moléculaires n'ont pas permis de résoudre ces problèmes de manière satisfaisante. L'objectif de ce travail de thèse est de contribuer à améliorer notre compréhension de l'évolution des cténophores à différentes échelles taxonomiques. A l'échelle des métazoaires, la position phylogénétique des cténophores a été abordée par une approche phylogénomique. Un effort significatif a été réalisé pour améliorer l'échantillonnage taxonomique à travers le séquençage et l'assemblage des transcriptomes de 22 espèces de non-Bilateria (cténophores, cnidaires, spongiaires). Deux jeux de données indépendants ont été analysés, le premier représentant une mise à jour d'une supermatrice existante de 128 gènes ; le second (4235 gènes) ayant été entièrement construit de novo via un protocole original comportant la mise au point de nouvelles méthodes pour traiter de manière semi-automatisée les principales sources potentielles d'artéfact (contaminations, paralogies, données manquantes). Les résultats contredisent certaines études récentes en montrant que les spongiaires et non les cténophores représentent le groupe-frère des autres métazoaires. La position exacte de ces derniers reste à ce stade incertaine (trois options se présentant suivant les analyses). A l'échelle intra-phyllétique, l'analyse d'un jeu de données comprenant les marqueurs ADNr 18S et Internal Transcribed Spacers (ITS), associée à des analyses de gènes dupliqués chez les cténophores et aux analyses phylogénomiques précédentes, a permis de résoudre une grande partie des relations phylogénétiques entre les ordres et les familles de cténophores, tout en permettant de préciser la position de la racine. Enfin, à une échelle taxonomique plus fine, une comparaison approfondie entre deux espèces du genre *Pleurobrachia* au moyen de marquages immunohistochimiques montre le potentiel de ces techniques comme source de nouveaux caractères structuraux « micro-anatomiques » à valeur diagnostique pour la délimitation et l'identification des espèces de cténophores. En conclusion, ce travail se veut une contribution au progrès de la systématique d'un embranchement encore méconnu et d'une grande importance pour la compréhension des événements anciens de l'évolution animale.

Abstract

Ctenophores are one of the four animal phyla positioned outside from the Bilateria. Most ctenophore species are planktonic and gelatinous, and are easily recognisable by their eight comb rows used for swimming. Ctenophore systematics remains nowadays poorly understood. Anatomical characters do not help much in placing them within the animal tree, and similarly the grounds for establishment of their internal phylogeny as well as delimitation of species are notoriously weak due to a paucity of informative morpho-anatomical characters. Until now, the use of molecular data has failed to improve significantly this situation. The aim of this PhD thesis was to bring a contribution to our understanding of ctenophore evolution at different taxonomic scales. At the metazoan level, the position of ctenophores was addressed using a phylogenomic approach. Taxonomic sampling was significantly improved through sequencing and assembly of transcriptomes of 22 non-bilaterian species (belonging to ctenophores, cnidarians and sponges). Two independent datasets were analysed, one consisting in an update of an existing supermatrix of 128 genes, the other one (4235 genes) having been entirely built *de novo*, thanks to a newly-devised semi-automated protocol intended to eliminate all major potential causes of artefacts (contaminations, paralogies, missing data). Results clearly contradict recent phylogenomic studies which claimed Ctenophora to be the most early-diverging animal lineage, our analyses instead supporting Porifera as the sister-group to other metazoans. The exact position of ctenophores remains however uncertain at this stage, different conditions of analyses yielding three contradictory hypotheses as open possibilities. At the intra-phyletic level, analyses of a ADNr 18S and ITS (internal transcribed spacers) dataset, together with study of duplicated genes and the results of phylogenomic analyses, allowed resolving most phylogenetic relationships between ctenophore orders and families, including the placement of the ctenophore tree root. Finally, at smaller taxonomic scale, in-depth comparison between two species of the genus *Pleurobrachia* using immunohistochemistry demonstrates the potential of these techniques for uncovering new structural “micro-anatomical” characters useful for diagnosis and identification of ctenophore species. It is our hope that this work will contribute to improving the systematics of a poorly known phylum of great importance for understanding early animal evolution.

Remerciements

Mes plus profonds remerciements vont, évidemment, à Michaël, qui m'a inconditionnellement prêté sa confiance du début à la fin de cette thèse. Au-delà de ses (exceptionnelles) compétences scientifiques, je tiens à souligner ici ce qui est important : son admirable empathie. Attentif et présent, il a su détecter l'encadrement nécessaire à ma rigueur et la liberté nécessaire à mon humeur. C'était parfait.

Mes remerciements les plus profonds (du coup, *grosso modo* à égalité avec Michaël...) vont à Muriel, dont la bienveillance, l'énergie et la disponibilité forcent l'humilité. Elle m'a aidé aux moindres besoins, et il est difficile d'imaginer la montagne de difficultés que je n'ai jamais eu à affronter grâce à elle. Mais je la remercie surtout parce qu'elle géniale. Muriel, si vous la diluez dans l'eau, ça en fait dix normales.

Je tiens à tout particulièrement remercier Hervé, d'abord pour son accompagnement tout au long de ces trois ans, ses conseils précieux et les biscuits bretons, mais également parce qu'en tant que responsable de l'école doctorale qui m'a accordé une bourse de thèse et président de mon comité de suivi, il a permis l'existence de ces recherches. Il pourra trouver dans ce manuscrit tous les arbres que j'ai trop tardé à lui montrer, je les lui doit bien !

Je remercie également, et chaleureusement, Éric pour son accompagnement durant cette thèse, ses discussions éclairées, ses questionnements stimulants, sa critique salubre, et notre haine commune des imprimantes. Les conversations avec lui sont aussi intéressantes et variées que la Clairette de Die qu'il ramène est bonne.

Je veux transmettre à Alicia, Karen, Joao et Thomas, tous mes remerciements du fond du cœur, à la fois pour les foutitudes de coups de mains dont ils m'ont fait l'honneur, mais surtout pour m'avoir continuellement accompagné dans la joie et la tendresse. Je tiens à leur dire tout le bonheur qu'ils m'apportent tous les jours depuis bientôt trois ans. Tous les jours, sans exception.

Je remercie chaleureusement tous les membres du laboratoire, et du A4 en général, avec des mentions spéciales pour les membres de l'équipe AIRE, dont l'aide m'a été précieuse, et à Dominique, Philippe et Danielle pour leur soutien administratif. Je remercie également Muriel Umbhauer et Elisabeth Clément de l'école doctorale 515 pour leur aide, et enfin je remercie bien sûr Gaëlle Boutin dont la sympathie et la compétence surpassent mes défauts administratifs, c'est dire.

Merci à Hervé Philippe et à Gert Wörheide, qui m'ont donné l'occasion de travailler avec eux, de profiter de leurs savoirs et de leur expérience. Merci de m'avoir accueilli dans vos laboratoires, ces voyages ont été pour moi extrêmement enrichissants. Merci également à Tetyana Nosenko bien sûr, ma toute première professeur de bio-informatique.

Je remercie également Alexander Ereskovsky et Frédéric Delsuc pour leur accompagnement durant ma thèse, leur participation à mon comité de suivi et les différents services qu'ils n'ont jamais hésité à me rendre.

Un grand merci à Evelyn Houliston et Olivier Gros pour leur accueil dans leurs laboratoires respectifs, et bien entendu merci aussi à Pascal, Carine, Tsuyoshi et Antonella. Je remercie

également les pêcheurs de la station biologique de Villefranche-sur-Mer pour les petits tours en bateau, et les membres de l'association de plongée Eden à Port-Louis pour les magnifiques ballades sous-marines.

Je tiens à remercier aussi les « anciens jeunes », Cyrielle, Alex et Roxanne, pour leurs visites sympathiques et leurs encouragements. Un merci tout particulier à Cyrielle qui m'a, il y a trois ans, encouragé et (facilement) convaincu de postuler pour cette thèse : c'était une drôlement bonne idée.

Mes pensées vont maintenant à Hannah, qui me fait l'honneur d'être la principale responsable de mon bonheur extra-professionnel. Sa force, sa gentillesse et son intelligence sont sans égal, et une source perpétuelle d'inspiration. Je t'aime ma Hannah.

Je remercie bien sûr mes amis, toujours présents quand il faut, et dont la compagnie est un enrichissement constant : Jérémie, colocataire de génie, David, Simon, Baptiste, Kevin, Renaud, Jean-Marc (ça c'est les garçons), Ingrid, Anaïs, Célia, Alix, Elise (ça c'est les filles), ainsi que tous les petits copains doctorants, les petits copains du muséum, les petits copains de Rennes, Les petits copains du Havre, les petits copains de la faille de l'invocateur, les petits copains du Finnegan's, les petits copains du hand et bien entendu tous ceux que j'aurais, honteusement, oublié ici. Si jamais vous appartenez à plusieurs des précédents groupes (par exemple, Kevin est un garçon, de rennes, de Lol), et bien je vous remercie plusieurs fois. Si-si, c'est cumulatif.

Enfin, comment ne pas finir par ma famille. Je remercie infiniment mes parents, mes frères et mes « soeurs », pour avoir toujours été à mes côtés et veillé sur moi. Je n'ai l'impression de n'être que ce que vous m'avez donné. Je vous doit tout, et je vous aime. Je pense également très fort à Yvonne, mes oncles et mes tantes adorées, mes cousin(e)s préféré(e)s (tous, en fait), et toute la nouvelle génération, Joshua en tête !

Merci à vous tous.

Je souhaiterais dédier ce travail à mon grand-père, un homme bon au souvenir duquel j'essaierai toujours de faire honneur.

Table des matières

Table des matières	1
1 Introduction	3
1.1 L'anatomie des cténophores et sa diversité	3
1.1.1 Introduction à l'anatomie des cténophores	3
1.1.2 Présentation des ordres du phylum Ctenophora	16
1.2 Évolution des cténophores	26
1.2.1 Les relations phylogénétiques entre cténophores	26
1.2.2 La position phylogénétique des cténophores au sein des métazoaires	31
1.2.3 Difficultés méthodologiques liées aux particularités des cténophores	39
1.3 Objectifs de la thèse et structure du manuscrit	51
2 La position phylogénétique des cténophores dans l'arbre des métazoaires abordée par une approche phylogénomique	53
2.1 Introduction	53
2.2 Article 1 : The position of ctenophores in the animal tree investigated using two independent large-scale phylogenomic datasets	54
3 Reconstruction des relations phylogénétiques au sein des cténophores	97
3.1 Introduction	97
3.2 Article 2 : Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora.	99
3.3 Compléments	114
3.3.1 Tentative d'analyse cladistique de caractères morpho-anatomiques	114
3.3.2 Apports des analyses phylogénomiques de l'article 1 pour la phylogénie des cténophores	116
3.3.3 Protocole dédié à la résolution de la position de la racine des cténophores	117
3.3.4 Bilan de la phylogénie moléculaire et retour sur les caractères anatomiques	122

4	Recherche de caractères à l'échelle "micro-anatomique" par réalisation de marquages immunohistochimiques	127
4.1	Introduction	127
4.2	Article 3 : "New species-level diagnostic characters identified in ctenophores using simple fluorescence microscopy techniques : a case study comparing two <i>Pleurobrachia</i> species (Ctenophora : Pleurobrachiidae)"	129
4.3	Compléments - Micro-anatomie comparée à une échelle inter-ordinale : étude de <i>Beroe gracilis</i> KÜNNE, 1939	168
5	Conclusions générales et perspectives	175
5.1	Recherche de nouveaux caractères morpho-anatomiques utiles pour la taxonomie des cténophores (chapitre 4)	175
5.1.1	Synthèse des résultats	175
5.1.2	Perspectives	176
5.2	La reconstruction des relations phylogénétiques au sein des cténophores (chapitre 3)	176
5.2.1	Synthèse des résultats	176
5.2.2	Perspectives	177
5.3	Recherche de la position des cténophores dans l'arbre des métazoaires (chapitre 2) .	177
5.3.1	Synthèse des résultats	177
5.3.2	Perspectives immédiates	178
5.3.3	Autres réflexions, implications et perspectives suscitées par ce travail	179
	Bibliographie	185
	A Appendice : article 1	199
	B Appendice : article 2	257

Chapitre 1

Introduction

1.1 L’anatomie des cténophores et sa diversité

1.1.1 Introduction à l’anatomie des cténophores

Afin de simplifier l’appréhension de l’anatomie d’un embranchement complet au lecteur, je commencerai ici par la description de l’espèce *Beroe gracilis* KÜNNE, 1939. Il sera ensuite fait mention de la disparité des structures chez d’autres taxons de cténophores. De la sorte, le problème de la définition d’un “plan d’organisation”, qui consiste en une vision idéalisée et fixe de la diversité biologique, sera contourné et, je l’espère, sera remplacé par une vision moins réductrice de ce que représente l’appartenance au groupe des cténophores.¹

Description générale d’un cténophore, *Beroe gracilis* KÜNNE, 1939

Initialement décrite en Mer du Nord (Künne 1939), l’espèce *Beroe gracilis* a depuis été identifiée dans l’océan pacifique en Californie et à Friday Harbor (Tamm and Tamm 1993), dans la Manche (Müller 2004), sur les côtes de L’Île de Ré (observations personnelles) et très récemment sur les côtes du Chili (Oliveira et al. 2014). Cette espèce est sujette aux phénomènes de blooming (apparition ponctuelle et massive d’individus due aux conditions océanographiques et planctoniques favorables). Bien que considérée comme la plus fragile des espèces du genre *Beroe*, elle fait partie des cténophores qu’il est aisé de collecter et de conserver.

Morphologie externe Le corps de *Beroe gracilis* est comparable à un sac de forme cylindrique, d’une taille variant de quelques millimètres à 25 mm. Son corps très transparent est généralement d’un blanc pur. Sa forme allongée est très régulière, en ce que le diamètre de l’animal reste presque constant du pôle oral jusqu’au pôle aboral. Le corps est légèrement compressé dans le plan tentaculaire, l’un des deux plans de symétrie de l’animal (Figure 1.1 A, B). Au pôle oral se trouve la

1. Le nom latin de l’embranchement est Ctenophora ; en français, ils peuvent également être appelés cténaïres.

bouche, large et extensible, tandis que le pôle aboral présente l'organe aboral, dont l'un des rôles est l'équilibration.

Cet organe aboral se compose d'un statolithe reposant sur quatre balanciers (Figure 1.1 C) et est encadré dans le plan pharyngien par deux champs polaires, qui sont des zones ciliées et entourées de papilles (Figure 1.1 D). Des quatre balanciers partent quatre couples de sillons ciliés, qui vont rejoindre chacun des quatre quadrants. Les couples de sillons ciliés se séparent alors et chaque sillon va se connecter au premier peigne de chaque rangée de peignes. Les deux champs polaires sont des structures anatomiques complexes, dont le rôle exact n'est pas bien connu. Leur surface densément ciliée permet entre autre la création d'un courant d'eau vers l'organe aboral. Les champs polaires sont bordés de nombreuses papilles dont la fonction, si elle existe, est elle aussi inconnue.²

Le long de l'animal courent huit rangées de peignes (Figure 1.1 A, B, E), qui peuvent parfois être légèrement teintées de rose ou de bleu. Les vagues de battement de ces peignes permettent la nage de l'animal. La structure des cils agglomérés qui forment ces peignes a pour propriété de diffracter la lumière (voir Welch et al. 2006), créant ainsi, à chaque vague de battement des peignes, une vague iridescente aux couleurs de l'arc-en-ciel (Figure 1.1 F).

Chaque peigne est constitué de cellules ciliées, hautes et prismatiques, appelées "polster cells" dont l'accolement forme le bourrelet médian. Ce bourrelet est recouvert d'un revêtement épithélial, et sa base est pénétrée d'éléments nerveux (Figure 1.1 G, Hernandez-Nicaise and Franc 1994). Les cils de ces cellules sont assemblés en une large palette. C'est le statolithe situé dans l'organe d'équilibration qui va, par gravité, appuyer sur les balanciers qui, en réponse, modifieront la fréquence de battement des rangées de peignes grâce aux sillons ciliés. Une augmentation de pression sur l'un des balanciers va diminuer la fréquence des battements des rangées de peignes correspondantes, et augmenter la fréquence des battements des rangées de peignes opposées (Horridge 1974).

Ecologie et comportement *Beroe gracilis* est une des espèces prédatrices du zooplancton marin. Comme tous les *Beroe*, les *Beroe gracilis* sont de bons nageurs mais restent assujettis aux courants d'eau. Ils ont été décrits comme des prédateurs exclusifs de *Pleurobrachia pileus* (O. F. MÜLLER, 1776) (Greve 1970), un autre cténophore. Néanmoins, cette préférence n'est pas une exclusivité car *Beroe gracilis* peut se nourrir de *Mnemiopsis* (un autre genre de Cténophore) et de Physonectae (un sous-ordre des siphonophores, Tamm and Tamm 1993). Ils chassent leurs proies en nageant la bouche en avant. Au contact d'une proie éventuelle, leur bouche s'élargit rapidement et si la taille de la proie le permet, l'avale totalement. Si la proie est trop grosse pour être engloutie par *Beroe gracilis* en une fois, ce qui est souvent le cas pour des individus juvéniles, ces derniers

2. Chez une autre espèce de cténophore, *Pleurobrachia pileus*, les travaux de l'équipe ont montré que les champs polaires possèdent des structures nerveuses particulières, comme les *corps Z* et de nombreuses sous-populations de cellules nerveuses immunoréactives avec des anticorps contre différents neurotransmetteurs (comme la sérotonine, la mélatonine la noradrénaline ou encore l'adrénaline) et contre la L-dopamine, l'enzyme Tyrosine-Hydroxylase, le récepteur Vmat, suggérant ainsi une fonction de type neuro-sensorielle (voir Jager et al. 2011, 2013).

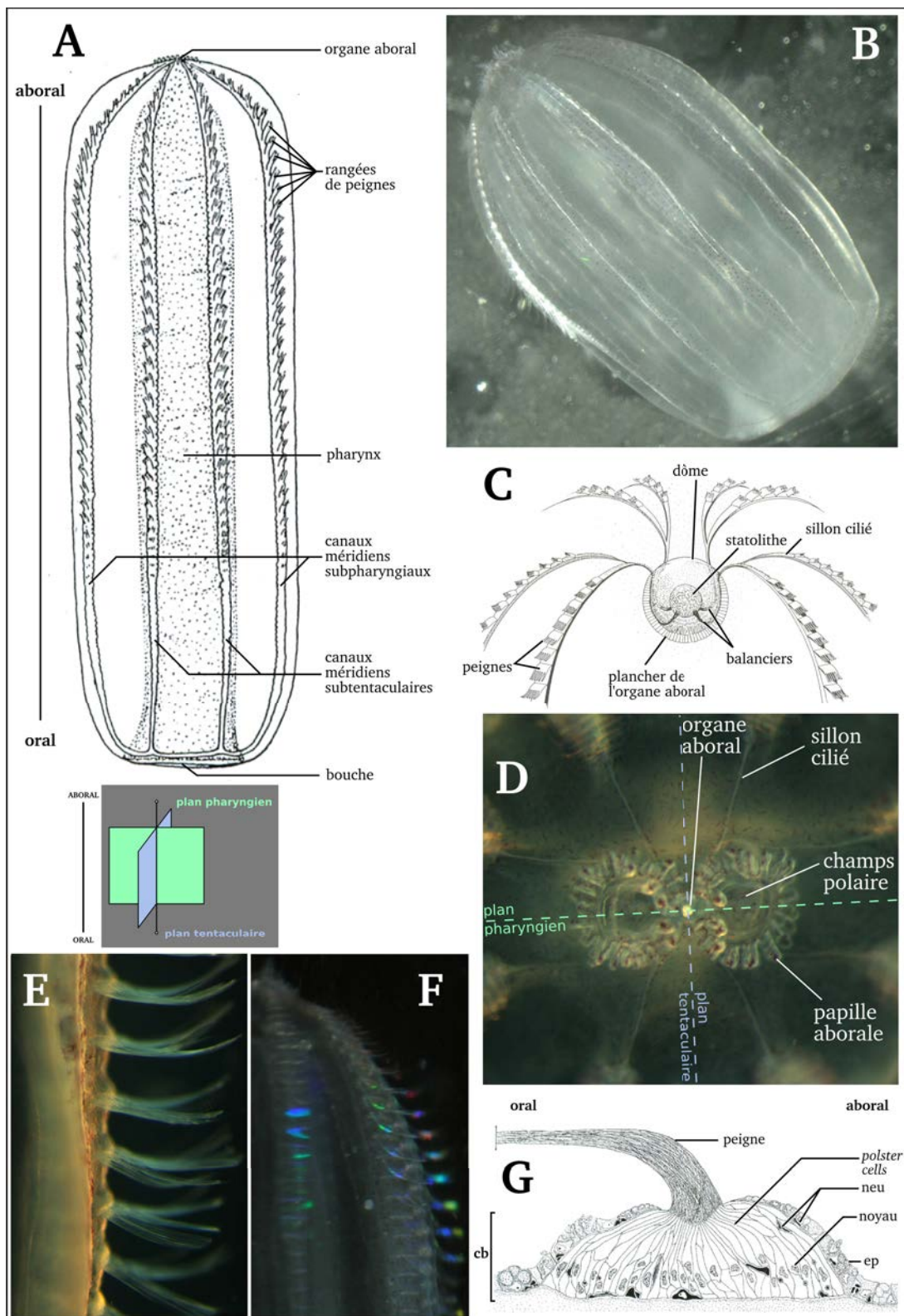


FIGURE 1.1 – Morphologie générale de *Beroë gracilis*. A) Dessin de l'anatomie générale de *Beroë gracilis*, d'après Künné (1939). B) Photo montrant l'aspect générale de *Beroë gracilis* (photo originale). C) Dessin de l'organisation de l'organe aboral, figurant sa relation avec les rangées de peignes via les sillons ciliés, d'après Brusca and Brusca (2003). D) Photo de l'organe aboral, des champs polaires et des sillons ciliés, en vue aborale (photo originale). E) Photo d'une rangée de peignes en vue latérale (photo originale). F) Photo de deux rangées de peignes pendant une vague de battements, chaque peigne diffractant différemment la lumière selon son inclinaison (photo originale). G) Schéma de la structure d'un peigne, vue en coupe transversale, d'après Hernandez-Nicaise (1973). Légendes : cb : coussinet basal ; ep : cellule épithéliale ; neu : neurone ;

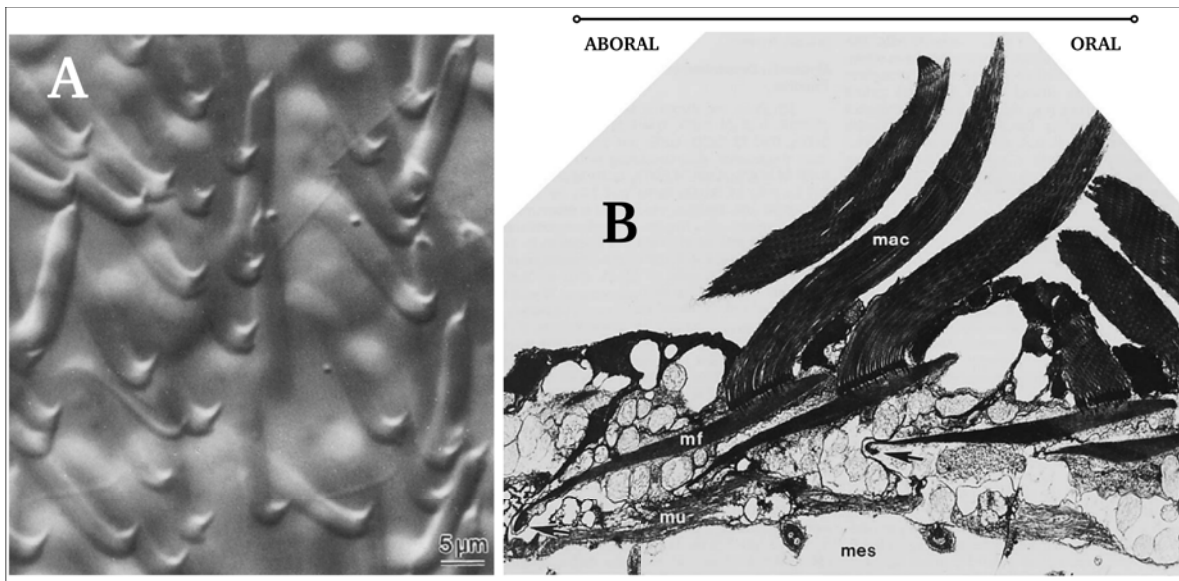


FIGURE 1.2 – Dents macrociliaires de *Beroe gracilis*. A) Forme et disposition des dents macrociliaires, vues en DIC, la bouche vers le bas, d’après Tamm and Tamm (1993). B) Cellules macrociliaires en vue longitudinale, d’après Tamm and Tamm (1987). Légendes : mac : macrocils ; mf : microfilaments ; mes : mésoglée ; mu : muscle lisse.

pourront mordre et arracher des bouts de la proie. Parfois les petits *Beroe* restent collés la bouche ouverte sur la proie un certain temps, technique ressemblant alors à de l’ectoparasitisme (Tamm and Tamm 1991b).

Bouche et lèvres La large bouche de *Beroe gracilis* est un outil de chasse remarquable. Ses lèvres présentent des cellules à la fois mécanoréceptrices et chimioréceptrices (Tamm and Tamm 1991a). Les mécanismes d’ouverture et de fermeture des lèvres impliquent des structures complexes : des bandes adhésives de cellules de part et d’autres de la bouche assurent le maintien fermé des lèvres, tandis que des contractions musculaires et des influx nerveux permettent leur ouverture extrêmement rapide grâce à un effet de “dézippage” (Tamm and Tamm 1991b).

L’entrée de cette bouche est ornée de dents macrociliaires (Figure 1.2 A). Ces dents sont disposées en une bande à l’entrée de la bouche, parallèle aux lèvres. Ce sont en réalité des cellules ciliées, ancrées dans l’épithélium du pharynx, de forme allongée, remplies d’assemblages d’actine, et reliées à des fibres musculaires (Figure 1.2 B). Les cils de ces cellules sont agglomérés en plusieurs pointes qui peuvent battre, tout comme les rangées de peignes, et aident ainsi à l’ingestion des proies et/ou à la fragilisation des tissus ingérés (Tamm and Tamm 1993, 1987). Ces structures dentaires ne sont probablement pas les seuls éléments complexes de la bouche de *Beroe gracilis* puisque des fibres nerveuses géantes ont été décrites chez des *Beroe* non-identifiés (Tamm and Tamm 1995). Cette bouche donne sur un très large pharynx représentant une grande partie du volume de l’animal. Ce pharynx est extrêmement compressé dans le plan tentaculaire mais est très extensible et permet l’ingestion de proies aussi grosses que le prédateur.

Système gastrovasculaire Le pharynx de *Beroe gracilis* débouche, en sa partie aborale, sur un petit infundibulum, élément central du réseau de canaux gastrovasculaires (Figure 1.3). De cette cavité gastrique émergent cinq canaux différents. Le premier est le canal aboral, extrêmement court à cause de la taille du pharynx chez *Beroe gracilis*, qui part vers l'organe aboral et se scinde ensuite en quatre, pour donner deux pores anaux et deux canaux clos. L'ensemble de la bouche, du pharynx, de l'infundibulum et du canal aboral forme la partie axiale du système gastrovasculaire.

Les quatre autres canaux émergeant de l'infundibulum sont les canaux interradiaires, qui partent perpendiculairement à l'axe oral-aboral de l'individu, chacun vers un quadrant de l'animal³. Ces quatre canaux interradiaires se divisent ensuite chacun en deux canaux adradiaires qui vont ensuite connecter les huit canaux méridiens. La connexion entre un canal adradiaire et son canal méridien se fait à la partie la plus aborale du canal méridien, ce qui donne l'impression que le canal adradiaire change simplement de direction et "devient" le canal méridien. Les huit canaux méridiens courent le long des huit rangées de peignes et continuent jusqu'à la bouche (les rangées de peignes s'arrêtent plus tôt). À l'endroit même où les canaux interradiaires connectent l'infundibulum, deux canaux paragastriques partent en direction de la bouche, en longeant le pharynx. Autour de la bouche, les quatre canaux méridiens et le canal paragastrique de chaque côté de l'animal fusionnent, formant ainsi deux canaux semi-circumoraux (Figure 1.3). Les canaux gastriques des cténophores possèdent une double ciliature interne. D'un côté du canal, la ciliature bat dans un sens, tandis que la ciliature de l'autre côté bat dans le sens opposé, créant ainsi une double circulation des liquides, ce qui permet de distribuer les nutriments ingérés à toutes les régions du corps.

Reproduction *Beroe gracilis* est une espèce hermaphrodite. Les gonades males et femelles sont situées dans la paroi des canaux méridiens, les gonades femelles d'un côté et les gonades mâles de l'autre (Figure 1.4). La disposition de ces gonades semble donc correspondre à la disposition des deux types de ciliature dans les canaux (Tamm 2014). Les individus pondent souvent dans les 48h après avoir suffisamment mangé. La ponte, jamais observée chez *Beroe gracilis*, se fait probablement comme chez la plupart des autres cténophores, par de petits canaux génitaux qui connectent les canaux méridiens au milieu extérieur. La fécondation est externe, et l'embryogenèse de *Beroe gracilis* suit un développement commun à celui des autres cténophores, avec un clivage unilatéral des cellules et une certaine mosaïcité (Martindale and Henry 1997a,

3. La présence ou l'absence des canaux interradiaires (voir Figure 1.3 C et Figure 1.7 A'-C') n'est pas évidente chez *Beroe gracilis*, à cause de la fusion relativement simultanée des quatre larges canaux adradiaires ensemble. La présence ou l'absence de canaux perradiaires, qui ne sont pas représentés ici (voir Figure 1.7 A', annotation 6), pourrait elle aussi être débattue. Cette fusion "brouille" la structure réelle de ces canaux et rend parfois possible plusieurs interprétations. Harbison décrira par exemple les *Beroida* comme ne possédant ni canal perradiaire ni canal interradiaire (Harbison 1985), tandis que les autres auteurs considèrent que les *Beroe* possèdent des canaux interradiaires (Chun 1880; Hyman 1940; Hernandez-Nicaise and Franc 1994). J'adopte ici leur point de vue. Je pense qu'il est même possible de considérer que des canaux perradiaires existent (au moins chez *Beroe gracilis* et *Beroe forskalii*), mais cela dépasse le cadre de la description de *Beroe gracilis*.

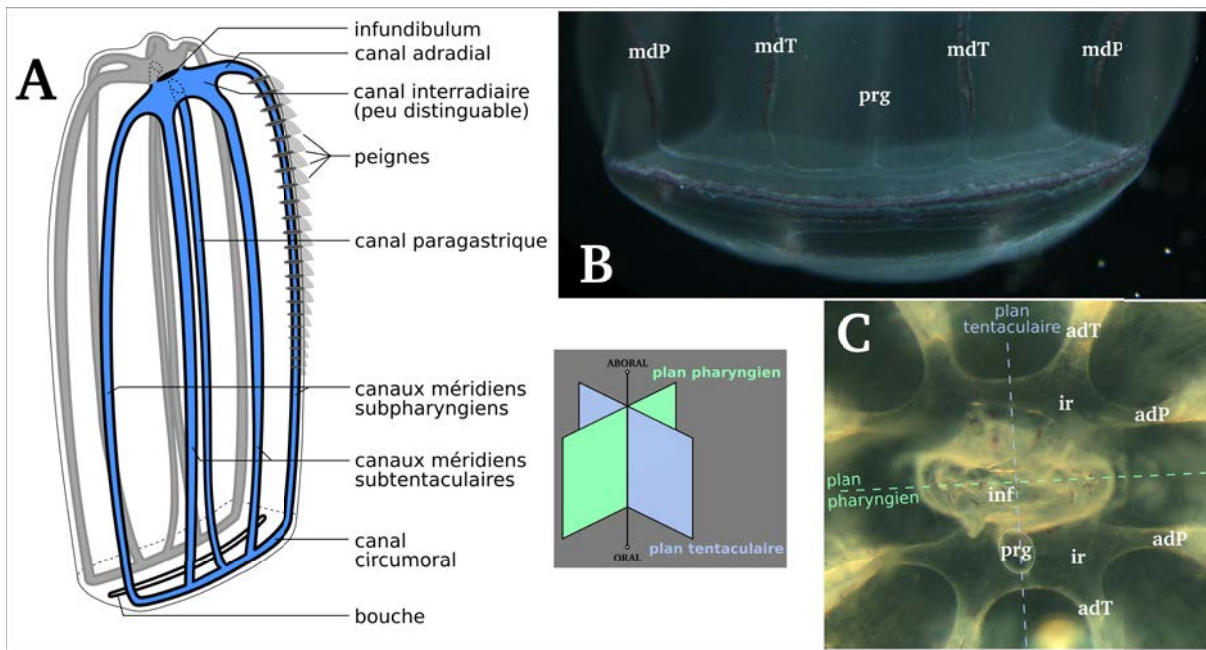


FIGURE 1.3 – Structure du système gastrovasculaire de *Beroë gracilis*. A) Schéma du système gastrovasculaire périphérique, en vue latérale (dessin original). Les canaux méridiens et paragastriques de chaque moitié (bleu ou grise) fusionnent autour de la bouche, donnant naissance à un complexe circumoral, composé de deux canaux semi-circumoraux. Les éléments du système gastrovasculaire axial ne sont pas représentés, à l’exception de la bouche. B) Canal semi-circumoral, composant le complexe circumoral, en vue latérale. Chacun des deux canaux semi-circumoraux est constitué de la fusion des terminaisons orales d’un canal paragastrique (prg), de deux canaux méridiens subtentaculaires (mdT) et de deux canaux méridiens subpharyngiens (mdP). C) Connexions des canaux adradiaux aux canaux méridiens, en vue orale interne. Les canaux interradiaires (ir) ne sont pas faciles à distinguer. Légendes : inf : infundibulum ; adT : canal adradiale subtentaculaire ; adP : canal adradiale subpharyngien.

1999). Le développement précoce de *Beroë gracilis* peut être observé sur le site http://gvondassow.com/Research_Site/Videos_of_cells_and_embryos.html. L’oeuf éclot un à deux jours après la fécondation, et le juvénile ressemble à une version miniature de l’adulte. La croissance de l’animal se fait sans métamorphose. La rapidité de cette croissance est variable et le juvénile peut passer de 1 mm à 17 mm en un mois (Greve 1970). La maturité sexuelle est atteinte entre 10 et 15 mm, et est le point de départ du ralentissement de la croissance, les plus grand individus de *Beroë gracilis* ne dépassant pas les 25 mm. Leur plus grande durée de vie observée en élevage est de six mois (Greve 1970).

Principales variations de l’anatomie des cténophores

La plupart des structures anatomiques présentes chez *Beroë gracilis* existent aussi chez d’autres cténophores, mais on peut observer des exceptions pour chacune d’entre elles. Même si l’on considère des organes complexes et importants comme les rangées de peignes ou encore l’organe aboral, il existe des cténophores qui ne les possèdent pas. Les *Beroë* eux-mêmes sont des exceptions, car ils ne possèdent ni tentacules ni appareil tentaculaire. En réalité, considérer qu’il existe un état “normal” chez les cténophores, c’est considérer un état fictif, traditionnellement appelé *plan d’organisation*. Il

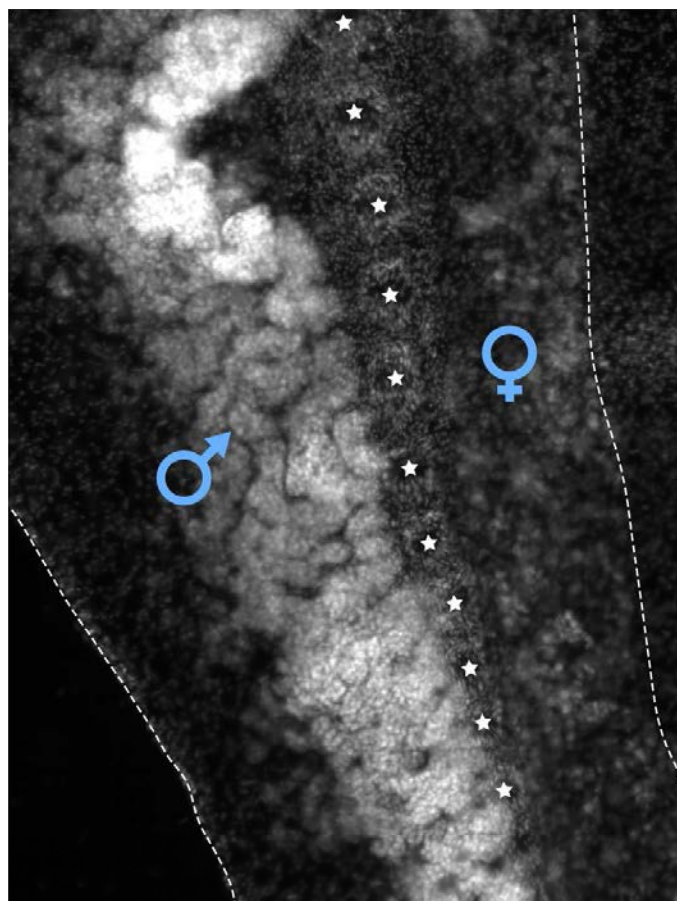


FIGURE 1.4 – Disposition asymétrique des gonades de *Beroe gracilis* au niveau d’un canal méridien, vue par marquage en DAPI. Les peignes sont marqués par les étoiles blanches, et le contour du canal méridien est montré par les pointillés blancs. Les gonades mâles sont à gauche de la photo et les gonades femelles à droite.

faudrait plutôt parler d’états ancestraux, et aucun organisme ne possède l’ensemble des états ancestraux des cténophores, même si certains en possèdent plus que d’autres, comme les “Cydippida”. Ici seront donc décrites les grandes lignes de la variabilité morphologique des cténophores.

Appareil tentaculaire La grande majorité des cténophores utilisent leur paire de tentacules pour capturer leurs proies et se nourrir. Les représentants des genres *Beroe* et *Neis* sont les seules exceptions, ne possédant d’appareil tentaculaire à aucun stade de leur développement. Chez les autres cténophores, cet appareil est composé d’une gaine tentaculaire, qui s’ouvre vers l’extérieur de l’animal, et au fond de laquelle se trouve une racine tentaculaire. Cette racine va produire un tentacule rétractable qui sort par l’orifice de la gaine tentaculaire. Ce tentacule peut porter des extensions appelées tentilles (Figure 1.5 A, B). Tentacules et tentilles portent des colloblastes, cellules gluantes qui permettent la capture des proies, et qui sont un type cellulaire unique aux cténophores (Figure 1.5 C).

Certains des éléments de cet appareil tentaculaire sont absents chez certains cténophores. Ainsi, les tentilles sont absentes chez les Haeckeliidae (par exemple chez *Haeckelia rubra* (KÖLLIKER,

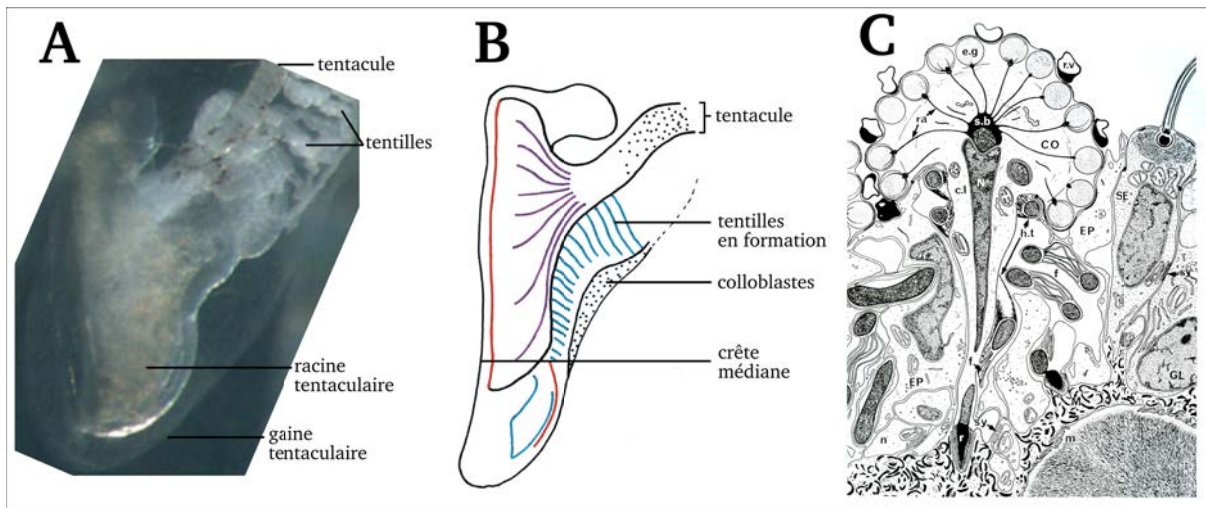


FIGURE 1.5 – Appareil tentaculaire des cténophores. A) Photo d’une racine tentaculaire de *Pleurobrachia pileus* (photo originale). B) Schéma d’une racine tentaculaire de *Pleurobrachia pileus* vue en coupe longitudinale latérale, d’après Dayraud (2011). En rouge sont représentées les cellules souches des composants non-épithéliaux, en bleu sont représentées les tentilles ainsi que les cellules progénitrices des muscles des tentilles et en violet sont représentées les fibres musculaires. C) Schéma d’un colloblaste de *Pleurobrachia* situé dans l’épithélium d’un tentille, vue en coupe longitudinale, d’après Hernandez-Nicaise (1991) (observation faite en MET). Légendes : CO : colloblaste; c.l : lamelle de connexion; e.g : granule éosinophile; EP : cellule épithéliale; f : replis membranaires du filament hélicoïdal; GL : cellule glandulaire; h.t : filament hélicoïdal; m : cellule musculaire lisse; M : mésoglée; n : neurite; r : racine; ra : radius; r.v : vésicule de réfraction; s.b : corps sphérique; SE : cellule ciliée sensorielle; sy : synapse.

1853), Chun 1880). La gaine tentaculaire disparaît durant la métamorphose des Lobata (comme chez *Mnemiopsis gardeni* L. AGASSIZ, 1860⁴; Hyman 1940). L’ensemble de l’appareil tentaculaire est manquant chez les Beroida (dont *Beroe gracilis* fait partie). L’existence de l’appareil tentaculaire complet chez la majorité des taxons et l’existence de la gaine tentaculaire chez les Lobata juvéniles sont deux arguments qui laissent penser que les absences de certaines parties de l’appareil tentaculaire sont des pertes secondaires au sein des cténophores (voir section chapitre 4).

Lobes et Auricules Les Lobata n’utilisent pas seulement leurs tentacules pour se nourrir, mais également des structures additionnelles appelées lobes et auricules (les premiers donnant son nom au groupe). Ces structures sont des extensions de la partie orale du corps, plus ou moins prononcées, dans le prolongement des rangées de peignes, et qui entourent la bouche (Figure 1.6). Les deux lobes portent chacun deux rangées de peignes subpharyngiennes, et les quatre auricules portent chacun une rangée de peignes subtentaculaire. Ensemble, les lobes, les auricules et les tentacules, vont attraper, coller ou piéger les proies. La morphologie de ces structures varie et permet différentes

4. L’espèce *Mnemiopsis leidyi* est la plus connue des cténophores, car elle est étudiée à la fois pour son importance écologique (elle est invasive, voir par exemple Ghabooli et al. 2013) et pour son génome assemblé disponible (Ryan et al. 2013). Une révision taxonomique de Seravin (1994) a établi la synonymie des trois espèces *M. leidyi*, *M. mccradyi* et *M. gardeni* et a attribué la priorité à *M. leidyi*. En réalité, la priorité de l’épithète spécifique revient à *M. gardeni*, décrit par Agassiz en 1860 (Agassiz 1860, page 269), tandis que *M. leidyi* a été décrit en 1865 (Agassiz 1865, page 20). J’utiliserai donc dans la présente thèse le nom *M. gardeni* comme valide, considérant *M. leidyi* comme un synonyme plus récent et en attendant que l’acte nomenclatural correspondant soit officiellement publié.

méthodes de chasse dont voici trois exemples :

1. Les membres de l'espèce *Ocyropsis crystallina* (RANG, 1828) ne possèdent pas de tentacules. Ils ramènent leurs vastes lobes musclés l'un contre l'autre pour capturer mécaniquement et individuellement leurs proies, qui sont ainsi bloquées et apportées au niveau de la bouche préhensile de l'animal (Figure 1.6 A). Ils sont ainsi capables de capturer des proies bien plus grosses que la plupart des autres cténophores (Harbison et al. 1978; Haddock 2007). Les auricules d'*Ocyropsis* sont courts et rigides, et aident peut-être au maintien de la proie.
2. Les individus de l'espèce *Leucothea multicornis* (QUOY AND GAIMARD, 1824) nagent rapidement, la bouche en avant et les grands lobes étendus à plat. Dans cette position, les tentacules secondaires longent la surface interne des lobes (Figure 1.6 B). Les lobes créent un espace dans lequel les auricules vont bouger en ondulant. Les proies (des copépodes, par exemple) vont passer entre les lobes, être dérangées par les mouvements des longs et fins auricules, et, dans leur tentative de fuite, se retrouver collées sur la face interne des lobes. Cela va déclencher le repli d'une moitié du lobe, ce qui assurera le contact entre la proie et les tentacules secondaires qui longent la surface interne des lobes (Harbison et al. 1978; Haddock 2007).
3. Enfin, chez l'espèce *Mnemiopsis gardeni*, qui chasse de manière statique ou en nage lente, deux processus simultanés et complémentaires permettent l'ingestion des proies (Figure 1.6 C). Le premier est le battement continu des cils bordant les quatre auricules courts et larges, qui créent un courant lent transportant passivement les petites proies (des larves nauplius, par exemple) vers les tentacules secondaires. Le second consiste en la collision des proies plus grosses contre les surfaces internes des lobes, recouvertes de mucus (Waggett and Costello 1999, voir Figure 1.6 D). Les proies sont ensuite transportées jusqu'à la bouche par des mouvements ciliaires.

Système gastrovasculaire Une fois digérés dans la partie la plus profonde du pharynx, les aliments sont transportés dans le système gastrovasculaire, qui a pour charge de distribuer les nutriments à tous les organes. L'architecture de ce système varie entre les ordres et les familles de cténophores, et représente une source de caractères utilisée pour la systématique de ces animaux. La liste exhaustive des composants possibles du système gastrovasculaire des cténophores est la suivante : la bouche, le pharynx, l'infundibulum, le canal aboral, les pores anaux et les canaux paragastriques (situés le long de l'axe oral-aboral) ; les canaux perradiaires, interradiaires, adradiaires, méridiens et tentaculaires (éléments périphériques du système). Beaucoup de ces composants peuvent être absents et, quand ils sont présents, ils peuvent présenter une grande variabilité de formes et de tailles relatives (Figure 1.7).

L'existence ou l'absence des canaux perradiaires et interradiaires chez les cténophores est particulièrement utilisée pour leur classification. La plupart des Pleurobrachiidae possèdent des canaux

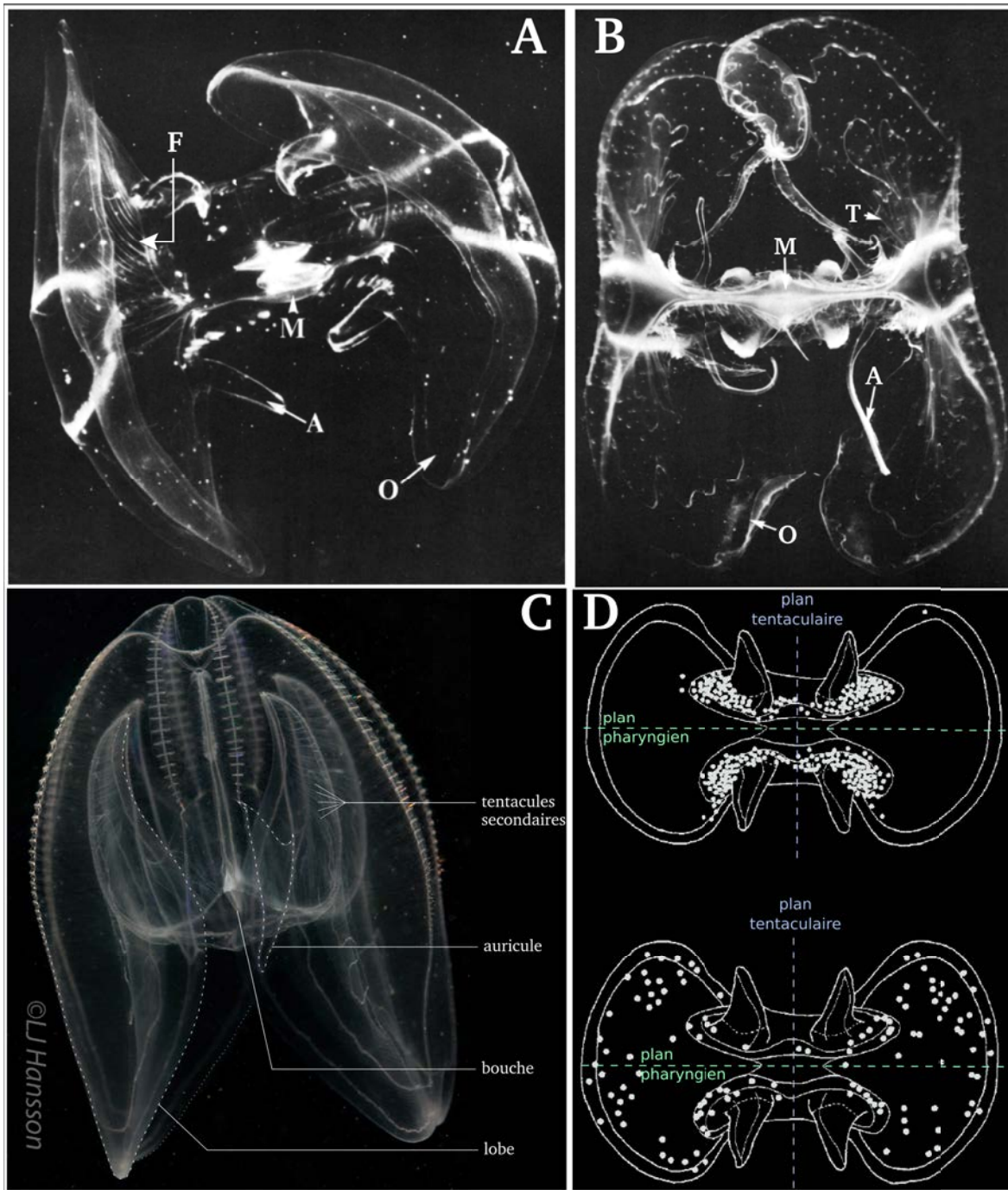


FIGURE 1.6 – Morphologie et utilisation des lobes et auricules lors de la chasse chez les Lobata. A) *Ocyropsis crystallina* chassant, en vue orale (Harbison et al. 1978). Les lobes possèdent des fibres musculaires très visibles. B) *Leucothea multicornis* chassant, en vue orale (Harbison et al. 1978). Les longs auricules ondulent entre les lobes largement étendus, recouverts de tentacules secondaires. C) *Mnemiopsis gardeni*, en vue latérale tentaculaire (photo de Lars Hansson). D) Localisation de différentes proies capturées par *Mnemiopsis gardeni* (d'après Waggett and Costello 1999). En haut, les petites proies sont capturées au niveau des auricules et des tentacules secondaires; en bas, les proies plus grosses sont capturées par les lobes. Légendes : F : fibres musculaires; T : tentacules secondaires; M : bouche; A : auricule; O : lobe.

perradiaires et interradiaires (Figure 1.7 A'), tandis que les Lobata et Beroida ne possèdent que des canaux interradiaires (Figure 1.7 B'), et que les Platyctenida, Mertensiidae et Bathyctenidae sont réputés ne posséder que les canaux perradiaires (Figure 1.7 C').

Toutefois, les opinions sur la présence ou l'absence de chacun de ces types de canaux diffèrent selon les auteurs. Par exemple, chez les Platyctenida, Hyman (1940) note la présence de canaux perradiaires, tandis que Harbison (1985) note leur absence. Cela est dû à l'absence de définition explicite de la nature de ces canaux. Les noms de ces canaux ont été définis à partir de la structure du système gastrovasculaire de certains Cydippida (comme *Pleurobrachia*, Figure 1.7 A'), mais n'ont jamais été pensés pour être adaptables au système gastrovasculaire des autres cténophores. La figure 1.7 C' est en accord avec la pensée de Hyman. Je définis la nature des canaux radiaires de la manière suivante :

- canal perradiaire : canal connecté à l'infundibulum qui, par éventuellement plusieurs ramifications intermédiaires, aboutira aux quatre canaux adradiaires présents dans chaque moitié tentaculaire de l'organisme.
- canal interradiaire : canal connecté à uniquement deux canaux adradiaires appartenant au même quadrant de l'organisme (un canal adradiaire subtentaculaire et un canal adradiaire subpharyngien).
- canal synradiaire (nouveau terme) : canal connecté à un canal perradiaire d'un côté, et de l'autre côté, à deux canaux adradiaires subtentaculaires (ainsi qu'à un canal tentaculaire), voir Figure 1.7 C', annotation 25.

D'autre part, la structure des terminaisons orales des canaux méridiens et paragastriques peut également varier. Chez les Cydippida, tous ces canaux sont clos en leur partie orale (Figure 1.7 A). Chez d'autres cténophores, les terminaisons orales des canaux méridiens et paragastriques peuvent s'anastomoser, reliant ainsi les canaux en un complexe circumoral. Plusieurs patrons d'anastomoses sont possibles, selon le nombre et la nature des canaux se connectant. Chez *Beroë cucumis* FABRICIUS, 1780, tous les canaux méridiens et paragastriques de chaque moitié tentaculaire de l'animal s'anastomosent en un canal semi-circumoral (Figure 1.7 D), de la même manière que chez *Beroë gracilis* (voir section 1.1.1 page 7). Le complexe circumoral est dans ce cas composé de deux moitiés de canal circumoral. Dans le genre *Mnemiopsis*, le patron d'anastomose est plus compliqué. Les canaux méridiens subpharyngiens adjacents sont connectés entre eux, tandis que les canaux méridiens subtentaculaires opposés sont connectés entre eux. Chacun des canaux paragastriques se scindent en deux en sa partie la plus orale. Les quatre terminaisons des canaux paragastriques connectent ensuite les deux couples de canaux méridiens subtentaculaires, créant ainsi un canal circumoral complet (Figure 1.7 B).

En plus de ces éléments constitutifs, certaines structures additionnelles peuvent exister. Certains canaux, comme les canaux méridiens, peuvent posséder des diverticules, c'est-à-dire des ramifications latérales (Figure 1.7 D). Ces ramifications peuvent parfois fusionner entre elles, créant ainsi un réseau complet d'anastomoses gastrovasculaires (Figure 1.7 E). Un tel réseau existe également

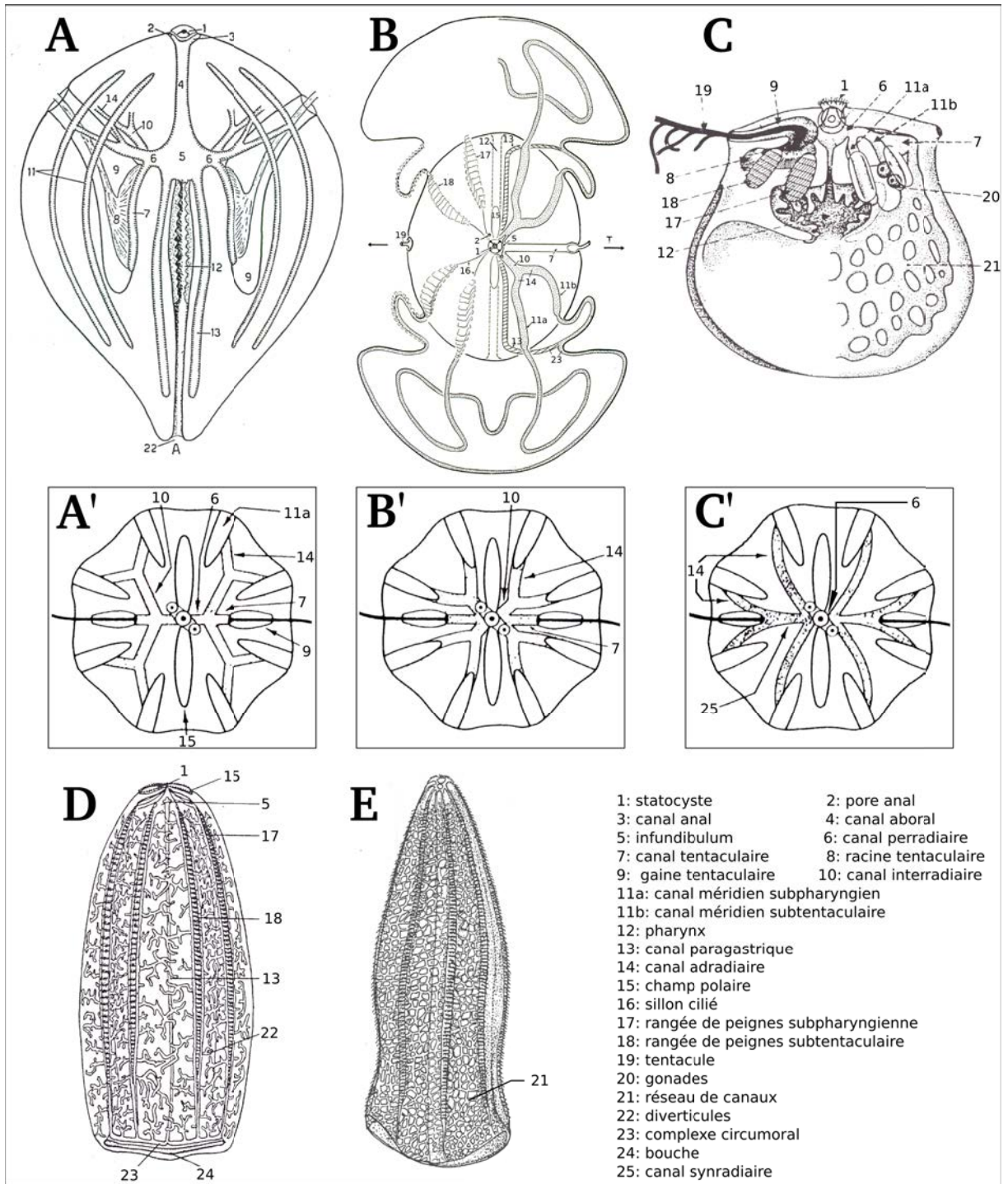


FIGURE 1.7 – Variations des systèmes gastrovasculaires chez les cténophores. A) Schéma du système gastrovasculaire d'un "Cydippida", *Pleurobrachia pileus*, en vue latérale pharyngienne, d'après Hyman (1940). B) Schéma du système gastrovasculaire d'un Lobé, *Mnemiopsis gardeni*, en vue aborale, d'après Hernandez-Nicaise and Franc (1994). C) Schéma du système gastrovasculaire d'un genre de Platyctenida, *Ctenoplana*, en vue latérale pharyngienne, d'après Hernandez-Nicaise and Franc (1994). A', B' et C') Schémas des connexions des canaux adradiaires à l'infundibulum en vue aborale chez les Pleurobrachiidae (A'), les Lobata ou les Pleurobrachiidae (B'), et les Platyctenida, Mertensiidae et Bathyctenidae (C'), tous d'après Harbison (1985). D) Dessin de la morphologie générale de *Beroe cucumis*, en vue latérale tentaculaire, d'après Seravin (1995). E) Dessin de la morphologie générale de *Beroe forskalii*, en vue latérale tentaculaire, d'après Seravin (1995).

chez les Platyctenida (Figure 1.7 C).

Forme du corps et modes de vie L’immense majorité des cténophores sont planctoniques à toutes les étapes de leur cycle de vie. Ils sont de formes subsphérique, plus ou moins ovale car plus ou moins compressés dans le plan tentaculaire ou pharyngien. Seuls les Cestida font exception (comme *Cestum veneris* LESUEUR, 1813), qui sont tellement compressés dans le plan tentaculaire et allongés dans le plan pharyngien que leur forme est celle d’un ruban (voir Figure 1.13).

Les Platyctenida ont quant à eux un mode de vie alternant des phases pélagiques avec des phases benthiques. Chez ces espèces, le mode de vie benthique est dû à leur capacité d’extension buccale, qui a pour conséquence, lorsque la bouche s’étend sur le substrat, de les aplatis complètement (voir section 1.1.2). L’applatissage des Platyctenida est plus ou moins prononcé selon les espèces (voir Figure 1.11), et la majorité d’entre elles peuvent “refermer” leur sole buccale et retrouver transitoirement une forme arrondie qui leur permet de dériver avec le courant. En dehors des Platyctenida, seule l’espèce *Lampea panzerina* (CHUN, 1879), de l’ordre des “Cydippida” (famille des Lampeidae) peut étendre sa bouche et s’applatir complètement. Cette espèce, qui se nourrit exclusivement de salpes, s’applatit en effet sur sa proie lorsque celle-ci est trop grosse. Comme les juvéniles de *Lampea panzerina* sont toujours plus petits que leurs proies, ils ont toujours été trouvés dans cette position d’adhésion à la proie, et sont qualifiés d’ectoparasites (Komai 1922; Hyman 1940; Harbison et al. 1978).

Rangées de peignes Synapomorphie du phylum des Ctenophora, leurs huit rangées de peignes (Figure 1.8) représentent leur moyen de locomotion principal, sinon exclusif pour la grande majorité des espèces planctoniques. La longueur des rangées de peignes peut être très variable. On observe le plus souvent entre 10 et 20 peignes par rangées, mais plusieurs dizaines de peignes par rangées peuvent exister chez certaines espèces, et, à l’inverse, chez *Paracelsia quadriloba* DAWYDOFF, 1946, les rangées de peignes sont composées chacune d’un unique peigne (Dawydoff 1946). Certains Platyctenida, dont les individus sont principalement benthiques (voir section 1.1.2), ne possèdent aucune rangée de peignes à l’état adulte. C’est, entre autres, le cas de *Tjalfiella tristomata* MORTENSEN, 1910. La découverte que les juvéniles de *Tjalfiella tristomata* étaient des larves possédant des rangées de peignes a non seulement permis de comprendre que l’adulte était un cténophore, mais suggère également que l’absence de rangées de peignes est une perte secondaire chez ces Platyctenida (Hyman 1940).

Chez divers groupes de cténophores, les rangées de peignes subtentaculaires et les rangées de peignes subpharyngiennes ont des tailles différentes, les premières pouvant être plus courtes ou plus longues que les secondes. Cette différence de taille est souvent corrélée avec le sens de compression générale du corps (par exemple, les rangées de peignes subtentaculaires des taxons compressés dans le plan tentaculaire sont souvent plus petites que les rangées subpharyngiennes, comme chez les cestidés).

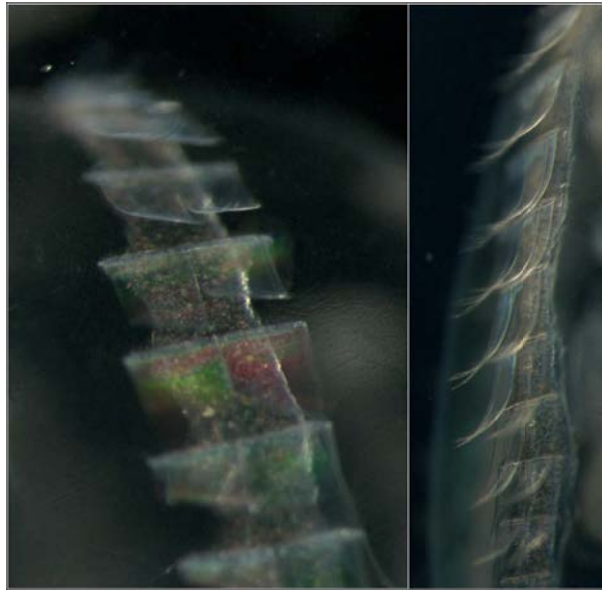


FIGURE 1.8 – Photos de deux rangées de peignes de *Pleurobrachia pileus* (photos originales).

Lors de la croissance des individus, des peignes sont rajoutés aux deux extrémités de chaque rangée. Les modalités de mise en place de ces peignes sont différentes selon qu’il s’agisse de la partie orale ou aborale de la rangée (Alié et al. 2011). En ce qui concerne la partie aborale de la rangée, ces modalités sont également différentes entre taxons (Tamm 2012).

Certains cténophores ont des modes de locomotion qui n’impliquent pas les rangées de peignes mais la musculature. Les Cestida ondulent comme des serpents, tandis que les Ocyropsidae utilisent la fermeture rapide de leurs lobes comme moyen de propulsion, à la manière de méduses. La reptation des Platyctenida lorsqu’ils se trouvent en phase benthique met elle aussi en jeu des muscles.

Les synapomorphies des cténophores Malgré la diversité anatomique qui vient d’être décrite, les cténophores possèdent plusieurs synapomorphies très claires, qui les rendent facilement reconnaissables en tant qu’embranchement. Ces synapomorphies sont notamment la présence de huit rangées de peignes, d’un complexe aboral neuro-sensoriel, de colloblastes, de deux tentacules munies de tentilles.

1.1.2 Présentation des ordres du phylum Ctenophora

Le phylum Ctenophora contient 9 ordres, 26 familles et 51 genres reconnus. La liste des espèces valides tenue à jour par Claudia Mills fait état d’environ 150 espèces décrites (Mills 1998-présent). Le nombre exact d’espèces est incertain car nombre d’entre elles sont suspectées de synonymie. Le nombre de caractères morphologiques utilisables pour classer les cténophores est particulièrement faible au niveau des espèces, où certaines ne diffèrent que par de subtiles variations de proportions,

de colorations ou encore de répartition géographique. En revanche, les distinctions morphologiques séparant les différents ordres sont très nettes, et parfois spectaculaires.

Les “Cydippida”

L'ordre des Cydippida a été montré paraphylétique par Podar et al. (2001), et est le plus grand chez les cténophores : il contient 9 familles et 25 genres. On trouve entre autres, parmi ses membres, les espèces *Pleurobrachia pileus* (famille des Pleurobrachiidae, Figure 1.9 A), *Mertensia ovum* (FABRICIUS, 1780) (famille des Mertensiidae, Figure 1.9 C) ou encore *Euplokamis dunlapae* MILLS, 1987 (famille des Euplokamididae, voir un individu du genre *Euplokamis* Figure 1.9 E). Cet ordre regroupe les cténophores chez qui les adultes sont planctoniques, avec un corps plus ou moins arrondi (rond ou ovale), où les canaux adradiaires connectent les canaux méridiens loin de leur extrémité aborale, l'extension aborale des canaux méridiens étant ainsi close, et dont les individus possèdent deux tentacules qui peuvent se retracter complètement dans leurs gaines tentaculaires (Figure 1.9). Ils ne possèdent ni lobe, ni auricule, ni aucune forme de complexe circumoral (ç.-à-d. de fusion des canaux méridiens et/ou paragastriques autour de la bouche, voir texte page 13).

Les tentacules servent à la capture des proies. Lorsqu'ils sont positionnés proches de la bouche, la rétractation des tentacules y amène directement les proies (comme chez les Lampeidae). En revanche, lorsque les tentacules sont plus éloignés de la bouche, l'animal adopte une technique à la fois complexe et surprenante : la moitié de ses rangées de peignes vont battre dans le sens opposé à la nage, tandis que l'autre moitié continue de battre dans le sens habituel, ce qui a pour effet de faire entrer l'animal en rotation. Une fois en rotation, l'animal va “s'enrouler” dans son propre tentacule de manière à ce que le tentacule touche la bouche. Il avale ensuite le tentacule afin que la proie se décroche dans le pharynx (Figure 1.9 F). Ce comportement est appelé le “spinning” (Tamm 1982).

Lorsque l'on évoque le “plan d'organisation” des cténophores (voir par exemple Brusca and Brusca 2003), il s'agit en réalité des aspects généraux de la morphologie des Cydippida. La morphologie de la phase “larvaire” (en réalité, juvénile) de beaucoup d'autres cténophores appartenant aux autres ordres correspond beaucoup à la morphologie des Cydippida adultes, ce qui a poussé à voir les Cydippida comme les cténophores “de base”, ou “ancestraux”, et les représentants des autres ordres comme des Cydippida modifiés. Comme le groupe des “Cydippida” est paraphylétique (Podar et al. 2001, et voir article 2), leurs états de caractères communs peuvent être compris comme des simplésiomorphies de l'ensemble des cténophores.

Les Lobata

Le second ordre le plus important en nombre d'espèces chez les cténophores est l'ordre des Lobata (Figure 1.10) qui contient 7 familles et 12 genres. Ses représentants les plus connus sont *Mnemiopsis gardeni* A. AGASSIZ, 1860 (famille des Bolinopsidae), *Bolinopsis infundibulum* (O.F.

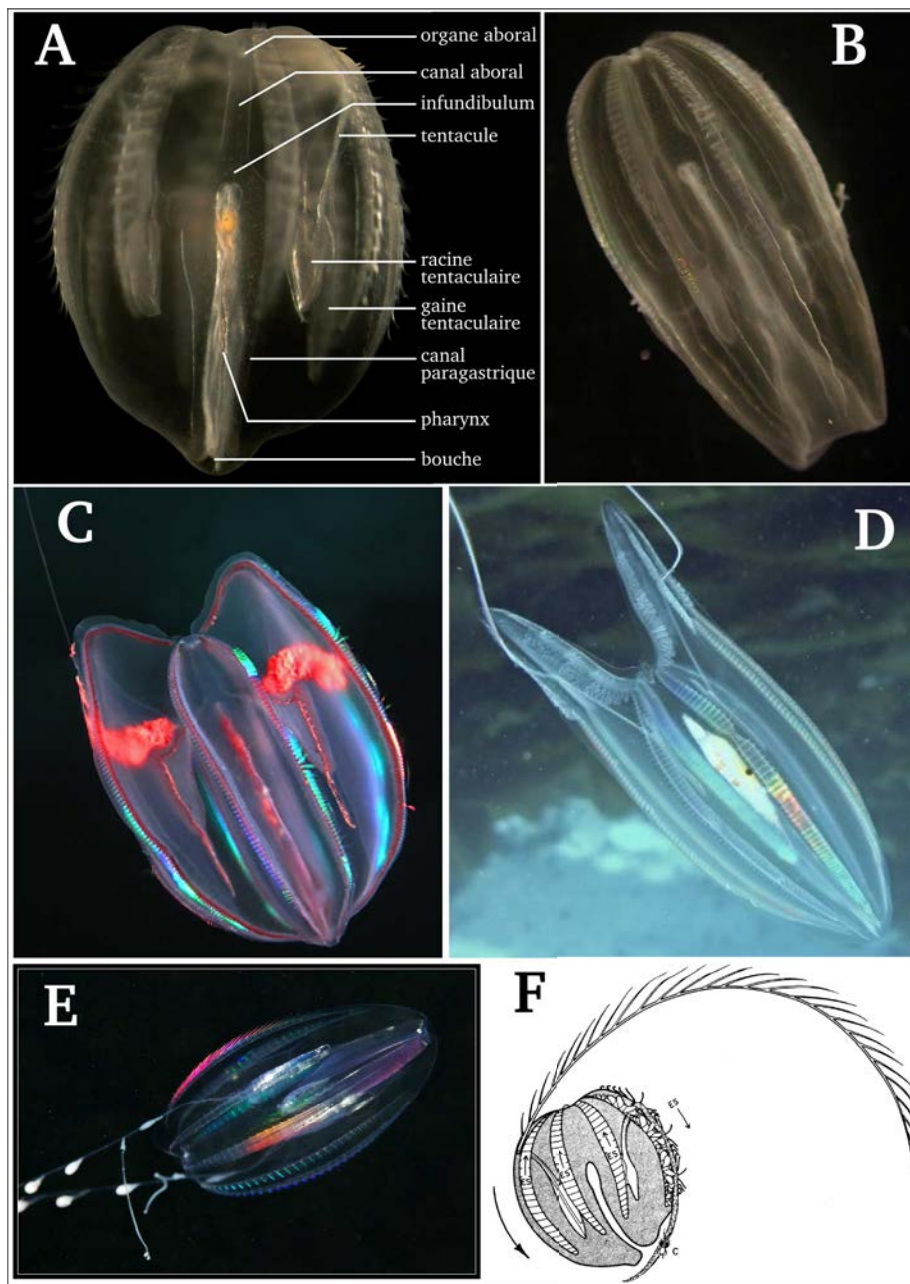


FIGURE 1.9 – Morphologie générale de différents représentants de l’ancien ordre des “Cydippida”. A) *Pleurobrachia pileus*, en vue latérale pharyngienne (photo originale). B) *Lampea pancerina*, en vue latérale (photo internet, auteur inconnu). C) *Mertensia ovum* (photo de Thormar). D) *Callianira antarctica* (photo de David Cothran). E) *Euplokamis* sp. (photo de Alexander Semenov). F) *Pleurobrachia* tournant sur lui-même afin d’apporter son tentacule à sa bouche (Tamm 2014). Légendes : c : copépodes ; es : sens de propulsion des peignes.

MÜLLER, 1776) (famille des Bolinopsidae), *Ocyropsis crystallina* (famille des Ocyropsidae) ou encore *Leucothea multicornis* (famille des Leucotheidae). Les Lobata sont parmi les cténophores les plus fragiles qui soient. Ils partagent un certain degré de compression dans l'axe tentaculaire et d'élargissement dans l'axe pharyngien. Cet élargissement est parfois très visible car exacerbé par la présence de deux lobes musculeux qui entourent la bouche, toujours dans le plan pharyngien. Ces lobes oraux, qui donnent son nom à l'ordre des Lobata, portent les quatre rangées de peignes subpharyngiennes. Les quatre rangées de peignes subtentaculaires sont portées, en leur partie la plus orale, par quatre auricules, extensions plus petites que les lobes mais qui entourent également la bouche (Figure 1.10, voir aussi Figure 1.6 et Figure 1.7 B). Les deux lobes et les quatre auricules servent à la capture des proies. Les tentacules des lobés sont situés très près de la bouche et des tentacules secondaires sont logés dans un sillon buccal. Les tentacules capturent ainsi les proies prises au piège par les lobes et auricules (voir détails dans la section 1.1.1 page 10). Les Lobata juvéniles ressemblent beaucoup aux "Cydippida" et subissent ensuite une métamorphose durant laquelle les lobes et auricules apparaissent, la racine tentaculaire migre à proximité de la bouche et la gaine tentaculaire qui entoure le tentacule disparaît. Chez l'adulte, les terminaisons orales des canaux méridiens subtentaculaires fusionnent avec les parties orales des canaux paragastriques, formant ainsi un complexe circumoral en forme d'anneau. Les canaux méridiens subpharyngiens fusionnent par deux, dans chacun des lobes (Hyman 1940, voir Figure 1.7 B).

Les Platyctenida

L'ordre des Platyctenida comprend 5 familles et 6 genres, dont les représentants sont les seuls cténophores qu'il peut être difficile d'identifier en tant que cténophores. En effet, l'immense majorité des autres groupes sont transparents, planctoniques, de forme arrondie et nagent en utilisant le battement de leurs peignes (voir section 1.1.1). Les Platyctenida peuvent en revanche être opaques, colorés, benthiques, plats et la plupart ne possèdent pas de rangées de peignes à l'état adulte. Ils sont compressés dans l'axe oral-aboral et vivent sur le substrat, la bouche tellement ouverte et étendue qu'elle représente la sole de reptation de l'animal (Figure 1.11). Certaines espèces peuvent refermer cette "bouche", ce qui leur confère alors une forme arrondie et leur permet de dériver selon les courants d'eau. Seul le genre *Ctenoplana* possède des rangées de peignes à l'état adulte, et peut alterner de véritables épisodes de vie pélagique avec des épisodes de vie benthique (Figure 1.11 E). Les Platyctenida possèdent une paire de tentacules munis de tentilles qu'ils utilisent pour attraper leurs proies et rétractent ensuite vers leur bouche. Ils ne possèdent pas de canaux interradiaires, et leurs canaux méridiens sont souvent fortement anastomosés entre eux, créant un véritable réseau gastrovasculaire s'étendant sur une grande partie du corps de l'animal (voir Figure 1.7 B, C). Chacun des deux canaux tentaculaires se scinde en deux avant de rejoindre sa gaine tentaculaire respective (Rankin 1956). Certains Platyctenida comme les *Coeloplana* possèdent une structure unique chez les cténophores, les poches de gestation. Ces dernières, formées

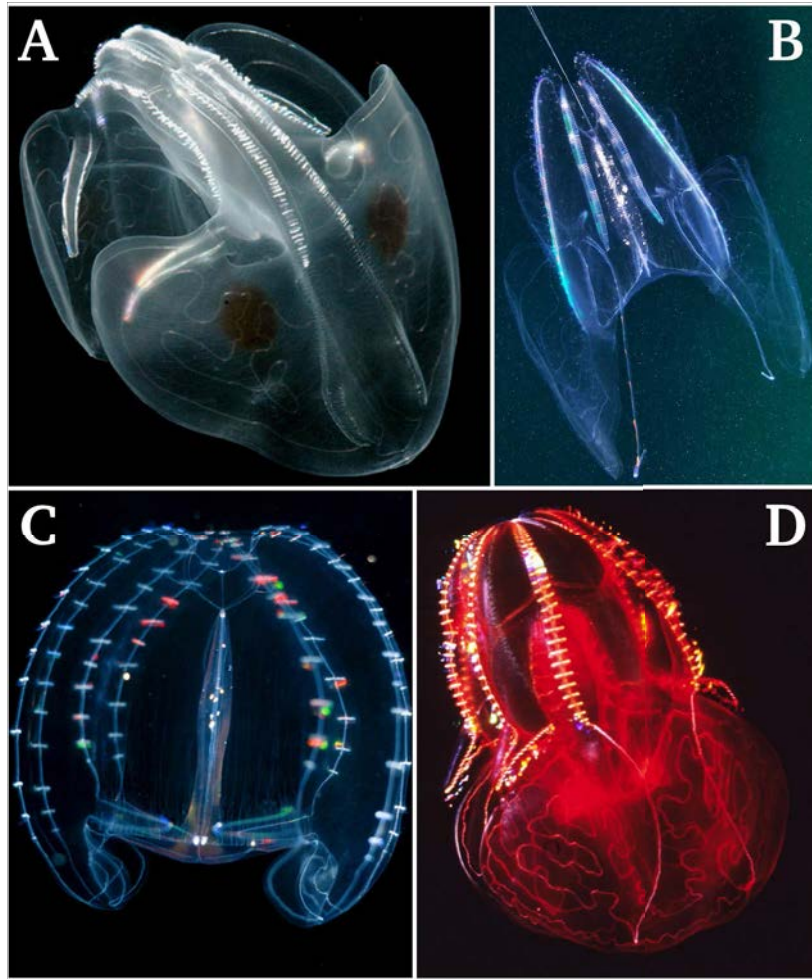


FIGURE 1.10 – Apparence générale d’espèces appartenant à l’ordre des Lobata. A) *Ocyropsis maculata* (photo de Larry Madin, WHOI). B) *Leucothea* sp. (photo d’Alexander Semenov). C) *Deiopea kaloktenota* (photo de l’internaute “Stylistvn8”). D) *Lampocteis cruentiventer* (Kevin Raskoff, MBARI).

par des invaginations dorsales au niveau des ovaires, servent de réceptacle séminale et abritent les embryons lors de leur développement (Komai 1963), qui peut ainsi être considéré comme interne.

Les Beroida

L’ordre des Beroida contient une seule famille et deux genres, *Beroe* et *Neis*. Les *Beroe* sont parmi les plus connus des cténophores car leurs tissus sont relativement résistants, leur distribution géographique couvre tous les océans et ils peuvent apparaître lors de *blooms* sur les côtes (voir section 1.1.1). Plus ou moins comprimés dans l’axe tentaculaire, ils ne possèdent, à aucun stade de leur développement ou de leur croissance, de tentacule ou de gaine tentaculaire. Ils possèdent un très grand pharynx qui occupe une large partie de leur corps. Leurs champs polaires sont relativement courts et sont bordés de papilles aborales (Figure 1.1 D). Selon les espèces, les canaux méridiens peuvent posséder ou non des extensions latérales (les diverticules) et celles-ci peuvent ou non s’anastomoser, créant ainsi un dense réseau gastrovasculaire (Figure 1.12 A, B). Les canaux

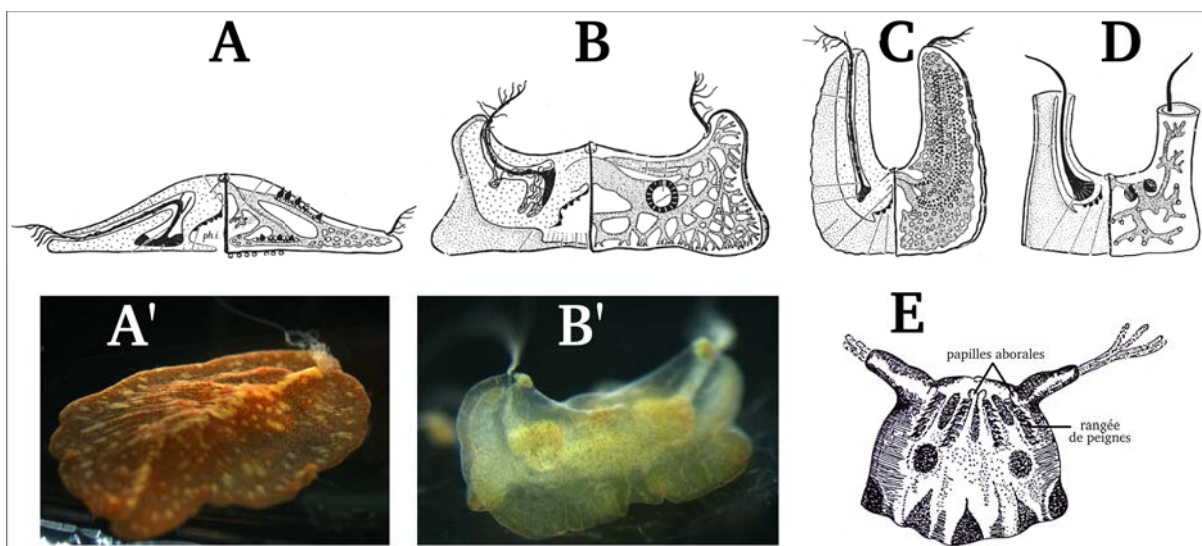


FIGURE 1.11 – Morphologie générale des Platyctenida. A, B, C et D) Anatomie interne, vue dans le plan pharyngien de *Coeloplana* (A), *Vallicula* (B), *Lyrocteis* (C) et de *Tjalfiella* (D) d'après Rankin (1956). Pour chaque dessin, la partie gauche représente l'appareil tentaculaire, et la partie droite représente le système gastrovasculaire et les gonades. A' et B') *Coeloplana* cf. *meteoris* et *Vallicula multiformis* (photos originales). E) Morphologie de *Ctenoplana* en posture de nage (d'après Hyman 1940).

interradiaires peuvent sembler absents comme chez *Beroe gracilis* (voir Figure 1.3 C), mais sont nettement distinguables chez *Beroe ovata* BRUGUIÈRE, 1789 (Figure 1.12 D). Les canaux méridiens et paragastriques se rejoignent au niveau de la bouche et fusionnent en un complexe circumoral plus ou moins complet : deux canaux semi-circumoraux chez les membres du genre *Beroe*, et un canal circumoral complet chez les membres du genre *Neis* (Lendenfeld 1885). Lorsque le complexe circumoral n'est pas complet mais composé de deux canaux semi-circumoraux, ce sont les canaux méridiens subpharyngiens adjacents qui ne fusionnent pas. Certaines zones de la bouche des Beroïda sont tapissées de dents macrociliaires (voir section 1.1.1 page 6 et Figure 1.2).

Les Cestida

Les représentants des Cestida appartiennent tous à la famille des Cestidae, qui contient deux genres, *Cestum* et *Velamen*, contenant chacun une espèce, respectivement *Cestum veneris* et *Velamen parrallelum* (FOL, 1869). Ils sont extrêmement comprimés dans l'axe tentaculaire à tel point que leur corps prend la forme d'un ruban (*Cestum veneris* est d'ailleurs appelée "Ceinture de Vénus"). Leurs rangées de peignes subtentaculaires sont très réduites et les rangées de peignes subpharyngiennes très longues, parcourant toute la longueur de la face aborale de l'animal. Les Cestida se meuvent par ondulation musculaire de leur corps, à la manière de serpents ou d'anguilles. Leurs rangées de peignes sont fonctionnelles et servent probablement à améliorer la stabilité et la pénétration dans l'eau (Figure 1.13). Comme chez les Lobata, leurs tentacules sont très proches de la bouche, mais sont en revanche entourés d'une gaine tentaculaire chez les adultes. Des tentacules secondaires sont logés dans un sillon buccal qui parcourt toute la face orale de l'animal. Chaque

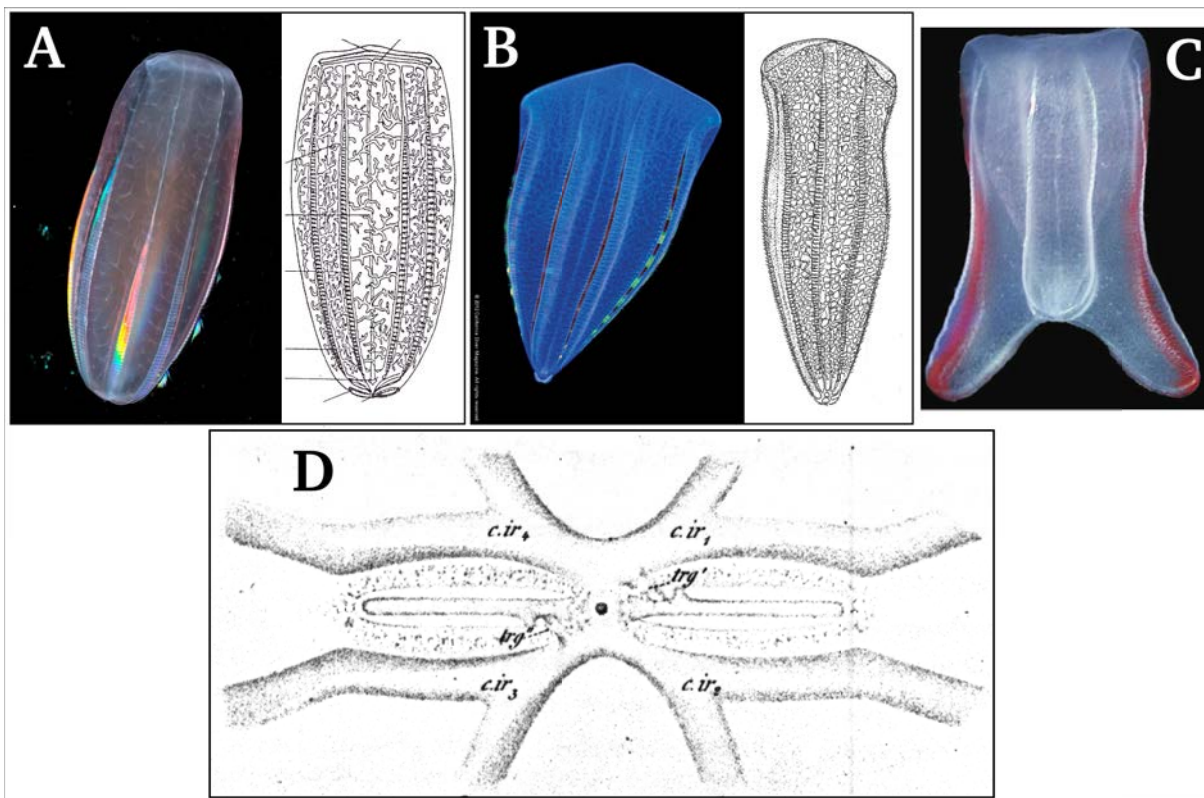


FIGURE 1.12 – Morphologie générale de plusieurs espèces de l'ordre des Beroida. A) *Beroe cucumis* (photo de Edith Widder). B) *Beroe forskalii* (photo de Kevin Lee). C) *Neis cordiger* (dans Gerschwin et al. 2010). D) Organisation du système gastrovasculaire périphérique de *Beroe ovata*, en vue aborale, d'après Chun (1880). Légendes : c.ir : canal interradaire; trg : pore anal.

canal paragastrique se scinde en deux aux abords du pharynx avant que chacune des deux parties, au niveau de la bouche, ne se réoriente vers une extrémité latérale de l'animal. À cette extrémité, chaque canal paragastrique fusionne avec les deux canaux méridiens (un subtentaculaire et un subpharyngien) de son quadrant. Les Cestida possèdent ainsi un complexe circumoral constitué de deux canaux semi-circumoraux, comme les Beroida. Les Cestida sont parmi les cténophores les plus délicats, et leur grande taille (jusqu'à 1m50 pour *Cestum veneris*) les rend vulnérables aux mouvements d'eau.

Les Thalassocalycida

L'ordre des Thalassocalycida est monospécifique. L'espèce qu'il contient, *Thalassocalyce inconstans* MADIN & HARBISON, 1978, a une apparence unique, puisque sa forme de cloche la fait fortement ressembler à une méduse (Figure 1.14). Au centre de la cloche, un pédoncule conique porte la bouche. Les Thalassocalycida possèdent une paire de tentacules avec des filaments latéraux, mais ces tentacules ne sont pas entourés d'une gaine tentaculaire. Les canaux paragastriques se terminent, clos, près de la bouche. Deux canaux perradiaires et quatre canaux interradiaires sont présents et tous les canaux méridiens sont clos en leur extrémité aborale (Figure 1.14 A et B).

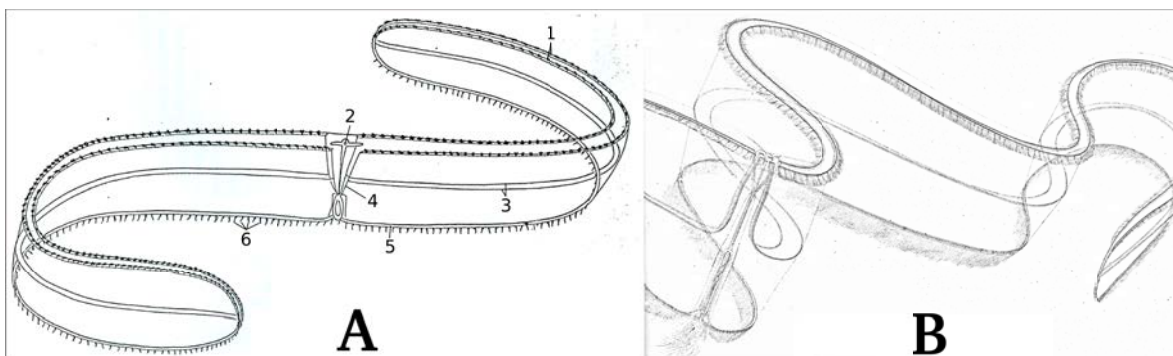


FIGURE 1.13 – Anatomie générale des deux espèces de Cestida. A) Dessin d’observation de *Velamen parallelum* (Hyman (1940), d’après Mayer (1912)). B) Dessin d’observation de *Cestum veneris* d’après Chun (1880). Légendes : 1 : rangées de peignes subpharyngienne ; 2 : organe aboral ; 3 : canaux méridiens subtentaculaires ; 4 : canal interradiare ; 5 : canal paragastrique ; 6 : tentacules secondaires.

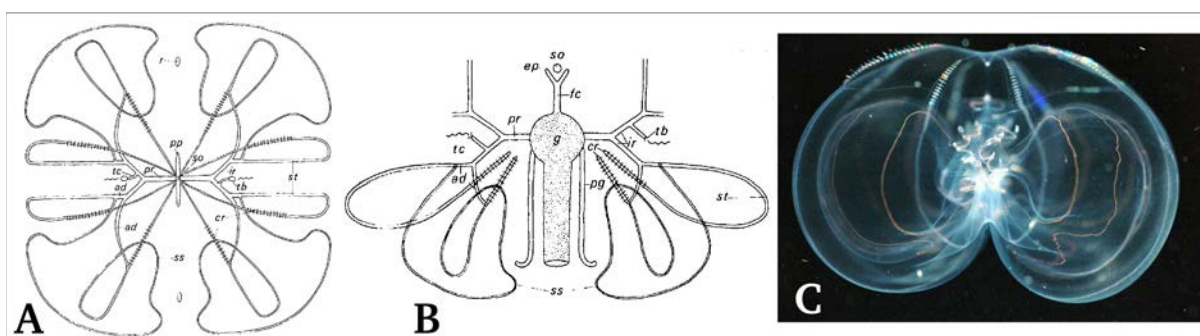


FIGURE 1.14 – Anatomie de l’unique espèce de l’ordre des Thalassocalycida, *Thalassocalyce inconstans*. A et B) Schémas de la structure du système gastrovasculaire de *Thalassocalyce inconstans*, d’après Madin and Harbison (1978), en vue aborale (A) et en vue latérale pharyngienne. C) Individu qui, par contraction, prend une forme “pincée” (photo par Jesse Claggett). Légendes : ad : canal adradiaire ; cr : rangées de peignes ; ep : pore anal ; fc : canal aboral ; g : infundibulum ; ir : canal interradiare ; pg : canal paragastrique ; pp : champs polaire ; pr : canal perradiare ; r : crête ; so : organe aboral ; ss : canal méridien subpharyngien ; st : canal méridien subtentaculaire ; tb : racine tentaculaire ; tc : canal tentaculaire.

Lorsque les Thalassocalycida contractent leur corps, cela a pour effet de “pincer” l’animal dans le plan tentaculaire, ce qui dessine alors deux moitiés de cloche dans le plan pharyngien (Figure 1.14 C). Ces deux moitiés de cloche rappellent fortement les lobes oraux des Lobata. Les Thalassocalycida possèdent donc à la fois des caractéristiques de Cydippida et de Lobata.

Les Ganeshida

Deux espèces composent l’ordre des Ganeshida : *Ganescha elegans* (MOSER, 1903) et *Ganesha annamita* DAWYDOFF, 1946. Elles ont en commun un pharynx très développé qui occupe une grande partie du corps et qui est très particulier : il est séparé en deux parties, applaties chacune dans un plan différent. La partie interne du pharynx est, comme chez tous les autres cténophores, applatie dans le plan pharyngien. La partie externe élargie est quand à elle applatie dans le plan tentaculaire (Figure 1.15). Une telle séparation du pharynx en deux parties ne se retrouve que

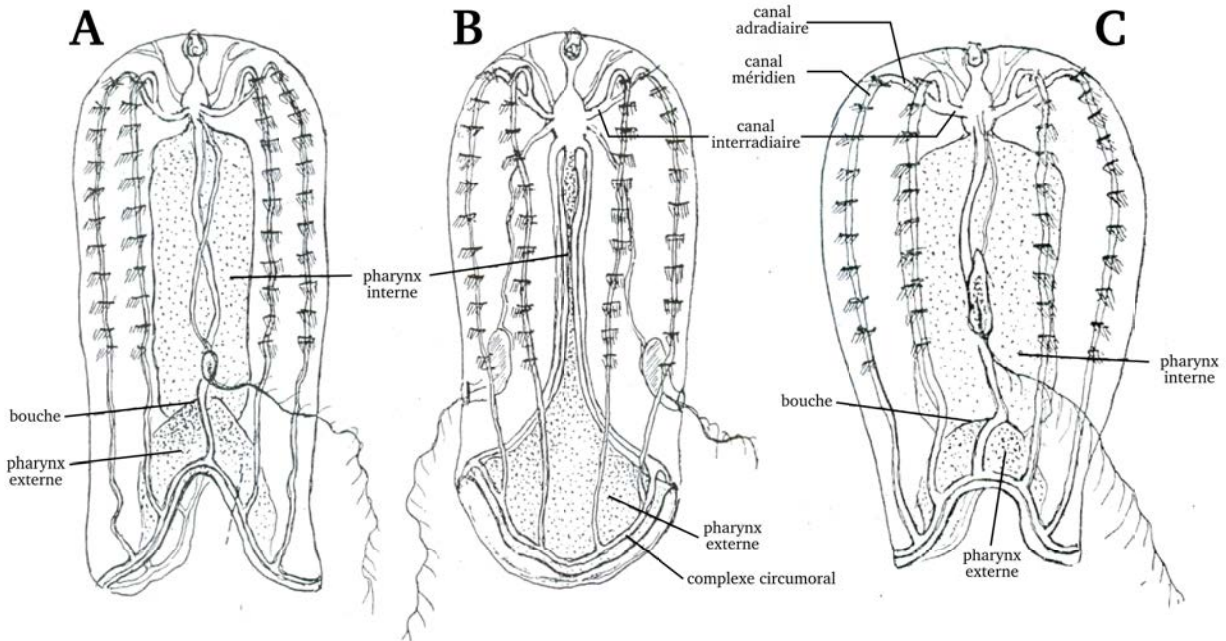


FIGURE 1.15 – Anatomie des deux espèces de l'ordre des Ganeshida, d'après Dawydoff (1946). A et B) *Ganeshia elegans* en vue latérale tentaculaire (A), et en vue latérale pharyngienne (B). C) *Ganeshia annamita*, en vue latérale tentaculaire.

chez les Platyctenida, où l'élargissement de la partie externe du pharynx existe aussi, aplatie également dans le plan tentaculaire. Chez les Ganeshida adultes, les gaines tentaculaires sont réduites, et les canaux interradiaires émergent directement de l'infundibulum. Enfin, les canaux adradiaires connectent les canaux méridionaux en leur extrémité aborale, de la même manière que chez les Lobata (Dawydoff 1946).

Les Cambojiida

L'ordre des Cambojiida⁵ est constitué d'une unique espèce, *Cambodgia elegantissima* DAWYDOFF, 1946, décrite d'après l'observation de deux individus pêchés dans le Golfe de Siam en mars 1930. Ils possèdent de grands champs polaires de forme presque rectangulaire. Leur large bouche extensible leur permet de capturer des proies aussi grosses qu'eux, comme des *Beroe* (Dawydoff 1946). Cette bouche donne sur un vestibule aplati dans le plan pharyngien. Des canaux perradiaires et interradiaires sont présents, et la connexion des canaux adradiaires aux canaux méridiens se fait au milieu de ces derniers. Les canaux méridiens possèdent des diverticules sur leur moitié la plus orale. De chaque côté du corps, les canaux méridiens subtentaculaires et le canal paragastrique fusionnent en leur partie la plus orale pour donner un semi-canal circumoral. Les deux canaux semi-circumoraux forment ainsi le complexe circumoral (Figure 1.16). L'auteur signale qu'il n'a pas pu démontrer que les deux individus décrits sont bien des adultes, on peut donc penser

5. Cet ordre contient l'unique genre *Cambodgia* DAWYDOFF, 1946, mais lors de sa création par Osipov (1985a), l'auteur mentionne le genre "*Cambojia*", utilisant donc une mauvaise orthographe. Le nom correct de cet ordre aurait probablement dû être *Cambodgiida*.

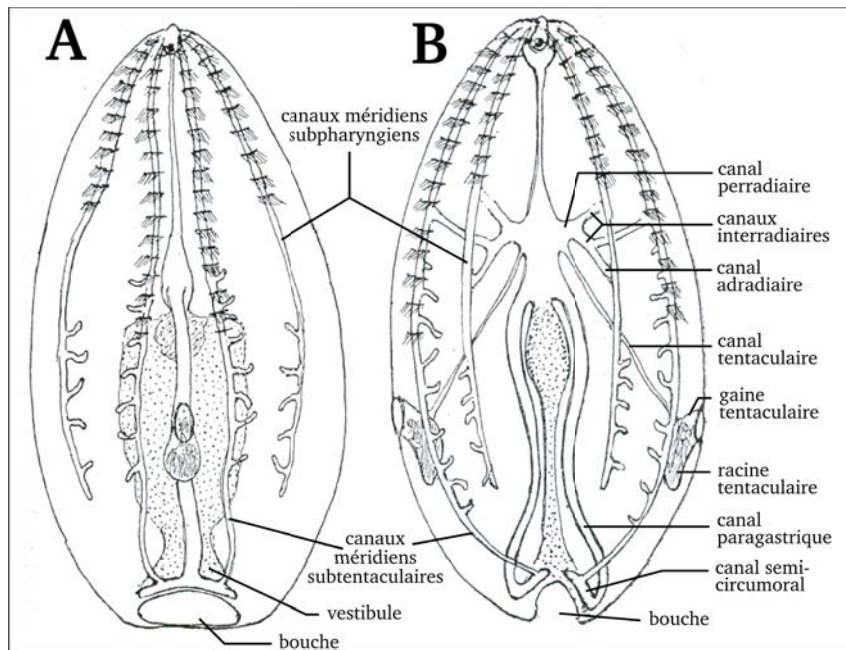


FIGURE 1.16 – Anatomie de *Cambodgia elegantissima*, d'après Dawydoff (1946). A) Individu en vue latérale tentaculaire. B) Individu en vue latérale pharyngienne.

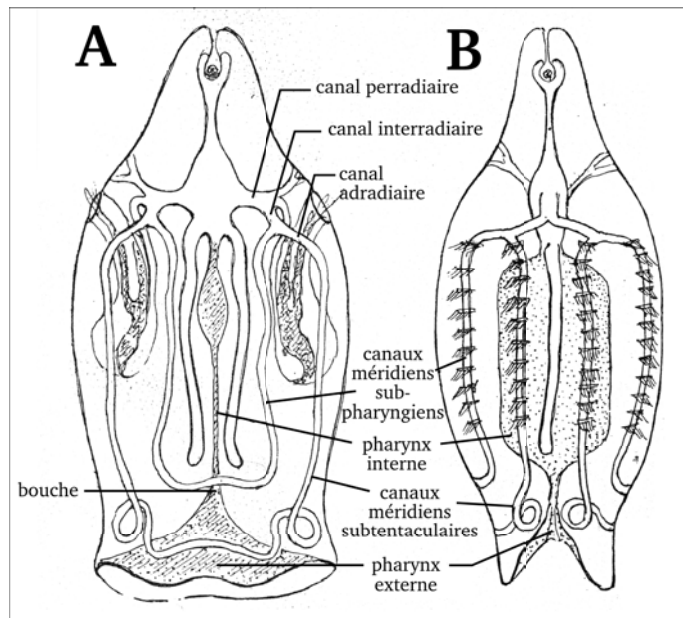


FIGURE 1.17 – Anatomie de *Lobocrypta annamita*, d'après Dawydoff (1946). A) Individu en vue latérale pharyngienne. B) Individu en vue latérale tentaculaire.

que la différence de longueur entre les canaux méridiens subpharyngiens et subtentaculaires ainsi que la structure du complexe circumoral pourraient changer lors de la croissance des individus.

Les Cryptolobiferida

L'ordre des Cryptolobiferida est composé de deux espèces : *Cryptolobata primitiva* MOSER, 1909, et *Lobocrypta annamita* DAWYDOFF, 1946. La première a été décrite en Indonésie, la seconde en mer de Chine. Tout comme les Ganeshida, ils possèdent une extension pharyngienne aplatie dans le plan tentaculaire (Figure 1.17). Leur appareil tentaculaire, ainsi que la structure de leurs canaux perradiaires, interradiaires et paragastriques rappellent en tout points ce que l'on peut observer chez les Pleurobrachiidae. En revanche, les canaux adradiaires des Cryptolobiferida connectent les canaux méridiens en leur extrémité la plus aborale, de la même manière que chez la plupart des Lobata. Un complexe circumoral est présent et est constitué de quatre "couples" : d'une part les canaux méridiens subpharyngiens adjacents fusionnent ensemble, formant ainsi deux "couples", et d'autre part les canaux méridiens subtentaculaires opposés fusionnent ensemble, créant deux couples supplémentaires (Figure 1.17). Les canaux paragastriques ne participent pas au complexe circumoral. L'existence de ces quatre couples déconnectés entre eux est diagnostique des Cryptolobiferida.

1.2 Évolution des cténophores

1.2.1 Les relations phylogénétiques entre cténophores

Hypothèses historiques sur l'évolution des cténophores

Bien qu'un certain nombre de monographies sur les cténophores existent et proposent parfois des éléments de réflexion quant à leurs relations de parenté, seulement trois études ont été dédiées à la reconstruction de leur évolution. Il s'agit de deux articles publiés en 1985, l'un en anglais (Harbison 1985), l'autre en russe (Ospovat 1985a) et d'un article publié en 2001 qui sera l'objet de la prochaine section (Podar et al. 2001). Les deux études publiées en 1985 sont en net désaccord : elles n'utilisent ni les mêmes caractères morphologiques, ni la même philosophie de reconstruction de l'évolution, ni n'aboutissent aux mêmes conclusions.

Évolution selon Harbison Pour Harbison (1985), certains caractères morphologiques comme la présence de diverticules ou d'anastomoses entre différents canaux sont inutilisables pour reconstituer les relations de parenté entre cténophores : "*Je pense que les complexes circumoraux ainsi que les diverticules représentent simplement des moyens pour obtenir une meilleure circulation plutôt que des indicateurs de relations phylogénétiques.*" (Harbison 1985). De la même manière, il ignore les caractères relatifs à l'applatissage du corps, justifiant que "*l'applatissage ne représente qu'une manière d'acquiescer une forme hydrodynamique*", et que le hasard suffit à expliquer que plusieurs groupes de cténophores soient aplatis dans le même plan (tentaculaire ou pharyngien). En revanche, il considère importants les comportements de chasse, et ainsi il privilégie les caractères

associés à l'appareil tentaculaire, aux lobes ou encore à la bouche. Il construit ensuite deux cladogrammes, probablement à l'aide d'un critère de parcimonie, mais sans réaliser d'analyse cladistique explicite. Ces deux hypothèses d'évolution correspondent en réalité aux mêmes relations de parenté, mais enracinées de manière différentes (Figure 1.18 A et B). En effet, Harbison considère que l'enracinement de l'arbre est la vraie difficulté, à cause du manque de données ontogénétiques et paléontologiques.

L'interprétation préférée d'Harbison suggère que les *Beroida* seraient le groupe-frère des autres cténophores (Figure 1.18 A), que l'ancêtre des cténophores ne possédait donc aucun appareil tentaculaire, et qu'il se nourrissait de plancton gélatineux. Ensuite, un appareil tentaculaire simple se serait développé, comme chez les *Haeckeliidae*, qui se nourrissent de cnidaires. L'apparition des tentilles sur les tentacules aurait permis ensuite une radiation adaptative, car elles permettent la capture efficace de proies de tailles très variables (Figure 1.18 A). La seconde hypothèse d'Harbison implique un ancêtre des cténophores possédant un appareil tentaculaire complet (Figure 1.18 B). Ensuite, une lignée composée des *Lampeidae*, *Mertensiidae*, *Haeckeliidae*, *Platyctenida* et *Beroida*, se serait spécialisée dans l'ingestion de grosse proies gélatineuses. On y assiste ensuite à la réduction progressive de l'appareil tentaculaire, jusqu'à son absence totale chez *Beroe*. Une seconde lignée composée des *Pleurobrachiidae*, *Lobata*, *Cestida*, *Thalassocalycida* et *Ganeshida*, se serait quant à elle spécialisée dans la capture de petites proies (Figure 1.18 B).

Évolution selon Ospovat Contrairement à Harbison, Ospovat (1985a) ignore les caractères relatifs à l'appareil tentaculaire pour reconstruire les relations de parenté des cténophores. En fait, il se concentre exclusivement sur les caractères morphologiques en lien avec la formation et la structure du complexe circumoral, ces mêmes caractères qui sont complètement ignorés par Harbison. Ospovat reconstitue une évolution dans le cadre théorique de la systématique évolutionniste, c'est-à-dire dans un gradisme qui autorise des taxons à descendre les uns des autres. Aussi, la figure principale de cette étude n'est pas un cladogramme, mais une généalogie des cténophores (Figure 1.18 C). Pour faciliter la comparaison, mais également pour mettre en avant certains résultats du travail d'Ospovat, sa figure a été transformée et présentée sous forme de cladogramme (Figure 1.18 D). Pour former cette figure, tous les groupes "ancestraux" chez Ospovat ont été transformés en groupe-frère de tous leurs "descendants".

L'hypothèse d'Ospovat consiste à présenter les *Cydippida* et leur absence de complexe circumoral comme ancestraux aux cténophores. Une lignée comprenant le genre *Lampea* aurait donné naissance aux *Platyctenida* (ces deux groupes partagent la capacité d'éversion de leur pharynx, qu'ils peuvent utiliser comme sole pedieuse, Komai 1963). Une seconde lignée émerge des *Mertensiidae* pour donner naissance à un groupe de cténophores où le complexe circumoral apparaît. Dans ce groupe, les *Cambojiida* donnent naissance aux *Cestida* d'une part, et aux *Beroida*, *Ganeshida*, *Cryptolobiferida*, *Lobata* et *Thalassocalycida* d'autre part. Ces ordres sont caractérisés par des complexes circumoraux différents.

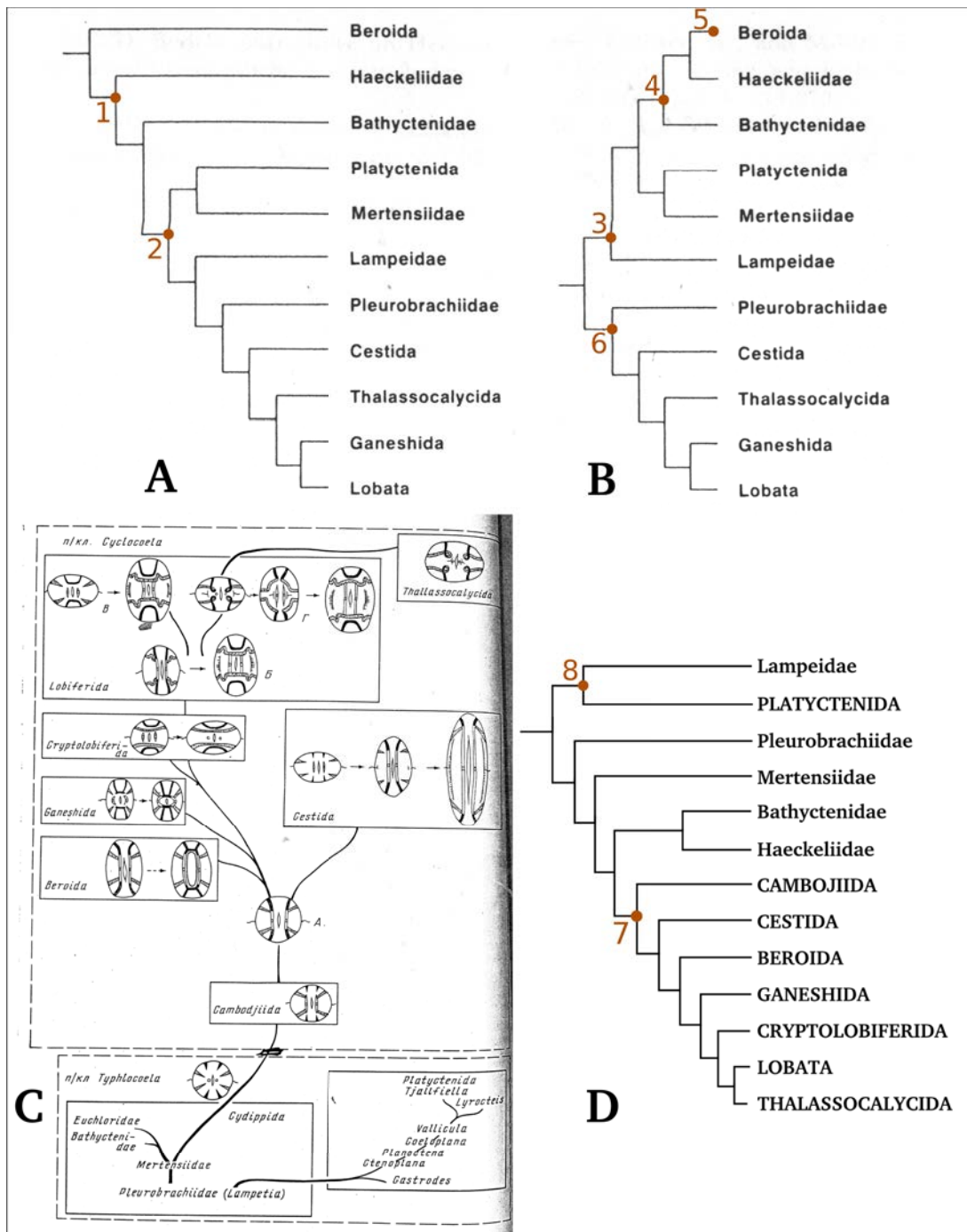


FIGURE 1.18 – Hypothèses des relations phylogénétiques des cténophores basées sur leur morphologie. A) Evolution des cténophores si leur ancêtre ne possédait pas de tentacules, selon Harbison (1985). B) Evolution des cténophores si leur ancêtre possédait des tentacules, selon Harbison (1985). C) Evolution des cténophores dans un contexte gradiste, selon Ospovat (1985a). D) Tentative de transposition des idées d'Ospovat dans un contexte cladistique. Chaque groupe “ancestral” a été considéré comme le groupe-frère de ses “descendants”. Cette figure, bien qu’adaptée des travaux d'Ospovat, ne peut en être considérée comme le résultat direct. Les taxons en lettres capitales sont des ordres, les autres sont les familles de l’ordre des “Cydippida”. Synapomorphies : 1 : acquisition d’un appareil tentaculaire simple ; 2 : acquisition des tentilles ; 3 : ingestion de grosses proies gélatineuses ; 4 : perte des tentilles ; 5 : perte de l’appareil tentaculaire ; 6 : ingestion de petites proies ; 7 : acquisition d’un complexe circumoral ; 8 : éversion possible du pharynx.

Un exercice morphologique difficile Les deux études précédentes souffrent du même problème : il existe peu de caractères morphologiques phylogénétiquement informatifs chez les cténophores. Même Harbison, qui essaie de réunir le plus d'information qu'il peut, ne trouve qu'une dizaine de caractères (Figure 1.18 A, B). De plus, les deux auteurs adoptent une démarche de choix *a priori* des caractères à utiliser pour la reconstruction des relations de parenté. Enfin, aucun des deux auteurs n'utilise de méthodologie satisfaisante : Ospovat n'utilise pas le cadre de la cladistique, et Harbison, tout en présentant des cladogrammes, n'utilise aucune analyse explicite pour les construire.

Il résulte de ces deux études des histoires évolutives largement incompatibles, sauf pour l'existence d'un groupe comprenant les Ganeshida, Lobata et Thalassocalycida (Harbison ne parle pas des Cryptolobiferida). Ces hypothèses diffèrent même à haute échelle taxonomique : l'hypothèse d'Harbison est en accord avec la séparation des cténophores en deux classes Nuda (ç.-à-d. les Beroida) et Tentaculata (ç.-à-d. tous les autres cténophores), tandis qu'Ospovat argumente pour une autre séparation, en deux classes Typhlocoela (comprennant les Cydippida et les Platyctenida) et les Cyclocoela (comprennant tous les autres cténophores). La très forte incompatibilité de ces résultats s'explique très simplement par le choix des caractères étudiés : il n'y en a pas un seul en commun entre les deux études !

Indépendamment du fait que ces deux auteurs utilisent des méthodologies discutables, le faible nombre de caractères anatomiques informatifs chez les cténophores laisse suspicieux quant à la possibilité de reconstruire leurs liens de parenté avec la seule utilisation de leur morphologie.

Apports de l'étude de l'ADNr 18S

En 2001 a été publiée par Podar et ses collaborateurs la première analyse phylogénétique visant à résoudre les relations de parenté au sein de l'embranchement des cténophores (Podar et al. 2001). Cette étude est basée sur des données moléculaires, les séquences du gène ribosomique ADNr 18S. L'utilisation de ce type de données a l'avantage de contourner le problème du faible nombre de caractères morphologiques chez les cténophores, tout en utilisant des caractères qui n'auront pas été choisis *a priori*. Ce gène a été séquencé pour 27 cténophores, représentant 10 familles.

Le principal résultat de cette analyse est la démonstration de la paraphylie de l'ordre des Cydippida. En effet, la famille des Mertensiidae se trouve plus proche de l'ordre des Platyctenida, et la famille des Haeckeliidae se trouve plus proche de l'ordre des Beroida, que des autres familles de Cydippida (Figure 1.19 A). Ces rapprochements sont cohérents avec la morphologie des cténophores. Mertensiidae et Platyctenida partagent la même architecture de canaux gastriques périphériques (ç.-à-d. l'absence de canaux interradiaires, voir Figure 1.7 C'), et Haeckeliidae et Beroida partagent les mêmes comportements de nage, la présence de papilles aborales, la présence d'un large pharynx, d'une bouche large et extensible, et un certain niveau de réduction de l'appareil tentaculaire.

On observe également une quasi absence de diversité génétique entre les Lobata, Cestida et

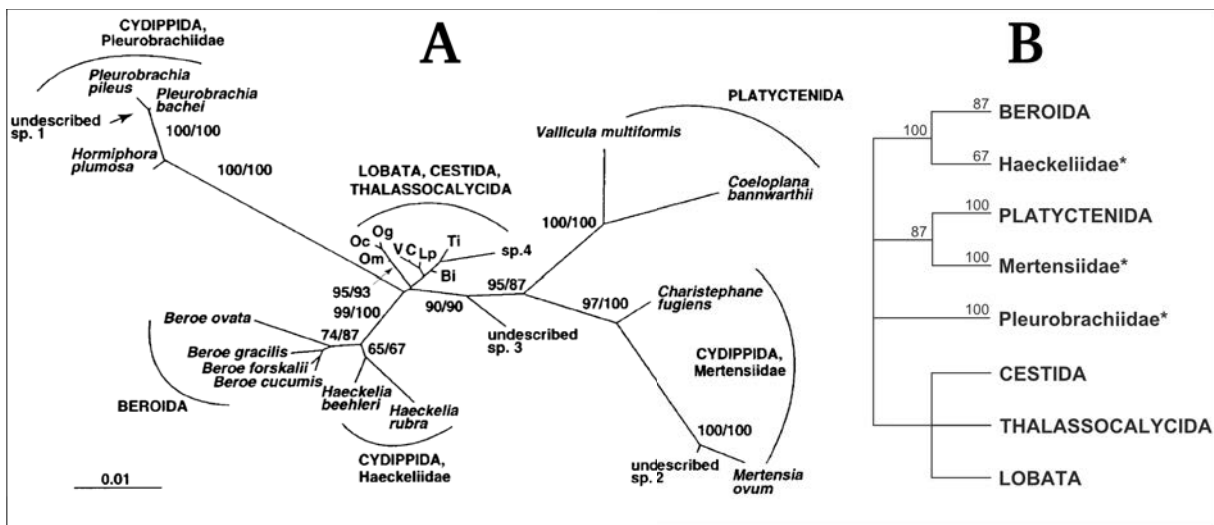


FIGURE 1.19 – Relations phylogénétiques entre cténophores, telles que reconstruites par Podar et al. (2001) avec des séquences du gène ribosomique ADNr 18S. A) Arbre phylogénétique non raciné. Les bootstraps obtenus en parcimonie et en maximum de vraisemblance supérieurs à 50 sont respectivement indiqués à chaque noeud. B) Arbre simplifié correspondant aux résultats des travaux de Podar et al. (2001). Les familles indiquées par des astérisques appartiennent à l'ordre paraphyletique des Cydippida.

Thalassocalycida, ce qui confirme l'idée, présente déjà chez Harbison, Ospovat, et de nombreux auteurs avant eux, que ces trois groupes sont très proches entre eux. Un fort aplatissement du corps dans le plan pharyngien rapproche Lobata et Cestida, ainsi que l'élargissement de leur bouche en une fente buccale bordée de tentacules secondaires. Les Thalassocalycida, eux possèdent un complexe circumoral extrêmement similaire à celui des Lobata.

L'analyse échoue en revanche à deux niveaux : la position de la racine chez les cténophores est irrésolue (la famille des Mertensiidae semble être le groupe-frère de tous les autres, mais sans soutien statistique), et les relations entre les Beroïda + Haeckeliidae, les Pleurobrachiidae, et les Lobata + Cestida + Thalassocalycida sont totalement irrésolues (Figure 1.19 A).

De manière plus générale, cette étude montre la très faible variabilité du gène ADNr 18S à l'échelle de l'embranchement des cténophores, avec un niveau de divergence des séquences inférieur à 5% (Podar et al. 2001). Cette faible diversité suggère que l'ancêtre commun des cténophores actuel est relativement récent. Ce résultat est surprenant car les disparités morphologiques élevées entre les cténophores laissent envisager une plus grande diversité moléculaire. La faible diversité moléculaire du gène ADNr 18S est également la cause des irrésolutions des relations de parenté entre les cténophores : il n'y a probablement pas assez de signal phylogénétique dans l'alignement des séquences de 18S pour les résoudre. Néanmoins, les résultats de cette étude rigoureuse apportent un point de départ crucial pour comprendre l'évolution des cténophores (Figure 1.19 B).

1.2.2 La position phylogénétique des cténophores au sein des métazoaires

La position phylogénétique des cténophores au sein des métazoaires est aujourd’hui encore très débattue. Ce problème est d’ailleurs loin de laisser la communauté scientifique, qui a profité de l’essor des nouvelles technologies de séquençage pour assembler des jeux de données de taille suffisante pour tenter de résoudre cette question. Depuis six ans, c’est ainsi huit études phylogénomiques qui ont été publiées sur le sujet, dont 3 dans les journaux *Nature* et *Science* (Dunn et al. 2008; Hejnal et al. 2009; Philippe et al. 2009; Schierwater et al. 2009; Pick et al. 2010; Nosenko et al. 2013; Ryan et al. 2013; Moroz et al. 2014). Ce grand intérêt pour la position des cténophores au sein de l’arbre des métazoaires s’explique par ses implications sur l’évolution de caractères importants à l’échelle des animaux (Figure 1.20).

En effet, les cténophores présentent des attributs morphologiques “complexes”, comme la présence de deux systèmes nerveux (un épithélial, l’autre mésogléen) associés à deux systèmes musculaires (dans les mêmes tissus). Les cténophores présentent une symétrie bi-radiale avec des éléments de symétrie tétra-radiale (notamment au niveau de l’organe aboral) et octo-radiale (comme les sillons ciliés et les rangées de peignes). Enfin, ils possèdent également un système digestif différencié, ramifié et ouvert aux deux bouts ainsi qu’un feuillet embryonnaire “mésodermique” (Martindale and Henry 1997b).

Les cténophores partagent donc avec les cnidaires et les bilatériens la présence de cellules nerveuses, musculaires, ainsi que l’existence de symétrie d’ordre défini, et ces trois lignées sont traditionnellement regroupées au sein des eumétazoaires. Les différentes positions phylogénétiques retrouvées pour les cténophores, par les différentes approches, ne sont pas toutes compatibles avec la monophylie des eumétazoaires (Figure 1.20 B et C) et impliquent donc des scénarios évolutifs très différents pour ces caractères anatomiques : dans le cas où les métazoaires ne sont pas monophylétiques, ils devront avoir été soit secondairement perdus chez certains embranchements, soit acquis indépendamment chez les cténophores d’une part, et les bilatériens et cnidaires d’autre part.

Les sections qui suivent passeront en revue l’apport de différentes sources de données quant à la position phylogénétique des cténophores.

Apports de la paléontologie

Les fossiles de cténophores sont rares, puisque pour conserver les parties molles des animaux, les conditions de fossilisation doivent être exceptionnelles, et que les cténophores ne possèdent pas de tissus rigides. Parmi ces fossiles, certains semblent posséder des rangées de peignes et une structure aborale et sont ainsi considérés sans grande ambiguïté comme apparentés à la lignée des cténophores (ç.-à-d. “stem-group” en anglais, Figure 1.21 A-F). Il s’agit des fossiles suivants :

- *Paleoctenophora brasseli* STANLEY & STÜRMER, 1983 (Stanley and Sturmer 1983, Figure

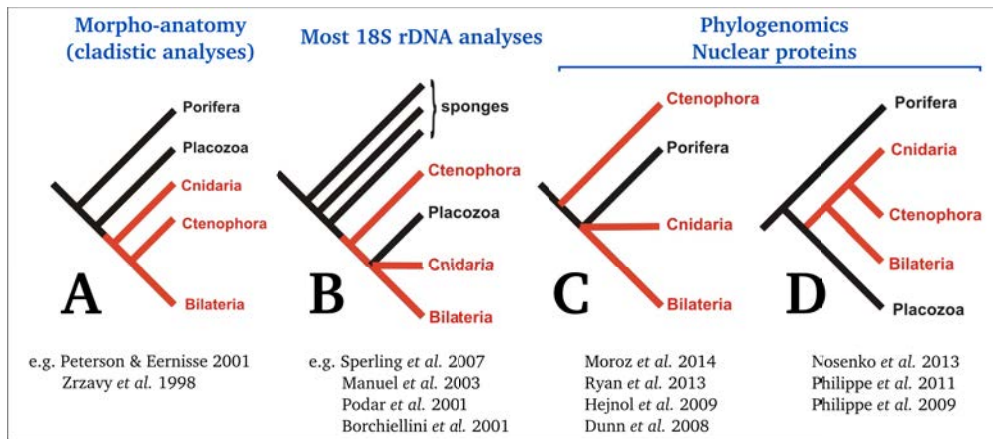


FIGURE 1.20 – Evolution des cellules nerveuses et musculaires selon différentes hypothèses de relations phylogénétiques chez les métazoaires. En rouge sont représentés les groupes possédant des cellules nerveuses et musculaires. Notez que l’analyse de Pick et al. (2010) a retrouvé, contrairement à ce qui est suggéré ici, les éponges monophylétiques.

1.21 A, daté à environ 400 Ma).

- *Archeocydippida hunsrueckiana* STANLEY & STÜRMER, 1987 (Stanley and Sturmer 1987, Figure 1.21 B, daté à environ 400 Ma).
- *Xanioascus canadensis* CONWAY MORRIS & COLLINS, 1996 (Conway Morris and Collins 1996, Figure 1.21 C, daté entre 505 et 515 Ma).
- *Ctenorhabdotus capulus* CONWAY MORRIS & COLLINS, 1996 (Conway Morris and Collins 1996, Figure 1.21 D, daté entre 505 et 515 Ma).
- *Maotianoascus octonarius* CHEN & ZHOU, 1998 (Chen and Zhou 1997, Figure 1.21 E, daté entre 515 et 521 Ma).
- Embryon de cténophore indéterminé (Chen et al. 2007, Figure 1.21 F, daté à environ 540 Ma).

Les interprétations de ces fossiles comme appartenant aux Ctenophora sont convaincantes, même si la preuve de l’existence d’une ciliature au niveau des rangées de peignes manque parfois. Tous ces fossiles partagent une forme globulaire et ainsi un mode de vie probablement pélagique. *Paleoectenophora*, *Archeocydippida*, *Maotianoascus* et l’embryon fossile découvert par Chen et al. (2007) possèdent tous 8 rangées de peignes, tandis que *Xanioascus* et *Ctenorhabdotus* en possèdent 24. Cette différence spectaculaire semble très bien se comprendre grâce à *Ctenorhabdotus*, où l’on observe trois rangées de peignes se rejoindre avant d’atteindre le pôle aboral, où seulement 8 éléments se connectent en un anneau (Figure 1.21 D en haut à droite). Il existait donc des cténophores possédant 8 rangées de peignes au début du Cambrien (*Maotianoascus* et le fossile d’embryon), au début du Dévonien (*Paleoectenophora*, *Archeocydippida*) comme de nos jours (tous les cténophores actuels possèdent strictement 8 rangées de peignes, voir section 1.1.1). Cependant, on trouve au milieu du Cambrien des cténophores qui possèdent 24 rangées de peignes. La découverte de nouveaux fossiles de cténophores devraient permettre de mieux comprendre et peut-être d’orienter l’évolution du

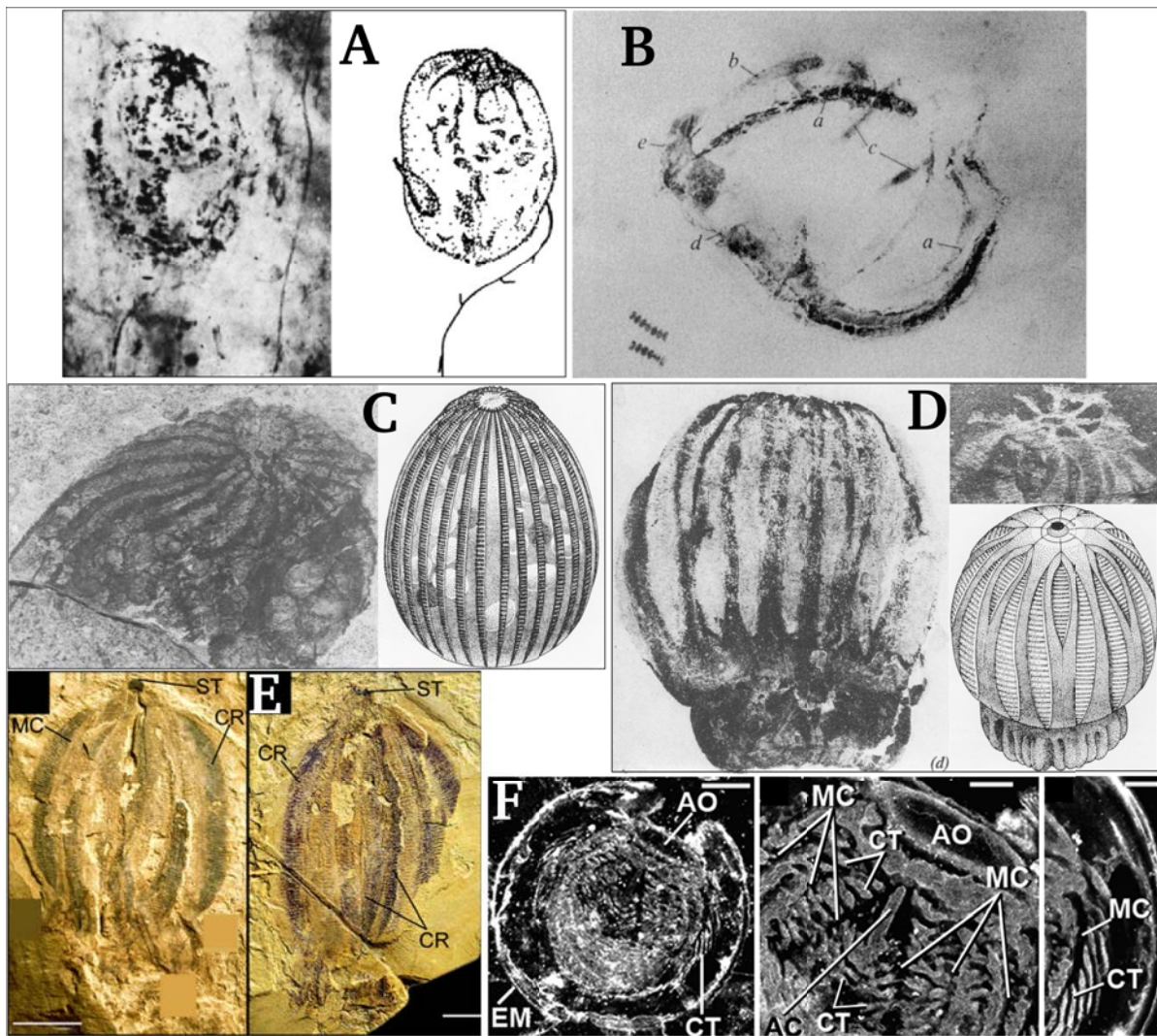


FIGURE 1.21 – Espèces fossiles de cténophores connus (liste exhaustive). A) *Paleoctenophora brasseli* (Stanley and Sturmer 1983). B) *Archeocydippida hunsrueckiana* (Stanley and Sturmer 1987). C) *Xanioascus canadensis* (Conway Morris and Collins 1996). D) *Ctenorhabdotus capulus* (Conway Morris and Collins 1996). E) *Maotianoascus octonarius* (Chen and Zhou 1997). F) Embryon de cténophore indéterminé (Chen et al. 2007). Légendes : MC : canal méridien ; CR : rangée de peignes ; ST : statolithe ; AO : organe aboral ; AC : canal aboral.

nombre de rangées de peignes au cours du temps.

Malheureusement, ces fossiles n'apportent pas de caractères morphologiques nouveaux qui pourraient aider à comprendre de quels autres animaux actuels les cténophores sont proches. Plusieurs études ont cependant cherché à rapprocher les cténophores d'autres fossiles de Metazoa, et certains d'entre eux ont été ré-interprétés comme faisant partie de la lignée des Ctenophora :

- *Fasciculus vesanus*, de la faune de Burgess, a été présenté comme un cténophore possédant 16 longues rangées de peignes et 64 courtes (Figure 1.22 A, Conway Morris and Collins 1996), soit 80 au total, c'est-à-dire un multiple de 8. L'interprétation de *Fasciculus* comme apparenté aux cténophores s'inscrit dans une hypothèse évolutive où les cténophores ont vécu une réduction progressive du nombre de rangées de peignes. Parmi les faiblesses de cette

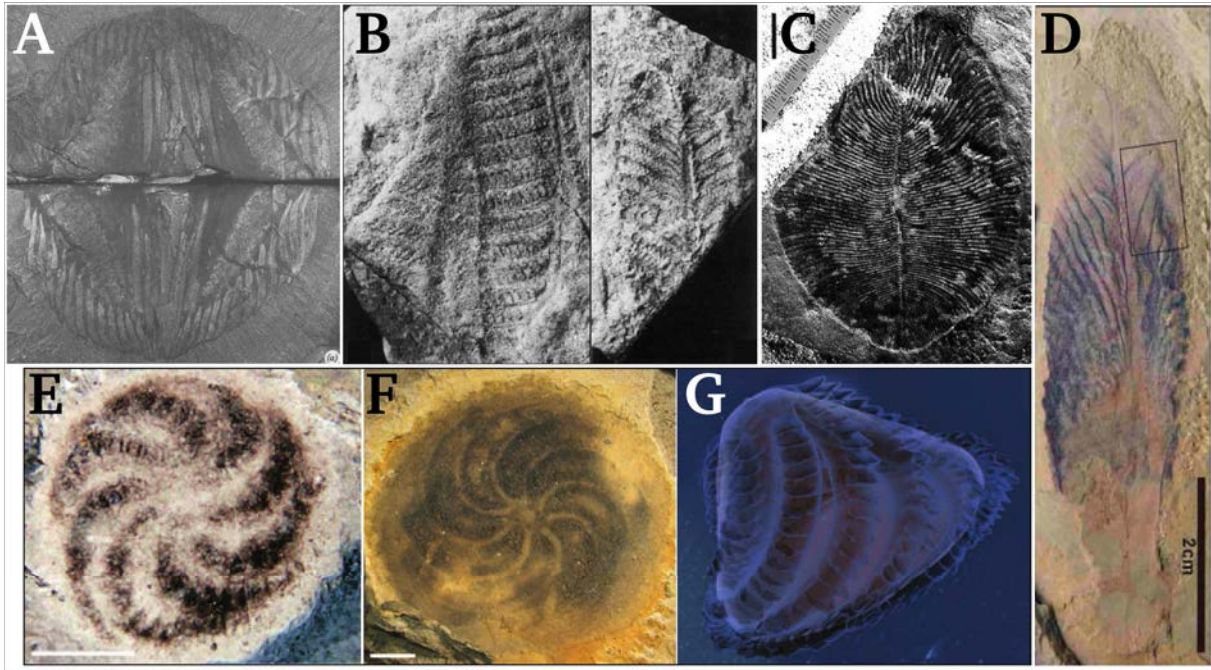


FIGURE 1.22 – Fossiles ayant été interprétés comme possiblement apparentés aux cténophores. A) *Fasciculus vesanus* (Conway Morris and Collins 1996). B) *Rangea* (Dzik 2002). C) *Dickinsonia* (Zhang and Reitner 2006). D) *Stromatoveris psygmoglena* (Shu et al. 2006). E, F et G) *Eoandromeda octobrachiata* (Tang et al. 2011). G) reconstruction artistique d'un adulte *Eoandromeda octobrachiata*.

interprétation, il faut noter que le fossile est incomplet, l'organe aboral n'est pas observé, les "rangées de peignes" ne convergent pas vers une zone aborale centrale, la présence de cils est hypothétique, et enfin, des cténophores plus anciens ne possédant que 8 rangées de peignes ont été décrits (?Chen et al. 2007).

- *Rangea*, daté de la fin de l'Ediacarien, a été comparé aux cténophores, et notamment à *Maotianoascus*, par Dzik (2002) (Figure 1.22 B). Cette interprétation repose sur 4 hypothèses d'homologies. L'une d'entre elle est convaincante et fait le lien entre les éléments tétraradiaires de *Rangea* et des cténophores. Les autres hypothèses d'homologies le sont moins (voir Table 1 dans Dzik 2002).
- *Dickinsonia*, daté de la fin de l'Ediacarien, a été ré-interprété comme possiblement apparenté aux cténophores (Zhang and Reitner 2006) (Figure 1.22 C). Les auteurs décrivent une structure gastrovasculaire très similaire à ce qui existe chez les cténophores actuels. Les deux limitations majeures de cette interprétation de *Dickinsonia* sont d'une part l'impossibilité de détecter bouche et anus, et d'autre part l'absence d'argument pour la présence de rangées de peignes. La conclusion de cette étude est donc logique : "Nous ne pouvons pas prouver de manière concluante une affinité avec les Ctenophores" (Zhang and Reitner 2006).
- *Stromatoveris psygmoglena*, daté du début du Cambrien, a été décrit par Shu et al. (2006) (Figure 1.22 D). Les auteurs associent ce fossile en "feuille" ("frond-like", en anglais) aux cténophores car il possède un certain nombre de branches, elles-même ornementées de struc-

tures transverses, comparables aux rangées de peignes des cténophores. La principale limite de cette comparaison est l'absence de preuve d'une ciliature associée à ces branches. En réalité, l'interprétation de ce fossile comme apparenté aux cténophores semble plus être une conséquence de l'impossibilité de l'interpréter comme proche des Cnidaria (ce qui représentait l'hypothèse traditionnelle en paléontologie), que la conséquence d'une forte ressemblance entre *Stromatoveris* et les cténophores.

- *Eoandromeda octobrachiata* TANG ET AL., 2008, un autre fossile de la faune de l'Ediacarien décrit par Tang et al. (2008) a été interprété comme apparenté aux cténophores (Tang et al. 2011) (Figure 1.22 E, F et G). Cette hypothèse est soutenue par l'existence de huit "bras" possédant de nettes structures transverses, comparables aux rangées de peignes. Ces huit bras se rejoignent en un anneau, au niveau d'un pôle de l'organisme, de la même manière que chez les cténophores fossiles *Ctenorhabdotus* et *Xanioascus* (voir Figure 1.21 C et D). Bien que plutôt convaincante, l'hypothèse de Tang souffre d'une difficulté majeure : l'absence de symétrie biradiaire chez *Eoandromeda*. En effet, chez ce fossile, les huit bras "courent" en spirale le long de l'animal.

Il y a ainsi parmi les paléontologistes deux points de vue qui s'opposent. Le premier, initié par Dzik et repris par Shu et Zhang, propose une émergence des cténophores à partir d'un ancêtre benthique en "feuille", en s'appuyant sur une gradation partant de *Dickinsonia*, et passant par *Rangea*, *Stromatoveris*, *Fasciculus*, pour arriver aux cténophores (voir Figure 1.22 A, B, C et D, ainsi que Dzik 2002, figure 8). Le second, suggéré par Tang, propose une émergence des cténophores à partir d'un ancêtre benthique spiralé, en s'appuyant sur la ressemblance entre *Eoandromeda* et les cténophores actuels et fossiles (voir Figure 1.22 E, F et G). Je ne me permettrais pas de prendre parti dans ce débat passionnant, n'en ayant pas les compétences, mais je ne peux m'empêcher de trouver, à titre personnel, l'hypothèse de Tang plus convaincante.

Apports de l'anatomie des cténophores actuels

Bien que les cténophores soient très facilement identifiables en tant qu'embranchement grâce à plusieurs synapomorphies aisément reconnaissables (comme les rangées de peignes, les colloblastes, ou l'organe aboral), il est extrêmement difficile de déterminer à quels autres métazoaires ils s'apparentent le plus. Dans sa revue sur le sujet, Harbison (1985) énumère la liste des groupes qui ont été considérés comme pouvant être les plus proches des cténophores. Cette liste comprend les Porifera, les Hydrozoa, les Scyphozoa, les Anthozoa, les Platyhelminthes (Polycladida), et les Bilateria. Parmi ces hypothèses, deux ont été considérées comme les mieux argumentées :

Parenté avec les Cnidaires La première de ces hypothèses place les Ctenophora groupe-frère des Cnidaria au sein du groupe des Coelenterata, et a été soutenue le plus récemment par Hyman (1940) et Komai (1963). Ce regroupement est soutenu par le partage entre ces deux groupes d'une symétrie tétraradiaire, d'un système gastrovasculaire ramifié, d'un système nerveux diffus, d'une

absence de coelome, d'une absence de système excréteur, d'une absence d'appareil génital, d'une mésoglée gélatineuse, de la présence d'un statocyste et d'une absence générale de système d'organe. Trois autres arguments ont été utilisés pour soutenir l'existence des Coelenterata : la présence de cnidocytes chez le cténophore *Haeckelia rubra* (Chun 1880) ; la description de *Hydroctena salenskii*, un Cnidaria dont la morphologie ressemble beaucoup à un Ctenophora (Dawydoff 1903) ; la ressemblance entre la larve de *Lampea pancerina* et la larve planula des Cnidaria. Dans cette hypothèse, l'ancêtre des cténophores est vraisemblablement pélagique.

Parenté avec les Polycladida La seconde hypothèse est favorisée principalement par Lang (1884), Mortensen (1912) et Harbison (1985). Elle repose sur le partage de certains caractères entre les Platyctenida et les Polycladida. Les représentants de ces deux groupes possèdent une forme ovale aplatie avec une différenciation dorso-ventrale, deux tentacules dorsaux, une reptation sur toute la surface ventrale, une bouche centrale du côté ventral, des canaux gastriques aux extrémités closes, un système nerveux en réseau, des gonoductes, une musculature mésodermique et un clivage déterminé. De plus, certaines larves de Polycladida possèdent huit lobes ciliés qui ressemblent un peu aux rangées de peignes des cténophores. Dans cette hypothèse, l'ancêtre des cténophores est vraisemblablement benthique.

Des arguments morphologiques faibles Aucune de ces ressemblances morphologiques ne suffit, seule, à choisir pour l'une des deux hypothèses. Certains de ces caractères ont été démontrés comme irrecevables, et les ressemblances restantes portent sur des caractères dont l'homologie est extrêmement douteuse. Parfois, c'est la description de l'état du caractère qui est fausse.

1. La qualification de la symétrie des cténophores comme généralement "tétraradiaire" est très discutable, car de très nombreux organes vont par paires, comme l'appareil tentaculaire, les canaux paragastriques, les champs polaires et la structure interne de l'organe aboral. De plus, le pharynx est aplati dans un sens, l'infundibulum dans l'autre. Quatre des caractères supposés unir les Cnidaria aux Ctenophora reposent sur l'absence partagée de structures existantes chez les bilatériens (ç.-à-d. coelome, système excréteur, appareil génital et système d'organes), et sont donc des simplésiomorphies inutilisables pour soutenir des groupes monophylétiques (voir section 1.2.3). D'autres organismes marins très différents des cténophores sont gélatineux (comme les salpes) ou possèdent des statocystes (comme certains céphalopodes), ce qui constitue une preuve que ces caractères peuvent être issus d'évènements de convergence évolutive. Il a été démontré que les cnidocytes de *Haeckelia rubra* sont issus de la kleptocnidie de cette espèce qui se nourrit de Cnidaires (Carré et al. 1989). La ressemblance d'*Hydroctena salenskii* avec les cténophores n'est que superficielle (Hyman 1940), et la ressemblance entre la larve planula des Cnidaria et la larve de *Lampea pancerina* n'est pas plus convaincante qu'avec la stéréoblastula des Platyhelminthes (Harbison 1985). Enfin,

les systèmes gastrovasculaires et nerveux des cténophores sont simultanément utilisés comme arguments pour les deux hypothèses.

2. Les Platyhelminthes sont maintenant connus pour appartenir au groupe des Spiralia, au sein des Bilateria (Hausdorf et al. 2007; Struck and Fisse 2008; Struck et al. 2014). Une relation de parenté entre les Polycladida et les Ctenophora impliquerait donc que les cténophores soient des Bilateria, extrêmement dérivés, qui auraient perdus secondairement toute trace de bilatéralité. De plus, il est clair que le mode de vie benthique des Platyctenida a été secondairement acquis au sein des cténophores (voir l'argumentaire de Rankin 1956). L'hypothèse d'un apparentement entre les Polycladida et les Ctenophora est donc extrêmement improbable et a été rejetée par toutes les phylogénies moléculaires existantes. La ressemblance perçue entre les Polycladida et les Ctenophora doit être comprise comme une convergence évolutive, liée au mode de vie benthique de ces deux groupes.

Pour conclure, il n'existe actuellement aucun argument morphologique qui permettrait de déterminer de quels autres métazoaires les cténophores seraient les plus proches.

Parenté avec *Dendrogramma* Une étude très récente a décrit une nouvelle famille de métazoaires, les Dendrogrammatidae (Just et al. 2014), qui comprend les deux espèces *Dendrogramma enigmatica* JUST ET AL., 2014 et *Dendrogramma discoïdes* JUST ET AL., 2014. Cette découverte est exceptionnelle puisque ces deux espèces ne semblent appartenir ni aux Bilateria, ni à aucun des quatre embranchements actuels extérieurs aux Bilateria. Néanmoins les auteurs citent deux ressemblances entre les Dendrogrammatidae et les embranchements des cnidaires et cténophores : 1) présence de deux feuillettes (ectoderme et endoderme) encadrant une mésoglée ; 2) une bouche ouvrant sur un système gastrovasculaire ramifié. Ces ressemblances sont insuffisantes pour attribuer ces nouvelles espèces à l'un de ces deux embranchements, mais pourraient suggérer que les Dendrogrammatidae soient le groupe-frère de l'un d'eux. L'absence d'observation de structures ciliées chez *Dendrogramma* pourrait affaiblir l'hypothèse d'une relation de proximité phylogénétique avec les cténophores, mais *a contrario*, le mode de ramification du système gastrovasculaire de *Dendrogramma* fait beaucoup penser à celui des cténophores (Just et al. 2014). Néanmoins, compte tenu du faible nombre de caractères morphologiques pour l'instant observé chez ces nouveaux animaux, la résolution de leur position phylogénétique au sein des métazoaires devra nécessairement attendre la récolte de nouveaux spécimens, et probablement leur séquençage.

Analyses cladistiques de caractères morphologiques Les analyses formelles de matrices de caractères morphologiques à l'échelle des métazoaires dans un contexte cladistique suggèrent une proche parenté entre les cténophores et les bilatériens (Schram 1991; Nielsen et al. 1996; Zrzavy et al. 1998; Sorensen et al. 2000; Peterson and Eernisse 2001; Brusca and Brusca 2003; Glenner et al. 2004), mais cette parenté repose sur le codage de quelques caractères seulement, comme la morphologie des spermatozoïdes, le mésoderme ou les orifices anaux du tube digestif,

dont l'hypothèse d'homologie primaire pourraient être débattue. Ces nombreuses études sont malheureusement limitées par le faible nombre de caractères morphologiques comparables avec les autres embranchements, et la difficulté de chacune de ces comparaisons.

Apports de la phylogénie moléculaire et de la phylogénomique

L'apparition de séquences moléculaires représente, à la fin des années 90, l'espoir de remplacer le manque de caractères morphologiques informatifs par l'abondance de caractères moléculaires qui permettraient de reconstituer les relations de parenté à l'échelle des métazoaires. De nombreuses analyses phylogénétiques basées sur le marqueur ADNr 18S ont supporté un arbre des métazoaires en nette contradiction avec les analyses basées sur des matrices morphologiques : les éponges y étaient paraphylétiques par rapport aux autres métazoaires (Zrzavy et al. 1998; Kim et al. 1999; Borchiellini et al. 2001; Medina et al. 2001; Peterson and Eernisse 2001; Glenner et al. 2004). Seules exceptions, une analyse du marqueur ADNr 28S a positionné les cténophores en groupe-frère des cnidaires, et l'analyse combinée des marqueurs ARNr 18S et 28S a retrouvé la monophylie des éponges (Medina et al. 2001, voir aussi Manuel et al. 2003). Enfin, une analyse très récente du marqueur ADNr 18S à l'aide d'un modèle d'évolution tenant compte de la structure secondaire de cette molécule s'est révélée incapable de résoudre la position des cténophores (Mallatt et al. 2012). Ce résultat est représentatif de l'impossibilité désormais acceptée de reconstituer avec certitude les relations de parenté à l'échelle des métazoaires à l'aide d'un ou de quelques marqueurs moléculaires seulement, et particulièrement, la "base" de l'arbre des métazoaires avec les gènes ribosomiques.

L'utilisation de jeux de données phylogénomiques, constitués de la concaténation de plusieurs dizaines à plusieurs centaines de gènes alignés, est depuis devenu la méthode dans laquelle le plus d'espoir est placé pour résoudre ces relations, et notamment la position phylogénétique des cténophores. Les résultats de ces récentes analyses (et ré-analyses) phylogénomiques peuvent être classés en trois catégories, correspondant à trois hypothèses différentes pour la position des cténophores au sein des métazoaires (voir Figure 1.23)⁶ :

1. Quatre études soutiennent que les cténophores forment le groupe-frère de tous les autres métazoaires (Dunn et al. 2008; Hejnol et al. 2009; Ryan et al. 2013; Moroz et al. 2014), en plus de l'analyse d'une sous-matrice d'une cinquième étude (Nosenko et al. 2013).
2. Une ré-analyse du jeu de données utilisé par Dunn et al. (2008) suggère, malgré un soutien statistique faible des noeuds profonds, que les éponges sont le groupe-frère de tous les autres métazoaires et que les cténophores sont le second embranchement à émerger (après les éponges, donc) au sein des métazoaires (Pick et al. 2010).

6. Une étude publiée par Schierwater et al. (2009) n'est pas listée ici car elle repose sur des données hautement contaminées, et je considère que certains de ses choix méthodologiques sont injustifiés, que l'arbre phylogénétique qui en résulte est à la fois improbable et très peu soutenu, et que son interprétation par les auteurs est, au mieux, difficile à suivre. Elle suggère notamment que les bilatériens sont le groupe-frère des autres métazoaires. Dans une contre-analyse, Philippe et al. (2011) ont démontré la faible qualité de l'étude initiale.

3. Trois études soutiennent que les cténophores sont le groupe-frère des cnidaires, formant ensemble le clade des Coelenterata (Philippe et al. 2009, 2011 ; une sous-matrice de Nosenko et al. 2013).

Une étude particulière comprend une ré-analyse volontairement “sabotée” du jeu de données utilisé par Philippe et al. (2009), dans laquelle le modèle d'évolution utilisé est moins approprié que celui d'origine, et donne pour résultat les cténophores groupe-frère des autres métazoaires (Philippe et al. 2011). Cette étude, celle de Pick et al. (2010) et celle de Nosenko et al. (2013) montrent toutes les trois que l'obtention des cténophores en groupe-frère des autres métazoaires pourrait être expliquée par un artéfact de reconstruction bien connu en phylogénétique (il s'agit de l'artéfact d'attraction des longues branches, voir section 1.2.3). La position phylogénétique des cténophores est donc toujours débattue et la communauté scientifique semble caricaturalement coupée en deux, avec d'un côté ceux qui pensent que les cténophores sont le groupe-frère des autres métazoaires, et d'un autre ceux qui pensent que les cténophores peuvent être partout dans l'arbre, excepté en tant que groupe-frère des autres métazoaires.

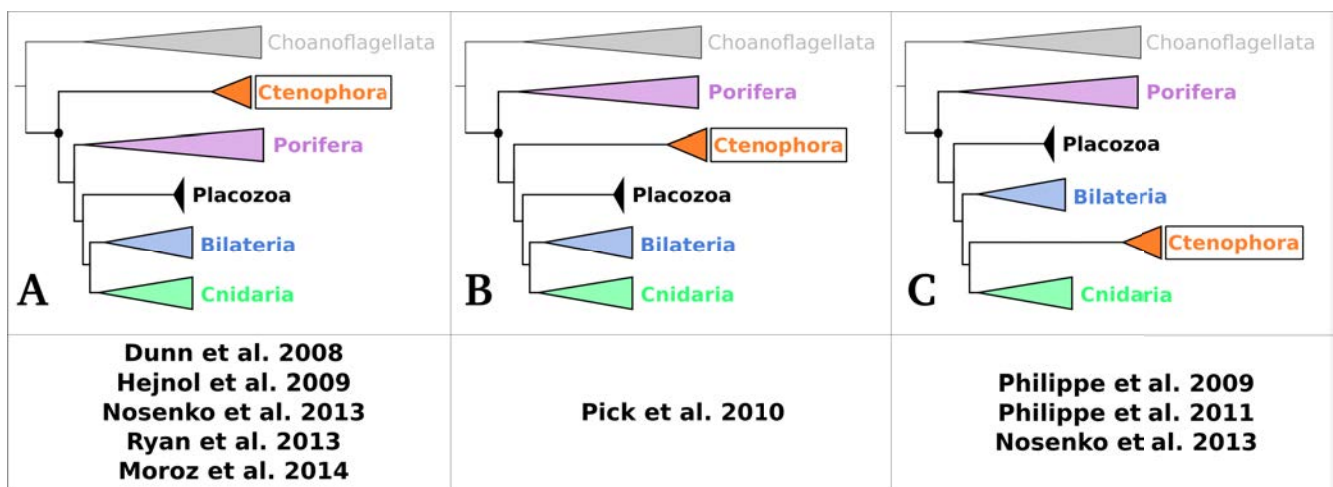


FIGURE 1.23 – Récapitulatif des trois hypothèses évolutives récemment supportées par des analyses phylogénomiques. A) Les cténophores représentent le groupe-frère des autres métazoaires. B) Les éponges représentent le groupe-frère des autres métazoaires, et les cténophores divergent en second. C) Les cténophores sont regroupés avec les cnidaires dans le clade des Coelenterata. La position des placozoaires peut être instable dans certaines études, leurs positions phylogénétiques schématisées ici sont donc à considérer avec précaution.

1.2.3 Difficultés méthodologiques liées aux particularités des cténophores

Afin de comprendre pourquoi il est difficile de résoudre les relations de parenté entre les cténophores et les autres métazoaires, il faut comprendre comment on infère ces relations phylogénétiques. Je vais tenter d'introduire ici un certain nombre de notions nécessaires à la compréhension de ce type d'analyses. Ensuite, les difficultés particulières posées par les cténophores

seront décrites. Enfin, des pistes méthodologiques seront proposées dans le but de résoudre certaines de ces difficultés.

Quelques principes de l'inférence phylogénétique

Les bases de la systématique phylogénétique ont été posées par Willi Hennig (Hennig 1966) et s'opposent à la "systématique évolutionniste", prévalente à l'époque, à différents niveaux. Dans le cadre de la systématique évolutionniste, on classifiait des espèces selon la ressemblance que pouvait présenter un certain nombre choisis de leur caractères morphologiques (ou comportementaux) que l'on décrivait comme homologues (ç.-à-d. caractères comparables car issus d'un ancêtre commun).

Hennig va objectiver cette méthode et tester les caractères morphologiques utilisés. L'idée est donc d'utiliser un grand nombre de caractères morphologiques que l'on suppose homologues, de calculer tous les arbres possibles et de garder celui qui fournit l'explication évolutive la plus parcimonieuse de tous ces caractères morphologiques (c.-à-d. qui demande le moins de transitions d'états de caractères). Le résultat est un arbre de relations phylogénétiques qui permet de dire *a posteriori* si chacun des états de caractères utilisés était bien homologue (c.-à-d. issu d'un unique événement évolutif), ou au contraire homoplasique (c.-à-d. issu d'événements de convergence évolutive ou de réversion). Les groupes représentés dans l'arbre sont dit monophylétiques et sont désormais définis par le partage d'un ensemble d'événements évolutifs uniques : ce sont les synapomorphies (c.-à-d. les états de caractères dérivés et partagés).

Dans la pratique, on construit une **matrice de caractères**. Pour chaque taxon (c.-à-d. les lignes de la matrice), on va décrire l'état observé dans la nature d'un certain nombre de **caractères** (c.-à-d. les colonnes de la matrice). Des espèces phylogénétiquement proches entre elles auront des états de caractères plus fréquemment identiques que des espèces phylogénétiquement plus éloignées. L'analyse phylogénétique consiste en l'exploration de tous les arbres possibles représentant les relations phylogénétiques entre les espèces de la matrice, et le choix de celui qui sera le plus parcimonieux.

Si les caractères ont été intelligemment choisis et leurs états correctement décrits par le morphologiste, une partie d'entre eux seront homologues et formeront des groupes monophylétiques. Ces caractères apportent du **signal phylogénétique** à l'analyse. D'autres caractères, dit homoplasiques, se révéleront "trompeurs" : il s'agira de structures anatomiques semblables, mais qui ne sont pas issues des mêmes événements historiques. Les yeux des mammifères et les yeux des céphalopodes sont un exemple de convergence évolutive, tandis que l'absence de pattes chez les Serpentes, appartenant pourtant au groupe des Tétrapodes, est un exemple de réversion. Ces caractères brouillent le signal phylogénétique, et apportent du **bruit** à l'analyse.

Phylogénie moléculaire

La phylogénie moléculaire est une adaptation aux données de séquences de la méthode décrite précédemment. Il s'agit de construire un **alignement de séquences** (ç.-à-d. une matrice de caractères). La séquence de chaque taxon (i.e lignes de la matrice) est un ensemble de positions qui, alignées verticalement, sont les caractères (ç.-à-d. les colonnes de la matrice). La nature de la base nucléotidique (ou de l'acide aminé) observée représente l'état du caractère.

Certaines positions de l'alignement porteront du signal phylogénétique, car plusieurs taxons posséderont une même base nucléotidique, différente chez d'autres taxons, et issue d'un évènement évolutif unique. D'autres positions en revanche apporteront du bruit à l'analyse, car plusieurs séquences posséderont la même base nucléotidique, mais issue de mutations historiquement différentes.

L'importance des taux de substitutions Les positions nucléotidiques du génome subissent des mutations plus ou moins fréquemment. Selon sa nature et là où elle a eu lieu, la mutation aura une probabilité de fixation au sein de la population variable. Les probabilités de mutation et de fixation à une position nucléotidique représentent son taux de substitution. Certaines positions possèdent des taux de substitutions plus élevés que d'autres. Puisqu'il n'y a que quatre nucléotides possibles dans l'ADN, plus une position subit des substitutions, plus il y a de chance qu'elle présente un nucléotide qu'elle avait historiquement déjà possédé. Cela rend indétectable les substitutions intermédiaires qui ont eu lieu durant l'évolution : on parle de **saturation** de la position. Pour la même raison, la probabilité pour deux taxons (même phylogénétiquement éloignés) de posséder par hasard le même nucléotide est plus élevée au niveau d'une position à fort taux de substitution qu'au niveau d'une position à faible taux de substitution.

De manière générale, certains organismes (ainsi que certaines portions d'ADN) possèdent des taux de substitution plus élevés que les autres. Cela pose deux problèmes lors de l'analyse de l'alignement : 1) Les bases nucléotidiques de ces séquences étant plus saturées, elle comportent un signal phylogénétique plus faible ; et 2) Si plusieurs de ces espèces sont présentes dans le même alignement, leurs bases nucléotidiques qui sont en moyenne plus saturées que les autres espèces, ont statistiquement plus de chance d'être identiques. Cela crée un signal non-historique qui aura tendance à regrouper ensemble dans l'arbre phylogénétique les espèces qui possèdent de forts taux de substitution. Cet artéfact de reconstruction est appelé **Attraction des Longues Branches** (LBA en anglais).

L'existence du phénomène de saturation rend très risqué de comptabiliser de manière identique, comme c'est le cas en parcimonie, toutes les substitutions visibles dans l'alignement : en effet, certaines positions auront historiquement subi beaucoup de substitutions, mais ces dernières seront indétectables à cause de la saturation de la position. C'est entre autre pour corriger ce défaut que les modèles d'évolution des séquences ont été développés.

Méthodes probabilistes et modèles d'évolution L'utilisation de données de séquences à la place de données morphologiques permet l'utilisation de modèles d'évolution lors du calcul de l'arbre phylogénétique, à la place de l'utilisation du critère de parcimonie. Comme les mutations appartiennent à des catégories facilement caractérisables (comme les transitions et les transversions), il est possible de les modéliser. Dans ce modèle, on peut alors faire intervenir des paramètres afin de **représenter les probabilités inégales des différentes substitutions** possibles. De nombreux modèles d'évolution de séquences existent et diffèrent par la complexité des mécanismes dont ils peuvent rendre compte (et du degré de liberté utilisé pour ce faire). L'utilisation de certains paramètres permet de **limiter en partie l'effet de la saturation** en explicitant que certaines positions, que certains gènes ou encore que certaines espèces vont plus probablement subir des mutations que les autres. Cela a également pour conséquence de réduire l'impact de l'artéfact d'attraction des longues branches. D'autres paramètres vont pouvoir expliciter que certains types de mutations sont plus fréquents que d'autres, et ainsi rendre plus réaliste le calcul de la vraisemblance de l'arbre. De manière générale, l'utilisation d'un modèle d'évolution représentant correctement les processus naturels de mutation et de substitutions est indispensable et permet une très forte **amélioration de la reconstruction phylogénétique** (pour une démonstration, voir par exemple Philippe et al. 2005).

Le modèle d'évolution "CAT" Contrairement à la majorité des autres modèles d'évolution, le modèle CAT (Lartillot and Philippe 2004) est un modèle dit "site-hétérogène". Cela lui donne l'avantage de pouvoir modéliser le fait que, dans la nature, la fréquence observée des différents acides aminés n'est pas la même à chaque position (exemple d'un alignement de protéines). La plupart du temps, on ne retrouve que quelques acide-aminés différents à une position donnée. Le modèle CAT peut donc attribuer différentes matrices de fréquences stationnaires (appelées *profils*) à différentes positions. Le résultat de cette modélisation de l'hétérogénéité de la composition en acides aminés des positions est que le modèle CAT détecte et reconstruit plus facilement les substitutions multiples que les modèles qui ne sont pas sites-hétérogènes. En effet, puisqu'à chaque position, un profil modélise le fait que seulement quelques acides-aminés y seront beaucoup plus probables que les autres, il sera donc également plus probable d'obtenir des remplacements multiples. Cette détection supérieure de la quantité de remplacements (ou de substitutions) multiples permet au modèle CAT de mieux gérer la saturation d'un jeu de données, et ainsi de réduire l'impact de l'attraction des longues branches. Ce modèle qui a été créé par Lartillot and Philippe (2004) est implémenté dans le logiciel Phylobayes.

Source potentielle de confusion, un autre modèle, implémenté cette fois dans le logiciel RAxML s'appelle également CAT, mais ne correspond absolument pas au modèle CAT que je viens de décrire.⁷

7. Le modèle CAT de RAxML correspond simplement à une méthode qui remplace la loi Γ par une approximation, dans le but d'accélérer le calcul de la meilleure topologie.

La phylogénomique Le développement de la phylogénomique est en lien avec l’augmentation exponentielle de la quantité de données moléculaires, elle-même due aux nouvelles technologies de séquençage (NGS, pour Next-Generation Sequencing en anglais). Il s’agit d’une application particulière des méthodes de la phylogénie moléculaire à des jeux de données beaucoup plus massifs, issus de génomes, transcriptomes ou protéomes pour résoudre les relations phylogénétiques entre taxons (voir Philippe and Telford 2006; Telford and Copley 2011; Telford 2013). L’idée en est plutôt simple : l’utilisation d’un ou de quelques alignements ne suffit parfois pas à obtenir suffisamment de signal phylogénétique afin de résoudre les relations de parenté entre taxons. La phylogénomique consiste donc à utiliser des dizaines ou centaines d’alignements en même temps. L’idée de cette démarche repose sur le fait que le signal phylogénétique (donc historique) est le même pour tous les alignements, et doit donc être additif. Le bruit dans les alignements est quant à lui stochastique, et sera donc différent à chaque fois et ne s’additionnera pas : donc si l’on analyse beaucoup d’alignements en même temps, le signal phylogénétique devrait pouvoir couvrir le bruit.

Les difficultés de la phylogénomique Malheureusement, une telle taille de matrice de caractères pose également un certain nombre de problèmes lors de l’analyse phylogénétique. Certains biais de reconstruction sont apparus beaucoup plus gênants que pour des alignements plus simples (pour une revue sur le sujet, voir Philippe et al. 2011).

- Les erreurs systématiques comme les artéfacts d’attraction des longues branches ou les biais de composition en nucléotides ou en acides aminés sont additifs et sont ainsi renforcés en phylogénomique (plus la matrice est grande, plus les mêmes biais sont répétés, renforçant le signal non-phylogénétique qu’ils créent).
- L’impact de la quantité et de la distribution des données manquantes dans les alignements est problématique pour la reconstruction, surtout lorsque les données sont manquantes dans un groupe en particulier (Roure et al. 2013).
- Les données utilisées en phylogénomique étant en plus grande quantité qu’en phylogénie “classique”, les processus évolutifs qui les régissent sont plus hétérogènes, et l’on peut s’attendre à ce que les modèles d’évolution ne les modélisent pas correctement (ces modèles ont été historiquement développés pour des alignements beaucoup plus petits). La violation des postulats du modèle d’évolution utilisé est particulièrement problématique pour la reconstruction d’une topologie correcte.

Enfin, et d’un point de vue pratique, la taille des jeux de données transcriptomiques ou génomiques utilisés rend impossible leur traitement “à la main”, et impose l’automatisation de toutes les étapes de la construction d’un jeu de données. Certaines étapes, comme la recherche d’un gène orthologue dans un transcriptome, peuvent être très complexes, et leur automatisation est alors un véritable défi.

Le cas problématique des cténophores

Si la position phylogénétique des cténophores n'est toujours pas résolue, c'est parce que cette question est une collection de tous les problèmes qu'il est possible de rencontrer en phylogénie. Les cténophores sont extrêmement divergents des autres animaux, d'un point de vue morphologique et moléculaire, et les cinq grandes lignées des Metazoa (Porifera, Placozoa, Bilateria, Cnidaria et Ctenophora) sont apparues relativement rapidement, entre il y a 600 et 780 millions d'années (Ma). Ce simple constat suffit, comme on va le voir, à faire de la position phylogénétique des cténophores un véritable défi méthodologique.

Un faible signal phylogénétique Les relations phylogénétiques unissant les métazoaires sont particulièrement anciennes. En effet :

- Un cténophore juvénile datant de 540 Ma a été décrit (Chen et al. 2007).
- Le fossile d'éponge le plus ancien consiste en des spicules datées à 535 Ma. Une révision stricte de fossiles plus anciens a invalidé leur attribution au phylum des porifères (Antcliffe et al. 2014).
- Les cnidaires sont connus dans le registre fossile grâce à des médusozoaires datés d'il y a 500 Ma (Cartwright et al. 2007) et certains auteurs suggèrent l'existence de cnidaires il y a 570 Ma (Xiao et al. 2000).
- Les bilatériens existent depuis au moins 530 Ma (âge du fossile *Anomalocaris*, Collins 1996), probablement même 558 Ma (ré-interprétation du fossile *Kimberella quadrata*, Ivantsov 2009) et des traces fossiles interprétées comme bilatériennes ont été datées à 560 Ma (Cartwright and Collins 2007).

L'apparition et la diversification de ces grands groupes de métazoaires a donc forcément eu lieu au moins avant le début du Cambrien (541 Ma), c'est-à-dire pendant l'Ediacarien. Cette ancienneté implique que les séquences utilisées pour reconstituer les relations de parenté entre les métazoaires auront eu énormément de temps depuis cette époque pour accumuler des substitutions. Les positions de ces séquences risquent donc d'être fortement saturées, réduisant d'autant la quantité de signal phylogénétique.

En plus d'être anciennes, les différentes lignées de métazoaires sont probablement apparues rapidement. Si l'on se base sur la présence stricte de fossiles identifiés comme métazoaires, aucun n'a été découvert avant l'Ediacarien, ce qui impliquerait une apparition et une diversification des métazoaires pendant cette période (entre 635 et 541 Ma). D'un autre côté, des datations moléculaires récentes ont estimé différentes dates d'apparition des métazoaires, variant entre 780 Ma (Erwin et al. 2011), 766 Ma (Peterson et al. 2008) et 664 Ma (Peterson et al. 2007). Ces dates très anciennes sont à considérer avec beaucoup de précaution car, 1) Il a été montré que le nombre de points de calibration d'âge maximum a un impact extrêmement fort sur les datations (Cartwright and Collins 2007; Duchêne et al. 2014) ; 2) certains points de calibration sont discutés,

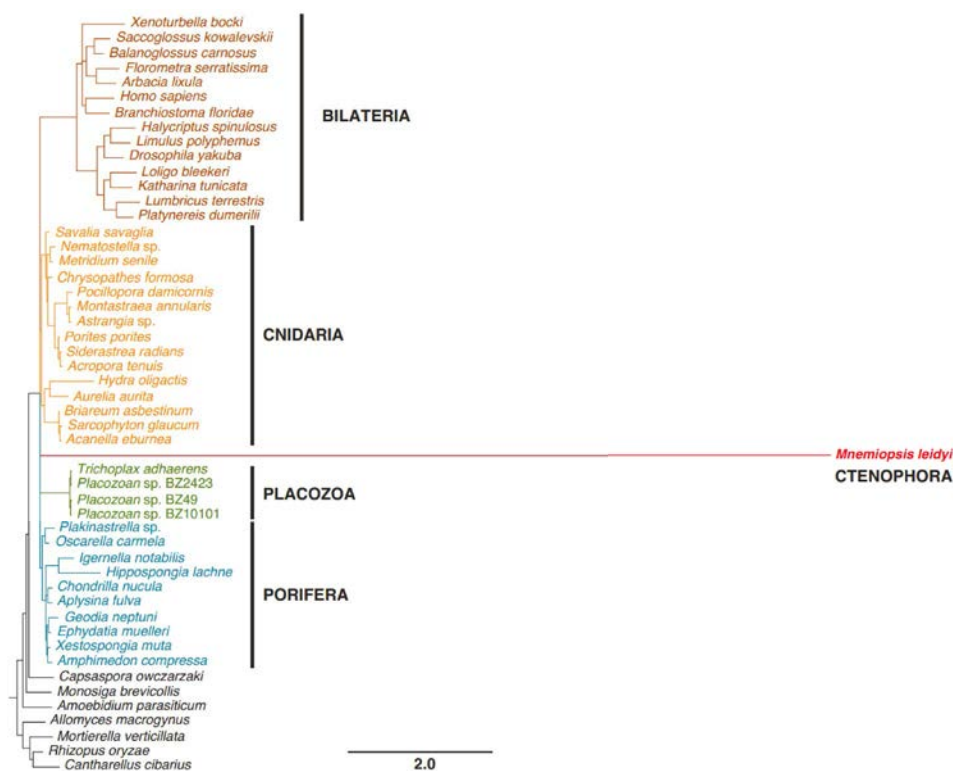


FIGURE 1.24 – Relations phylogénétiques entre métazoaires reconstruites avec des séquences mitochondriales, d’après Pett et al. 2011. La concaténation des cinq gènes mitochondriaux les plus conservés chez *Mnemiopsis*, représentant 1402 acides aminés au total, a été analysée avec Phylobayes, en utilisant le modèle CAT+Γ. La branche du cténophore *Mnemiopsis* est extrêmement longue, indiquant un très fort taux de substitution par rapport aux autres métazoaires.

car ils dépendent des interprétations taxonomiques de différents fossiles, elles-mêmes très débattues (notamment le point de calibration d’âge maximal des porifères); 3) ces datations moléculaires sont basées sur des relations de parenté encore inconnues et utilisent au maximum sept gènes pour estimer les vitesses de l’horloge moléculaire, ce qui est potentiellement trop peu (Erwin et al. 2011). Pour conclure, le registre fossile ainsi que les datations moléculaires actuelles indiquent que les métazoaires seraient apparus et se seraient diversifiés dans une période de temps comprise entre 100 et 200 Ma.

Pour reconstituer les relations phylogénétiques entre les cinq lignées de métazoaires, il faut donc retrouver, dans les données moléculaires d’aujourd’hui, les rares positions qui ont subi des substitutions pendant cette relativement courte période, et qui n’auraient pas été saturées pendant les plus de 541 millions d’années qui ont suivi : la “fenêtre de tir” est donc réduite (voir aussi Rokas et al. 2003).

Une forte divergence Les cténophores ne ressemblent à aucun des autres animaux. Tous leurs organes sont des innovations sans équivalent facilement détectable chez les autres métazoaires (voir section 1.2.2), et souvent leurs séquences moléculaires sont également difficilement comparable à

celles des autres métazoaires. Le taux de substitutions de ces séquences est extrêmement élevé, et de manière générale, les modifications du contenu génétique des cténophores sont bien plus nombreuses que pour beaucoup d'autres métazoaires. Le génome nucléaire de *Mnemiopsis gardeni* appartient aux 7% des plus petits génomes chez les métazoaires ; son génome mitochondrial possède 25 gènes de moins que les autres métazoaires, et est très fortement dérivé (Figure 1.24, Pett et al. 2011; Ryan et al. 2013).

Que le cténophore *Mnemiopsis gardeni* soit très divergent des autres métazoaires ne constitue pas nécessairement un problème pour la reconstruction des liens de parenté entre le phylum Ctenophora et les autres phylums. Le problème est que les cténophores possèdent une faible diversité génétique, probablement car issus d'un ancêtre commun récent (voir section 1.2.1). Cela signifie que *l'ensemble* des cténophores est très divergent des autres métazoaires. La conséquence est donc une quasi-impossibilité d'observer des animaux qui possèderaient à la fois des états de caractères de cténophores et des états de caractères d'autres métazoaires. Cela semblait déjà évident d'un point de vue morphologique, mais c'est également vrai d'un point de vue moléculaire.

Ce problème est d'ailleurs parfaitement visible sur les arbres de relations phylogénétiques reconstruits à partir d'alignements de protéines : la branche qui regroupe les cténophores est toujours très longue (Figures 1.24, 1.25). Comme la diversité des cténophores ne suffit pas à réduire la longueur de cette branche, à la "casser", il devient évident que les cténophores vont être extrêmement sensibles aux effets de l'artéfact d'attraction des longues branches.

Difficultés méthodologiques Aux difficultés biologiques issues de l'histoire naturelle des cténophores s'ajoutent un certain nombre de difficultés méthodologiques, qui compliquent indirectement la résolution du problème de la position phylogénétique des cténophores. Lors de la création d'alignements de séquences, des logiciels (par exemple, BLAST, pour Basic Local Alignment Search Tool) permettent de chercher et de sélectionner des gènes et des protéines dans des bases de données. Les outils de manipulation de séquences moléculaires comme le BLAST sont tous basés sur des mesures de similarité. Cela signifie que lorsque l'on va comparer différentes séquences, on va seulement calculer une ressemblance entre les séquences. Une telle mesure de distance ne peut pas prendre en compte la saturation des positions (ni les probabilités différentielles entre des types de substitutions, d'ailleurs). Ainsi, les BLASTs de séquences de cténophores, qui évoluent rapidement, échouent beaucoup plus souvent à détecter des séquences similaires que pour d'autres espèces, car les séquences de cténophores ne ressemblent pas assez à celles des autres métazoaires.

Puisque ces outils de comparaison ne font que mesurer la distance entre différentes séquences, ils rendent impossible la distinction entre différentes causes d'éloignement entre séquences. Une séquence peut être homologue à d'autres mais néanmoins distante car 1) il s'agit d'une séquence issue d'une **contamination** par un organisme très différent, 2) il s'agit d'une séquence appartenant à une autre sous-famille de gènes (c.-à-d. un **paralogue**) ou bien 3) il s'agit d'une séquence appartenant à la même sous-famille de gènes (c.-à-d. un **orthologue**), mais qui a subi beaucoup

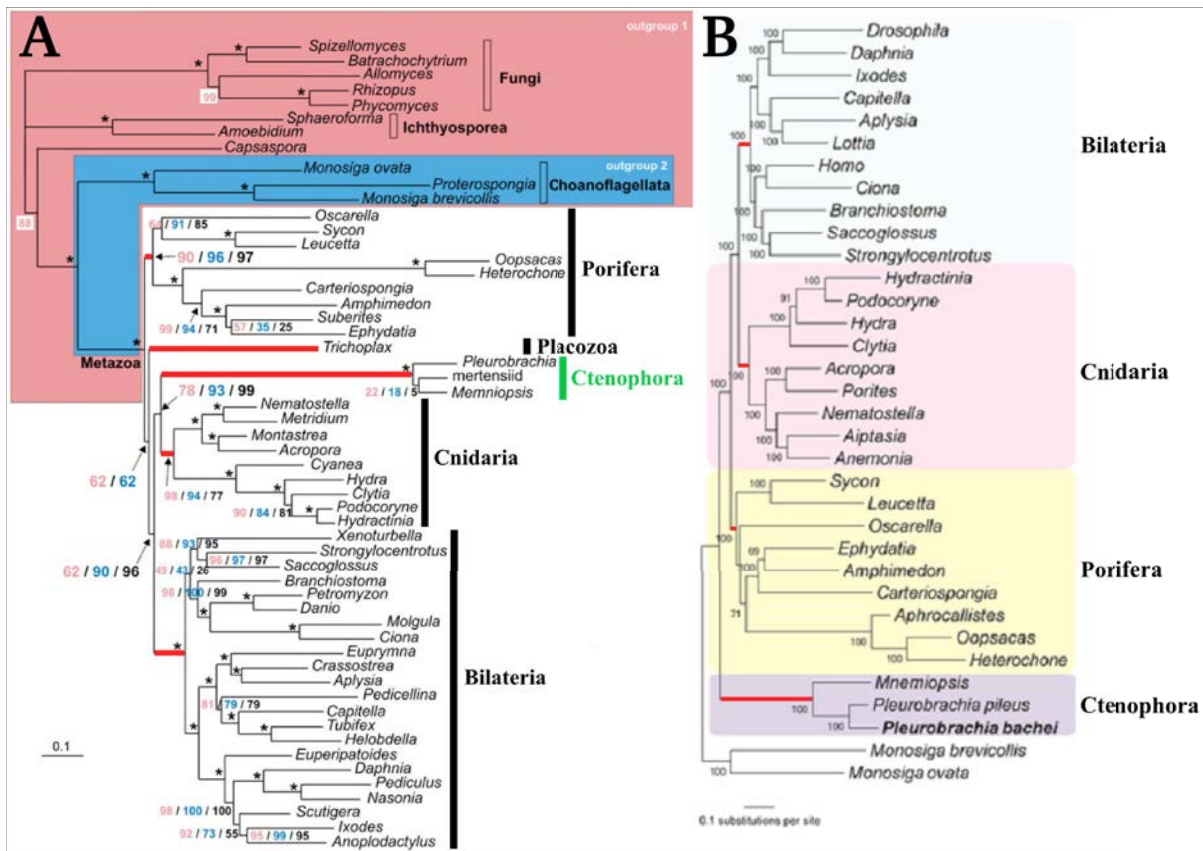


FIGURE 1.25 – Analyses phylogénétiques entre métazoaires basées sur des séquences protéiques. En rouge sont représentées les branches de chacune des cinq lignées principales de métazoaires. Celle des cténophores (et également celle des placozoaires) est extrêmement longue. A) Concaténation de 128 protéines (30257 acides aminés au total), analysée sous Phylobayes avec le modèle CAT+ Γ , d'après Philippe et al. (2009). Trois analyses ont été effectuées avec trois échantillonnages taxonomiques de groupes externes différents. Les scores de bootstrap avec 100 répliquats de ces analyses sont affichés pour chaque noeud, la couleur correspondant au groupe externe utilisé. B) Concaténation de 586 protéines (170871 acides aminés au total), analysée sous RAxML avec le modèle WAG+CAT+G, d'après Moroz et al. (2014). Ce modèle ne correspond pas au modèle CAT de Lartillot and Philippe (2004), voir remarque page 42. Les scores de bootstrap sont représentés à chaque noeud.

de substitutions. Lors de la construction d'alignements de séquences pour résoudre des relations phylogénétiques, il est crucial d'utiliser *uniquement* des séquences orthologues. La détection de séquences orthologues chez les cténophores correspond au cas n°3 ; il sera donc souvent très difficile de les différencier des séquences contaminées et des séquences paralogues.

Améliorations méthodologiques souhaitables

Augmentation de la quantité de signal phylogénétique et de l'échantillonnage taxonomique La résolution de la position phylogénétique des cténophores implique de construire une matrice de caractères moléculaires particulièrement grande. En effet, le signal phylogénétique correspondant à la radiation des métazoaires est difficile à détecter car les cladogénèses sont anciennes (voir texte page 44). Il faut donc une matrice riche en caractères afin de réunir le plus de signal phylogénétique possible. Une quantité suffisante de signal devrait permettre de couvrir

le “bruit” de l’analyse. Il faut également une diversité taxonomique suffisante afin de renseigner au mieux le modèle d’évolution sur la nature et la quantité des substitutions présentes dans la matrice. Cette diversité permet de mieux renseigner les états de caractères au niveau des noeuds (ç.-à-d. des ancêtres) de l’arbre. La plupart des analyses phylogénomiques récentes ont ainsi eu pour objectif de maximiser la taille de la matrice, tant en terme de nombre de gènes utilisés que de nombre d’espèces incluses.⁸ Puisque la quantité de données de séquençage disponible pour des organismes non-bilateriens est plus faible que pour les bilateriens, il faut trouver un compromis lors de la construction d’une supermatrice moléculaire à l’échelle des métazoaires : vaut-il mieux beaucoup de gènes séquencés pour peu d’espèces, ou bien peu de gènes séquencés pour beaucoup d’espèces ? Une grande quantité à la fois de gènes et de taxons est bien entendu la solution la plus satisfaisante.

Souvent, l’augmentation de la diversité taxonomique va de pair avec l’augmentation du nombre de données manquantes dans la matrice.⁹ L’impact de ces données manquantes peut être fort et fausser la reconstruction des relations de parenté (Laurin-Lemay et al. 2012; Roure et al. 2013). La quantité et la distribution des données manquantes doivent donc être surveillées attentivement, d’autant que leur impact a été relativement peu étudiée jusqu’à maintenant.

Réduire les biais systématiques Les modèles d’évolution utilisés pour reconstruire des phylogénies sont forcément limités dans la quantité des processus qu’ils peuvent modéliser. La réalisation d’une matrice de grande taille implique la présence de positions ayant été soumises à des processus évolutifs hétérogènes. Les biais systématiques correspondent à l’impuissance d’un modèle à représenter correctement les processus évolutifs qui ont abouti aux séquences de l’alignement.

Le taux de substitution peut varier selon les positions, qui peuvent être plus ou moins contraintes, donc plus ou moins conservées. Cette variation est le plus souvent modélisée par le paramètre α selon une loi Gamma (Γ). Ce taux peut également varier le long des branches de l’arbre, c’est-à-dire selon le temps, phénomène appelé hétérotachie (Lopez et al. 2002). Une méthode a été proposée pour prendre en compte ce processus (Pagel and Meade 2008). De forts taux de substitution aboutissent à la saturation de certaines positions, et si ces taux sont particulièrement élevés chez certaines espèces, ils entraînent l’apparition de LBA. Afin de réduire l’impact de cet artéfact, quatre approches (compatibles entre elles) peuvent être utilisées :

8. Seule exception à cette tendance, une analyses récente a largement sacrifié la diversité taxonomique utilisée et a pour résultat une topologie particulièrement aberrante (voir Ryan et al. 2013, Figure 4). Néanmoins, la réduction de l’échantillonnage ne peut être tenue pour seule cause de cet échec, car cette analyse repose non pas sur une matrice de séquences alignées, mais sur une matrice de présence/absence de gènes. Puisque seuls deux états de caractères sont possibles, la saturation de ces données est très élevée. De plus, puisque l’absence d’un gène est considéré comme état, cela biaise fortement l’analyse car certaines espèces sont mal séquencées, ou bien ont subi beaucoup de pertes de gènes, tandis que d’autres espèces sont suffisamment divergentes (comme les cténophores) pour que les blasts ne détectent pas la séquence orthologue. Dans cette analyse, les taxons divergents ou pour lesquels les données de séquençage sont lacunaires sont regroupés à la racine de l’arbre, partageant beaucoup d’absences de caractères.

9. Cette remarque est de moins en moins vrai, à mesure que des données NGS (notamment séquencées avec la technologie Illumina) sont produites et remplissent progressivement les bases de données publiques.

1. l'augmentation de l'échantillonnage taxonomique permet de mieux détecter les substitutions multiples présentes à une position donnée, ce qui aura pour effet de "casser" les longues branches.
2. construire un arbre grâce à une supermatrice, et pour l'alignement de chacun des gènes utilisés, confronter le nombre de substitutions observables dans l'alignement au nombre de substitution reconstruit selon l'arbre, ce qui permettra de supprimer les gènes qui présentent les plus fort taux de saturation.
3. utiliser un modèle d'évolution qui détecte plus facilement les substitutions multiples, comme le modèle CAT (voir 1.2.3 page 42).
4. Supprimer du jeux de données les espèces qui présentent les plus longues branches.

La composition en acides aminés des protéines peut varier. Certaines espèces possèdent une composition en acides aminés globalement différente des autres espèces (Foster et al. 1997), et on parle alors d'hétérogénéité de composition. Si plusieurs espèces (même phylogénétiquement éloignées) présentent le même biais de composition (ç.-à-d. qu'elles sont enrichies des mêmes acides aminés), alors elles posséderont par hasard plus d'états de caractères en commun et leurs branches seront attirées les unes par les autres dans l'arbre phylogénétique reconstruit (Foster and Hickey 1999). Des modèles ont été développés pour gérer la modification au cours du temps de la composition en acides aminés (ç.-à-d. le long des branches de l'arbre) (Tamura and Kumar 2002; Blanquart and Lartillot 2006; Dutheil and Boussau 2008; Jayaswal et al. 2011; Groussin et al. 2013), et l'un d'entre eux a notamment été intégré au modèle CAT pour donner le modèle CAT-BP¹⁰ (Blanquart and Lartillot 2008). Afin de réduire l'impact sur la topologie de ces biais de compositions, plusieurs approches (compatibles entre elles) sont possibles :

1. l'utilisation de tests postérieurs prédictifs tels qu'implémentés dans le logiciel Phylobayes permet la détection des espèces dont la composition en acide aminés est la plus biaisée et ces dernières peuvent ensuite être supprimées.
2. l'utilisation de ces mêmes tests permet de détecter les gènes dont la composition en acide aminés est la plus biaisée, et de les supprimer de la supermatrice.
3. l'utilisation de l'un des modèles d'évolution qui autorise la variation de la composition en acide aminés des séquences au cours du temps (voir plus haut).

Eviter les erreurs d'hypothèses d'homologies primaire Lors de l'ajout d'une séquence dans un alignement, ce qui correspond à une hypothèse d'homologie entre les séquences, deux types d'erreurs sont possibles. La première consiste à ajouter une séquence homologue, mais issue d'une duplication de gène plus ancienne : il s'agit alors d'une séquence paralogue, et non pas orthologue. Cette séquence appartient alors à la bonne famille multigénique, mais pas au bon

10. Celui-ci est particulièrement complexe et impose un temps de calcul extrêmement long.

groupe d'orthologie au sein de cette famille. Dans ce cas, plus la duplication du gène était ancienne, plus cette mauvaise séquence va empêcher la reconstruction des relations de parenté à un niveau profond de l'arbre. Le seul cas où l'insertion d'un paralogue n'est pas problématique est lorsque la duplication est suffisamment récente pour être espèce-spécifique : on parle alors d'in-paralogie. Dans tous les autres cas, la séquence paralogue doit être détectée et retirée de l'alignement. Cette détection peut uniquement se faire lors de l'inspection de l'arbre de gènes de toute la famille multigénique. Elle peut donc se faire "à la main", mais peut aussi être implémentée pour être automatique. Dans ce dernier cas, la méthode la plus rigoureuse est de ne conserver que le sous-ensemble de gènes monophylétique pour lequel chaque espèce n'est représentée que par une seule séquence.

Un script récemment publié permet de détecter ces sous-arbres (PhyloTreePruner, Kocot et al. 2013), mais souffre d'un défaut : cette recherche est faite sur un arbre raciné. Il est malheureusement difficile d'enraciner de manière certaine un arbre "simple-gène" à l'échelle des métazoaires, car le signal phylogénétique n'est pas assez puissant pour résoudre les relations les plus anciennes. Ce script n'est donc bien souvent capable que de retrouver un ensemble de séquences orthologues sous-optimal, en omettant une partie des séquences orthologues. Ce script élimine correctement les séquences paralogues, mais au prix de la suppression de nombreuses séquences orthologues.

La seconde erreur possible lors de l'ajout d'une séquence dans un alignement est l'ajout d'un orthologue qui appartient en réalité à une autre espèce : c'est une contamination. Il existe trois grandes sources de contaminations : 1) il est courant que de multiples organismes, surtout marins, vivent en association et il est parfois très difficile, voir impossible, de séparer ces organismes ; 2) de plus, des cross-contaminations lors des différentes manipulations en laboratoire, et notamment lors de l'extraction des ARNs de différentes espèces, sont possibles et doivent être surveillées ; 3) enfin, les nouveaux protocoles de biologie moléculaire et de bioinformatique utilisés par les prestataires de service de séquençage peuvent également aboutir à des cross-contaminations lors du traitement de différents échantillons. La transition de la technologie Sanger aux technologies NGS a démultiplié la profondeur du séquençage, démultipliant en même temps le séquençage des ARNs contaminants et donc leur présence dans les jeux de données. De plus, il a été montré que l'utilisation de *tags* identifiant différents échantillons présents dans une même ligne de séquençage Illumina a pour résultat une attribution de 0.3% des *reads* (en moyenne) à un mauvais échantillon (Kircher et al. 2011, 2012). Ces chiffres sont alarmants et devraient imposer à eux seuls l'existence d'un protocole de détection et de suppression systématique des cross-contaminations lors de l'utilisation de données NGS pour faire de la phylogénie.

L'utilisation de séquences paralogues ou contaminées en phylogénie pose un très lourd problème : la séquence possède un signal non-phylogénétique fort, possiblement additif et contradictoire avec l'histoire évolutive des taxons. Deux ré-analyses de jeux de données phylogénomiques récents ont montré en partie l'impact de tels erreurs sur la reconstruction des relations de parenté à l'échelle des métazoaires (Pick et al. 2010; Philippe et al. 2011).

1.3 Objectifs de la thèse et structure du manuscrit

Dans le cadre de cette thèse, j'ai tenté de contribuer à l'amélioration des connaissances sur la systématique des cténophores, en abordant leur évolution et leur diversité à trois échelles taxonomiques distinctes :

Détermination de la position phylogénétique des cténophores à l'échelle des métazoaires

Dans le chapitre 2, j'ai utilisé une approche phylogénomique afin de tenter de résoudre la position phylogénétique des cténophores au sein des métazoaires. L'objectif principal était de reconstituer ces relations de parenté grâce à deux jeux de données indépendants. Une attention toute particulière a été portée à la méthodologie, dans le but de maximiser la qualité et la quantité des jeux de données construits. Il a été mis au point une procédure semi-automatisée qui traite des données transcriptomiques afin de constituer une supermatrice moléculaire. Ce protocole est capable de détecter la contamination au sein des données et la présence de séquences paralogues afin de les supprimer. De nombreuses analyses phylogénétiques ont ensuite été menées afin de "disséquer" le contenu en information phylogénétique des supermatrices assemblées.

Reconstruction des relations phylogénétique au sein des cténophores

Dans le chapitre 3, j'ai tenté de résoudre les relations phylogénétiques unissant les ordres, les familles et les genres de cténophores. Pour cela, j'ai utilisé comme base le travail de Podar et al. (2001), amélioré à de multiples niveaux. Notamment, j'ai ajouté à leur alignement du gène ADNr 18S un second marqueur moléculaire (il s'agit des *internal transcribed spacers*, ITS). De plus, l'échantillonnage taxonomique des espèces représentées a été augmenté. Enfin, j'ai utilisé un modèle d'évolution capable de tenir compte de la structure secondaire du gène ADNr 18S, afin d'améliorer la qualité des analyses phylogénétiques. Dans ce chapitre seront aussi discutés les enseignements des analyses phylogénomiques du chapitre précédent en ce qui concerne les relations au sein des cténophores.

Identification de nouveaux caractères anatomiques entre deux espèces de cténophores proches

Dans le chapitre 4, je me suis intéressé à la taxonomie des cténophores à l'échelle des espèces. L'identification de ces dernières peut parfois être très difficile à cause d'un manque de caractères anatomiques diagnostiques. J'ai réalisé des marquages immunohistochimiques chez deux espèces proches du genre *Pleurobrachia*, afin de déterminer si leur "micro-anatomie" peut contenir de nouveaux caractères potentiellement intéressants pour la taxonomie des cténophores. J'ai ensuite étendu cette comparaison en réalisant les mêmes marquages chez une espèce appartenant à l'ordre

des Beroida, *Beroe gracilis*, afin de tester l'intérêt de tels caractères à un niveau taxonomique plus élevé.

Chapitre 2

La position phylogénétique des cténophores dans l’arbre des métazoaires abordée par une approche phylogénomique

2.1 Introduction

Les relations phylogénétiques à la base de l’arbre des métazoaires restent très débattues. En particulier, la position des cténophores est extrêmement instable, et l’utilisation récente de jeux de données phylogénomiques n’a pour l’instant pas pu la résoudre. Deux hypothèses contradictoires ont été suggérées suivant les études. L’hypothèse des Cténophores groupe-frère de tous les autres métazoaires a été soutenue par les jeux de données comportant le plus de gènes et construits par des méthodes automatisées (Dunn et al. 2008; Hejnol et al. 2009; Ryan et al. 2013; Moroz et al. 2014), tandis que l’hypothèse des “Coelenterata”, regroupant les cténophores avec les cnidaires, a été soutenue par des jeux de données plus petits, mais dont la qualité est probablement meilleure car garantie par des contrôles manuels multiples, assurant en particulier l’absence de séquences contaminantes ou paralogues (Philippe et al. 2009; Pick et al. 2010; Philippe et al. 2011; Nosenko et al. 2013).

Dans ce chapitre est présenté la version préliminaire d’un manuscrit dans lequel une procédure bioinformatique complète et inédite a été mise en place, afin d’assembler un jeu de données qui posséderait les avantages des deux types de données décrits plus haut : la présence d’une très grande quantité de gènes, et un rigoureux contrôle de leur qualité. En parallèle, l’un des jeux de données (128 gènes, Philippe et al. 2011) qui avait suggéré la monophylie du clade des “Coelenterata” a été amélioré et complété. L’analyse comparée de ces deux jeux de données indépendants permet d’examiner la corroboration (ou non) des résultats phylogénétiques obtenus. Dans ce cadre, de nouvelles données transcriptomiques ont été produites afin d’améliorer l’échantillonnage taxonomique des embranchements non-bilatériens (Porifera, Cnidaria et Ctenophora,

précédemment sous-représentés par rapport aux bilatériens). Il est donc également attendu de ces analyses une résolution des relations phylogénétiques au sein de ces embranchements. Bien que prometteurs, nos résultats sont préliminaires, et des modifications de certaines étapes de la procédure semi-automatisée sont susceptibles d'être apportées. Il est à noter que les relations au sein des cténophores ne seront pas discutées dans ce manuscrit, car elles le seront dans le chapitre suivant (section 3.3.2).

2.2 Article 1 : The position of ctenophores in the animal tree investigated using two independent large-scale phylogenomic datasets

Simion, P. ; Philippe, H. ; Baurain, D. ; Jager, M. & Manuel, M.

The position of ctenophores in the animal tree investigated using two independent large-scale phylogenomic datasets

PAUL SIMION¹, HERVE PHILIPPE², DENIS BAURAIN³, MURIEL JAGER¹, MICHAEL MANUEL¹

¹UPMC Univ Paris 06, Institut de Biologie Paris–Seine, UMR 7138 “Evolution Paris–Seine”
CNRS UPMC, Case 05, 7 quai St Bernard, 75005 Paris, France.

²CBTM Station d'Ecologie Expérimentale du CNRS - Centre for Biodiversity Theory and Modelling,
2 route du CNRS F-09200 Moulis

³Univ Liege, Department of Life Science, B-4000 Liege, Belgium

Corresponding author michael.manuel@snv.jussieu.fr

Abstract

The phylogenetic position of ctenophores (comb jellies) remains hotly debated, several recent phylogenomic studies having notably reached the surprising conclusion that ctenophores would be sister-group to all other metazoans. To address this problem, here we constructed an entirely original large-scale dataset of more than one thousand genes (*AUTO* dataset), semi-automatically built using a new protocol specifically aimed at eliminating as many of the potential sources of errors as possible (mainly: contaminations, paralogies and missing data). In parallel, we updated a manually-curated data set of 128 genes used in several previous phylogenomic studies (*MANU* dataset). In this context, we sequenced and assembled 22 transcriptomes of sponge, cnidarian and ctenophore species belonging to previously unsampled key taxa. To insure independency between both datasets, genes present in the *MANU* dataset were discarded from the *AUTO* dataset. They were analysed under various conditions as to explore the impact of taxon sampling, missing data, evolution models, amino-acid composition bias, heteropecilleous site removal, and fast-evolving site removal. Globally the results suggest that the “basal” ctenophore topology is a consequence of long-branch attraction and indicate instead that Porifera (whose monophyly is confirmed) are certainly the sister-group to all other animals. However, the exact position of ctenophores remains uncertain. Most *AUTO* dataset analyses place them sister to a “Placozoa + Cnidaria + Bilateria” clade, whereas in the *MANU* dataset there is some signal to group them with cnidarians. A third hypothesis is suggested by *AUTO* analyses after removal of a large proportion of fast-evolving sites, and after removal of the 28 most compositionally-biased taxa: ctenophores sister-group to Bilateria (the Acrosomata hypothesis). In conclusion, the extremely long branch of ctenophores hampers their correct positioning in the tree even with very large and high-quality supermatrices.

Key words: Ctenophora, Metazoa, Porifera, Phylogenomics, Contaminations, Orthology assignment, Heteropecilly

Introduction

The phylogenetic position of Ctenophora is of crucial importance for understanding metazoan evolution. These marine gelatinous animals indeed possess complex anatomical features such as a highly-elaborated neuro-sensory system and rather sophisticated muscular system, a combination of biradial, tetradial and octoradial symmetry, a "mesodermal" germ layer and a differentiated gastric system (Hernandez-Nicase and Franc 1994; Brusca and Brusca 2003; Pang and Martindale 2008; Jager et al. 2011; Dayraud et al. 2012). Morpho-anatomical evidence therefore supports the placement of ctenophores close to other eumetazoans (i.e. cnidarians and bilaterians), all these features being absent in poriferans and placozoans. Reconstruction of the origin and evolution of such important characters as nerve and muscle cells, mesoderm, as well as symmetry and polarity properties, is therefore strongly dependent upon where ctenophores are positioned in the phylogenetic tree of animals. However, relationships between early-diverging metazoan lineages remain highly debated, with different kinds of datasets and methods having supported different topologies (Figure 1).

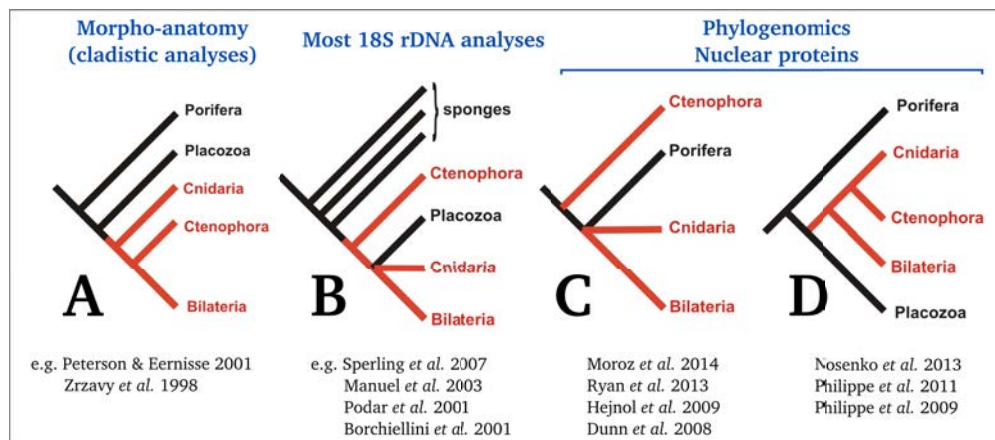


Figure 1: Examples of metazoan phylogenetic relationships as reconstructed by morphology, 18S rDNA and phylogenomics. In red are represented the clades presenting nerve and muscle cells.

Notably, two radically contradictory evolutionary hypotheses have been suggested by recent phylogenomic analyses. One of these places Ctenophora as the first phylum to diverge within Metazoa (Figure 1 C), in contradiction with the traditional conception that Porifera is the sister-group to all other metazoans. First obtained by Dunn et al. (2008), this unexpected result has been repeatedly supported in recent phylogenomic studies (Hejnal et al. 2009; Ryan et al. 2013; Moroz et al. 2014) and is regarded with both much excitement and scepticism (Marlow and Arendt 2014). Its corollary is indeed paraphyly (or even polyphyly) of Eumetazoa, implying either losses of many fundamental cellular and anatomical features in poriferans and placozoans (such as nerve and muscle cells), or, as very recently suggested by Moroz et al. (2014), their independent evolutionary origin between Ctenophora and the rest of the Eumetazoa (Cnidaria+Bilateria, see Figure 1 C). The second hypothesis, supported by other phylogenomic studies, places Ctenophora as the sister-group to Cnidaria, in a Coelenterata clade (Philippe et al. 2009, 2011, see Figure 1 D). A close relationship between ctenophores and cnidarians has been suggested for a long time based on a variety of morphological resemblances (e.g. tentacles, gelatinous body, tetradial symmetry, diffuse nerve net, branching gastrovascular system; see Hyman 1940) but none of these can

be regarded as solidly established homologies (Harbison 1985). The Coelenterata clade has also been recovered monophyletic in some of the phylogenomic analyses of Nosenko et al. (2013). This phylogenetic placement of Ctenophora with Cnidaria is consistent with the monophyly of Eumetazoa and thus with a single evolutionary origin of their shared complex cellular and anatomical features.

An understanding of how these two contradicting topologies have been constructed is needed in order to reach a critical assessment of their implications. Obviously, molecular trees in which eumetazoans are monophyletic are more satisfying from a morpho-anatomical point of view, but the alternative “Ctenophora-first” hypothesis is not inconceivable. The distribution of anatomical characters among metazoan lineages can be interpreted on any given topology, and modern phylogenetics offers many examples of complex patterns of morphological character evolution (involving convergent acquisitions and losses) that overtly contradict traditional conceptions (e.g., homoplastic evolution of coelomic cavities and metamerism in Bilateria, see Edgecombe et al. 2011; Telford 2013). Thus, the two topologies discussed here cannot simply be appreciated with respect to “credibility” of their implications, but should primarily be evaluated with respect to the quality of the corresponding molecular datasets and methodologies.

Phylogenomic analyses are prone to a variety of systematic biases depending on both the quantity and the quality of the data used (Delsuc et al. 2005; Philippe and Telford 2006; Telford and Copley 2011). For example, missing data can be deleterious when reconstructing phylogenetic relationships, even at low levels of about 20% if the missing data are concentrated in a particular clade (Roure et al. 2013). This configuration is met in current phylogenomic datasets where, for instance, most sampled poriferan species have high amounts of missing data (Laurin-Lemay et al. 2012). Another critical source of problems is the presence in phylogenomic datasets of sequences resulting from contaminations, which has a dramatic impact on phylogeny reconstructions because they bring a specific contradictory and additive phylogenetic signal (Philippe et al. 2011). Contaminations can be due to natural biological contaminants, or to artificial contaminations during preparation of the samples or sequencing (Kircher et al. 2011, 2012) and are found in practice in virtually any large-scale dataset issued from deep coverage transcriptome sequencing (NGS technologies). The presence of paralogous sequences brings a similarly contradictory signal. Multigenic families and clade-specific duplications are frequent and cannot be treated by using only distance comparison methods (e.g. BLAST). Correct orthology/paralogy assessments require inspection of single-gene trees, which nevertheless can also be insufficient, since single-gene phylogenies can be poorly resolved especially concerning ancient relationships. Lastly, large phylogenomic matrices are more sensitive than smaller matrices to biases with additive properties such as the long branch attraction artifact (LBA) and amino-acid compositional bias (Felsenstein 1978; Delsuc et al. 2005; Philippe et al. 2005; Philippe and Telford 2006; Nesnidal et al. 2010; Song et al. 2010; Telford and Copley 2011; Philippe et al. 2011; Philippe and Roure 2011; Nesnidal et al. 2013). This is of particularly critical importance as far as ctenophores as concerned, since this lineage has a strongly accelerated rate of substitutions and a high molecular divergence with respect to other metazoans.

The Coelenterata hypothesis has been supported by a manually-curated dataset dominated by ribosomal proteins, free of contaminations and paralogs (eliminated by systematic visual inspection of single gene trees, Philippe et al. 2009, 2011). This dataset corresponds to 128 highly expressed genes for

which wide taxonomic sampling is available at the metazoan scale. Single-gene trees with such detailed taxon sampling allow fine detection of paralogs and contaminations, each time new sequences are added to the alignments, through a protocol detailed in Baptiste et al. (2002). The supermatrix resulting from concatenation of these alignments contained 30,257 amino-acid positions with 27% of missing data (Philippe et al. 2009). Its analysis supported the Coelenterata hypothesis (support was only moderate when using distant outgroups such as Fungi). The analysis of a semi-automatically built dataset composed of 87 ribosomal proteins also supported the Coelenterata hypothesis and showed that these ribosomal proteins were less saturated than the non-ribosomal proteins used in the same study (Nosenko et al. 2013). However, the number of genes in these matrix was not very high, and since all ribosomal proteins are involved in translation, they might evolve in a concerted fashion.

The "basal" Ctenophora hypothesis has been obtained from very large automatically-built datasets with high amounts of missing data and of which precise gene content was not described. The automated clustering protocol designed by Dunn et al. (2008: 150 genes) has been re-used with minor changes in two subsequent studies (Hejnol et al. 2009: 1,487 genes; Ryan et al. 2013: 406 genes). At the first step of this protocol, species transcriptomes were BLASTed onto the Homologene Database. The BLAST results were used in order to clusterise the transcripts. Clusters were then evaluated according to various metrics such as the mean number of sequences per Operational Taxonomical Unit (OTU). If too many sequences were present, a tree was built from the cluster using parsimony criterion and this tree was used to detect and remove supernumerary sequences. These three studies recovered the same metazoan topology. The strategy used in Moroz et al. (2014: 586 genes) is comparable to the one described above, yet it differs in one major way : the transcript clustering step was replaced by the use of an existing cluster set of 1,032 Orthology Groups (OGs) (Ebersberger et al. 2009) and HaMStR software was used to incorporate a new transcript into a given OG if the transcript gave Best-Reciprocal-Hit (BRH) with the *Drosophila* protein present in that OG. Another noticeable difference is that Moroz et al. (2014) removed sequences from the clusters if they were too divergent with respect to the alignment consensus sequence. The two similar procedures used in these previous works (Dunn et al. 2008; Hejnol et al. 2009; Ryan et al. 2013; Moroz et al. 2014) resulted in the creation of large supermatrices with amounts of missing data ranging from 54.6% to 81%.

Current automated matrix building workflows are inefficient at detecting and therefore removing both sequence contaminations and paralogs. These protocols use as a starting point similarity clusters which contain orthologs as well as paralogs of multigenic families. By solely using a clustering of BLAST results, or a BRH criterion with a divergent reference taxon (i.e. *Drosophila*) to populate these clusters with transcripts, both orthologs and paralogs can be incorporated to the alignments. This is because similarity scores are poor indicators of evolutionary distance and relatedness (e.g. Koski and Golding 2001). Number of sequences per species is then invoked as a criterion to detect paralogy, but in fact this only works when both paralog(s) and ortholog(s) are present in a cluster. If the true ortholog is too divergent, has been secondarily lost through evolution, or has not been sequenced (three frequent situations), it will not be present in the cluster. Thus even if the metrics detect a single sequence for the species, it can be a paralog. Reciprocally, the presence of several sequences for a given species is not

problematic when these originated from recent, taxon-specific duplications (“in-paralogy”). Paralogy can be rigorously detected only from single-gene phylogenetic trees. Such inspections are implemented in the automated procedures described above, but were only applied to alignments not fulfilling the metrics requirements, leaving a large part of the dataset unchecked. In addition, none of these automated workflows (Dunn et al. 2008; Hejnol et al. 2009; Ryan et al. 2013; Moroz et al. 2014) can take into account the existence of contaminations in datasets issued from transcriptome sequencing. However, this problem is critical, as demonstrated by a re-analysis of Dunn et al. (2008) dataset, showing that it contains many paralogs and contaminations (see in Philippe et al. 2011, Table S2 and Figure S11) and that their removal significantly impacts the topology recovered from phylogenetic analyses (see also Struck 2013 for another example).

In an attempt to resolve the phylogenetic position of Ctenophora, we constructed two datasets, fully independent from one another, using respectively the two different approaches that led to the two currently competing hypotheses (i.e. automatic procedure vs. manual construction). At the same time, taxonomical sampling was greatly improved by sequencing 22 new non-bilaterian transcriptomes (representing poriferan, ctenophore and cnidarian subgroups that were until now absent or poorly represented in phylogenomic datasets). The first dataset (called “*MANU-128*”) is an update of the 128-genes supermatrix which supported the Coelenterata hypothesis (Philippe et al. 2009, 2011). To build the second dataset (“*AUTO*”) we designed a new automated procedure to select orthology groups at metazoan scale, with specific steps to eliminate paralogs and contaminations. This approach resulted in the creation of 4,235 initial alignments, from which supermatrices were generated in order to minimise the amount of missing data and eliminate redundancy with the *MANU-128* dataset, while optimising both the quality and the quantity of the data, taxonomic representation, and the number of positions. We compared these two different datasets and the outcomes of their phylogenetic analyses across various options of outgroup content, evolutionary model, amount of missing data, taxon sampling, and progressive removal of heteroprecilleous or fast-evolving sites. Our results demonstrate the complexity of phylogenetic signal detection in large phylogenomic supermatrices, in accordance with the conclusions of recent studies (Evans et al. 2010; Salichos and Rokas 2013; Sharma et al. 2014; Straub et al. 2014). While our analyses suggest that monophyletic Porifera, and not Ctenophora, are sister-group to all other metazoans, the exact position of ctenophores remains unresolved. There is some signal for monophyletic Coelenterata in the *MANU-128* Dataset, but the *AUTO* dataset generally supports ctenophora as sister-group to (Placozoa + Cnidaria + Bilateria), with some analyses instead grouping ctenophores with bilaterians in an Acrosomata clade.

Materials and methods

1. Protocol of semi-automated creation of a metazoan-scaled phylogenomic dataset

We designed a semi-automated bioinformatic procedure (summarised in Figure 2) to build a metazoan scaled phylogenomic dataset. A total of 80 non-bilaterian transcriptomes (22 of which are new and were sequenced in the context of this study) were assembled and cleaned out from contaminations. In parallel, a set of 4,412 similarity clusters were built from high-quality holozoan proteomes with OrthoMCL (Li et al. 2003). Homologous proteins from the decontaminated transcriptomes were incorporated into the clusters using a new orthology assignment software: “Forty 2”. Similarity clusters were then filtered using various metrics, and we subsequently used PhyloTreePruner (Kocot et al. 2013) in order to eliminate paralogous sequences from the clusters that passed the previous filters. The remaining alignments were later concatenated into various supermatrices.

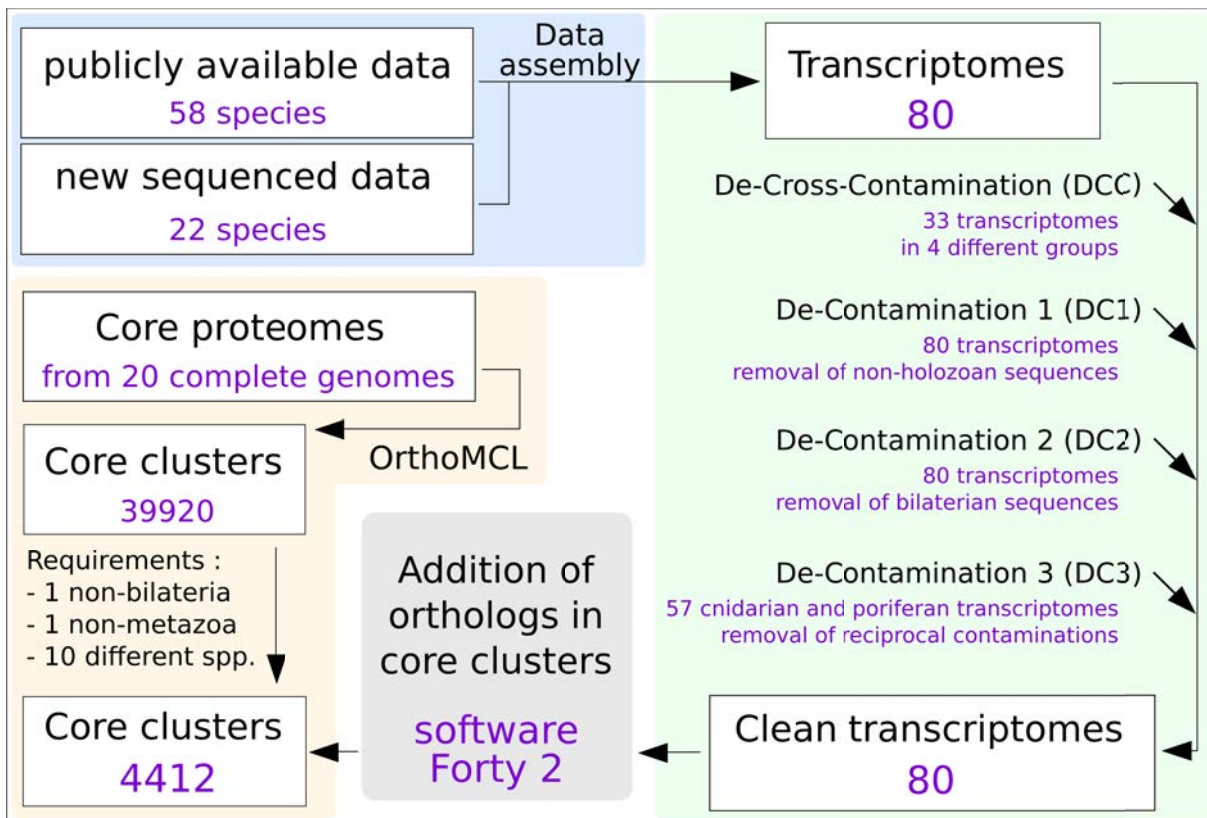


Figure 2: Summary of our new semi-automated procedure for construction of a metazoan-scaled phylogenomic dataset. Resulting Core clusters were subsequently treated for paralogy using PhyloTreePruner (not shown).

1.1 Creation of core clusters

Core proteome selection We used selected available metazoan proteomes and their closest outgroups as input for the OrthoMCL software to create clusters of similar sequences (taken as proxies for orthology groups). A total of 20 “core proteomes” have been chosen (see Table 1) according to the following criteria:

phylum	species	source	number of proteins
Filasterea	<i>Capsaspora owczarzaki</i>	NCBI	16878
Choanoflagellata	<i>Monosiga brevicollis</i>	NCBI	18141
Choanoflagellata	<i>Salpingoeca rosetta</i>	NCBI	22780
Porifera	<i>Amphimedon queenslandica</i>	Ensembl	25858
Placozoa	<i>Trichoplax adhaerens</i>	Ensembl	11217
Ctenophora	<i>Mnemiopsis gardeni</i>	NHGR	16030
Cnidaria	<i>Nematostella vectensis</i>	Ensembl	22084
Cnidaria	<i>Acropora digitifera</i>	Compagen	21433
Cnidaria	<i>Hydra magnipapillata</i>	Compagen	29647
Echinodermata	<i>Strongylocentrotus purpuratus</i>	Ensembl	27147
Hemichordata	<i>Saccoglossus kowalevskii</i>	JGI	28734
Cephalochordata	<i>Branchiostoma floridae</i>	UniProt	27293
Chordata	<i>Homo sapiens</i>	UniProt	24884
Mollusca	<i>Crassostrea gigas</i>	Ensembl	23542
Mollusca	<i>Aplysia californica</i>	NCBI	24436
Annelida	<i>Capitella teleta</i>	Ensembl	30014
Annelida	<i>Helobdella robusta</i>	Ensembl	21952
Arthropoda	<i>Daphnia pulex</i>	Ensembl	26539
Arthropoda	<i>Ixodes scapularis</i>	Ensembl	16659
Arthropoda	<i>Tribolium castaneum</i>	Ensembl	14942

Table 1: List of the 20 “core species”, corresponding sources, and sizes of proteomes used to build similarity clusters with OrthoMCL. The number of proteins indicated for each species correspond to proteins after removal of sequences smaller than 100 amino-acids or with more than 10% of ambiguous or missing characters.

1) proteomes derived from high quality, assembled and annotated complete genome sequence; 2) species with globally low substitution rate; and 3) species chosen as to optimise taxonomic diversity at the holozoan scale (metazoans and their unicellular relatives).

We will refer to those selected proteomes as “core proteomes” and to their corresponding species as “core species”. Since the determination of the root of the metazoan tree belongs to the aims of this study, three non-metazoans holozoans species were included and will be used as outgroups for phylogenetic analyses in the next steps.

Production of core clusters using OrthoMCL Clusters were created with OrthoMCL (Li et al. 2003). Within the OrthoMCL protocol, the BLAST step was replaced by a Usearch step (with parameters as follows: evaluate 1e-5; accel 1.0; Edgar 2010), and the granularity parameter of the MCL step was set to 1.5. This protocol resulted in the creation of 39,920 clusters of similar sequences. We will refer to these sets of similar sequences as “core clusters”.

Cluster filtering We then filtered the 39,920 core clusters in order to eliminate clusters evidently unsuitable for phylogenetic analyses of metazoan relationships due to insufficient taxonomic coverage. Initial core clusters contained from 1 to 20 of the core species and from 1 to a very large number of sequences per core species. Those with few represented core species are likely to correspond to proteins

	Taxonomy	Species	Location	sequencing technology	Species pooling	read number
PORIFERA	Calcarea	<i>Leucosolenia variabilis</i>	laboratory tanks	454	LAB_454	130,957
		<i>Leuconia nivea</i>	Roscoff, France	Illumina	LAB_I	27,051,000
		<i>Clathrina coriacea</i>	Roscoff, France	Illumina	LAB_I	24,118,600
		<i>Grantia compressa</i>	Roscoff, France	Illumina	LAB_I	32,598,500
	Homoscleromorpha	<i>Plakina jani</i>	Marseille, France	Illumina	LAB_I	28,462,400
	Demospongiae	<i>Pleraplysilla spinifera</i>	Marseille, France	Illumina	LAB_I	25,684,600
CTENOPHORA	Cestida	<i>Cestum veneris</i>	Villefranche-sur-Mer, France	454	LAB_454	162,577
	Platyctenida	<i>Coeloplana cf. meteoris</i>	Inland Aquatics, USA	Illumina	LAB_I	73,657,000
		<i>Vallricula multiformis</i>	Inland Aquatics, USA	Illumina	LAB_I	34,999,800
	« Cydippida »	<i>Lampea pancerina</i>	Villefranche-sur-Mer, France	Illumina	LAB_I	28,732,300
		<i>Pleurobrachia pileus</i>	Roscoff, France	454	none	1,260,258
CNIDARIA	Octocorallia	<i>Pennatula rubra</i>	Banyuls-sur-Mer, France	454	LAB_454	181,391
		<i>Alcyonium palmatum</i>	Banyuls-sur-Mer, France	Illumina	LAB_I	28,441,600
	Hexacorallia	<i>Parazoanthus axinellae</i>	Villefranche-sur-Mer, France	454	LAB_454	150,085
		<i>Pachycrianthus solitarius</i>	Etang de Berre, France	454	LAB_454	216,745
		<i>Antipathes caribbeana</i>	Port-Louis, Guadeloupe	Illumina	LAB_I	49,997,500
		<i>Plumapathes pennacea</i>	Port-Louis, Guadeloupe	Illumina	LAB_I	59,779,500
	Medusozoa	<i>Carybdea xaymacana</i>	Pointe-à-Pitre, Guadeloupe	454	LAB_454	161,900
		<i>Lucernariopsis campanulata</i>	Île Callot, Mortaix, France	Illumina	LAB_I	28,539,600
		<i>Pelagia noctiluca</i>	Villefranche-sur-Mer, France	Illumina	LAB_I	37,484,100
<i>Clytia hemisphaerica</i>		Villefranche-sur-Mer, France	454, Illumina	none	151,811,039	
		<i>Liriope tetraphylla</i>	Sainte-Rose, Guadeloupe	Illumina	LAB_I	30,934,800

Table 2: Summary of the 22 non-bilaterian species sequenced for this study: species, sampling location, sequencing method, species pooling for sequencing, number of reads. LAB_454 indicates species that have been pooled together during 454 sequencing and LAB_I indicates a similar pooling for Illumina sequencing.

poorly conserved at the metazoan scale. We used custom Perl scripts in order to keep only clusters gathering at least 10 different core species, and containing at least both 1 non-bilaterian metazoan core species and 1 non-metazoan core species. A number of 4,412 core clusters remained after this filtering step.

1.2 Preparation of transcriptomes

Additional transcriptomes to be included into the core clusters were prepared as follows.

RNA preparation and sequencing of new non-bilaterian transcriptomes Total RNA was extracted after careful cleaning of the biological samples (see collection data in Table 2) using either a Trizol-based protocol or the QIAGEN RNeasy Kit. In both cases, sample tissue was first reduced into powder in liquid nitrogen. In the first protocol, sample powder was then incubated for 5 min in trizol solution, before addition of chloroform for another 15 min of incubation. The solution was then centrifuged for 15 min at 4°C (12,000 G) in order to retain the superior aqueous phase only, which was subsequently incubated for 10 min in isopropanol. Samples were then centrifuged for 10 min at 4°C (12,000 G) and the supernatant was eliminated. Pellet was vortexed and centrifuged for 5 min at 4°C (7,500 G) in Ethanol 75%. After supernatant elimination, the dried pellet was lastly eluded in RNase-free water. The second protocol was as described in the QIAGEN RNeasy kit manual.

The cDNA libraries of these 22 new non-bilaterian species were constructed at GATC Biotech (Konstanz, Germany), which then used either 454 pyrosequencing or Illumina technology for sequencing (see Table 2). Among these species, six of them were pooled into a single 454 run and fourteen others were pooled into two Illumina lanes in the same run (see Table 2).

Selection of additional non-bilaterian transcriptomes Additional transcriptomes from 58 non-bilaterian species (see Table 3) present in publicly available databases were selected to improve the taxonomic sampling according to the following requirements: 1) transcriptomes helping to enhance the taxonomic diversity of either metazoans or outgroups ; 2) large transcriptomes of good quality (as far as possible); and 3) transcriptomes from slowly evolving species (as far as possible). Transcriptome assembly for *Euplectella* was kindly provided by Pr. Nori Satoh (OIST, Okinawa, Japan).

Assembly of transcriptomes All 454 data newly sequenced in the context of this study as well as raw 454 data retrieved from publicly available databases were assembled using MIRA (Chevreux 2005) or a combination of MIRA and CAP3 (Huang and Madan 1999). The data newly sequenced with Illumina technology as well as Illumina reads retrieved from publicly available databases were assembled using either Trinity (Haas et al. 2013) or SOAPdenovo-trans (Xie et al. 2014) in accordance with recent assessment of their better assembly quality in comparison with other softwares (Zhang et al. 2013). See Table 3 for details and the respective numbers of assembled contigs.

De-Cross-Contamination (DCC) Multiple deep sequencing of different samples from the same lab and/or sequencing facility implies multiple cross-contaminations of the samples. The existence of such cross-contaminations can easily be detected by constructing and inspecting single-gene trees with transcriptomic data. In our case, this was done for a number of ribosomal proteins after inclusion of sponge, cnidarian and ctenophore transcripts sequenced by our lab. It was not rare to see a transcript supposed to belong to a given species (e.g. a cnidarian) positioned in the tree at a “wrong” place (e.g. very close to a poriferan transcript), with 100% nucleotidic identity between these transcripts ascribed to unrelated species (in our aforementioned example, the cnidarian transcriptome is cross-contaminated by a transcript from a poriferan species). These cross-contaminations occurred despite multiple precautions during preparation of the RNA samples at the bench. Although their causes are not entirely clear, they might arise at many different steps, from RNA preparation, to library construction, addition of oligonucleotide tags, and demultiplexing of samples after sequencing (note that, in this study, all the steps following RNA preparation were done at GATC). In particular, sequencing using multiplexing strategies with sample-specific tags are known to suffer from a proportion of incorrect sample identification: as high as 0.3% of the reads can be assigned to the wrong sample (Kircher et al. 2012). Our analyses of multiple transcriptomes recently produced by other labs revealed that they are all affected by this cross-contaminations phenomenon. The existence and the extent of cross-contaminations in transcriptomic datasets is correlated with the increase of sequencing depth (i.e. from Sanger to 454 and to Illumina sequencing technologies).

Such contaminations are extremely deleterious if not eliminated from the transcriptomes used to infer phylogenetic relationships (Laurin-Lemay et al. 2012). Indeed they bring a strong false “phylogenetic” signal which can weaken support for some clades, or lead to wrong topologies even when using phylogenomic datasets (e.g. in a supermatrix assembled without prior de-cross-contaminations with 1,400 genes, we observed that *Liriope tetraphylla*, a cnidarian belonging to Trachymedusae, branched closely to another species that was sequenced at the same time : *Pleraplysilla spinifera*, a Demospongiae).

We designed a “De-Cross-Contamination” (DCC) protocol that detects contaminant transcripts within a series of several species sequenced by the same lab using a given sequencing technology (i.e. Illumina or 454 pyrosequencing) and removes them from the transcriptomes. This protocol is based on the idea that a cross-contaminant should be under-represented with respect to the genuine transcript of the sequenced species. Indeed, Illumina RNA-seq is supposed to be quantitative and is widely used to evaluate differential gene expression, the levels of expression of a gene being reflected by the relative numbers of reads mapping to this gene in different sequenced samples. A contaminant RNA should behave like a lowly expressed RNA and the genuine RNA should have many more reads (under the assumption that the expression level of orthologs is similar in the different species). Our DCC protocol is therefore composed of three steps.

1. All transcripts from several species produced by the same lab or sequencing facility were compared to each other (all-vs-all nucleotidic BLAST).
2. The number of reads participating to each transcript during their previous assembly was incorporated to the BLAST results.
3. When a given transcript was ascribed to several species (with nucleotide identity $> 98\%$), a comparison of the number of reads was made. If a sequence of one species had much more reads than all others (by a factor of 5), it was considered as the genuine species from which this transcript comes, and the transcript was eliminated from the transcriptomes of the other species. Otherwise (if not at least five times more represented in one of the species), the gene was eliminated from all datasets, since in this case its origin could not be determined with confidence. In case a transcript had no hit, even not to itself, it was eliminated from its transcriptome (this can only be the case for very low quality transcripts mostly containing low-complexity regions).

This method pre-supposes that the samples used for RNA extraction are biologically comparable (i.e. mixtures of adult tissues used in the case of all species newly sequenced in this study). Our protocol was validated by comparing the distribution of initial transcripts vs. de-cross-contaminated transcripts in single-gene trees for the 128 phylogenetic markers manually curated and maintained in Hervé Philippe’s lab (*MANU-128* dataset). Cross-contaminations are in most cases easily identified in these gene trees thanks to extensive taxonomic representation (e.g. if without DCC, a newly sequenced hydrozoan species has two genes in a given tree, one grouping with hydrozoans and the other falling within demosponges, and if a demosponge was part of the sequencing project, it is clear that the second gene is a cross-contamination). Visual inspection of these gene trees showed that the DCC procedure indeed removed all cross-contaminations, and that the gene selected by the DCC program was in all cases correctly positioned with respect to the taxonomy of the species. The major drawback of our protocol is the possible removal of correct sequences that were not sufficiently “overexpressed” in the right sample, or that were naturally identical in nucleotides to another (very closely-related) species sequenced and equally expressed (in our sampling of sequenced species, this situation might have potentially occurred only for the two Anthipatharia species).

We performed a total of 4 different DCCs for different groups of transcriptomes (for transcriptome groupings details, see Table 3). Note that the DCC for the “DX1” group used a nucleotide identity threshold of 95% instead of 98%. Indeed, this DCC group did not contain very closely related species, which enabled us to reduce the identity threshold with a lower risk of removing transcripts that would not be true cross-contaminations. This process resulted in the suppression on average of 8.9% of the transcripts per species, and ranged from a suppression of 0.16% of the transcripts (*Pleurobrachia pileus*, this study) to a suppression of 70.64% of the transcripts (ctenophore sp3 A, Moroz et al. 2014). The particular last case could theoretically be explained by a high-level of RNAs degradation of the sample which would inflate the relative quantity of cross-contaminant RNA sequenced. The occasionally extremely high number of cross-contaminations have strong implications for phylogenomic practice: the detection and subsequent removal of these cross-contaminations should be a mandatory step of future analyses based on transcriptomic data obtained using next-generation sequencing technologies.

Removal of contaminations coming from other sources When sequencing animal tissues with deep coverage, presence of contaminating life forms such as epibiotic organisms, symbionts, parasites (all these either small multicellular or unicellular organisms) is almost inevitable, regardless of the efforts deployed in “cleaning” the samples. Some animals are particularly prone to such contaminations (e.g. sponges, benthic colonial cnidarians, etc.). However, these contaminations bring into a dataset strong contradictory non-phylogenetic signal which, like cross-contaminations, will hamper correct reconstruction of the evolutionary history. Contaminations across metazoan phyla are the most deleterious because we specifically aim to resolve relationships between these phyla. To solve this problem, we designed a “De-Contamination” (DC) protocol which is composed of three successive steps:

- **Suppression of non-holozoan sequences (DC1):** a dataset of proteomes representing the molecular diversity of Eukaryota, Eubacteria and Archaea was compiled (listed in Supplementary Table 1 in Appendix A). We extracted the open-reading-frames (ORFs) of several transcriptomes for which we had no established proteomes available, using the `-findorfs` option of Usearch (with parameters as follows: `-xlat -orfstyle 7 -mincodons 60`). To this list have been added the 20 core proteomes in order to represent the molecular diversity of holozoans. The ORFs of every transcript from every holozoan DCC treated transcriptome were then “blasted” against this home-made database using the `-ublast` option of Usearch (with following parameters: `-evalue 1e-10 -id 0.2 -weak_id 0.1 -maxaccepts 1000 -maxrejects 3000 -maxhits 100`). If the best bitscore corresponded to a transcript that did not belong to the 20 core proteomes, the transcript was removed from the corresponding transcriptome.
- **Suppression of bilaterian contaminations (DC2):** a dataset of various proteomes and transcriptomes from taxa representing the molecular diversity of Bilateria down to the phylum level was compiled (listed in Supplementary Table 2 in Appendix A). We extracted the ORFs of bilaterian transcriptomes for which we had no proteomes available using the `-findorfs` option of usearch (with parameters as follows: `-xlat -orfstyle 7 -mincodons 160`). We also extracted the ORFs of the non-bilaterian transcriptomes to be decontaminated (same parameters as above), which we used

TAXONOMY	SPECIES	DATA ASSEMBLY				DECONTAMINATIONS					CLEAN CONTIG NB	OTUs			
		CORE	SOURCE	ASSEMBLER	DCC GROUP	RAW CONTIG NB	DCC	DC1	DC2	DC3					
OUTGROUPS	FILASTEREA	<i>Capsaspora owczarzaki</i>	yes	ncbi									65,94		
		<i>Ministeria vibrans</i>		SRA		13531		9055	927				94		
	ICHTHYOSPOREA	<i>Sphaeroforma arctica</i> J P610		SRR050206_SRR050211		Trinity (DB)	47232		41593	705				65,94	
		<i>Amoebidium parasiticum</i> JAP 7 2		ncbi			105237		83038	3361				65,94	
	CHOANOFAGELLATA	<i>Monosiga brevicollis</i>	yes	ncbi										65,94	
<i>Salpingoeca rosetta</i>		yes	ncbi										65,94		
<i>Monosiga ovata</i>			ncbi			76608		29988	12797				65,94		
<i>Stephanoea diplocostata</i>			ERS204956		Trinity (FD)	41143		30976	1760				65,94		
	<i>Acanthoea</i> sp. 10tr		SRA			91237		63697	3621				65,94		
PORIFERA	DEMOSPONGIAE	<i>Amphimedon queenslandica</i>	yes	ensembl									65,94		
		<i>Carteriospongia foliascens</i>		ncbi			3556		2329	81				94	
		<i>Ephydatia muelleri</i>		SRR1041944_ERR216192		SOAPdenovo trans + Trinity (FD)	167630		145097	2836	3548			16149	
		<i>Suberites domuncula</i>		ncbi			13672		4964	1013	936			6759	
		<i>Tethya wilhelma</i>		ERR216193		Trinity (FD)	26057		19201	620	751			5485	
		<i>Pachydictyum globosum</i>		ncbi			1719		989	40	19			671	
	HEXACTINELLIDA	<i>Lubomirskia baicalensis</i>		ncbi			2188		922	96	32			1138	
		<i>Asbestopoma hypogae</i>		ERR216190			17930		14013	489	436			2992	
		<i>Pteraplyllia spinifera</i>		this study		SOAPdenovo trans (FD)	114816	LAB_I	99891	1933	2732			7482	
		<i>Heterodone calyx</i>		ncbi			4790		2554	209	367			1660	
		<i>Oopscas minuta</i>		ncbi			1443		735	86	43			579	
		<i>Crateromorpha meyeri</i>		ncbi			1498		752	65	82			599	
	CALCAREA	<i>Euplectella aspergillum</i>		this study		MIRA + CAP3 (PS)	20332		13302	1560	2146			3324	
		<i>Leucetta chagosensis</i>		ncbi			5703		3417	170	190			1926	
		<i>Sycon raphanus</i>		ncbi			1629		806	57	7			759	
<i>Sycon dilatatum</i>			ERR (many)		Trinity (DB)	776523		669555	50083	20774			36111		
<i>Leucosolenia variabilis</i>			this study		MIRA (PS)	7331	LAB_454	5205	157	71			1749		
<i>Clathrina coriacea</i>			this study		SOAPdenovo trans (FD)	118488	LAB_I	2708	100217	2976			2306		
<i>Leuconia nivea</i>			this study		SOAPdenovo trans (FD)	172384	LAB_I	2017	154935	1607			623		
<i>Grantia compressa</i>			this study		SOAPdenovo trans (FD)	124273	LAB_I	1715	109333	966			504		
HOMOSCLEROMORPHA	<i>Oscarella lobularis</i>		ncbi			1502		708	143	201			450		
	<i>Plakina jani</i>		this study		SOAPdenovo trans (FD)	128893	LAB_I	869	113981	4172			6319		
CTENOPHORA	CYCLOCOELA	<i>Trichoplax adhaerens</i>	yes	ensembl									65,94		
		<i>Mnemopsis gardeni</i>	yes	NIHR										65,94	
		<i>Mnemopsis gardeni A</i>		SRR789900		SOAPdenovo trans (PS)	229809	DXM	65336	130918	11845			21710	
		<i>Bolinopsis infundibulum</i>		SRR786491		SOAPdenovo trans (PS)	199328	DXM	91050	93394	3656			11228	
		<i>Cestum veneris</i>		this study		MIRA (PS)	15003	LAB_454	1990	3964	425			8624	
	TYPHLOCOELA	<i>Beroe</i> sp.		LMU			22352		16262	383				5707	
		<i>Beroe abyssicola</i>		SRR777787		SOAPdenovo trans (PS)	44269	DXM	965	30110	876			12318	
		<i>Pleurobrachia pileus</i>		this study		MIRA (PS)	48882	LAB_454	80	30798	487			17517	
		<i>Pleurobrachia pileus A</i>		SRR789901		SOAPdenovo trans (PS)	242292	DXM	20426	199217	4626			18023	
		<i>Pleurobrachia bachei</i>		GeneModels			19523		9241	312				9970	
		<i>Lampea pancarina</i>		this study		SOAPdenovo trans (FD)	80639	LAB_I	759	65025	590			14265	
		<i>Codopiana cf. meteoris</i>		this study		SOAPdenovo trans (FD)	110525	LAB_I	4484	91806	452			13783	
		<i>Codopiana astericola</i>		SRR786490		SOAPdenovo trans (PS)	142939	DXM	21261	108098	1227			12353	
		<i>Vallidula multiiformis</i>		this study		SOAPdenovo trans (FD)	79493	LAB_I	3673	63556	212			12052	
		<i>Vallidula multiiformis A</i>		SRR786489		SOAPdenovo trans (PS)	194558	DXM	84130	96506	1809			12113	
		<i>Euplokamis dunlapae</i>		SRR777663		SOAPdenovo trans (PS)	170355	DXM	1474	152654	2238			13989	
		sp3		Dunn 2008			1434		843	23				568	
		sp3 A		SRR786492		SOAPdenovo trans (PS)	171855	DXM	121398	45748	1550			3159	
		<i>Dryodora glandiformis</i>		SRR777788		SOAPdenovo trans (PS)	55048	DXM	5156	36941	1157			11794	
		CNIDARIA	ANTHOZOA	<i>Nematostella vectensis</i>	yes	ensembl									65,94
<i>Metridium senile</i>				compagen			8284		4503	94	21			3666	
<i>Anemonia viridis</i>				compagen			10575		6695	134	43			3703	
<i>Hormathia digitata</i>				SRR504348		MIRA + CAP3 (PS)	40375		24064	496	43			15772	
<i>Bolocera tuedia</i>				SRR504347		MIRA + CAP3 (PS)	50694		34124	485	56			16029	
<i>Alptasia pallida</i>				SRA		MIRA (J F)	147665		124256	1362	395			21652	
<i>Urticina eques</i>				SRR942796		Trinity (FD)	38965		29075	499	42			9349	
<i>Edwardiaella lineata</i>				SRA			90440		61187	2105	142			27006	
<i>Parazoanthus axinellae</i>				this study		MIRA (PS)	10190	LAB_454	341	5180	305	128			4236
<i>Pachyzoanthus solitarius</i>				this study		MIRA (PS)	13375	LAB_454	1252	8055	484	73			3511
<i>Antipathes caribbeana</i>				this study		SOAPdenovo trans (FD)	126470	LAB_I	1495	105117	995	96			18767
<i>Plumapathes pennacea</i>				this study		SOAPdenovo trans (FD)	153927	LAB_I	2148	129279	1226	125			21149
<i>Acropora digitifera</i>				yes	compagen										65,94
<i>Montastraea faveolata</i>				compagen			13498		7662	268	59				5509
<i>Montastraea faveolata 2</i>				SRR863037		Trinity (DB ?)	158677		124138	798	247				33494
<i>Porites astreoides</i>			compagen			7245		5124	90	63				1968	
<i>Porites australiensis</i>			ncbi			74998		57096	481	164				17257	
<i>Acropora cervicornis</i>			ncbi			95681		69163	886	74				25558	
<i>Podilopora damicornis</i>			SRR448024_SRR448025		SOAPdenovo trans (FD)	190816		159229	3045	74				28468	
<i>Platygyria carnosus</i>			SRA			162149		132404	766	214				28765	
<i>Gorgonia ventalina</i>			SRA			169763		149814	2253	479				17217	
<i>Pennatula rubra</i>			this study		MIRA (PS)	12788	LAB_454	630	6541	280	129			5208	
<i>Alcyonium palmatum</i>			this study		SOAPdenovo trans (FD)	102069	LAB_I	3218	83038	978	158			14677	
MEDUSOZOA	<i>Hydra magnipapillata</i>		yes	compagen										65,94	
	<i>Hydractinia echinata</i>			ncbi			9464		5794	243	89			3338	
	<i>Podocoryne carnea</i>			ncbi			3466		1476	82	25			1883	
	<i>Cytia hemisphaerica</i>			this study		multiple (Tsuyoshi Momose)	36542		22604	805	403			12730	
	<i>Nanomia bijuga</i>			SRR09297_SRR871527		SOAPdenovo trans (FD)	376015		353922	3027	660			18406	
	<i>Craspedacusta sowerbyi</i>			SRR923472		SOAPdenovo trans (PS)	350418	DX1	5036	325398	2338	325			17321
	<i>Ectopleura larynx</i>			SRR923510		SOAPdenovo trans (PS)	327804	DX1	16908	285740	11399	834			12923
	<i>Hydractinia polydina</i>			SRR923509		SOAPdenovo trans (PS)	407627	DX1	6214	377891	4298	973			18251
	<i>Liriope tetraphylla</i>			this study		SOAPdenovo trans (FD)	184056	LAB_I	6333	152052	3495	503			21673
	<i>Cyanea capillata</i>			Holland 2007			2472		1315	306	4				847
	<i>Aurilia aurita</i>			SRR040477		MIRA (PS)	15139		9877	266	53				4943
	<i>Tripedalia cystophora</i>			SRR1182852			10256		6403	330	75				3448
	<i>Stomatophus melagris</i>			SRR1168418		Trinity (DB)	118942		84941	2080	427				31494
	<i>Carybdea xaymacana</i>		this study		MIRA (PS)	6281	LAB_454	13	3611	233	61			2363	
	<i>Lucernariopsis campanulata</i>		this study		SOAPdenovo trans (FD)	73120	LAB_I	1080	58845	1288	489			11418	
<i>Pelagia noctiluca</i>		this study		SOAPdenovo trans (FD)	111636	LAB_I	785	94825	1243	203			14580		

Table 3: Taxonomical sampling, genomic and transcriptomic data used in this study. Simplified taxonomy, core species (i.e. species used to build the core clusters), data source, method for assembly of reads when performed in the scope of this study, initial number of transcripts, number or transcript suppressions in successive DCC, DC1, DC2 and DC3, final number of transcripts used to complete core clusters. The last “OTUs” column indicates whether an OTU was used only in complete taxon sampling (“94”), or in both complete and reduced ones (“65,94”). “P*” means Placozoa.

to complete the molecular diversity representation of holozoans. The ORFs of every transcript from non-bilaterian transcriptomes (including outgroups) were then “blasted” against this home-made database. Transcript hits against themselves were ignored. If the best bitscore corresponded to a bilaterian transcript (or if the best non-bilaterian bitscore was equal to the best bilaterian bitscore), the transcript was regarded as a bilaterian contamination and subsequently removed from the corresponding transcriptome.

- **Suppression of contaminations between Porifera and Cnidaria (DC3):** preliminary observations of single-gene trees (mostly for ribosomal proteins) indicated a non-negligible level of contaminations of cnidarian origin in sponge transcriptomes, and reciprocally. These marine animals (particularly sponges and benthic cnidarians) are very difficult to clean from natural contaminants (e.g. small hydrozoans present on the surface and/or in the cavities of a sponge; sponge larvae on a cnidarian sample, etc.). It was therefore necessary to remove sponge contaminations in cnidarian transcriptomes and reciprocally. We extracted the ORFs of poriferan and cnidarian transcriptomes to be decontaminated using the `-findorfs` option of Usearch (with parameters as follows: `-xlat -orfstyle 7 -mincodons 100`). The resulting ORFs were used in combination with *Amphimedon* and cnidarian core proteomes to create a database of protein sequences. The ORFs of each transcript were then “blasted” against this database, ignoring transcripts hits against themselves. If the best bitscore for a transcript of a given species did not belong to the expected phylum (Porifera or Cnidaria), it was removed from its transcriptome. Note that contaminations between non-bilaterian phyla involving ctenophores or placozoans are much less likely, and were never observed during single gene tree inspection.

In addition to the removal of contaminations, our decontamination protocol implies the removal of other types of sequences as a side-effect. Indeed, highly-divergent sequences, taxonomically-restricted genes, low-quality transcripts as well as sequences resulting from horizontal gene transfers are also removed during this procedure (particularly at the DC1 step). These removals are harmless since they would have happened anyway during the following steps of supermatrix construction (i.e. insertion of putative orthologs in the core clusters will require a blast-hit against a Metazoa protein, see below). Detailed results of numbers of removed transcripts per species and at each step can be found in Table 3. These side-effect are illustrated by the large number of suppressed transcripts during the first suppression step (i.e. the suppression of non-Holozoan sequences) which ranged from 30,46% (*Cestum veneris*, this study) to 94.21% (*Craspedacusta sowerbyi*, SRR923472). Taking into account the three decontamination steps altogether, transcript suppression ranged from 33,73% (*Cestum veneris*, this study) to 97,23% (*Plakina jani*, this study), with a mean of 75,51%. Particularly high percentages of suppressed transcripts might in some cases reflect accelerated substitution rates (this can be suspected for instance for the hexactinellid sponge *Euplectella aspergillum*). In conclusion, the global side-effect of our particularly stringent DC step is loss of data. However, empirical observations of ribosomal single-gene trees showed that this protocol removed very few correct sequences while suppressing all contaminations. Therefore, we consider this data loss acceptable, given the gain in terms of quality of data.

1.3 Taxonomic enrichment of the core clusters

We completed the 4412 core clusters with the corresponding homologous sequences from the 80 species for which we had decontaminated the transcriptomes. In order to only add true homologous sequences and facilitate future treatment of paralogy in the clusters, this procedure includes a first preliminary assessment of orthology using a multiple Best Reciprocal Hit (multi-BRH) strategy against reference proteomes. The procedure allows importation of several sequences per species, to permit subsequent power detection and removal of paralogs. The procedure is summarised below, and in Figure 3 on the following page.

Forty 2 software For this purpose we designed a new software called “Forty 2” for updating a cluster with orthologous sequences from a new species, which works as follows:

1. Several cluster sequences (i.e. query sequences) belonging to specified query organisms are BLASTed against the new species transcriptome of which is extracted a list of **homologous transcripts**. Here, *Amphimedon*, *Monosiga*, *Lottia*, *Nematostella* and *Batrachochytrium* were used as query organisms.
2. The same query sequences are BLASTed against specified reference proteomes of which is extracted a **list of query best-hits**. Note that if a particular query sequence corresponded to several best-hits of which bitscores were nearly equal, all those hits were retained in the best-hit list. This feature is particularly important as such equal hits likely correspond to in-paralogs or to close out-paralogs. Retaining all those paralogs will later allow their detection and if necessary subsequent suppression. Here, we used proteomic data for *Amphimedon*, *Monosiga*, *Lottia*, *Homo*, *Nematostella* and *Batrachochytrium* as references.
3. The homologous transcripts of step 1 are then BLASTed against the same reference proteomes, resulting in the creation of a list of **homologous best-hits** (“3a” on Figure 3). If the best-hit is also found in the list of query best-hits from the previous step, then the homologous transcript is considered to be an ortholog (“3b” on Figure 3).
4. Orthologs are then BLASTed against the entire cluster in order to detect their closest relative sequences. Out of the five closest sequences, the one that aligned with the longest part of the ortholog was chosen as the most appropriate template to be used to align and incorporate the ortholog to the cluster.

For more detailed descriptions of this procedure, see the software’s manual. The program was conceived by Denis Baurain and Hervé Philippe, and was written by Denis Baurain.

During this step, different transcriptomic datasets corresponding to the same species have been combined into one Operational Taxonomical Unit (OTU). The 80 transcriptomic datasets therefore resulted in 75 different OTUs (for details, see Table 3). All sequences have been aligned using the *exonerate* aligner.

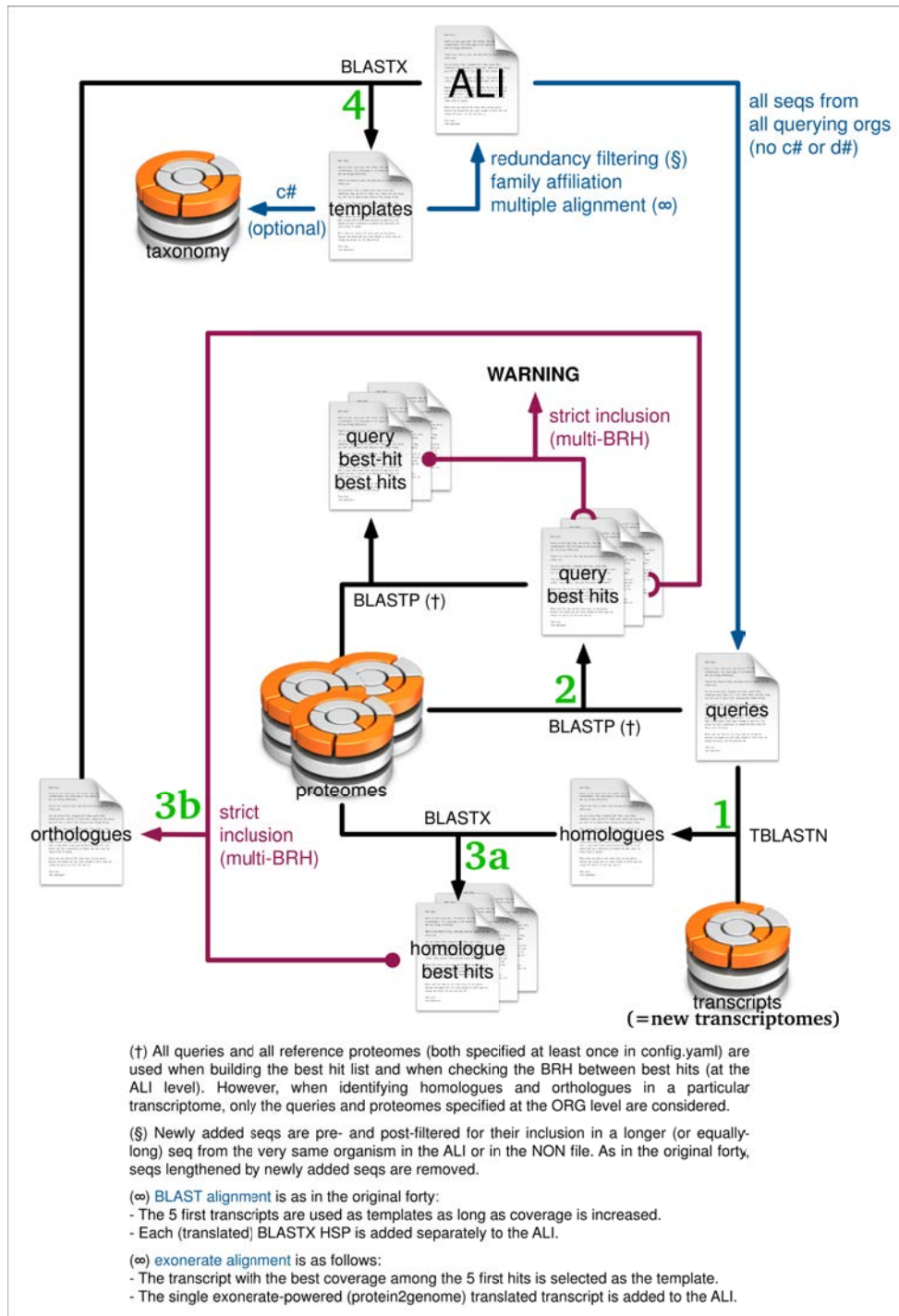


Figure 3: Summary of Forty 2 Software procedure for orthology assessment of sequences to be integrated into clusters. Contaminations of putative orthologs can optionally be screened based on taxonomic information.

1.4 Detection and removal of paralogs

Core clusters may contain paralogs for different species, especially when considering multigenic families. A treatment of these paralogies must differentiate duplications within a species (i.e. “in-paralogy”) and duplications within a group of species, creating different groups of orthology. In-paralogies are harmless when inferring large scale phylogenies, but earlier duplications must be detected, and only one of the resulting orthology group must be kept in the dataset. Otherwise, such ancient paralogs from different orthology groups would hamper correct reconstruction of phylogenetic relationships.

Pre-filtering of clusters Custom Perl scripts were used in order to detect and remove clusters which (after inclusion of all available transcriptomes) presented either a higher level of paralogy, or a lower taxonomical diversity, than thresholds based on various metrics. Each cluster was checked, and removed if: 1) the number of species represented was inferior to 40; 2) the number of species presenting several paralogs was superior to 14; and 3) the ratio “number of sequences / number of species” was superior to 1.6. These parameters have empirically been chosen to retain as much clusters as possible while removing those which would have presented too much difficulty for the next step of the protocol. This pre-filtering retained 1225 clusters.

Treatment of paralogs The existence of paralogs for a given gene can only be detected by analysing its phylogenetic tree in order to isolate different orthology groups. Although this can be done “by eye”, an automation of this analysis is required in order to handle 1225 alignments. The software PhyloTreePruner (Kocot et al. 2013) has been designed to detect in a given topology a sub-tree in which each species is represented by only one sequence, or by several monophyletic sequences (i.e. in-paralogies). To correctly detect and differentiate orthology groups, the phylogenetic tree must be correctly inferred, and as many species must be represented in as many orthology groups as possible. This strategy is possible thanks to the Forty 2 software property of adding several homologous sequences per species when relevant.

We therefore used the software PhyloTreePruner (Kocot et al. 2013) in order to remove all potential paralogs (or in other words, to select among each cluster a set of sequences representing only one and the same orthology group). However, the original algorithm was modified as follows: (i) the MAFFT alignment step was removed, since outputs from the Forty 2 software are already aligned clusters. (ii) The use of Gblocks to remove unconserved positions has been replaced by a custom Perl script which removes positions with an amount of missing data higher than 70% of the number of sequences. (iii) We suppressed the best-fitting evolution model detection step since during the next step, we systematically used the LGF evolution model in RAxML to reconstruct the phylogenetic tree (using the function “-f D”). Lastly, PhyloTreePruner was executed with the following parameters: minimal number of species = 10 and bootstrap cut-off = 0.5. This paralogy treatment resulted in 1156 pruned clusters (clusters with less than 10 species after PhyloTreePruner treatment were removed).

Note that contrary to what the PhyloTreePruner manual says, the software does not always detect and retrieve the “maximally inclusive sub-tree” free of out-paralogs, but often retrieves only a sub-optimal inclusive sub-tree free of out-paralogs. This is because the analysis of the bipartitions is made

on a rooted tree, instead of an unrooted one. However, this does not impact significantly our procedure since paralogs are still correctly removed, and the detection of sub-optimal sub-tree only results in some amount of data loss.

Insuring dataset independency In order to avoid redundancy and therefore to create truly independent matrices from the *MANU-128* dataset created by Philippe et al. (2009), we removed all clusters which corresponded to the *MANU-128* genes, by using custom scripts. All sequences from all clusters have been BLASTed again all *MANU-128* genes with an e-value threshold of 1e-10 and if the best-hit belonged to *MANU-128* sequences, the corresponding cluster was removed. This step resulted in the removal of 13 clusters, finally retaining 1143 pruned clusters with no overlap with the *MANU-128* dataset.

1.5 “*AUTO*” supermatrices assembly

The 1143 clusters (referred to as the *AUTO-1143* dataset) were then assembled into a supermatrix using ScaFos (see 4). In presence of multiple copies, ScaFos choses the least divergent sequence based on distances computed by IQPNNI. This approach was satisfactory here since at this stage, all remaining paralogs corresponded to in-paralogy cases. We also assembled the 4412 clusters as they were before treatment for paralogy into another supermatrix, also using ScaFos. It is to be noted that since the Treepuzzle version that was used cannot handle alignments with more than 256 sequences, only 4235 of these clusters were finally assembled into a supermatrix (referred to as the *AUTO-4235* dataset). Also, since ScaFos sorts out paralogs with only simple molecular distance estimates, this second supermatrix was not entirely free from ancient paralogs.

2. Update and assembly of manually-curated supermatrices

All newly sequenced, and assembled and cleaned transcriptomes as well as publicly available data (see Table 3) have been incorporated in the 128 alignments used in Philippe et al. 2009 (and referred to as the *MANU-128* dataset) with the protocol described in Baptiste et al. (2002). Single-gene tree inspection showed no trace of paralogy nor contamination. Since the alignments correspond to highly expressed genes, they were present in most of sequenced non-bilaterian transcriptomes and therefore presented a limited amount of missing data. These 128 alignments have been concatenated into a supermatrix using ScaFos from which ambiguously aligned positions were subsequently trimmed with Gblocks as in Philippe et al. (2009).

3. Phylogenetic analyses

Various sub-matrices (detailed in Table 4) have been created and analysed using various evolution models and softwares in order to study their impact on the resulting topology. These analyses are grouped in five categories of tests, each one focusing on a known phylogenomic bias that could potentially hinder the phylogenetic reconstruction from inferring metazoan relationships.

matrix name	Dataset	PTP	PB %	PIP %	Bin removal	length	nb taxa	% missing data
—	AUTO-1143	yes	—	—	—	457,095	95	75.3%
94_PB70	AUTO-1143	yes	70	—	—	151,067	94	55.7%
65_PB70	AUTO-1143	yes	70	—	—	151,067	65	38.5%
94_PB50	AUTO-1143	yes	50	—	—	34,509	94	44.2%
65_PB50	AUTO-1143	yes	50	—	—	34,509	65	24.3%
65_PB50_PIP10	AUTO-1143	yes	50	10	—	31,058	65	24.4%
65_PB50_PIP20	AUTO-1143	yes	50	20	—	27,608	65	24.4%
65_PB50_PIP30	AUTO-1143	yes	50	30	—	24,158	65	24.5%
65_PB50_PIP40	AUTO-1143	yes	50	40	—	20,708	65	24.5%
65_PB50_PIP50	AUTO-1143	yes	50	50	—	17,258	65	24.5%
65_PB50_Bin19	AUTO-1143	yes	50	—	Bin20	32093	65	24.6%
65_PB50_Bin18	AUTO-1143	yes	50	—	Bin20,19	22321	65	24.6%
65_PB50_Bin17	AUTO-1143	yes	50	—	Bin20,19,18	14216	65	24.3%
5_PB50_NOCOMP	AUTO-1143	yes	50	—	—	34,509	37	25.4%
—	AUTO-4235	no	—	—	—	5,621,026	95	89.2%
4235_65_PB60	AUTO-4235	no	60	—	—	379,855	65	34.7%
UPCB_116	MANU-128	—	—	—	—	21,736	116	12.1%
UPCB_55	MANU-128	—	—	—	—	20,994	55	4.9%
UPCB_55_PIP10	MANU-128	—	—	10	—	18,851	55	4.9%
UPCB_55_PIP20	MANU-128	—	—	20	—	16,755	55	4.9%
UPCB_55_PIP30	MANU-128	—	—	30	—	14,659	55	5.0%
UPCB_55_PIP40	MANU-128	—	—	40	—	12,563	55	4.9%
UPCB_55_PIP50	MANU-128	—	—	50	—	10,388	55	4.9%
UPCB_55_Bin19	MANU-128	—	—	—	Bin20	19,756	55	4.9%
UPCB_55_Bin18	MANU-128	—	—	—	Bin20,19	16,258	55	4.9%
UPCB_55_Bin17	MANU-128	—	—	—	Bin20,19,18	12,351	55	4.6%
UPCB_55_NOCOMP	MANU-128	—	—	—	—	20,994	34	7.1%
65_PB50+UPCB_116	AUTO-1143+MANU	—	—	—	—	56,245	65	16.2%

Table 4: Supermatrices used in this study, and their characteristics : set of clusters used, paralogy treatment (PTP = PhyloTreePruner), removal of positions by maximal percentage of missing data (PB %), removal of heteropecilleous positions (PIP % = percentage of the most heteropecilleous position removed), removal of fast-evolving positions (Bin, see text), number of positions, number of taxa and mean global amount of missing data. Matrices with no name have not been used for phylogenetic analyses and appear only to indicate their dimensions prior to any treatment.

Reducing the amount of missing data In order to study the impact of missing data and their distribution within the *AUTO* supermatrices (see Roure et al. 2013), as well as to remove parts of these matrices that were not correctly aligned or that corresponded to introns, we created sub-matrices in which we removed all positions that presented more than a given percentage of missing data (i.e. 70%, 60% or 50%, respectively coded as PB70, PB60 and PB50 in Tables 4 and 5). Since the missing data were mainly found in some poorly sequenced species for the *AUTO* datasets, we used two different species samplings: one broad, corresponding to the full taxonomic sampling (i.e. 95 species, but see the “*Crateromorpha* exception” below); and one reduced sampling in which the 30 less complete species have been discarded (i.e. 65 species, see the “OTUs” column in Table 3). Note that the species *Crateromorpha meyeri* (Porifera, Hexactinellida) has also been removed from the complete sampling due to its dramatic amount of missing data reaching 99%.

Progressive removal of heteropecilleous sites As previously shown by Roure and Philippe (2011), the substitution process can vary through time and therefore different taxa may present different substitution processes at a given position. Such positions are in clear violation of the site-specific CAT evolution model and would be deleterious for phylogenetic reconstruction. We therefore used the protocol described in Roure and Philippe (2011) to progressively remove the most heteropecilleous positions of two matrices (*AUTO-1143_65_PB50* and *MANU-128_UPCB-55*), similar in length and taxonomical sampling coverage (see “PIP %” in Table 4). This treatment could not be applied to all matrices because of computation time limit.

Progressive removal of fast-evolving sites Saturation of positions results in a loss of phylogenetic signal and increases the long-branch artifact. In order to optimise the quality/quantity ratio of the data and to assess whether the presence of fast-evolving sites could have an effect on the topology, we used the TIGER software (Cummins and McInerney 2011) to classify matrix positions into 20 categories (called “Bins”) corresponding to their relative evolution rate. We then created sub-matrices by progressively removing the most rapidly-evolving positions (they are coded as Bin19, Bin18 and Bin17, in Tables 4 and 5). This treatment could not be applied to all matrices because of computation time limit.

Accounting for compositional heterogeneity Shifts in amino-acid frequencies may occur during evolution of species. These shifts represent violations of the assumption by most evolution models that amino-acid frequencies are stationary. Recent phylogenomic analyses have showed that compositional heterogeneity can cause artificial phylogenetic groupings at many phylogenetic levels among animals (Song et al. 2010; Nesnidal et al. 2010, 2013) as well as in plants (Liu et al. 2014). We therefore tested the deviation from compositional homogeneity of our datasets using posterior predictions as implemented in Phylobayes. In order to reduce compositional heterogeneity between species, we removed the most deviating taxa (i.e. z-score > 4) from the matrices and re-analysed them. We made an exception for ctenophores in the *AUTO-128* dataset, in which they were all kept while being deviating, because otherwise ctenophores would not be present in this dataset. This exception is not supposed to severely impact the topology since compositional heterogeneity can only be a bias if other taxonomic groups are

also significantly deviating. This treatment resulted in suppression of 28 taxa from the *65_PB50* matrix, and 21 taxa from the *UPCB_55* matrix (see *65_PB50_NOCOMP* and *UPCB_55_NOCOMP* in Tables 4 and 5). It could not be applied to all matrices because of computation time limit.

Use of different reconstruction methods Reconstruction methods and models are of prime importance in a phylogenetic analysis. They must be able to account for the evolutionary processes that led to the molecular sequences we observe today (i.e. sufficiently complex), while their computational cost has to be kept reasonable. We used here four methods with different properties: 1) the distance method as implemented in BioNJ under a Poisson model, which performs better than traditional Neighbor-Joining method when substitution rates are heterogenous in the tree (Gascuel 1997); 2) the maximum likelihood method implemented in RAxML and using the LG matrix of amino-acid replacement which is more accurate than similar JTT and WAG matrices (Le and Gascuel 2008); 3) the site-heterogeneous CAT model implemented in Phylobayes which has been showed to perform better than other evolution models, notably with respect to multiple substitutions which can create LBA artifacts (Lartillot and Philippe 2004; Lartillot et al. 2007); 4) the more complex CAT-GTR model also implemented in PhyloBayes, which superimposes a GTR substitution matrix to the site-heterogeneous stationary frequencies.

Maximum likelihood analyses were conducted with RAxML under the LG+ Γ 4 evolution model, while for each Bayesian analysis we ran two independent MCMC using Phylobayes until satisfactory convergence was reached (i.e. meandiff < 0.02 and all deep nodes converged). Consensus trees were then obtained from the two chains after removal of roughly 50% of the trees as a burn-in. Computational time varied greatly depending on supermatrice sizes and complexity of the evolution model, and chains were generally stopped after 1,000 to 2,000 generations.

Results and discussion

An overview of the results from all analyses performed on both datasets (*AUTO* and *MANU*) is given in Table 5, with selected trees shown in Figs 4 to 9.

Which is the most early-diverging phylum within metazoans? Porifera were consistently found to be monophyletic, and were positioned as the first emerging phylum among metazoans in a majority of analyses (Table 5). This result was repeatedly supported with maximal support in rooted (i.e. with outgroups) analyses of the *AUTO-1143* and *AUTO-4235* datasets, with taxonomic sampling either complete or reduced and sizes of the supermatrices ranging from 379,855 to 14,216 positions (see Table 5 and Figures 4, 5, 6). This placement of Porifera was also supported by analyses in which variable quantities of supermatrices positions were removed in order to reduce the level of saturation (i.e. fast-evolving sites) or the level of CAT model violations (i.e. heteropecilly). Only two analyses of *AUTO* supermatrices found ctenophores placed as the sister-group to all other metazoans (Table 5). One of them was executed under the LG+ Γ evolution model, which has been proved to be less appropriate for such large-scale datasets, and prone to the Long Branch Attraction artifact (LBA, see Lartillot and Philippe

2004; Philippe et al. 2005). The other one used the complex CAT+GTR+ Γ evolution model, but the two chains were run for only 127 generations due to time constraint, which is clearly insufficient, and in the preliminary consensus, support for the “basal” position of ctenophores was not maximal (PP=0.93). Furthermore, chain monitoring showed that this support was progressively decreasing before the analysis was stopped (data not shown), suggesting that ctenophore position might have changed if the analysis had been continued. The *MANU-128* dataset also supports Porifera, and not Ctenophora, as the sister-group to other metazoans (see Table 5). Although several analyses based on this dataset yielded basal ctenophores (e.g. Figure 8), two experiments show that this is due to systematic biases. (i) The removal of 10% of the most heteropecilleous sites of the supermatrix weakened the support for “Ctenophora-first”, and the removal of 20% or more of these sites recovered the “Porifera-first” topology (see Table 5), (ii) the removal of fast-evolving sites (responsible for the saturation of the dataset) also resulted in the placement of Porifera as the first-emerging phylum among metazoans, even after a removal of only 1,238 of the fastest-evolving sites corresponding to the “Bin19” (see Table 5).

In conclusion, careful analyses of these two fully independent datasets under various conditions supports that Porifera is a monophyletic phylum and the sister-group to all other metazoans. This finding is in contradiction with several recent phylogenomic studies (Dunn et al. 2008; Hejnol et al. 2009; Ryan et al. 2013; Moroz et al. 2014) which instead obtained Ctenophora as the most early-diverging animal phylum. The automated procedure used here to build the large-scale *AUTO* dataset is similar to that employed in those previous studies but is more rigorous, especially with respect to treatment of paralogy and contaminations. Our conclusions are in line with previous studies which showed that the topology with ctenophores as the first lineage to diverge within metazoans, obtained in several phylogenomic studies, is in fact caused by methodological problems (Pick et al. 2010; Philippe et al. 2011; Philippe and Roure 2011; Nosenko et al. 2013) leading to exacerbation of LBA between ctenophores and non-metazoan outgroups.

Support for “Ctenophora-second” from *AUTO* datasets Both *AUTO* datasets (1,143 and 4,235 genes) repeatedly recovered with strong support ctenophores as the second-emerging metazoan phylum, after the divergence of Porifera (see Table 5 and Figures 4, 5, 6). This topology was resilient to many experiments: it was not altered by changes in the number of gene clusters used, matrix size, amount of missing data, taxon sampling, and removal of fast-evolving or heteropecilleous sites (see Table 5) (with three exceptions that will be discussed later on). In addition, the concatenation of *MANU* and *AUTO-1143* datasets (total evidence analysis) also supported this topology (although the two chains did not converge, due to demosponges being attracted within bilaterians in one of them, clearly an aberrant result).

This phylogenetic position of Ctenophora has already been recovered in two previous phylogenomic studies and has been considered in two different ways. First, Pick et al. (2010) interpreted this placement as due to LBA artefact still being present even after their corrections and improvements from a previous dataset (that of Dunn et al. 2008), substantiating this claim by the evaluated level of saturation of the corresponding dataset. Second, similar finding in the phylogenomic work of Ryan et al. (2013)

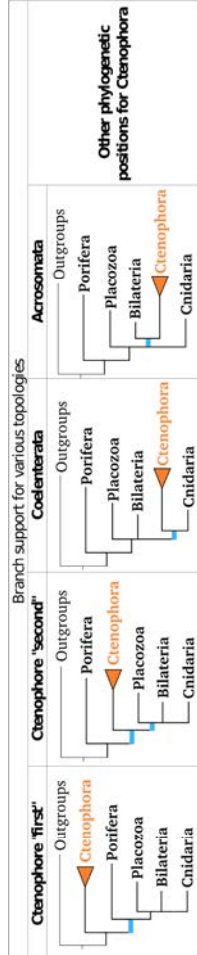
Dataset	Experiment	Matrix	length	% missing	Method	Outgroups	Branch support for various topologies			
							Ctenophore "first"	Ctenophore "second"	Coelenterata	Acrosomata
		65_PB50	34,509	24.3			0.99			
		94_PB70	34,509	44.2		yes	1			
	Variations of taxon sampling and % of missing data	65_PB70	151,067	38.5	CAT+Γ		1			
		94_PB70	151,067	55.7			1			
		65_PB50	34,509	24.3		no	I*			
		94_PB50	34,509	44.2			I*			
		65_PB70	151,067	38.5			I*			
		65_PB50_PIP10	31,058	24.4			1			
		65_PB50_PIP20	27,608	24.4			1			
		65_PB50_PIP30	24,158	24.5		yes	1			
		65_PB50_PIP40	20,708	24.5			1			
		65_PB50_PIP50	17,258	24.5			1			
	progressive removal of heteropithecous sites	65_PB50_PIP10	31,058	24.4	CAT+Γ		0.97*			
		65_PB50_PIP20	27,608	24.4			0.99*			
		65_PB50_PIP30	24,158	24.5		no	0.87*			
		65_PB50_PIP40	20,708	24.5			0.99*			
		65_PB50_PIP50	17,258	24.5			0.71*			
AUTO 1143		65_PB50_bin19	32,093	24.6			1			
		65_PB50_bin18	22,321	24.6		yes	0.99			
		65_PB50_bin17	14,216	24.3					0.91	
	progressive removal of fast evolving sites	65_PB50_bin19	32,093	24.6	CAT+Γ		I*			
		65_PB50_bin18	22,321	24.6		no	I*			
		65_PB50_bin17	14,216	24.3					0.83	
	removal of compositionally deviating taxa	65_PB50_NOCOMP	34,509	25.4	CAT+Γ	no				1
		34,509	24.3		CAT+GTR+Γ					
		34,509	24.3		LG+Γ	yes				
	Variations of evolution model	65_PB50	34,509	24.3	BiNJ					72
		34,509	24.3		LG+Γ	no				X*
		34,509	24.3		BiNJ					100*
AUTO 4235	larger supermatrix	4235_65_PB60	379,855	34.7	CAT+Γ	yes				1
AUTO 1143 + MANU	total evidence	65_PB50+URCS_116	56,245	16.2	CAT+Γ	yes				C1=1** C2=1**

* = equivalent topologies caused by the absence of outgroups
 ** = no convergence, demographies within bilaterians in one chain
 X = no bootstrap support has been calculated on the best topology

Table 5: Summary of metazoan phylogenetic relationships and support for various position of Ctenophora as reconstructed from various combinations of methods and datasets. For Bayesian analysis two independent Phylobayes chains have been run until convergence could be reached (see materials and methods). Support values correspond to blue branches in the upper tree (for the Ctenophora “second” hypothesis only the lowest value of the two is reported). Brown cases correspond to Bayesian analyses with very large matrix or using complex evolution model that were run for only less than 400 generations due to computational time constraints. Second part of the table, comprising results for the *MANU-128* dataset, is on the next page.

Dataset	Experiment	Matrix	length	% missing	Method	Outgroups	Branch support for various topologies			
							Ctenophore "first"	Ctenophore "second"	Coelenterata	Acrosomata
Variations of taxon sampling and evolution model	UPCB_116		21,736	12.1	CAT+I	0.9				
			21,736	12.1	LG+I	X				
	UPCB_55		21,736	12.1	BioNJ	100				
			20,994	4.9	CAT+GTR+I	0.64				
	UPCB_116		20,994	4.9	CAT+I	0.67				
			20,994	4.9	LG+I	X				
	UPCB_55		20,994	4.9	BioNJ	100				
			20,388	12.1	CAT+I	0.85*				
	UPCB_55		20,388	12.1	LG+I					
			20,388	12.1	BioNJ					
UPCB_55		19,072	4.9	CAT+I						
		19,072	4.9	LG+I						
UPCB_55_PIP10		19,072	4.9	BioNJ						
		18,851	4.9							
MANU 128	UPCB_55_PIP20		18,851	4.9						
			16,755	4.9						
	UPCB_55_PIP30		16,755	4.9						
			14,659	5.0						
	UPCB_55_PIP40		14,659	4.9						
			12,563	4.9						
	UPCB_55_PIP50		12,563	4.9						
			10,388	4.9						
	UPCB_55_PIP10		10,388	4.9						
			18,851	4.9						
UPCB_55_PIP20		18,851	4.9							
		16,755	4.9							
UPCB_55_PIP30		16,755	4.9							
		14,659	5.0							
UPCB_55_PIP40		14,659	4.9							
		12,563	4.9							
UPCB_55_PIP50		12,563	4.9							
		10,388	4.9							
UPCB_55_bin19		10,388	4.9							
		19,756	4.9							
UPCB_55_bin18		19,756	4.9							
		16,258	4.9							
UPCB_55_bin17		16,258	4.9							
		12,351	4.6							
UPCB_55_bin19		12,351	4.6							
		19,756	4.9							
UPCB_55_bin18		19,756	4.9							
		16,258	4.9							
UPCB_55_bin17		16,258	4.9							
		12,351	4.6							
UPCB_55_NOCOMP		12,351	4.6							
		20,994	7.1	CAT+I						

* = equivalent topologies caused by the absence of outgroups
 ** = no convergence, demosponges within bilaterians in one chain
 X = no bootstrap support has been calculated on the best topology



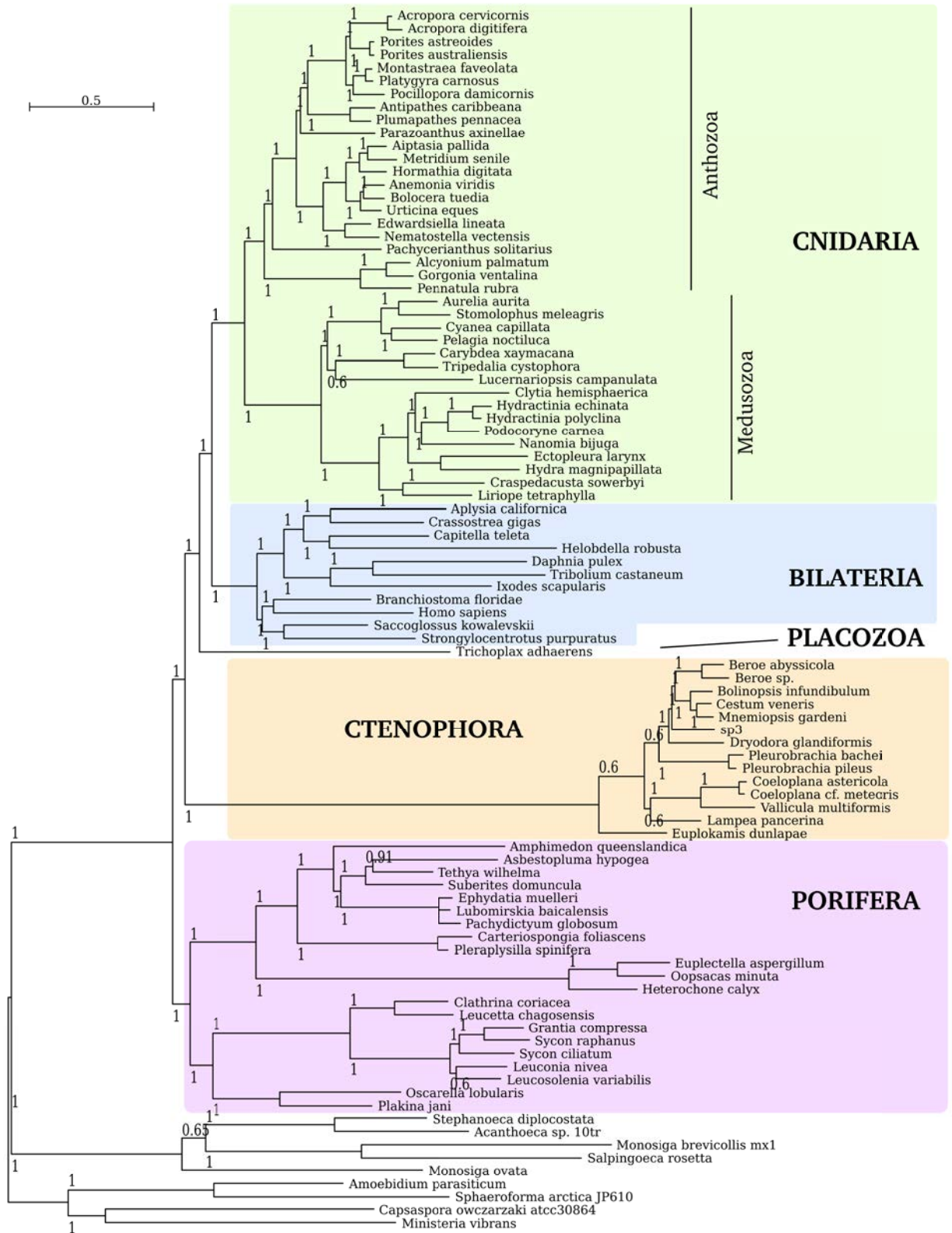


Figure 4: Phylogenetic relationships based on the *AUTO-1143_94-PB70* matrix and reconstructed with CAT+ Γ evolution model. Posterior probabilities of each branch are noted on the tree.

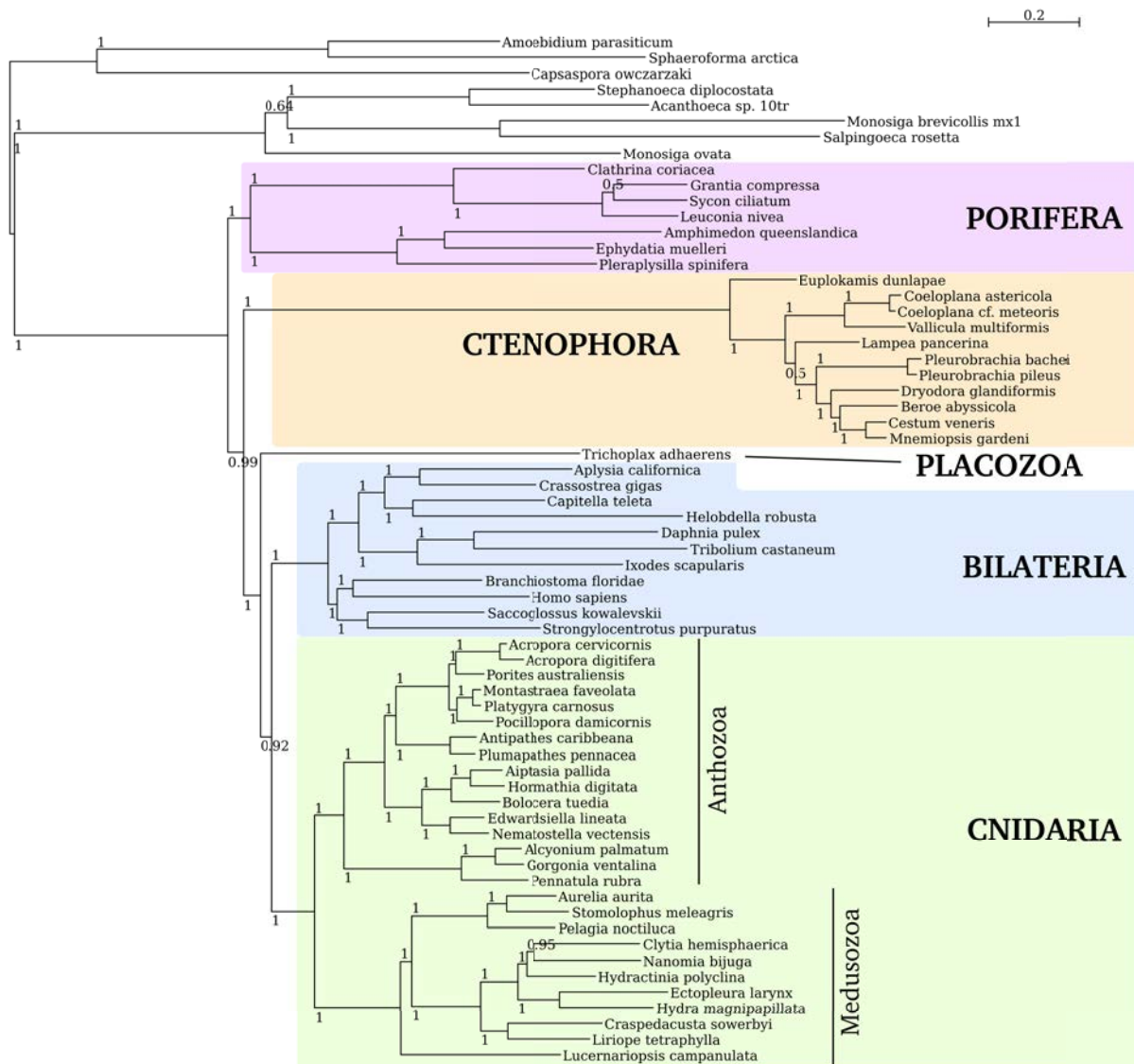


Figure 5: Phylogenetic relationships based on the *AUTO-1143_65_PB50* matrix and reconstructed with CAT+ Γ evolution model. Posterior probabilities of each branch are noted on the tree.

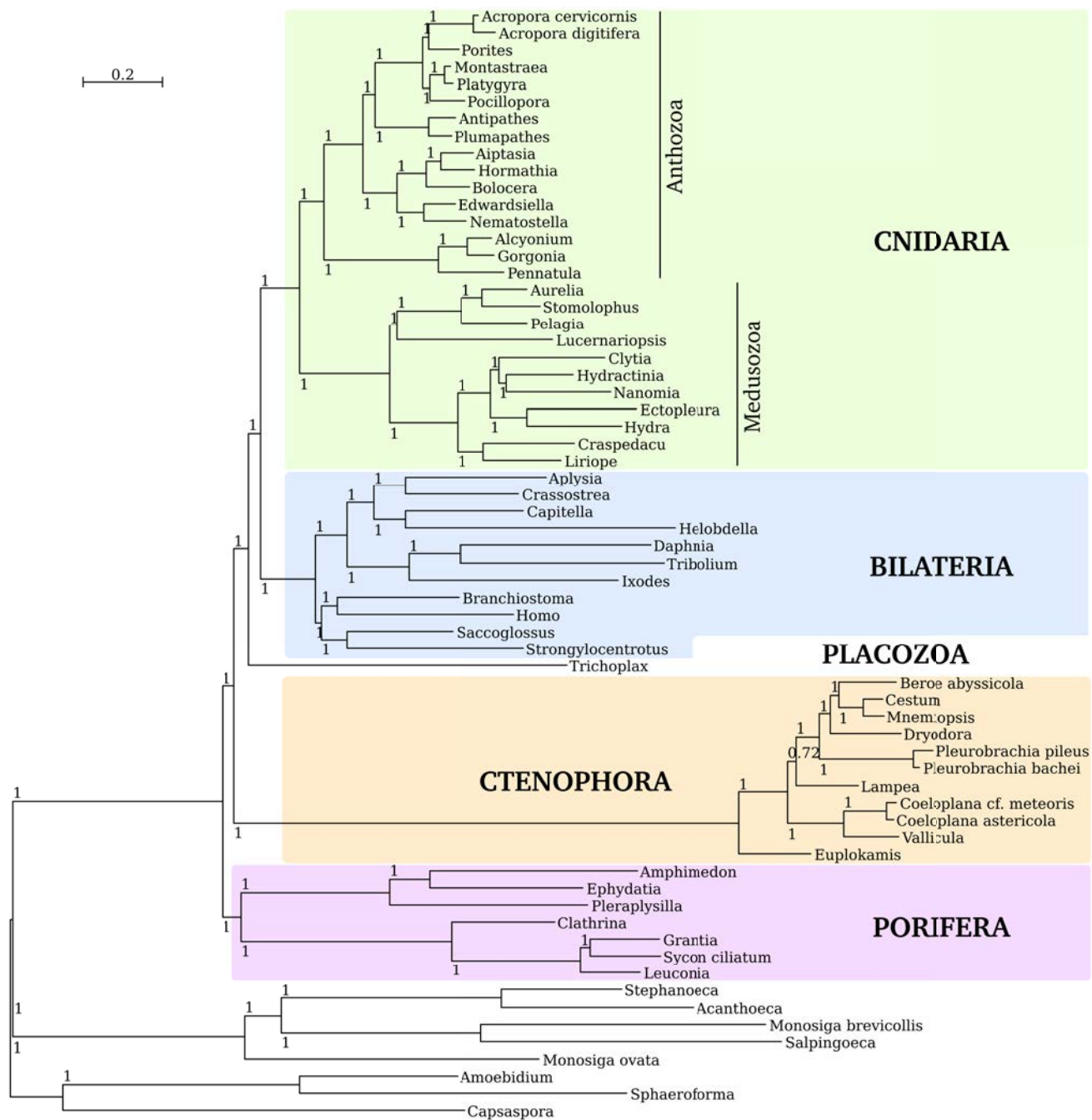


Figure 6: Phylogenetic relationships based on the *AUTO-4235_65_PB60* matrix and reconstructed with CAT+Γ evolution model. Posterior probabilities of each branch are noted on the tree.

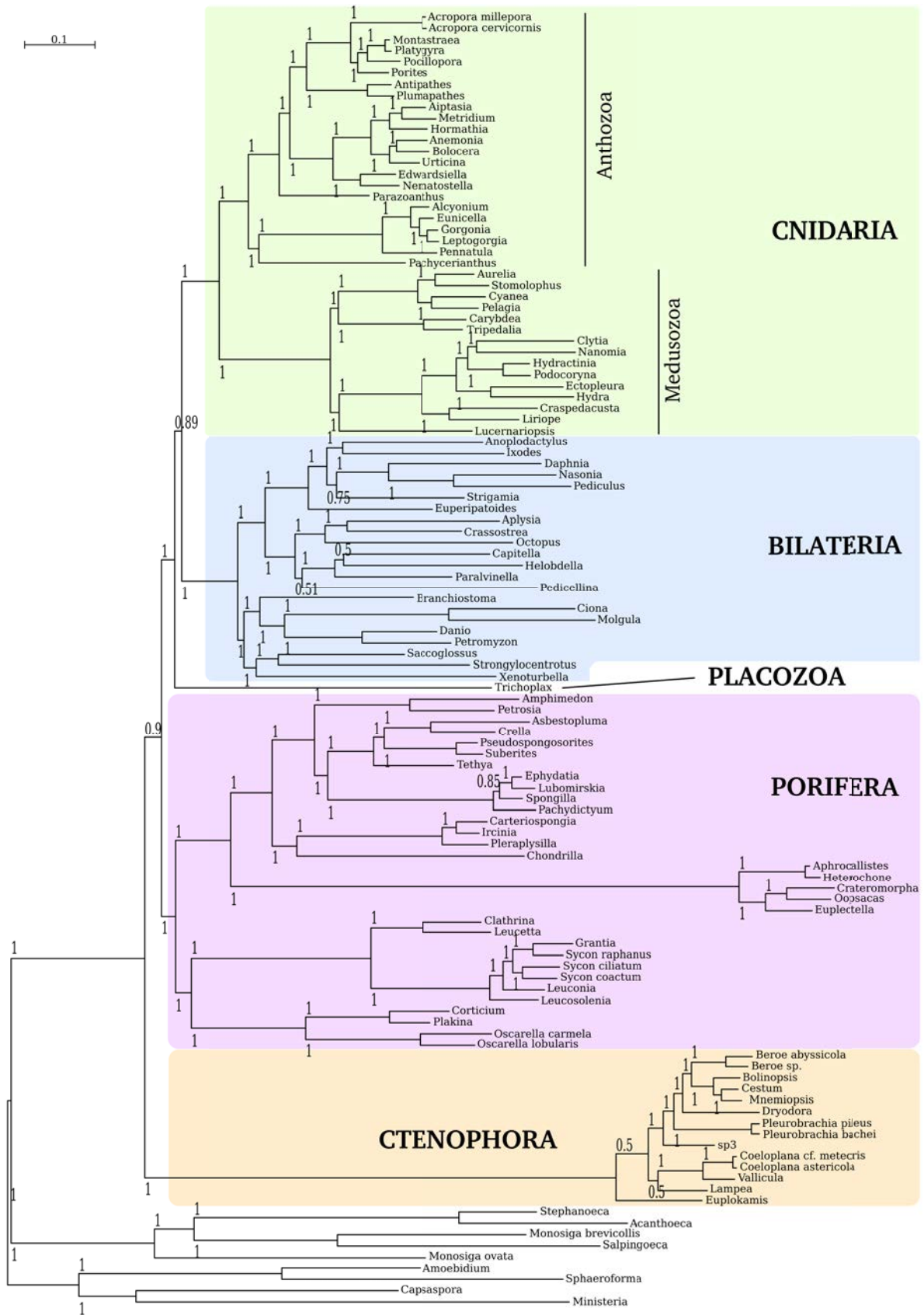


Figure 8: Phylogenetic relationships based on the *MANU-128_UPCB_116* matrix and reconstructed with CAT+Γ evolution model. Posterior probabilities of each branch are noted on the tree.

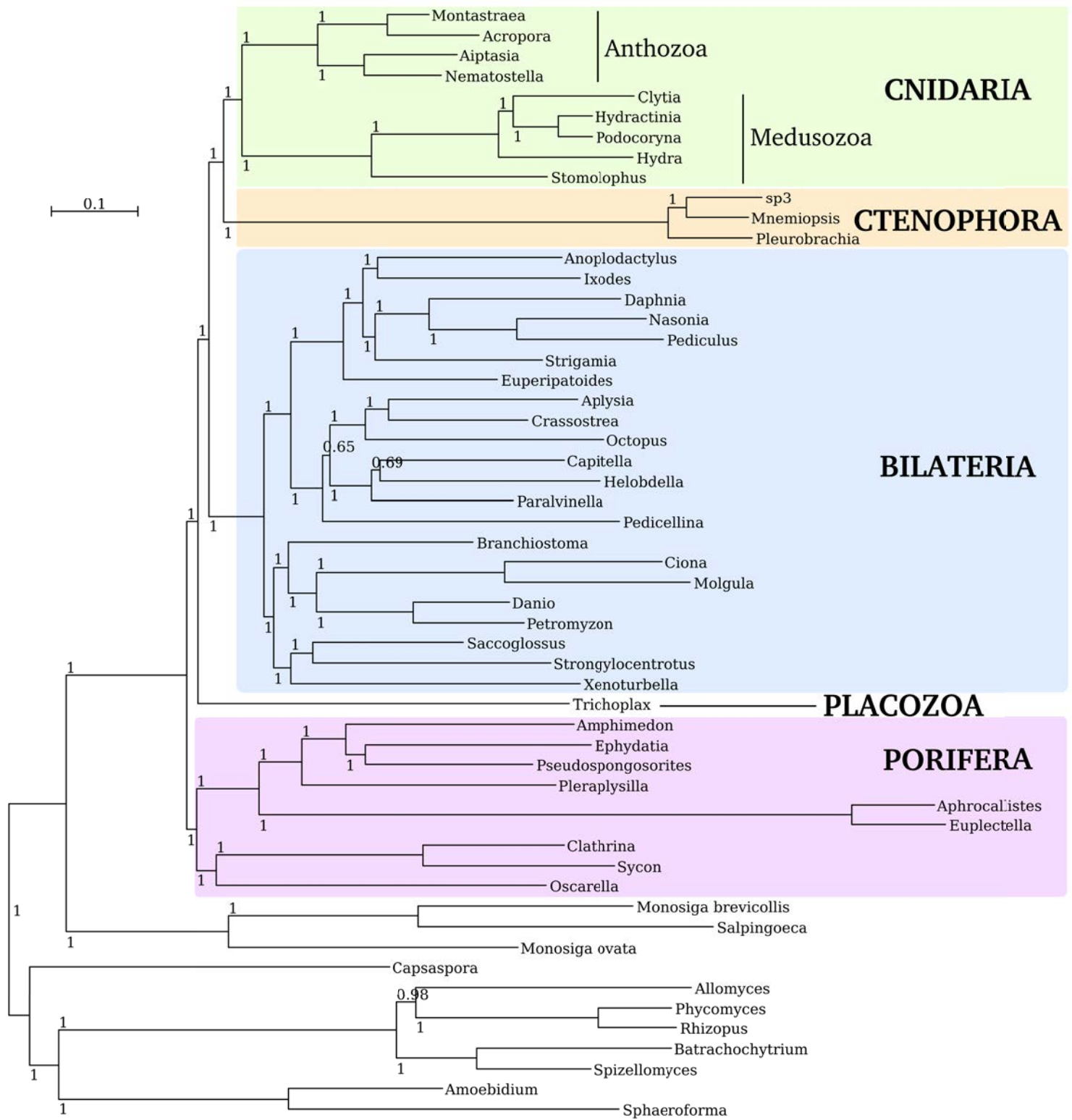


Figure 9: Phylogenetic relationships based on the *MANU-128_UPCB_55_PIP30* matrix and reconstructed with CAT+ Γ evolution model. Posterior probabilities of each branch are noted on the tree.

when using the CAT evolution model (see trees o, p, q, r in Fig. S1 of Ryan et al. 2013) was not really taken into account by the authors, who invoked failure of these analyses to converge. In light of these previously-reported results, and of our own analyses using *AUTO* supermatrices, we regard the “Ctenophora-second” hypothesis as a well-supported topology worth to be considered as a creditable scenario of basal metazoans relationships.

The “Coelenterata” hypothesis obtains limited support from the *MANU-128* dataset Analyses based on the smaller but more complete (in terms of both taxon sampling and matrix completeness) *MANU-128* dataset suggest a different view in which ctenophores and cnidarians are grouped in a Coelenterata clade (see Figure 9). As mentioned above, various experiments showed that “basal” ctenophores obtained in some of *MANU-128* analyses is probably the consequence of LBA. Three particular experiments (done with the *MANU-128* supermatrix with reduced taxon sampling), aiming at enhancing the phylogenetic signal otherwise “masked” by systematic biases, recovered monophyletic Coelenterata: (i) removal of outgroups; (ii) progressive removal of the most heteropecilleous sites (>30% of sites); (iii) progressive removal of the fastest-evolving sites (>41% of sites) (see Table 5). The removal of compositionally deviating taxa also resulted in the ctenophore placement as sister-group to cnidarians, but since all outgroups had to be removed in this experiment, this result is not independent with regard to the “outgroup removal” experiment.

Some support from the *MANU-128* dataset for the Coelenterata hypothesis was expected, since this gene set is the same as in Philippe et al. (2009) where Coelenterata were already retrieved. Since then, another phylogenomic study has recovered the same topology with a dataset strictly composed of ribosomal genes, which have a rather low level of saturation (Nosenko et al. 2013). Since a large proportion of the genes from *MANU-128* dataset encode ribosomal proteins, it is safe to assume that these genes are responsible for the signal grouping ctenophores with cnidarians. However, it is difficult to determine whether these genes recover monophyletic Coelenterata because their low level of saturation allows a better detection of the phylogenetic signal, or whether they collectively carry some sort of artifactual signal (of yet unknown nature) leading to attraction between ctenophores and cnidarians.

Intriguing occasional support for the “Acrosomata” hypothesis Three analyses of *AUTO* supermatrices under rather “extreme” conditions supported a third potential phylogenetic position of ctenophores, as sister-group to Bilateria (Table 5 and Figure 7). This corresponds to the “Acrosomata” hypothesis (Ax 1995). These three analyses are based on the *AUTO-1143* dataset and correspond to the largest amount of fast-evolving sites removal (59% of sites), with and without outgroups, and to the removal of the 28 taxa that presented the highest level of compositional heterogeneity when compared to the others (Table 5). Although these results might be interpreted as due to loss of phylogenetic information in such depleted supermatrices, the overall level of phylogenetic resolution obtained in these three analyses remains good. The two supermatrices resulting from the removal of 59% of the fastest-evolving sites removed still contain 14,216 amino-acid positions and recover perfectly coherent relationships at the intra-phyletic level (within sponges, cnidarians, ctenophores) and within bilaterians. The same holds true for the supermatrix in which 28 compositionally heterogeneous taxa were deleted. It still comprises

37 taxa representing ctenophores, bilaterians, cnidarians and demosponges, and reconstructed internal relationships within these taxa were satisfactory (Figure 7). The minimal numbers of taxa and positions needed to recover metazoan inter-phyletic relationships could well be fairly low if the quality of characters is high. The three analyses that recovered the Acrosomata hypothesis correspond to stringent treatments aimed at "cleaning" the alignments and expected to enhance the phylogenetic signal concentrated in a limited number of positions. In other words, in the original large-sized supermatrices this signal might be masked by systematic biases (e.g. LBA, or compositional heterogeneity), in which case the "ctenophore-second" topology would represent a compromise between LBA attraction of ctenophores towards the base and their true position sister-group to the Bilateria.

Interestingly, the grouping of ctenophores with bilaterians is in agreement with metazoan phylogenetic relationships reconstructed with matrices of morpho-anatomical characters (Schram 1991; Nielsen et al. 1996; Zrzavy et al. 1998; Peterson and Eernisse 2001; Brusca and Brusca 2003; Glenner et al. 2004). Possible synapomorphies for a ctenophore + bilaterian clade include the existence of an acrosome in spermatozooids (Ax 1995; Nielsen 2001), similarities between ctenophore sperm centrosomes and bilaterian cytasters (Salinas-Saavedra and Vargas 2011), presence of a mesodermal cell lineage individualised during gastrulation (Martindale and Henry 1997), and multinucleated muscle fibres whose differentiation involves fusion between mononucleate myoblasts (Dayraud et al. 2012).

Phylogenetic placement of Placozoa Several phylogenetic positions for placozoans (represented by the single species *Trichoplax adhaerens*) were recovered in this study. Most often they were placed as sister-group to a Bilateria + Cnidaria clade by analyses based on *AUTO* and *MANU-128* datasets, regardless of ctenophore position, as in Figures 4, 5, 6, 8, 9. However, two alternative positions were also recovered when using trimmed datasets. On one hand, the removal of fast-evolving sites and heteropecilleous sites on the *AUTO-1143* dataset resulted in a placement of placozoans as sister-group of cnidarians, whether outgroups were used or not (see Supplementary Figure 17 in Annexe A). This is reminiscent of old views interpreting *Trichoplax* as a modified planula larva, which were prevalent before the works of K.G. Grell (for details, see Grell 1981). On the other hand, the removal of fast-evolving sites on the *MANU-128* dataset placed placozoans closer to bilaterians (see Supplementary Figure 50 in Annexe A). It is difficult to determine which of these three positions is the most likely. For the time being, the position of Placozoa therefore remains unresolved.

Implications for evolution of key features in animals Dismissal of the "basal ctenophore" hypothesis weakens the grounds for independent evolutionary origins of the complex features they share with cnidarians and bilaterians, notably the presence of a nervous system (Moroz et al. 2014). However, in our analyses, animal lineages with nerve cells, muscle cells and a gastric cavity (Eumetazoa *sensu stricto*, i.e. Ctenophora, Cnidaria and Bilateria) were most often not monophyletic due to the position of placozoans as discussed above (see Table 5). Convergent origin of these characters therefore cannot be firmly ruled out. However their absence in placozoans seems easier to explain by multiple character losses in this lineage, a hypothesis that is furthermore consistent with recent interpretations of paleontological data regarding modern placozoans as secondarily simplified, with respect to their stem group

lineages of the Ediacarian period (Sperling and Vinther 2010).

Phylogenetic relationships within Porifera Porifera were recovered monophyletic in all analyses, based on both *AUTO* and *MANU-128* datasets. The support for this monophyly was usually maximal (i.e. PP=1), except for some analyses in which many fast-evolving or heteropecilleous sites were removed, in which support was slightly lower. Previous suggestions of sponge paraphyly based on various small-scaled molecular datasets (e.g. Borchiellini et al. 2001; Medina et al. 2001; Manuel et al. 2003; Sperling et al. 2007, 2009; Nielsen 2008) seem therefore definitely ruled out. The branching pattern among the four poriferan classes (namely, Demospongiae, Hexactinellida, Homoscleromorpha and Calcarea) was consistently the same across all analyses: one clade comprising Demospongiae and Hexactinellida (Silicea) was sister-group to a second clade comprising Homoscleromorpha and Calcarea (Figures 4, 8, 9). This topology is in agreement with previous studies (e.g. Dohrmann et al. 2008; Philippe et al. 2009, 2011; Pick et al. 2010; Nosenko et al. 2013; Ryan et al. 2013).

The topology obtained within demosponges from analyses with extended taxon sampling (94 taxa with the *AUTO-1143* supermatrix; 116 taxa with the *MANU-128* supermatrix, Figures 4 and 8) is robust and mostly consistent with previous molecular phylogenies (Borchiellini et al. 2004; Cárdenas et al. 2012; Hill et al. 2013). A clade of Keratosa (here represented by *Carteriospongia foliascens*, *Ircinia fasciculata* and *Pleraplysilla spinifera*) and Myxospongiae (here only represented by *Chondrilla nucula*) is sister-group to other demosponges. Within the latter, “marine haplosclerids” or Haploscleromorpha (Cárdenas et al. 2012) (*Amphimedon queenslandica*, *Petrosia ficiformis*) diverge first, followed by the freshwater spongillids, sister to a clade comprising the rest of demosponges (Democlavia, Sperling et al. 2009, here represented by species of the genera *Tethya*, *Suberites*, *Pseudospongopores*, *Crella* and *Asbestopluma*). This position of Spongillidae obtained from both datasets and well-supported contradicts a recent multiloci study that found them instead sister-group to Haploscleromorpha (Hill et al. 2013).

Phylogenetic relationships within Cnidaria Our analyses resolve most relationships within the phylum Cnidaria except the position of Staurozoa. Cnidarians were recovered monophyletic in all analyses with strong support (PP>0.99), and the same holds true for the two major subclades, Anthozoa and Medusozoa (see all Figures). In particular, monophyly of Anthozoa is unambiguously supported here, in strong contradiction with recent mitogenomic studies which suggested their paraphyly due to recovery of a sister-group relationship between Octocorallia and Medusozoa (Kayal and Lavrov 2008; Zou et al. 2012; Park et al. 2012; Kayal et al. 2013). It is likely that highly heterogeneous substitution rates in the mitogenomic datasets of these studies hampered reconstruction of deep cnidarian relationships.

Within Anthozoa, Octocorallia was always recovered monophyletic, as well as Hexacorallia *sensu stricto* (i.e. without cerianths). However, the phylogenetic position of Ceriantharia (for which a new transcriptome has been sequenced and assembled) is still undetermined since the two *AUTO* datasets repeatedly placed them as sister-group of all other Hexacorallia, whereas the analyses based on the *MANU-128* dataset placed them as the sister-group of Octocorallia. Within Hexacorallia, Scleractinia and Antipatharia formed a clade. Zoanthinaria was generally recovered as the sister-group to other

hexacorallians (to the exclusion of Ceriantharia) but was placed as sister-group to the Scleractinia + Antipatharia clade in one analysis (Figure 4).

Within Medusozoa, two groups were consistently recovered with maximal support: Hydrozoa on one side, and a clade grouping Scyphozoa and Cubozoa on the other side. This is also in contradiction with a recent mitogenomic study which recovered a clade grouping Scyphozoa and Hydrozoa (Kayal et al. 2013), but in agreement with some previous rRNA-based molecular phylogenies (e.g. Collins et al. 2006). However, the position of Staurozoa (represented in our dataset by *Lucernariopsis campanulata*) varied across analyses. The *AUTO* datasets usually recovered *Lucernariopsis* as sister-group to other medusozoans (as in Figures 5 and 7), but the two largest matrices analysed (namely, *AUTO-1143_94_PB70* and *AUTO-4235_65_PB60*) placed them closer to scyphozoans (and cubozoans when present) than to hydrozoans (Figure 4 and 6). By contrast, the *MANU-128* dataset supported *Lucernariopsis* as closely related to hydrozoans (Figure 8).

Conclusions Recent studies have documented various types of signal conflict in large phylogenomic datasets, such as LBA and conflict between fast- and slow-evolving sites (Telford et al. 2014; Struck et al. 2014; Straub et al. 2014; Sharma et al. 2014; Meiklejohn et al. 2014), incongruence between genes (Leigh et al. 2008; Salichos and Rokas 2013) or influence of compositional bias (Jeffroy et al. 2006; Nesnidal et al. 2010, 2013). In the present work, numerous phylogenomic analyses based on two independent large supermatrices yielded more than four different positions for ctenophores, in addition to some other topological discrepancies at lower taxonomical level, therefore representing another illustration of the diversity of signals nested in large concatenations of genes. Simply adding more genes is not enough to reach answers to the most difficult phylogenetic problems (Philippe et al. 2011). Our study is another illustration of the high sensitivity of large supermatrices to multiple contradictory signals.

Nevertheless, the analyses reported here bring an important contribution to understanding early animal evolution by supporting that sponges and not ctenophores are sister-group to all other metazoans, and by resolving most of the high-level relationships within Porifera, Cnidaria and Ctenophora. Finally, we expect that the new procedure and tools devised here to build a high-quality large-scale phylogenomic dataset will prove useful to a large community of researchers.

Acknowledgments We would like to thank Alexander Ereskovsky and Pascal Lapébie who provided us with sponge and cnidarian samples. Many thanks to Nori Satoh who kindly gave us access to the *Euplectella* transcriptome and to Frédéric Delsuc for his help with transcriptome assembly. We thank Philippe Lopez, Eric Baptiste and the members of their research team for the access to their machines and for their help with Perl scripting. We thank UPMC biological stations of Banyuls-sur-Mer and Villefranche-sur-Mer, as well as Olivier Gros (UAG, Guadeloupe) for providing lab facilities for field sampling.

Author contributions MM and HP initially conceived the project. PS, MM and HP collectively designed the protocol of *AUTO* dataset construction. MJ, MM and PS participated in sample collection and MJ and PS performed all RNA preparations. PS wrote the DCC, DC1, DC2 and DC3 scripts as well

as many other custom scripts, assembled some of the transcriptomes, performed most steps of *AUTO* dataset assembly, built the various sub-matrices, conceived the analysis strategy, performed about half of the phylogenetic analyses, made all tables and figures and wrote a draft of the manuscript. HP updated the *MANU-128* dataset, designed validation method for DC steps, did preliminary single gene analyses, and performed about half of the phylogenetic analyses. DB assembled some of the transcriptomes and wrote the Forty 2 program. PS, MM, HP and DB contributed to the final manuscript text.

References

- Ax, P. (1995). *Das System der Metazoa*, volume Vol. 1. Gustav Fischer, Stuttgart.
- Baptiste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., Duruffé, L., Gaasterland, T., Lopez, P., and Müller, M. (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proceedings of the National Academy of Sciences*, 99(3):1414–1419.
- Borchiellini, C., Chombard, C., Manuel, M., Alivon, E., Vacelet, J., and Boury-Esnault, N. (2004). Molecular phylogeny of demospongiae: implications for classification and scenarios of character evolution. *Mol Phylogenet Evol*, 32(3):823–837.
- Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J., and Le Parco, Y. (2001). Sponge paraphyly and the origin of Metazoa. *Journal of Evolutionary Biology*, 14(1):171–179.
- Brusca, R. and Brusca, G. (2003). *Invertebrates. 2nd Edition*. Sinauer, Sunderland, Massachusetts.
- Cárdenas, P., Pérez, T., and Boury-Esnault, N. (2012). Sponge systematics facing new challenges. *Adv Mar Biol*, 61:79–209.
- Chevreur, B. (2005). *MIRA: An Automated Genome and EST Assembler*. PhD thesis, Ruprecht-Karls-University.
- Collins, A., Schuchert, P., Marques, A., Jankowski, T., Medina, M., and Schierwater, B. (2006). Medusozoan phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. *Systematic Biology*, 55(1):97–115.
- Cummins, C. A. and McInerney, J. O. (2011). A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Systematic Biology*, 60(6):833–844.
- Dayraud, C., Alié, A., Jager, M., Chang, P., Le Guyader, H., Manuel, M., and Quéinnec, E. (2012). Independent specialisation of myosin II paralogues in muscle vs. non-muscle functions during early animal evolution: a ctenophore perspective. *BMC Evolutionary Biology*, 12(1):107.
- Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5):361–375.
- Dohrmann, M., Janussen, D., Reitner, J., Collins, A. G., and Worheide, G. (2008). Phylogeny and evolution of glass sponges (porifera, hexactinellida). *Syst Biol*, 57(3):388–405.
- Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sørensen, M. V., Haddock, S. H. D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M., Wheeler, W. C., Martindale, M. Q., and Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452(7188):745–749.

- Ebersberger, I., Strauss, S., and von Haeseler, A. (2009). HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evolutionary Biology*, 9(1):157.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than blast. *Bioinformatics*, 26(19):2460–2461.
- Edgecombe, G. D., Giribet, G., Dunn, C. W., Hejnol, A., Kristensen, R. M., Neves, R. C., Rouse, G. W., Worsaae, K., and Sørensen, M. V. (2011). Higher-level metazoan relationships: recent progress and remaining questions. *Organisms Diversity & Evolution*, 11(2):151–172.
- Evans, N. M., Holder, M. T., Barbeitos, M. S., Okamura, B., and Cartwright, P. (2010). The phylogenetic position of myxozoa: Exploring conflicting signals in phylogenomic and ribosomal data sets. *Molecular Biology and Evolution*, 27(12):2733–2746.
- Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Biology*, 27(4):401–410.
- Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, 14(7):685–695.
- Glenner, H., Hansen, A. J., Sørensen, M. V., Ronquist, F., Huelsenbeck, J. P., and Willerslev, E. (2004). Bayesian inference of the metazoan phylogeny: a combined molecular and morphological approach. *Current Biology*, 14(18):1644–1649.
- Grell, K. (1981). *Trichoplax adherens* and the origin of Metazoa. *Atti dei convegni Lincei (Roma)*, 49:107–121.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N., and Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8):1494–1512.
- Harbison, G. R. (1985). On the classification and evolution of the ctenophora. In *The Origins and Relationships of lower Invertebrates*, pages 78–100. Oxford University Press, Oxford.
- Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., Martinez, P., Baguna, J., Bailly, X., Jondelius, U., Wiens, M., Muller, W. E. G., Seaver, E., Wheeler, W. C., Martindale, M. Q., Giribet, G., and Dunn, C. W. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings of the Royal Society B: Biological Sciences*, 276(1677):4261–4270.
- Hernandez-Nicaise, M. and Franc, J. (1994). Embranchement des cténaïres. In *Traité de Zoologie*, volume Tome III, pages 943–1075. Masson et Cie, Paris, p.p. grassé edition.

- Hill, M. S., Hill, A. L., Lopez, J., Peterson, K. J., Pomponi, S., Diaz, M. C., Thacker, R. W., Adamska, M., Boury-Esnault, N., Cárdenas, P., Chaves-Fonnegra, A., Danka, E., Laine, B.-O. D., Formica, D., Hajdu, E., Lobo-Hajdu, G., Klontz, S., Morrow, C. C., Patel, J., Picton, B., Pisani, D., Pohlmann, D., Redmond, N. E., Reed, J., Richey, S., Riesgo, A., Rubin, E., Russell, Z., Rützler, K., Sperling, E. A., di Stefano, M., Tarver, J. E., and Collins, A. G. (2013). Reconstruction of family-level phylogenetic relationships within demospongiae (porifera) using nuclear encoded housekeeping genes. *PLoS One*, 8(1):e50437.
- Huang, X. and Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. *Genome Research*, 9(9):868–877.
- Hyman, L. (1940). *The Invertebrates*, volume I. Protozoa through Ctenophora, chapter VIII The Radiate Phyla – Phylum Ctenophora, pages 662–695. McGraw Hill.
- Jager, M., Chiori, R., Alié, A., Dayraud, C., Quéinnec, E., and Manuel, M. (2011). New insights on ctenophore neural anatomy: Immunofluorescence study in *Pleurobrachia pileus* (müller, 1776). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 316B(3):171–187.
- Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends in Genetics*, 22(4):225–231.
- Kayal, E. and Lavrov, D. V. (2008). The mitochondrial genome of *Hydra oligactis* (cnidaria, hydrozoa) sheds new light on animal mtDNA evolution and cnidarian phylogeny. *Gene*, 410(1):177–186.
- Kayal, E., Roure, B., Philippe, H., Collins, A. G., and Lavrov, D. V. (2013). Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evolutionary Biology*, 13(1):5.
- Kircher, M., Heyn, P., and Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*, 12.
- Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the illumina platform. *Nucleic Acids Research*, 40(1):e3.
- Kocot, K. M., Citarella, M. R., Moroz, L. L., and Halanych, K. M. (2013). PhyloTreePruner: A Phylogenetic Tree-Based Approach for Selection of Orthologous Sequences for Phylogenomics. *Evolutionary Bioinformatics*, 9:429–435.
- Koski, L. B. and Golding, G. B. (2001). The closest blast hit is often not the nearest neighbor. *Journal of Molecular Evolution*, 52(6):540–542.
- Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evolutionary Biology*, 7 Suppl 1:S4.
- Lartillot, N. and Philippe, H. (2004). CAT: un modèle phylogénétique bayésien permettant de prendre en compte l’Hétérogénéité des processus de substitution entre sites dans les alignements protéiques. *Biosystema*, 22:97–104.

- Laurin-Lemay, S., Brinkmann, H., and Philippe, H. (2012). Origin of land plants revisited in the light of sequence contamination and missing data. *Current Biology*, 22(15):R593–R594.
- Le, S. Q. and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7):1307–1320.
- Leigh, J. W., Susko, E., Baumgartner, M., and Roger, A. J. (2008). Testing congruence in phylogenomic analysis. *Systematic Biology*, 57(1):104–115.
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). Orthomcl: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13(9):2178–2189.
- Liu, Y., Cox, C. J., Wang, W., and Goffinet, B. (2014). Mitochondrial phylogenomics of early land plants: Mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Systematic Biology*.
- Manuel, M., Borchiellini, C., Alivon, E., Le Parco, Y., Vacelet, J., and Boury-Esnault, N. (2003). Phylogeny and Evolution of Calcareous Sponges: Monophyly of Calcinea and Calcaronea, High Level of Morphological Homoplasy, and the Primitive Nature of Axial Symmetry. *Systematic Biology*, 52(3):311–333.
- Marlow, H. and Arendt, D. (2014). Evolution: ctenophore genomes and the origin of neurons. *Curr Biol*, 24(16):R757–R761.
- Martindale, M. Q. and Henry, J. Q. (1997). *Embryology: constructing the organism*, chapter Ctenophorans, the comb jellies, pages 87–111. Sinauer, Sunderland, MA.
- Medina, M., Collins, A. G., Silberman, J. D., and Sogin, M. L. (2001). Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proceedings of the National Academy of Sciences*, 98(17):9707–9712.
- Meiklejohn, K. A., Danielson, M. J., Faircloth, B. C., Glenn, T. C., Braun, E. L., and Kimball, R. T. (2014). Incongruence among different mitochondrial regions: A case study using complete mitogenomes. *Molecular Phylogenetics and Evolution*, 78(0):314 – 323.
- Moroz, L. L., Kocot, K. M., Citarella, M. R., Dosung, S., Norekian, T. P., Povolotskaya, I. S., Grigorenko, A. P., Dailey, C., Berezikov, E., Buckley, K. M., Ptitsyn, A., Reshetov, D., Mukherjee, K., Moroz, T. P., Bobkova, Y., Yu, F., Kapitonov, V. V., Jurka, J., Bobkov, Y. V., Swore, J. J., Girardo, D. O., Fodor, A., Gusev, F., Sanford, R., Bruders, R., Kittler, E., Mills, C. E., Rast, J. P., Derelle, R., Solovyev, V. V., Kondrashov, F. A., Swalla, B. J., Sweedler, J. V., Rogaev, E. I., Halanych, K. M., and Kohn, A. B. (2014). The ctenophore genome and the evolutionary origins of neural systems. *Nature*, 510(7503):109–114.

- Nesnidal, M. P., Helmkampf, M., Bruchhaus, I., El-Matbouli, M., and Hausdorf, B. (2013). Agent of whirling disease meets orphan worm: Phylogenomic analyses firmly place myxozoa in Cnidaria. *PLoS ONE*, 8(1):e54576.
- Nesnidal, M. P., Helmkampf, M., Bruchhaus, I., and Hausdorf, B. (2010). Compositional Heterogeneity and Phylogenomic Inference of Metazoan Relationships. *Molecular Biology and Evolution*, 27(9):2095–2104.
- Nielsen, C. (2001). *Animal evolution : interrelationships of the living phylums. 2nd Edition*. Oxford University Press, New York.
- Nielsen, C. (2008). Six major steps in animal evolution: are we derived sponges larvae? *Evol. Dev.*, 10:241–257.
- Nielsen, C., Scharff, N., and Eibye-Jacobsen, D. (1996). Cladistic analyses of the animal kingdom. *Biological Journal of the Linnean Society*, 57(4):385–410.
- Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., Müller, W. E., Nickel, M., Schierwater, B., Vacelet, J., Wiens, M., and Wörheide, G. (2013). Deep meta-zoan phylogeny: When different genes tell different stories. *Molecular Phylogenetics and Evolution*, 67(1):223–233.
- Pang, K. and Martindale, M. Q. (2008). Ctenophores. *Current Biology*, 18(24):R1119–R1120.
- Park, E., Hwang, D.-S., Lee, J.-S., Song, J.-I., Seo, T.-K., and Won, Y.-J. (2012). Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Molecular Phylogenetics and Evolution*, 62(1):329–345.
- Peterson, K. J. and Eernisse, D. J. (2001). Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. *Evolution & Development*, 3(3):170–205.
- Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., and Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS biology*, 9(3):e1000602.
- Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D. J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G., and Manuel, M. (2009). Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, 19(8):706–712.
- Philippe, H. and Roure, B. (2011). Difficult phylogenetic questions: more data, maybe; better methods, certainly. *BMC biology*, 9(1):91.
- Philippe, H. and Telford, M. J. (2006). Large-scale sequencing and the new animal phylogeny. *Trends in Ecology & Evolution*, 21(11):614–620.

- Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., and Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. *BMC evolutionary biology*, 5(1):50.
- Pick, K. S., Philippe, H., Schreiber, F., Erpenbeck, D., Jackson, D. J., Wrede, P., Wiens, M., Alie, A., Morgenstern, B., Manuel, M., and Worheide, G. (2010). Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Molecular Biology and Evolution*, 27(9):1983–1987.
- Roure, B., Baurain, D., and Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, 30(1):197–214.
- Roure, B. and Philippe, H. (2011). Site-specific time heterogeneity of the substitution process and its impact on phylogenetic inference. *BMC Evolutionary Biology*, 11(1):17.
- Ryan, J. F., Pang, K., Schnitzler, C. E., Nguyen, A.-D., Moreland, R. T., Simmons, D. K., Koch, B. J., Francis, W. R., Havlak, P., NISC Comparative Sequencing Program, Smith, S. A., Putnam, N. H., Haddock, S. H. D., Dunn, C. W., Wolfsberg, T. G., Mullikin, J. C., Martindale, M. Q., and Baxevanis, A. D. (2013). The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science*, 342(6164):1242592–1242592.
- Salichos, L. and Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, 497(7449):327–331.
- Salinas-Saavedra, M. and Vargas, A. O. (2011). Cortical cytasters: a highly conserved developmental trait of Bilateria with similarities to Ctenophora. *Evodevo*, 2:23.
- Schram, F. (1991). *The early evolution of Metazoa and the significance of problematic taxa*, chapter Cladistic analysis of metazoan phyla and the placement of fossil problematica, pages 35–46. Cambridge Univ. Press, WY.
- Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C., and Giribet, G. (2014). Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution*.
- Song, H., Sheffield, N. C., Cameron, S. L., Miller, K. B., and Whiting, M. F. (2010). When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. *Systematic Entomology*, 35(3):429–448.
- Sperling, E., Peterson, K., and Pisani, D. (2009). Phylogenetic-signal dissection of nuclear housekeeping genes supports the paraphyly of sponges and the monophyly of eumetazoa. *Mol. Biol. Evol.*, 26(10):2261–2274.
- Sperling, E., Pisani, D., and Peterson, K. (2007). *The rise and fall of the Ediacaran biota*, chapter Poriferan paraphyly and its implications for Precambrian palaeobiology, pages 355–368. Geological Society, London, special publications 286.

- Sperling, E. A. and Vinther, J. (2010). A placozoan affinity for *Dickinsonia* and the evolution of late proterozoic metazoan feeding modes. *Evolution & Development*, 12(2):201–209.
- Straub, S. C., Moore, M. J., Soltis, P. S., Soltis, D. E., Liston, A., and Livshultz, T. (2014). Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Molecular Phylogenetics and Evolution*, 80(0):169 – 185.
- Struck, T. H. (2013). The impact of paralogy on phylogenomic studies : A case study on annelid relationships. *PLoS ONE*, 8(5):e62892.
- Struck, T. H., Wey-Fabrizius, A. R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., Kück, P., Herlyn, H., and Hankeln, T. (2014). Platyzoan paraphyly based on phylogenomic data supports a non-coelomate ancestry of Spiralia. *Molecular Biology and Evolution*.
- Telford, M. J. (2013). Evolution. the animal tree of life. *Science*, 339(6121):764–766.
- Telford, M. J. and Copley, R. R. (2011). Improving animal phylogenies with genomic data. *Trends in Genetics*, 27(5):186–195.
- Telford, M. J., Lowe, C. J., Cameron, C. B., Ortega-Martinez, O., Aronowicz, J., Oliveri, P., and Copley, R. R. (2014). Phylogenomic analysis of echinoderm class relationships supports asterozoa. *Proceedings of Biological Science*, 281(1786).
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.-W., Li, Y., Xu, X., Wong, G. K.-S., and Wang, J. (2014). Soapdenovo-trans: de novo transcriptome assembly with short rna-seq reads. *Bioinformatics*, 30(12):1660–1666.
- Zhang, J., Ruhlman, T., Mower, J., and Jansen, R. (2013). Comparative analyses of two geraniaceae transcriptomes using next-generation sequencing. *BMC Plant Biology*, 13(1):228.
- Zou, H., Zhang, J., Li, W., Wu, S., and Wang, G. (2012). Mitochondrial genome of the freshwater jellyfish *Craspedacusta sowerbyi* and phylogenetics of medusozoa. *PLoS ONE*, 7(12):e51465.
- Zrzavy, J., Mihulka, S., Kepka, P., Bezdek, A., and Tietz, D. (1998). Phylogeny of the metazoa based on morphological and 18S ribosomal DNA evidence. *Cladistics*, 14(3):249–285.

Chapitre 3

Reconstruction des relations phylogénétiques au sein des cténophores

3.1 Introduction

Les relations phylogénétiques entre les différents groupes de cténophores sont encore en grande partie méconnues, car il existe trop peu de caractères anatomiques informatifs. Les cténophores sont en effet morphologiquement très disparates à l'échelle inter-ordinale. Certains organes sont conservés à l'échelle des cténophores, comme les rangées de peignes, l'organe aboral ou bien une paire de tentacules, mais ils présentent assez peu de variabilité. Les autres organes sont, quant à eux, majoritairement incomparables entre cténophores : les Lobata possèdent des lobes et des auricules, mais aucun autre cténophore n'en possède. Les Beroida ne possèdent pas de tentacules, tandis que tous les autres cténophores en ont.¹ Ces caractères sont des autapomorphies des différents ordres et familles de cténophores, et n'apportent pas d'information sur les relations de parenté entre ces groupes.

Bien que certains caractères pourraient se révéler tout à fait intéressants, une autre facette du problème est que leurs états n'ont pas été décrits dans la littérature pour tous les taxons. Il est très facile de comprendre qu'un auteur ne puisse pas décrire toutes les structures d'un organisme, surtout lorsqu'il a peu de spécimens à sa disposition et que ces derniers sont transparents et très fragiles. De plus, comme la conservation en collections des cténophores est extrêmement difficile (peu de cténophores résistent aux fixateurs chimiques), il est impossible de ré-observer ces structures anatomiques, d'en vérifier les descriptions ou encore d'en préciser les observations. Ce problème majeur aboutit lui aussi à un manque de caractères informatifs pour comprendre les relations de parenté entre les cténophores.

Néanmoins, une analyse cladistique sur la base d'une matrice de caractères morfo-anatomiques

1. En réalité, une espèce du genre *Ocyropsis* (appartenant à l'ordre des Lobata) ne possède pas non plus de tentacules, y compris pendant son développement.

a été réalisée au sein de notre équipe de recherche. Elle sera présentée et discutée dans la section des compléments du présent chapitre.

Difficulté de l'établissement d'un jeu de données moléculaires Dans ce contexte de carence en caractères anatomiques, l'utilisation de données moléculaires semble être la seule solution pour résoudre les relations phylogénétiques entre les cténophores. Une importante limitation est alors l'échantillonnage taxonomique, nécessitant la récolte et l'identification préalable des espèces qui seront séquencées. Sans contexte taxonomique solide chez les cténophores (voir article 3), l'identification de certaines espèces peut être très difficile, voir impossible dans certains cas lorsque le statut de certaines espèces est indéterminé (c'est le cas de *Beroe ovata*, voir discussion dans Mills et al. 1996). L'échantillonnage des espèces, quant à lui, est rendu difficile à la fois par la fragilité extrême de nombreux cténophores, par leur mode de vie pélagique et par leur apparition plus ou moins fugace dans le plancton. Enfin, de nombreuses espèces décrites par le passé n'ont pas été ré-observées depuis et doivent être soit rares, soit très localisées, soit disparues. La constitution d'un jeu de données moléculaires représentant toute la diversité des cténophores est donc un objectif difficile à atteindre.

Une première avancée dans la phylogénie des cténophores Une seule étude a jusqu'à présent tenté de résoudre les relations phylogénétiques entre cténophores grâce à des données moléculaires (Podar et al. 2001). En utilisant un alignement du gène ribosomique ADNr 18S échantillonné chez 10 familles de cténophores (sur les 27 actuellement reconnues), les auteurs ont pu conclure à la paraphylie de l'ordre des Cydippida, en montrant l'existence de deux nouveaux clades : les Beroida + Haeckeliidae et les Platyctenida + Mertensiidae. Le reste de la topologie ne résout que les relations déjà connues grâce à la morphologie des cténophores, c'est-à-dire la monophylie des Platyctenida, des Beroida ou encore des Pleurobrachiidae. Autre résultat intéressant, leur étude montre des écarts de taux de substitution très importants entre certains groupes : les Platyctenida évoluent très vite, tandis que les Lobata, Thalassocalycida et Cestida accumulent les substitutions très lentement. Enfin, leurs analyses ne permettent pas de déterminer la position de la racine des cténophores.

Tentative d'amélioration du jeu de données et de la méthodologie Les résultats obtenus par Podar et al. (2001) sont malgré tout encourageants et suggèrent que l'utilisation de données moléculaires permettra la résolution des relations phylogénétiques entre les cténophores, tout comme elle a permis l'amélioration de notre compréhension de l'évolution de très nombreux autres groupes d'organismes. La stratégie la plus simple est donc de continuer le travail commencé par ces auteurs, en améliorant plusieurs aspects : 1) l'échantillonnage taxonomique des cténophores doit être augmenté le plus possible, jusqu'à une hypothétique complétude ; 2) les précédents auteurs ayant séquencé d'autres marqueurs moléculaires (ç.-à-d. les *internal transcribed spacers* -ITS- 1 et

2 ainsi que le gène ribosomique ADNr 5.8S) mais sans les utiliser, une amélioration de l'alignement des séquences est souhaitable pour profiter de ces caractères moléculaires supplémentaires²; 3) les méthodes de reconstruction phylogénétique, et notamment les modèles d'évolution utilisés, ont été largement améliorées depuis 2001. Ainsi, une méthodologie plus avancée pourrait permettre d'améliorer la résolution des relations entre cténophores.

C'est dans cet esprit que nous avons mené l'étude phylogénétique qui fera l'objet de la prochaine section (article 2). Dans la section complémentaire qui clôture ce chapitre, je présenterai une analyse cladistique des caractères morpho-anatomiques, puis les indications tirées des analyses phylogénomiques du chapitre 2 en ce qui concerne les relations intra-cténophores, et enfin quelques analyses phylogénétiques supplémentaires sur des gènes dupliqués chez les cténophores afin de tenter de préciser la position de leur racine. Enfin, je tenterai de faire une synthèse de l'état actuel de nos connaissances sur l'évolution de l'embranchement des Ctenophora, en discutant la distribution de certains caractères morphologiques sur un arbre bilan issu des différentes analyses moléculaires.

3.2 Article 2 : Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora.

Simion, P. ; Bekkouche, N. ; Jager, M. ; Quéinnec, E. & Manuel, M. *Zoology*, 2014, e-pub ahead of print, DOI : 10.1016/j.zool.2014.06.004

2. Dans leur section Matériels et Méthodes, Podar et al. (2001) précisent qu'ils n'ont pas réussi à aligner correctement les séquences des ITS1 et 2, et qu'ils n'ont donc pas utilisé ces données.



Contents lists available at ScienceDirect

Zoology

journal homepage: www.elsevier.com/locate/zool

ZOOLOGY

Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora

Paul Simion^a, Nicolas Bekkouche^{a,b}, Muriel Jager^a, Eric Quéinnec^a, Michaël Manuel^{a,*}

^a Institut de Biologie Paris-Seine (IBPS), Université Pierre et Marie Curie, UMR 7138 UPMC CNRS, 7 quai St Bernard, F-75005 Paris, France

^b Department of Biology, Marine Biological Section, University of Copenhagen, Strandpromenaden 5, DK-3000 Helsingør, Denmark

ARTICLE INFO

Article history:

Received 13 December 2013

Received in revised form 13 June 2014

Accepted 17 June 2014

Available online xxx

Keywords:

Ctenophora

18S rRNA

ITS

Doublet model

ABSTRACT

Ctenophores are a phylum of non-bilaterian marine (mostly planktonic) animals, characterised by several unique synapomorphies (e.g., comb rows, apical organ). Relationships between and within the nine recognised ctenophore orders are far from understood, notably due to a paucity of phylogenetically informative anatomical characters. Previous attempts to address ctenophore phylogeny using molecular data (18S rRNA) led to poorly resolved trees but demonstrated the paraphyly of the order Cydippida. Here we compiled an updated 18S rRNA data set, notably including a few newly sequenced species representing previously unsampled families (Lampeidae, Euryhamphaeidae), and we constructed an additional more rapidly evolving ITS1 + 5.8S rRNA + ITS2 alignment. These data sets were analysed separately and in combination under a probabilistic framework, using different methods (maximum likelihood, Bayesian inference) and models (e.g., doublet model to accommodate secondary structure; data partitioning). An important lesson from our exploration of these datasets is that the fast-evolving internal transcribed spacer (ITS) regions are useful markers for reconstructing high-level relationships within ctenophores. Our results confirm the paraphyly of the order Cydippida (and thus a “cydippid-like” ctenophore common ancestor) and suggest that the family Mertensiidae could be the sister group of all other ctenophores. The family Lampeidae (also part of the former “Cydippida”) is probably the sister group of the order Platyctenida (benthic ctenophores). The order Beroida might not be monophyletic, due to the position of *Beroe abyssicola* outside of a clade grouping the other *Beroe* species and members of the “Cydippida” family Haeckeliidae. Many relationships (e.g. between Pleurobrachiidae, Beroida, Cestida, Lobata, Thalassocalycida) remain unresolved. Future progress in understanding ctenophore phylogeny will come from the use of additional rapidly evolving markers and improvement of taxonomic sampling.

© 2014 Elsevier GmbH. All rights reserved.

1. Introduction

Ctenophora is a metazoan phylum of great interest for many fields of biology. Exclusively marine, they play an important role mostly in pelagic ecosystems as planktonic predators. Ctenophores are the focus of intense research in marine ecology notably addressing “blooming” phenomena and the invasive potential of some species (Lucic et al., 2011; Costello et al., 2012; Purcell, 2012). Our understanding of the morphological diversity and anatomical complexity of ctenophores has been renewed by several recent studies (Jager et al., 2008, 2011; Pang and Martindale, 2008a,b; Yamada et al., 2010; Alié et al., 2011; Pang et al., 2011; Dayraud et al., 2012; Borisenko and Ereskovsky, 2013).

Most ctenophores share several easily recognisable characters: (i) eight comb rows; (ii) biradial symmetry; (iii) a pair of tentacles covered with colloblasts (adhesive cells); (iv) an aboral sensory organ associated with a pair of polar fields; (v) a distinctive unilateral cleavage during early embryogenesis. However, these features can sometimes be missing or highly modified as in adult Coeloplanidae without comb rows, Beroida without tentacles at any life stage, or Haeckeliidae with pseudocolloblasts instead of colloblasts (Carré and Carré, 1989). Ctenophores possess nerve cells and muscle cells, the latter being partly derived from a mesodermal lineage individualised during gastrulation (Martindale and Henry, 1999). Since these two cell types are the classical landmarks of Eumetazoa, their comparison with counterparts in cnidarians and bilaterians is of great interest to understanding early metazoan evolution (Pang and Martindale, 2008a,b; Philippe et al., 2009; Telford and Copley, 2011; Ryan et al., 2013).

Ctenophores are currently classified into 9 orders, 27 families, and roughly 150 species (Mills, 2012). Four of these orders contain

* Corresponding author. Tel.: +33 144273469.
E-mail address: michael.manuel@snv.jussieu.fr (M. Manuel).

only one or two species and have rarely been observed (Cambiojida; Ganeshida; Cryptolobiferida and Thalassocalocida), whereas the five other orders (“Cydippida”, a paraphyletic assemblage – see below; Platyctenida; Lobata; Cestida; Beroida) are more diversified and comprise many widely distributed species.

The classical morphology-based taxonomy of ctenophores (Harbison, 1985; Hernandez-Nicaise and Franc, 1994) heavily relies on the branching structure of the gastrovascular system, composed of two parts: an axial portion parallel to the ctenophore oral–aboral axis (mouth, pharynx, infundibulum, infundibular canal, anal canals and anal pores) and various peripheral components supplying nutrients to organs such as comb rows or the tentacle roots (perradial, interradial, adradial, meridional, tentacular and paragastric canals). The nature of these gastrovascular canals and their branching pattern greatly varies between ctenophore orders and families; furthermore, some canals can also bear diverticulae, which, in some cases, create anastomoses with other canals. Other characters traditionally used in ctenophore systematics include the size and position of organs (e.g., relative size of comb rows, position of tentacle roots, body shape, absence/presence of tentillae on the tentacle; Harbison, 1985).

No consensus has yet been reached concerning the phylogenetic position of Ctenophora within Metazoa. This hampers the reconstruction of the early evolution of metazoan key body plan characters (Dohrmann and Wörheide, 2013; Maxmen, 2013; Pennisi, 2013). Since the second half of the 19th century, the traditional morphology-based classification has united ctenophores and cnidarians under the Coelenterata, based on the presence in both taxa of a gelatinous mesoglea, tentacles, a branching gastrovascular system, statocysts, tetramerous symmetry, and a general lack of organ systems (Hyman, 1940). An alternative view related ctenophores to polyclad flatworms, based on the flattened anatomy of benthic ctenophores (i.e., order Platyctenida; Lang, 1884; Mortensen, 1912). Both views are based on highly debatable hypotheses of homology (Harbison, 1985). For example, the statocysts present in some cnidarian medusae have been acquired by multiple convergence events (Collins et al., 2006) and are therefore clearly not homologous with the ctenophore statocyst (apical organ); likewise, cnidarian and ctenophore tentacles are not homologous (Harbison, 1985), and this probably also holds true for the tetradial symmetry (Manuel, 2009).

Molecular phylogenies based on ribosomal RNA have most often placed ctenophores as a sister group to cnidarians + bilaterians (Cavalier-Smith et al., 1996; Abouheif et al., 1998; Collins, 1998; Zrzavy et al., 1998; Borchiellini et al., 2001; Peterson and Eernisse, 2001; Glenner et al., 2004; Halanych, 2004). Some recent studies using supermatrices of nuclear proteins from high-throughput sequencing data (phylogenomics; Philippe and Telford, 2006) have suggested a “basal position” of Ctenophora as sister group to all other metazoans (Dunn et al., 2008; Hejnol et al., 2009; Ryan et al., 2013), implying either two independent acquisitions of nervous systems and muscular cells or their loss in sponges and placozoans. However, this result is contentious, as ctenophores suffer from a strongly accelerated rate of molecular substitutions, and therefore their “basal” placement is likely to be due to long branch attraction (LBA) (Philippe et al., 2011; Dohrmann and Wörheide, 2013), a source of systematic error which tends to be exacerbated in large-scale datasets. Other phylogenomic studies came up with trees more in line with morphological evidence, suggesting notably Eumetazoan monophyly (Ctenophora + Cnidaria + Bilateria), and ctenophores either grouped with Cnidaria in a Coelenterata clade (Schierwater et al., 2009; Philippe et al., 2009, 2011; Nosenko et al., 2013), or falling in a polytomy with cnidarians and bilaterians (Pick et al., 2010). Since ctenophores are represented by only 2–4 species in these molecular datasets, their unresolved

position might be due to insufficient sampling of ctenophore diversity in addition to LBA and other potential methodological problems (Philippe et al., 2011; Dohrmann and Wörheide, 2013).

Similarly, the intraphyletic relationships between ctenophores, which represent the focus of the present study, are currently poorly understood. So far, anatomical characters have been of limited value for phylogenetic inference, notably due to strong morphological disparity (i.e., character discontinuity between taxa) particularly at the interordinal level, and also due to the lack of detailed anatomical knowledge for many taxa. Many characters are autapomorphies of orders, such as the presence of lobes and auricles in Lobata, or the extreme compression in the tentacular axis in Cestida. Other features are strongly homoplastic, such as the presence of diverticulae and anastomoses in the gastrovascular system (Harbison, 1985). For example, in *Ocyropsis maculata* (Rang, 1828), a Lobata species, subpharyngeal meridional canals show diverticulae, whereas they are absent in its sister species *Ocyropsis crystallina* (Rang, 1828) (Harbison, 1985). In the order Beroida, all possible levels of connections between meridional canals are observed, from independent meridional canals without diverticulae in *Beroe gracilis* Künne, 1939, to diverticulae without anastomoses as in *Beroe cucumis* Fabricius, 1780, and completely anastomosed meridional canals in *Beroe ovata* Bruguière, 1789 and *Beroe forskalii* Milne Edwards, 1841. This situation led Harbison (1985) to write that “... the diverticulae are simply means of achieving an improved circulation rather than indicators of a phylogenetic relationship”. From a methodological point of view, classifications to date based on anatomical data either did not use a phylogenetic systematics framework (Ospovat, 1985) or relied on that theoretical framework but without performing an explicit cladistic analysis of morphological characters (Harbison, 1985).

Only one molecular phylogeny study entirely dedicated to ctenophore internal relationships has been published so far and this work was based on 18S rRNA sequences (Podar et al., 2001). This study analysed 27 ctenophore sequences using maximum likelihood and maximum parsimony methods. As important results, the authors demonstrated the non-monophyly of the order Cydippida. Indeed, two cydippid families formed well-supported clades with other orders: Mertensidae with Platyctenida and Haeckeliidae with Beroida. Unfortunately, the position of the root of the ctenophore tree could not be determined and most nodes throughout the tree were unsupported. These uncertainties certainly resulted partly from a lack of phylogenetic information in the dataset (i.e., very low level of molecular diversity between the ctenophore 18S rRNA sequences), but might also be due to the limits of reconstruction methods and models used at that time, and to insufficient taxon sampling.

In the present study, we re-investigated the phylogenetic relationships of ctenophores using two molecular datasets (18S rRNA, and an alignment comprising internal transcribed spacer 1 (ITS1), 5.8S rRNA and internal transcribed spacer 2 (ITS2)). The species sampling covers 14 families, including the first sequenced members of Lampeidae (*Lampea pancarina*), Euramphaeidae (*Deiopea kaloktenota*) and Lampoctenidae (*Lampocteis cruentiventer*). This data set was explored through different combinations of taxon sampling, reconstruction methods and evolutionary models. Along with a confirmation of some previous findings (Podar et al., 2001) such as non-monophyly of the order Cydippida, the analyses provide evidence for the paraphyly of additional orders and families and furthermore support several interesting novel clades. We discuss potential strategies which future studies might follow to overcome the remaining lack of resolution for many of the interordinal ctenophore relationships and the persisting problem of the rooting of the ctenophore tree.

2. Materials and methods

2.1. Production of new sequence data

In the present study we determined the complete 18S rRNA, internal spacer 1 (ITS1), 5.8S rRNA and internal spacer 2 (ITS2) sequences for 9 ctenophore species sampled at various locations (Table 1). Collections and identifications were made by the authors based on various identification keys and descriptions (Trégouboff and Rose, 1957; Hernandez-Nicaise and Franc, 1994; Matsumoto, 1999; Mills, 1987, 2012; Oliveira et al., 2007). Samples were frozen in liquid nitrogen and kept at -80°C prior to DNA extraction.

DNA extraction of entire individuals (devoid of any visible parts of other animals or parasites), or portions of about 200 mg for very large specimens, was performed using a modified CTAB (hexadecyltrimethylammonium bromide) protocol (as described in Doyle and Doyle, 1987), without RNase A treatment. The dried pellet was resuspended in Tris buffer (1 M, pH 8). Universal eukaryotic primers used for polymerase chain reaction (PCR) amplification of 18S rDNA were 18SA (5' AACCTGGTTCATCTGCCAGT 3') and 18SB (5' TGATCCTTCYGCAGGTTACCTAC 3') (Medlin et al., 1988). 40 cycles were performed and the annealing temperature was set at 57°C . The ITS region (ITS1 + 5.8S rRNA + ITS2) was amplified using primers anchored towards the 3' end of 18S rRNA (1400F = 5' TGYACACACCGCCCGTC 3') and the 5' end of 28S rRNA (28Sr = 5' CTTAAGTTCAGCGGTAGTCTCG 3') (Podar et al., 2001). The annealing temperature was set at 55°C . Expected sizes of 18S rRNA and ITS region fragments were 1800 bp and 850 bp, respectively. Single PCR bands were obtained in all cases. PCR products were purified using the QIAquick gel extraction kit (Qiagen, Venlo, Netherlands) and sent for sequencing at GATC Biotech (Konstanz, Germany). The resulting chromatograms were checked by eye using Chromas (Technelysium, South Brisbane, Australia). New sequences generated in this study are available in GenBank under the accession numbers KJ754153–KJ754170.

2.2. Construction of alignments

A subset of 73 aligned 18S rRNA sequences (see Table S1 in the supplementary online Appendix A) as well as the corresponding information about secondary structure was extracted from the existing alignment of Mallatt et al. (2010) using a Perl script (available from the corresponding author on demand). Stem-loop secondary structure in this study strictly follows the designations of Mallatt et al. (2010). Additional 18S rRNA sequences from 7 Porifera species available in GenBank were added by hand to the alignment (using other sponge sequences as references for the alignment) in order to improve the outgroup sampling (Table S2 in the supplementary online Appendix A). The newly generated ctenophore sequences, as well as all ctenophore sequences available from GenBank at the date of the search (March 2014; see Table 1) were also added by hand to the alignment. Ctenophore sequences originally present in the alignment of Mallatt et al. (2010) were removed to avoid redundancy (in their study, Mallatt et al. used the sequences from Podar et al., 2001). Ambiguously aligned sites were then removed using G-blocks (block parameters as follows: 65–65–5–2–all; Castresana, 2000), but all sites corresponding to a stem in the secondary structure as annotated in the Mallatt et al. (2010) alignment were kept, as well as all sites that were correctly aligned among ctenophore sequences. The resulting 1883 base-pair long alignment contains 47 ctenophore sequences as well as sequences from 76 other species representing the taxonomic diversity of Holozoa, and is referred to as the 18S dataset.

Newly sequenced ctenophore ITS regions were aligned together with ctenophore ITS sequences available from GenBank (Table 1) and many of them originally from Podar et al. (2001). When

several sequences from the same species were identical, we only selected the longest one to represent the species. The alignment, which encompasses ITS1 and ITS2 (around 250 nucleotides each) as well as 5.8S rDNA (157 nucleotides) in between, was automatically generated using ClustalW (Larkin et al., 2007) with the following parameters: BLOSUM matrix; gap open cost = 2; gap extension cost = 0.05; gap distance = 1; no end gaps. Both ITS1 and ITS2 sequences were surprisingly more easily alignable among ctenophores than stated in Podar et al. (2001). A few explicit mistakes in the ClustalW alignment were corrected by eye, and ambiguously aligned sites were removed using G-blocks (block parameters: 19–19–6–2–half). The resulting alignment is referred to as the ITS dataset. The alignments (18S and ITS datasets) before and after treatment by G-blocks are available from the Dryad Digital Repository (<http://doi.org/10.5061/dryad.n63q1>).

The concatenation of these two datasets is naturally referred to as the 18S-ITS dataset. A specific problem arises when 18S rRNA and ITS sequences assigned to a given species have been produced by different laboratories and thus from different specimens and localities. We deliberately adopted a cautious approach to species-level assignments of the sequences, and never concatenated 18S and ITS sequences coming from different laboratories/localities even when assigned to the same species. In the particular cases of the genus *Thalassocalyce* and for the species *Beroe abyssicola* and *Bolinopsis infundibulum* (no complete pair of 18S + ITS sequences generated by the same laboratory), this led us to keep only the 18S sequence in the combined dataset. Otherwise, the genus *Thalassocalyce* and the species *B. abyssicola* would be each represented in the combined dataset by two non-overlapping (18S and ITS) sequences, a potential cause of topological artefacts.

2.3. Outgroup samplings

For the 18S and 18S-ITS analyses, we used three datasets of non-ctenophore taxa by progressively removing outgroup sequences without changing the alignment: (i) extended outgroup of 76 species representing Fungi, Choanoflagellata, the four classes of Porifera, Cnidaria, Placozoa and Bilateria (see details of species used for each group in Tables S1 and S2 in the supplementary online Appendix A); (ii) a reduced dataset comprising cnidarians only, considering that some recent phylogenomic studies have suggested a possible sister group relationship of cnidarians and ctenophores (Philippe et al., 2009, 2011; Nosenko et al., 2013); (iii) a third dataset with no outgroup (ctenophores only). Indeed, ctenophores have been shown to be extremely divergent from other metazoan lineages whatever the molecular markers analysed (Philippe et al., 2009, 2011; Pick et al., 2010; Pett et al., 2011; Kohn et al., 2012; Nosenko et al., 2013), implying that the outgroups are necessarily much distant from the ingroup, which could affect the accuracy of the phylogenetic reconstruction of relationships within ctenophores.

Since ITS regions are not alignable between ctenophores and other metazoan phyla, there is no possible outgroup for the ITS dataset. This notably implies that in analyses of the 18S-ITS dataset with outgroups, the ITS sequences of all outgroups were represented by missing data.

2.4. Phylogenetic analyses

All datasets containing 18S sequences were analysed in two different ways: taking and not taking into account the 18S rRNA secondary structure. Additionally, all datasets that include the ITS markers were analysed with and without partitioning the data (either in three partitions: ITS1, 5.8S and ITS2; or in four partitions: 18S, ITS1, 5.8S and ITS2).

Table 1

List of ctenophore sequences used in the present study (classification, source, accession numbers). Classification is from Mills (2012).

Taxon, species	Source	18S	ITS
Order Cydippida			
Fam. Pleurobrachiidae			
<i>Pleurobrachia pileus 1</i>	Villefranche-sur-mer (FRANCE)	This study	This study
<i>Pleurobrachia pileus 2</i>	Podar et al. (2001)	AF293678	AF293678
<i>Pleurobrachia bachei</i>	Podar et al. (2001)	AF293677	AF293677
<i>Pleurobrachia brunnea</i>	St Rose (GUADELOUPE)	This study	This study
<i>Hormiphora plumosa</i>	Podar et al. (2001)	AF293676	AF293676
<i>Hormiphora</i> sp. undescrbed sp. 1	Collins (1998)	AF100944	–
	Podar et al. (2001)	AF293675	–
Fam. Euplokamididae			
<i>Euplokamis</i> sp.	Granhag et al. (2012)	387597972	387597972
Fam. Lampeidae			
<i>Lampea pancerina</i>	Villefranche-sur-mer (FRANCE)	This study	This study
Fam. Mertensiidae			
<i>Mertensia ovum 1</i>	Podar et al. (2001)	AF293679	–
<i>Mertensia ovum 2</i>	GenBank, unpublished	FJ668937	FJ668937
<i>Charistephane fugiens 1</i>	Podar et al. (2001)	AF293682	–
<i>Charistephane fugiens 2</i>	Collins (2002)	AF358113	–
undescrbed sp. 2	Podar et al. (2001)	AF293680	–
Fam. Haeckeliidae			
<i>Haeckelia beehleri</i>	Podar et al. (2001)	AF293673	AF293673
<i>Haeckelia rubra</i>	Podar et al. (2001)	AF293674	AF293674
Order Beroida			
Fam. Beroidae			
<i>Beroe forskalii 1</i>	Villefranche-sur-mer (FRANCE)	This study	This study
<i>Beroe forskalii 2</i>	Podar et al. (2001)	AF293698	AF293698
<i>Beroe forskalii 3</i>	Podar et al. (2001)	AF293697	–
<i>Beroe cucumis 1</i>	Podar et al. (2001)	AF293695	AF293695
<i>Beroe cucumis 2</i>	Podar et al. (2001)	AF293699	AF293699
<i>Beroe cucumis 3</i>	GenBank, unpublished	D15068	–
<i>Beroe ovata</i>	Podar et al. (2001)	AF293694	AF293694
<i>Beroe gracilis</i>	Podar et al. (2001)	AF293696	–
<i>Beroe abyssicola 1</i>	Markova et al. (2012)	JN673817	–
<i>Beroe abyssicola 2</i>	GenBank, unpublished	–	AB377608
Order Platyctenida			
Fam. Coeloplanidae			
<i>Coeloplana</i> cf. <i>meteoris</i>	Inland aquatics, IN (USA)	This study	This study
<i>Coeloplana bannwarthi</i>	Podar et al. (2001)	AF293683	AF293683
<i>Coeloplana bocki</i>	Song et al. (2011)	HQ435813	HQ435814
<i>Coeloplana anthostella</i>	Song et al. (2011)	HQ435810	HQ435812
<i>Coeloplana agniae</i>	Collins (2002)	AF358112	–
<i>Vallicula multififormis</i>	Podar et al. (2001)	AF293684	AF293684
Order Lobata			
Fam. Bolinopsidae			
<i>Mnemiopsis leidyi 1</i>	Villefranche-sur-mer (FRANCE)	This study	This study
<i>Mnemiopsis leidyi 2</i>	Podar et al. (2001)	AF293700	AF293700
<i>Bolinopsis infundibulum 1</i>	Podar et al. (2001)	AF293687	–
<i>Bolinopsis infundibulum 2</i>	GenBank, unpublished	–	EF175465
Fam. Leucotheidae			
<i>Leucothea multicornis</i>	Villefranche-sur-mer (FRANCE)	This study	This study
<i>Leucothea pulchra</i>	Podar et al. (2001)	AF293688	AF293688
Fam. Eurhampheidae			
<i>Deoipea kaloktenota</i>	Villefranche-sur-mer (FRANCE)	This study	This study
Fam. Ocyropsis			
<i>Ocyropsis maculata maculata</i>	Podar et al. (2001)	AF293689	AF293689
<i>Ocyropsis crystallina crystallina</i>	Podar et al. (2001)	AF293690	AF293690
<i>Ocyropsis crystallina guttata</i>	Podar et al. (2001)	AF293691	AF293691
Fam. Lampoctenidae			
<i>Lampocteis cruentiventer</i>	GenBank, unpublished	KF202290	–
Unclassified Lobate species undescrbed sp. 4	Podar et al. (2001)	AF293686	AF293686
Order Cestida			
Fam. Cestidae			
<i>Cestum veneris 1</i>	Villefranche-sur-mer (FRANCE)	This study	This study
<i>Cestum veneris 2</i>	Podar et al. (2001)	AF293692	–
<i>Cestum veneris 3</i>	Chow et al. (2008)	–	AB377605
<i>Cestum veneris 4</i>	Chow et al. (2008)	–	AB377606
<i>Velamen parallelum</i>	Podar et al. (2001)	AF293693	AF293693
Order Thalassocalycida			
Fam. Thalassocalycidae			
<i>Thalassocalyce inconstans</i>	Podar et al. (2001)	AF293685	–
<i>Thalassocalyce</i> sp. 1	Chow et al. (2008)	–	AB377603
<i>Thalassocalyce</i> sp. 2	Chow et al. (2008)	–	AB377604
Unclassified Ctenophore species undescrbed sp. 3	Podar et al. (2001)	AF293681	–

Please cite this article in press as: Simion, P., et al., Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora. Zoology (2014), <http://dx.doi.org/10.1016/j.zool.2014.06.004>

The three datasets (18S, ITS, and 18S-ITS) were each analysed with Jmodeltest (Darriba et al., 2012) to determine the model of sequence evolution which best fits the data. For practical reasons, we then used the first next-to-best model which was available in both RAxML (Stamatakis, 2006) and MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), which was GTR + Γ + I in all cases.

For analyses taking into account the secondary structure of the 18S rRNA, we used a doublet model of sequence evolution, which partitions the data in two groups of positions corresponding to stems and loops, respectively. Bases in stems were paired according to the helix structure (as determined by Mallatt et al., 2010) and the S16 doublet model was used in order to represent the dependency between the two nucleic acids of each pair. Different base-pair substitution models are implemented in RAxML, but since MrBayes only implements the S16 doublet model to take into account the 18S rRNA secondary structure, we also used this model in RAxML in order to allow comparisons of results from the two algorithms. Also, it was not possible to use RAxML to analyse the 18S-ITS dataset taking into account both the 18S rRNA secondary structure and the ITS partitioning since this is not implemented in RAxML. Substitutions at loop positions were handled by the best-fit model (i.e. GTR + Γ + I) as mentioned above.

Maximum likelihood analyses were run with RAxML under a GTR + Γ + I model of sequence evolution, performing 100 topology searches in order to ensure that the best topology found corresponds to the true maximum likelihood. Branch support was estimated by bootstrapping (100 replicates). Bayesian analyses were run with MrBayes under the GTR + Γ + I model, with four parallel chains of 1,000,000 steps with trees sampled every 1000 steps. We verified chain convergence by checking that (i) the log probability plot had reached a plateau; (ii) the potential scale reduction factor was close to 1; (iii) the average standard deviation of split frequencies reached a sufficiently low level (usually between 0.01 and 0.02). Posterior probabilities were calculated after exclusion of a burn-in of 50% of sampled trees.

2.5. Character optimisation

The mapping of six chosen morphological characters was carried out by using Mesquite (Maddison and Maddison, 2010) without using the majority rule for character optimisation in polytomies. We used the topology that best summarises our phylogenetic analyses results. However, we decided to add a clade grouping the orders Cestida, Thalassocalycida and Lobata due to their very strong morphological similarities.

3. Results and discussion

The phylogenetic relationships of ctenophores were investigated using two molecular datasets: 18S rRNA (slowly evolving dataset or “18S dataset”) and ITS1 + 5.8S rRNA + ITS2 (rapidly evolving dataset or “ITS dataset”), analysed independently and in combination, under different rooting options and models. The taxonomic sampling covers 14 different ctenophore families (more than 50% of the total number of ctenophore families) representing 6 orders (Table 1). An overview of these molecular analyses is presented in Table 2, with selected trees shown in Figs. 1–3 (all trees given in Figs. S1–S25 in the supplementary online Appendix B).

In our analyses using 18S rRNA and the extended outgroup sampling of 76 species, ctenophores form a sister clade to a cnidarian + bilaterian + placozoan clade and sponges were found to be paraphyletic (Fig. 1A). This is in accordance with most previous studies based on ribosomal RNAs (e.g., Cavalier-Smith et al.,

1996; Collins, 1998; Zrzavy et al., 1998; Borchellini et al., 2001; Peterson and Eernisse, 2001; Halanych, 2004). This topology likely reflects the limits of this molecular marker at deep phylogenetic levels rather than the true relationships between early diverging metazoan lineages (Abouheif et al., 1998; Rokas et al., 2003). In particular, recent studies based on large phylogenomic data sets found strong support for sponge monophyly (Philippe et al., 2009, 2011; Pick et al., 2010; Nosenko et al., 2013). In the following, we will focus exclusively on the relationships within ctenophores.

3.1. Usefulness of rapidly evolving markers to resolve internal ctenophore relationships

Internal transcribed spacers (ITS1 and ITS2) are usually used to infer relationships at the species or genus level (e.g., Zhao et al., 2012; Ghada et al., 2013). However, and rather unexpectedly, ITS1 and ITS2 sequences happen to be easily alignable even between distantly related ctenophore taxa (see alignments). The ITS dataset provides 200 parsimony informative characters, whereas the 18S rRNA alignment at the ctenophore level contains only 140 parsimony informative characters. Many branches that are strongly supported by the 18S rRNA dataset alone were also retrieved with the ITS dataset. In the latter the contribution of 5.8S is negligible, with only 9 parsimony informative characters (Table 2; Figs. 1–3). A major limitation of using ITS data only is that analyses are necessarily unrooted, since ctenophore ITS sequences cannot be aligned with their counterparts from other phyla. However, with respect to 18S analyses, addition of this rapidly evolving dataset led to a significant increase of statistical support from combined analyses for the monophyly of, e.g., Mertensiidae, Ocyropsidae, or the order Cestida (Table 2). In line with previous suggestions that present-day lineages of ctenophores might result from a recent radiation following a bottleneck (Podar et al., 2001), an important lesson from these observations is that future attempts to resolve ctenophore phylogeny should consider integrating more rapidly evolving markers than those typically used at the phylum level in metazoans (such as, for example, nucleotide sequences of the cytochrome oxidase subunits or the cytochrome b mitochondrial genes).

3.2. Causes of topological inconsistencies and possible perspectives for improvement

Not surprisingly, several higher-level ctenophore relationships remain unresolved. A first area of difficulty is the rooting of the ctenophore tree. When non-ctenophore sequences were included (either extended or reduced outgroup sampling, see Section 2.3), five alternative positions were obtained for the ctenophore root depending on tree reconstruction options (Table 2; summarised in Fig. 4). Five of these analyses yielded a topology in which Mertensiidae are sister group to all other ctenophores, in two cases with moderate support (Bayesian analyses of 18S rRNA with and without doublet model) (see Figs. 1–3, Table 2, and Figs. S1–S25 in the supplementary online Appendix B). None of the four following alternative rootings came with significant support values, and some are clearly rejected by morphology (Table 2, Figs. 4 and 6): (i) Lampeidae as sister group to other ctenophores; (ii) the lobate genus *Ocyropsis* as sister group to other ctenophores (apparently an aberrant result as it would imply independent acquisition of the Lobata apomorphies: lobes and auricles); (iii) Cestida as sister group to other ctenophores (equally aberrant given strong morphological similarities between Cestida and Lobata; Harbison, 1985); (iv) two large sister clades, one containing (Lampeidae + Platyctenida + Mertensiidae + sp3), and the other (Lobata + Cestida + Thalassocalycida + Pleurobrachiidae + Beroidae + Haeckeliidae). An additional element of topological instability across analyses using various models and outgroups was the

Table 2

Overview of results across multiple analyses with different datasets, taxon samplings and evolution models. For each analysis (see Figs. S1–S25 in the supplementary Appendix), a dot indicates an inexistent clade, a grey or black square indicates the clade is present in the tree, a number indicates the support value (either bootstrap value or posterior probability, according to the reconstruction method used) between 50 and 99, and a black square indicates maximum support (bootstrap value or posterior probability of 100%). All outgroups: Fungi, Choanoflagellates, Porifera, Placozoa, Bilateria and Cnidaria.

Dataset	18S						ITS				18S-ITS														
	all outgroups		Cnidaria		none		None		all outgroups		Cnidaria		none												
Model	●	○	●	○	●	○	▲	◇	■	□	■	□	■	□											
Method	ML	B	ML	B	ML	B	ML	B	ML	B	ML	B	ML	B											
All ctenophores except Mertensia + Charistephane	50	84	88	.	.	.	No outgroup	No outgroup	50	No outgroup											
All ctenophores except Lampeidae	No outgroup	No outgroup	.	.	.	52	.	No outgroup											
All ctenophores except Ocyropsis	No outgroup	No outgroup	No outgroup											
Mertensiidae + Lampea + Platyctenida + sp3	50	.	No outgroup	No outgroup	No outgroup											
All ctenophores except Cestida	No outgroup	No outgroup	No outgroup											
Mertensia + Euplokamis sp.	■	■	■	■	■	■	■	■	■	■	■	■	■	■											
Mertensiidae (including Euplokamis sp.)	91	93	91	81	97	98	98	■	■	99	No ITS for sp2, Charistephane	88	97	98	97	98	■	■	■	■					
Lampea + Platyctenida	■	59	■	57	■	■	■	■	55	53	53	53	.	.	.	■	■	55	60						
Pleurobrachia (including Hormiphora sp.)	93	■	78	■	92	■	85	■	96	■	92	■	73	94	66	92	■	■	99	■	■	97	■	96	
Pleurobrachiidae	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
Thalassocalyce + sp4 + Lampoctenidae	■	98	■	98	■	94	■	92	50	98	51	96	No ITS for Lampocteis	54	95	90	50	93	92	■	■	91	92		
Lobata	
Leucotheidae + Cestida	■	52	■	57	■	■	■	■	■	55	■	56	
Leucotheidae + Bolinopsis + deoipea + Mnemiopsis	87	87	■	■	■	■	54	60	■	■	59	62	
Leucotheidae + Deoipea	80	76	■	■	73	58	■	■	67	69	.	.	
Bolinopsidae	
Lobata + Cestida + Thalassocalycida	.	.	.	■	■	.	■	63	.	■	■	56	.	.	.	
Cestida	51	92	■	88	58	93	■	94	63	98	■	98	70	■	70	■	67	55	63	74	94	95	75	■	■
Ocyropsidae	75	97	.	■	.	94	.	91	97	■	95	■	73	98	81	98	73	97	88	81	99	71	■	■	■
Haekeliidae	70	■	77	■	77	■	87	■	66	97	73	81	99	■	■	95	■	■	96	■	■
Beroe
Haekeliidae + Beroe (except Beroe abyssicola)	78	86	85	91	81	95	76	97	75	90	67	93	73	81	87	67	88	93	69	73	77
Haekeliidae + all Beroe	84	■	84	■	87	■	96	■	85	■	85	99	88	99	99	87	99	99	74	97	■
Pleurobrachiidae + Cestida + Beroidea + Haekeliidae + Lobata + Thalassocalycida	65	86	.	■	.	55	.	65	96	■	92	■	54	.	■	84	.	95	■	■	■

ML: maximum likelihood; B: Bayesian inference Symbols used to indicate evolutionary models: ● GTR + Γ + I without 18S rRNA structure; ▲GTR + Γ + I without ITS partitioning; ■ GTR + Γ + I without 18S rRNA structure, without ITS partitioning; ○ GTR + Γ + I with 18S rRNA structure (S16 doublet model); ◇ GTR + Γ + I with ITS partitioning; □ GTR + Γ + I with 18S rRNA structure (S16 doublet model), with ITS partitioning.

430 arrangement of the various families of Lobata with respect to
431 Cestida and Thalassocalycida.

432 These topological instabilities are probably in part due to lack
433 of phylogenetic signal in the small molecular datasets analysed
434 here, but also certainly arise from strong inequalities in substitution

435 rates across the tree. The ctenophore clade has a very long branch
436 within the metazoan tree (Fig. 1) and sequence divergence for 18S
437 rRNA between ctenophores and other phyla is huge compared to
438 sequence divergence for the same marker among ctenophores (this
439 is also observed for conserved nuclear protein markers; see trees

Please cite this article in press as: Simion, P., et al., Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora. Zoology (2014), <http://dx.doi.org/10.1016/j.zool.2014.06.004>

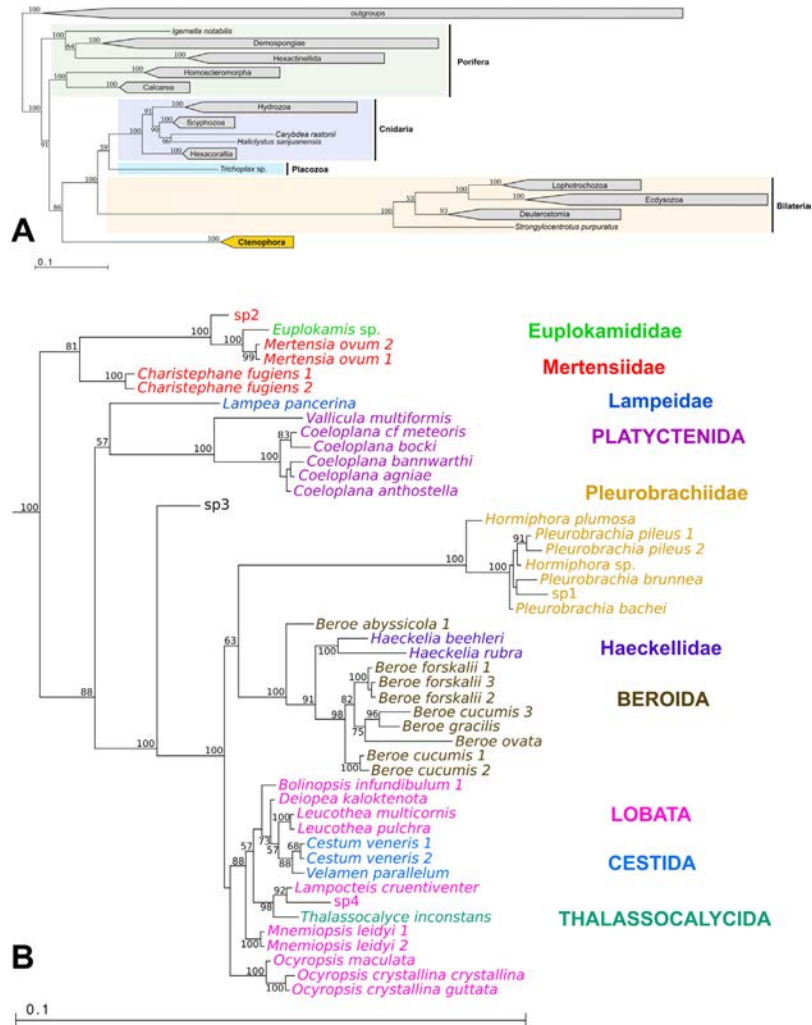


Fig. 1. Ctenophore phylogenetic relationships reconstructed with 76 outgroup species using Bayesian inference (MrBayes) on the 18S dataset. A doublet model was used for 18S rRNA to represent its secondary structure. Posterior probabilities above 50% are indicated on the branches. (A) Simplified tree at the metazoan level. The horizontal length of each grey box corresponds to the maximum branch length found within the corresponding clade. (B) Detailed phylogenetic relationships within ctenophores in the same analysis. Scale bar: number of inferred substitutions per site.

440 in Philippe et al., 2009, 2011; Pick et al., 2010; Pett et al., 2011;
 441 Nosenko et al., 2013; Ryan et al., 2013). This situation hampers
 442 correct rooting of the ctenophore clade, because identification of
 443 a subset of ctenophore species (A) as sister group to the others (B)
 444 requires that a sufficient number of molecular characters are shared
 445 as plesiomorphies between A and the outgroups, but derived in B.
 446 Positions potentially relevant to place the ctenophore root must
 447 therefore be variable at an intra-ctenophore level, but then this
 448 kind of positions has a high probability to have undergone multi-
 449 ple substitutions along the long common branch of ctenophores,
 450 causing non-homology of character states between (A) and the
 451 outgroup.

452 Furthermore, substitution rates are highly unequal among the
 453 ctenophore lineages. This holds true for both data sets analysed
 454 here, with similar tendencies: very long branches for Pleurobrachi-
 455 idae and for the Platyctenida, contrasting with very short branches

456 for species belonging to Lobata, Cestida and Thalassocalycida, while
 457 the other taxa stand in between (Figs. 1–3).

458 These problems of unequal substitution rates (between
 459 ctenophore groups; between ctenophores and other animal lin-
 460 eages) not only weaken tree reconstructions using a small number
 461 of molecular markers like in this study, but it can be predicted that
 462 they will also affect future tree reconstructions based on multiple
 463 markers, such that simply adding more and more genes might not
 464 solve the problem but instead might reinforce topological artefacts
 465 due to addition of systematic errors (Philippe and Roure, 2011).
 466 Suggestions to overcome this bias include combining slowly and
 467 more rapidly evolving genes in multi-marker datasets, removing
 468 compositionally biased sites and/or long-branched taxa, and the
 469 use of models able to handle multiple substitutions such as the
 470 CAT model (Lartillot and Philippe, 2004), which was not used here
 471 because to be effective this statistically complex model requires

Please cite this article in press as: Simion, P., et al., Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora. Zoology (2014), <http://dx.doi.org/10.1016/j.zool.2014.06.004>

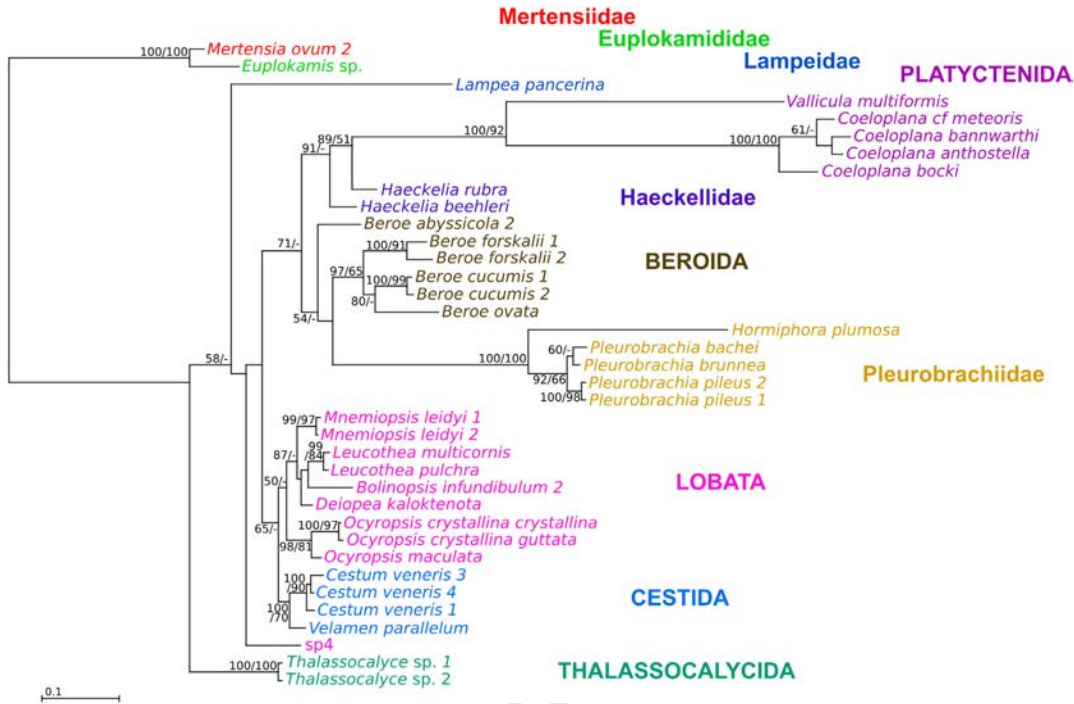


Fig. 2. Bayesian inference (MrBayes) analysis of ctenophore phylogenetic relationships without outgroups and using ITS1, 5.8S rRNA and ITS2 markers (partitioned GTR + Γ + I model). Numbers on branches indicate the Bayesian posterior probabilities (first number) and bootstrap values obtained in a maximum likelihood (RAxML) analysis of the same dataset (second number) when above 50%. The tree is unrooted. Scale bar: number of inferred substitutions per site.

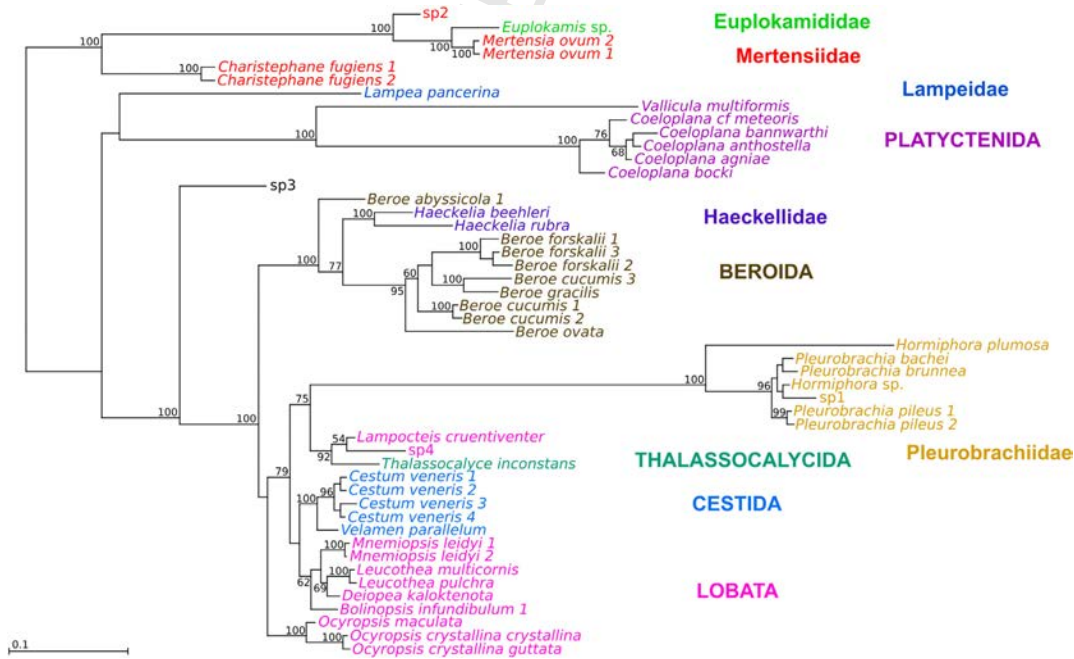


Fig. 3. Bayesian inference (MrBayes) analysis using the combined 18S-ITS dataset (with doublet model for 18S rRNA and a partitioned GTR + Γ + I model for the ITS dataset). Posterior probabilities higher than 50% are indicated above the branches. The tree is unrooted. Scale bar: number of inferred substitutions per site.

Please cite this article in press as: Simion, P., et al., Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora. Zoology (2014), <http://dx.doi.org/10.1016/j.zool.2014.06.004>

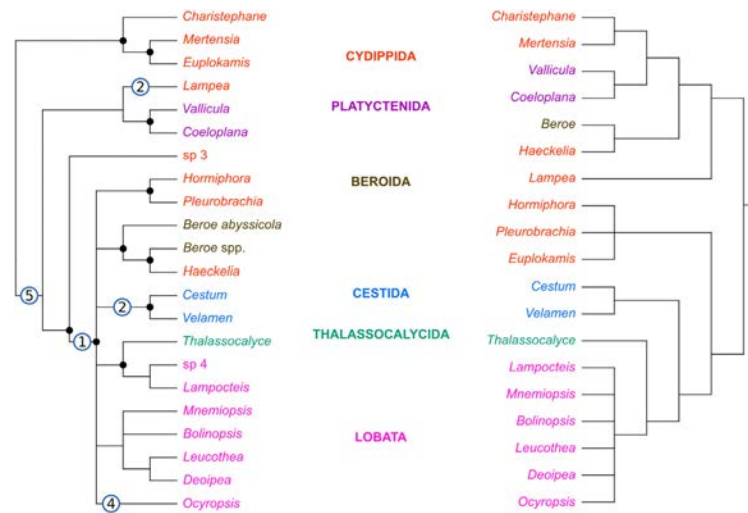


Fig. 4. Consensus tree summarising the results of the present study (left part) compared to a morphology-based phylogeny (right part, adapted from Harbison, 1985). The tree recapitulates stable clades (i.e., present in all unrooted topologies from analyses of the 18S–ITS dataset). Black dots represent supported nodes (a bootstrap value $\geq 80\%$ or a posterior probability $\geq 90\%$ in at least one of the unrooted analyses of the 18S–ITS dataset). Blue circles represent the different positions obtained for the ctenophore root and the numbers inside indicate the number of phylogenetic analyses that support each rooting.

472 alignments with a much larger number of positions (at least several
 473 thousand) than was used here.

474 Finally, whereas including 18S rRNA secondary structure infor-
 475 mation has been previously shown to improve accurate inference
 476 of transitions during phylogenetic reconstruction (Kjer, 1995;
 477 Gillespie et al., 2005; Erpenbeck et al., 2007; Voigt et al., 2008),
 478 in our analyses it did not significantly improve the resolution of
 479 ctenophore relationships and node support values (see Table 2).
 480 The same applies to data partitioning, intended to account for the
 481 distinctive modalities of molecular evolution between different
 482 markers (here 18S rRNA, ITS1, 5.8S rRNA, ITS2). Our analyses using
 483 partitioned models failed to improve the results with respect to
 484 non-partitioned analyses (Table 2). These observations hold true
 485 for both methods used here (ML and Bayesian), which generally
 486 gave congruent results for a given model. Clearly these strategies
 487 (doublet model and data partitioning) were insufficient to over-
 488 come the limitations discussed above (lack of phylogenetic signal
 489 and unequal substitution rates).

490 3.3. Newly observed groupings with interesting taxonomic 491 implications

492 Even if some higher-level ctenophore relationships remain
 493 unresolved, different areas of the tree appear stable across analyses
 494 (bootstrap support values $>75\%$ in most ML analyses, and posterior
 495 probabilities >90 in most Bayesian analyses, Table 2). A summary
 496 tree representing all clades repeatedly found when using the com-
 497 bined dataset, confronted with a “traditional” morphology-based
 498 ctenophore tree (Harbison, 1985), is given in Fig. 4. The present
 499 study confirms previous findings (Harbison, 1985; Podar et al.,
 500 2001) concerning the non-monophyly of the order Cydippida (here
 501 represented by five families instead of three as in Podar et al., 2001,
 502 namely Mertensiidae, Lampeidae, Haeckeliidae, Pleurobrachiidae
 503 and Euplokamiidae) and the close relationship of Beroidea and
 504 Haeckeliidae. In addition, our analyses yield several new results
 505 of taxonomic interest (detailed below) with respect to previous
 506 molecular phylogenies of ctenophores (both based on the 18S rRNA
 507 marker: Podar et al., 2001; Granhag et al., 2012), notably thanks

508 to the addition of the families Lampeidae (represented by *Lampea*
 509 *pancerina*), Eurhampheidae (represented by *Deiopea kaloktenota*)
 510 and Lampocteniidae (represented by *Lampocteis cruentiventer*), and
 511 of *Beroe abyssicola*.

512 Compared with the morphology-based ctenophore tree of
 513 Harbison (1985) (right part of Fig. 4; see comments in the legend),
 514 our summary tree notably differs in terms of the relative positions
 515 of Mertensiidae, Platyctenidae and Lampeidae, and with respect
 516 to the placement of the “Beroidea + Haeckeliidae” clade (close to
 517 Mertensiidae + Platyctenidae according to Harbison (1985); closer
 518 to lobates and pleurobrachiids in our analyses) and of the genera
 519 *Euplokamis* and *Thalassocalyce* (discussed below).

520 The order Lobata (as currently defined; Mills, 2012; Table 1)
 521 was never found monophyletic with any dataset. This is surprising
 522 because all Lobata members share the presence of two oral
 523 lobes and four auricles. Analyses including 18S rRNA never obtained
 524 the monophyly of Lobata *sensu stricto*, but sometimes yielded a
 525 larger clade including lobates, cestids and *Thalassocalyce*, in con-
 526 sistent with morphology (Harbison, 1985; see Fig. 4, right part).
 527 In all these trees, *Thalassocalyce* forms a clade (often well sup-
 528 ported) with the lobates *Lampocteis* and “sp.4”. This suggests that
 529 *Thalassocalycidae* is a family of lobates rather than representing a
 530 separate order (*Thalassocalycida*) as in the current classification
 531 (Mills, 2012). This is in accordance with the description of the
 532 medusoid bell of *Thalassocalyce* by Madin and Harbison (1978) who
 533 wrote that “clearly, the medusoid bell is morphologically equiva-
 534 lent to two fused lobes”. Other important observations concerning
 535 Lobata include non-monophyly of the family Bolinopsidae (gen-
 536 era *Mnemiopsis* and *Bolinopsis*), observed in all analyses (Table 2),
 537 and monophyly of Leucotheidae + Eurhampheidae (*Deiopea*) sug-
 538 gested by all combined 18S–ITS analyses (Table 2). Finally, the
 539 monophyly of Cestida is obtained in most analyses with high sup-
 540 port, and this order always groups within or close to Lobata, in line
 541 with previous studies (Harbison, 1985; Podar et al., 2001), but its
 542 exact position either as a sister group of Lobata, or as derived Lobata,
 543 remains undetermined.

544 The only sampled species of the genus *Euplokamis* Chun, 1879
 545 (family Euplokamiidae in Mills, 2012) is robustly placed within

Mertensidae and very close to *Mertensia ovum* in all analyses, in consistence with [Granhag et al. \(2012\)](#). The main diagnostic character of *Euplokamis*, the presence of striated muscles in the tentillae ([Mills, 1987](#)), is an autapomorphy among ctenophores and therefore is not informative with respect to the phylogenetic position of this genus. Otherwise, *Euplokamis* species are very similar to members of Pleurobrachiidae (and have previously been included in this family; [Harbison, 1985](#)), but the described absence of interradiar canals in *Euplokamis dunlapae* [Mills, 1987](#) (while they are present in pleurobrachiids) is consistent with an affinity with mertensids. However, this shared feature is likely a symplesiomorphy since some other ctenophore taxa like Platyctenida and Lampeidae also lack interradiar canals ([Harbison, 1985](#)) (see optimisation of this character in Section 3.4). Of note is that the only available sequence of *Euplokamis* was produced from larval-stage specimens ([Granhag et al., 2012](#)); furthermore, the presence of striated muscle in tentillas was apparently not checked in the adults said to correspond to these larvae ([Granhag et al., 2012](#)), thus misidentification cannot be excluded. Sequencing of additional *Euplokamis* species is therefore mandatory before confirming the inclusion of this genus within the family Mertensidae.

A result of strong taxonomic importance is the paraphyly of the genus *Beroe*, observed in all analyses owing to the position of *Beroe abyssicola*. When 18S sequences were included, *Beroe abyssicola* fell outside a strongly supported clade grouping Haeckeliidae and the other *Beroe* species ([Figs. 1 and 3; Table 2](#)). This placement was unexpected because Beroida are uniquely characterised by the absence of tentacles (present in Haeckeliidae), clearly an apomorphic feature as beroids are nested within the ctenophore tree in all analyses. Because secondary re-acquisition of tentacles in haeckeliids seems unlikely, the position of *Beroe abyssicola* suggests that this species has lost the tentacular apparatus independently from the other beroids (see optimisation of this character in Section 3.4). If confirmed in the future, this result will not only imply the creation of a new genus for *abyssicola* (the type species of *Beroe*

being *B. ovata* [Eschscholtz, 1828](#)) but will also call for a new definition of the order Beroida. Historically, ctenophores have been subdivided into two classes, Tentaculata (tentaculate ctenophores) and Nuda (ctenophores lacking tentacles even in juvenile stages, i.e. Beroida) ([Harbison, 1985; Mills, 2012](#)). Non-monophyly of the former was already shown by [Podar et al. \(2001\)](#). The paraphyly of the genus *Beroe* obtained in this study also invalidates the concept of monophyletic Nuda.

The present study is the first to include a member of the Lampeidae (*Lampea pancarina*) in a molecular dataset. This species does not show any particular affinity with other “Cydippida” families, but instead branches as a sister group to the benthic order Platyctenida in almost all analyses using 18S rRNA ([Table 2](#)), albeit with non-significant (<80%) support values in all cases. This grouping has been previously suggested by [Ospovat \(1985\)](#) and [Harbison \(1985\)](#). The latter author pointed out the remarkable creeping behaviour of *Lampea pancarina* adult specimens (originally noticed by [Chun, 1880](#)), which can expand their mouth and pharynx onto a substrate, as do platyctenids ([Fig. 5](#)). Furthermore, [Komai \(1922\)](#) had assigned the parasitic “*Gastrodes parasiticum*” (in fact later proved to represent the larval and juvenile stages of *L. pancarina*) to the Platyctenida due to its flat creeping morphology and similarities with the platyctenid gastrovascular canal system (see [Fig. 5](#)). It is highly probable that the acquisition of the creeping capabilities observed in both lampeids and platyctenids represents a unique evolutionary event during the evolution towards a benthic life style. We also note on the drawings of “*Gastrodes*” ([Fig. 5A](#)) and a platyctenid ([Fig. 5B](#)) an additional similarity, the unusual position of anal pores in the pharyngeal plane. In sum, the existence of a Lampeidae + Platyctenida clade makes sense from the point of view of comparative anatomy. In most trees that included outgroups, this clade emerged as the sister group to all other ctenophores except Mertensidae (see summary tree, [Fig. 4](#)).

Finally, our results disagree with recent phylogenomic studies ([Nosenko et al., 2013; Ryan et al., 2013](#)) which found *Pleurobrachia*

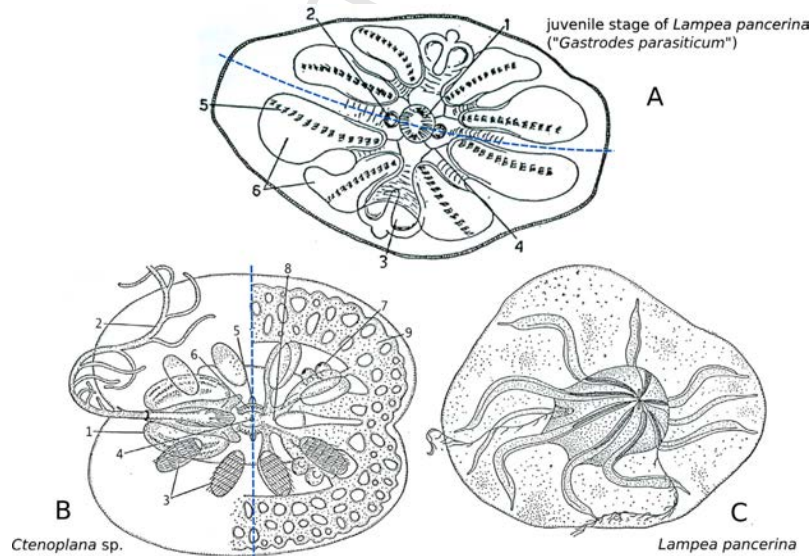


Fig. 5. Published drawings of *Ctenoplana* sp. (Platyctenida) and *Lampea pancarina*. The dashed lines in (A) and (B) indicate the pharyngeal symmetry plane. (A) Drawing of a comb-bearing stage of “*Gastrodes parasiticum*”, the juvenile stage of *Lampea pancarina* (from [Hyman, 1940](#), after [Komai, 1922](#)). Labels: 1, statocyst; 2, anal pores; 3, tentacle sheath; 4, ciliated furrows; 5, comb rows; 6, meridional canals. (B) Drawing of *Ctenoplana* in aboral view (from [Hyman, 1940](#), after [Komai, 1934](#)). Left side: external view; right side: internal structures. Labels: 1, tentacle sheath; 2, tentacle; 3, comb rows; 4, statocyst; 5, anal pores; 6, pharynx; 7, gonads; 8, gastrovascular canals; 9, peripheral network of gastrovascular canals. (C) Drawing of an adult *Lampea pancarina* stuck onto a substrate with the pharynx everted (from [Hyman, 1940](#), after [Komai, 1934](#)).

Please cite this article in press as: Simion, P., et al., Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora. *Zoology* (2014), <http://dx.doi.org/10.1016/j.zool.2014.06.004>

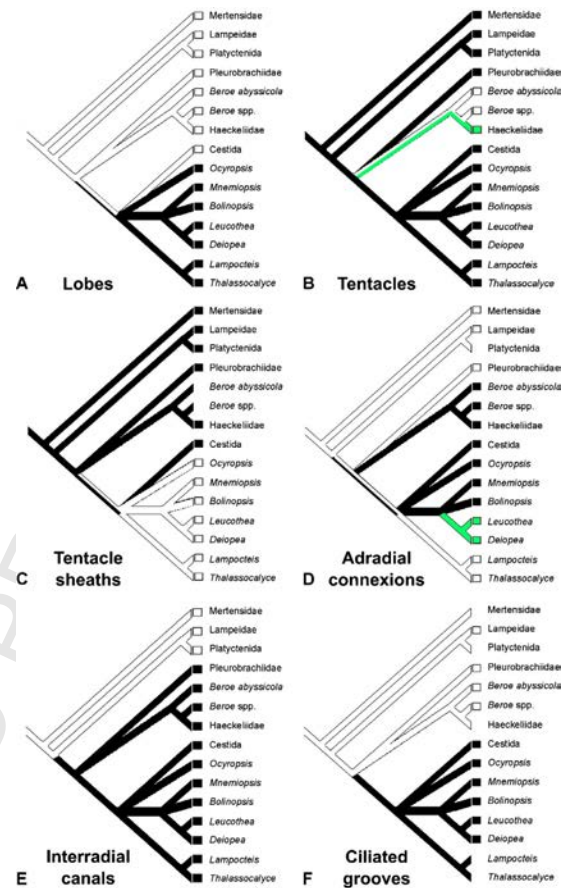
616 as sister group to other ctenophores, followed by a mertensiid
617 species sister group to *Cestum* and *Beroe*. However, ctenophore taxo-
618 nomic diversity is very low in these phylogenomic datasets (3
619 species in Ryan et al., 2013 and 4 species in Nosenko et al., 2013).
620 The long branch of pleurobrachiids might be responsible for the
621 basal attraction of *Pleurobrachia* in these studies.

622 3.4. Towards an understanding of ctenophore evolution

623 This study strongly confirms the hypothesis of a “Cydippida-
624 like” ancestor of Ctenophora. This hypothesis, previously suggested
625 based on the paraphyly of the order Cydippida (Harbison, 1985;
626 Podar et al., 2001), is reinforced by the recurrent position of Mertensi-
627 siidae and (Lampeidae + Platyctenida) close to the ctenophore root
628 in our analyses (Table 2, Figs. 1–3). The fact that Mertensiidae,
629 Lampeidae and an undescribed cydippid species (i.e. “sp. 3”) are
630 three of the four first emerging clades within Ctenophora (Fig. 4)
631 represents strong evidence that their shared anatomical features
632 (e.g., rounded planctonic bodies; two extensible tentacles; pres-
633 ence of tentacle sheaths; aboral meridional canals ending blindly)
634 are symplesiomorphies of ctenophores. This hypothesis implies
635 that anatomically divergent features of Ctenophore orders such as
636 platyctenids (e.g., flat body, benthic life style), lobates (e.g., lobes
637 and auricles) or cestids (e.g., greatly elongated body) are apomor-
638 phies of these clades, as previously proposed based on morphology
639 (Komai, 1922; Harbison, 1985).

640 Persisting irresolution of relationships, particularly between
641 Pleurobrachiidae, Beroida + Haeckeliidae, Cestida, Lobata and Tha-
642 lassocalycida (Figs. 1–4) makes reconstruction of evolutionary
643 changes impossible for most anatomical characters. We neverthe-
644 less tentatively present results of parsimony optimisation for six
645 characters that provide mostly unambiguous changes when plotted
646 onto our currently most realistic tree topology (Fig. 4). Note, how-
647 ever, that the topology used for these optimisations comprises an
648 additional clade (well-supported by morphology) grouping lobates,
649 cestids and *Thalassocalyce* (without addition of this branch all
650 reconstructed character changes would be ambiguous). Accord-
651 ingly, a pair of oral lobes would have been acquired only once
652 (Fig. 6A), but currently it is not possible to determine whether
653 their absence in cestids is plesiomorphic or secondary. The ten-
654 tacles devoid of tentillae which characterise haeckeliids represent
655 either a tentacle reduction on the path to complete loss of ten-
656 tacles (which then occurred once in *Beroe abyssicola* and once in
657 the other beroids) or a secondarily evolved tentacle apparatus (if
658 lost in a common ancestor of beroids and haeckeliids) (Fig. 6B).
659 Loss of the tentacular sheath could be a synapomorphy for a clade
660 grouping lobates and *Thalassocalyce* but not cestids (Fig. 6C).
661 A character of the gastro-vascular architecture (mode of connection
662 adradial and meridional canals) shows an interesting distribution
663 (Fig. 6D). Clearly the ancestral state is adradial canals fused to
664 meridional canals far from their aboral extremity, while fusion at
665 the aboral extremity of meridional canals could potentially be a
666 synapomorphy for a clade grouping beroids, haeckeliids, lobates,
667 cestids and *Thalassocalyce* (Fig. 6D). Exceptions within that clade
668 (fusion far from the aboral extremity for all – *Lampocteis* and
669 *Thalassocalyce* – or only four – *Leucothea* + *Deiopea* – of the
670 meridional canals) could be the result of reversion (Fig. 6D). The
671 acquisition of interradial canals is a synapomorphy for the clade
672 grouping all ctenophores except Mertensiidae, Lampeidae and
673 Platyctenida (Fig. 6E). Finally, a ciliated groove running all
674 along the comb rows is clearly a derived feature of lobates and
675 cestids (but its presence or absence is undetermined in
676 *Lampocteis* and *Thalassocalyce*) (Fig. 6F).

677 Improvement of the taxonomic sampling should be a priority for
678 future studies of ctenophore relationships, as three orders (Cambio-
679 jida, Ganeshida and Cryptolobiferida) and no less than 13 families
680 are still missing in our molecular dataset. Since ctenophores are



680 Fig. 6. Parsimony optimisation of selected morpho-anatomical characters using
681 Mesquite. (A) Lobes (white: absent; black: present). (B) Tentacles (white: absent;
682 green: present but without tentillae; black: present and with tentillae). (C) Ten-
683 tacle sheaths (white: absent; black: present). (D) Connexions between adradial and
684 meridional canals (white: all far from the aboral extremity of meridional canals;
685 green: subtentacular connexions far from the aboral extremity and subpharyn-
686 geal connexions at the aboral extremities; black: all connexions at or very close
687 to the aboral extremity of meridional canals). (E) Interradial canals (white: absent;
688 black: present). (F) Ciliated grooves along comb rows (white: absent; black: present).
689 Bars containing more than one colour (A–D) indicate ambiguous character recon-
690 struction. In this tree, the Mertensiidae are taken *sensu lato*, i.e. including the genus
691 *Euplokamis* as suggested by our molecular analyses.

680 hard to collect, extremely delicate and challenging to preserve,
681 a significant enrichment of taxon sampling will only be possible
682 through a global collecting effort from the scientific community. In
683 parallel, increasing the number of informative positions with care-
684 ful addition and analysis of variable molecular markers should help
685 solve the above-mentioned persisting uncertainties.

686 Acknowledgments

687 We thank our collaborators who helped us with field sam-
688 pling: Evelyn Houlston for hosting us in the Station Biologique de
689 Villefranche-sur-mer, Jean-Yves Carval for sea access in the bay
690 of Villefranche-sur-mer and Olivier Gros for providing laboratory
691 facilities at the Université des Antilles et de la Guyane. We thank
692 Claudia Mills for discussions on ctenophore identification as well as

Oliver Voigt for his advice on doublet models and rRNA secondary structure. Many thanks to Cyrielle Dayraud and Alexandre Alié who contributed to sampling. This study was funded by grants from the Institut Universitaire de France and by the Université Pierre et Marie Curie to M.M.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.zool.2014.06.004>.

Appendix B. Supplementary phylogenies

Supplementary phylogenetic trees associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.zool.2014.06.004>.

References

- Abouheif, E., Zardoya, R., Meyer, A., 1998. Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. *J. Mol. Evol.* 47, 394–405.
- Alié, A., Leclère, L., Jager, M., Dayraud, C., Chang, P., Le Guyader, H., Quéinnec, E., Manuel, M., 2011. Somatic stem cells express Piwi and Vasa genes in an adult ctenophore: ancient association of germline genes with stemness. *Dev. Biol.* 350, 183–197.
- Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J., Le Parco, Y., 2001. Sponge paraphyly and the origin of Metazoa. *J. Evol. Biol.* 14, 171–179.
- Borisenko, I., Ereskovsky, A.V., 2013. Tentacular apparatus ultrastructure in the larva of *Bolinopsis infundibulum* (Lobata: Ctenophora): tentacular apparatus ultrastructure in Ctenophora. *Acta Zool.* 94, 193–202.
- Carré, C., Carré, D., 1989. *Haeckelia bimaculata* sp. nov., une nouvelle espèce méditerranéenne de ctenophore (Cydippida Haekelidae) pourvue de cnidocystes et de pseudocolloblastes. *Comptes Rendus Acad. Sci.* 308 (Series III), 321–327.
- Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.
- Cavalier-Smith, T., Allsopp, M., Chao, E.E., Boury-Esnault, N., Vacelet, J., 1996. Sponge phylogeny, animal monophyly, and the origin of the nervous system: 18S rRNA evidence. *Can. J. Zool.* 74, 2031–2045.
- Chow, S., Ueno, Y., Toyokawa, M., Oohara, I., Takeyama, H., 2008. Preliminary analysis of length and GC content variation in the ribosomal first internal transcribed spacer (ITS1) of marine animals. *Mar. Biotechnol.* 11, 301–306.
- Chun, C., 1880. Die Ctenophoren des Golfes von Neapel und der angrenzenden Meeres-Abschnitte: eine Monographie. W. Engelmann, Leipzig.
- Collins, A.G., 1998. Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence. *Proc. Natl. Acad. Sci. U.S.A.* 95, 15458–15463.
- Collins, A.G., 2002. Phylogeny of Medusozoa and the evolution of cnidarian life cycles. *J. Evol. Biol.* 15, 418–432.
- Collins, A.G., Schuchert, P., Marques, A.C., Jankowski, T., Medina, M., Schierwater, B., 2006. Medusozoa phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. *Syst. Biol.* 55, 97–115.
- Costello, J.H., Bayha, K.M., Mianzan, H.W., Shiganova, T.A., Purcell, J.E., 2012. Transitions of *Mnemiopsis leidyi* (Ctenophora: Lobata) from a native to an exotic species: a review. *Hydrobiologia* 690, 21–46.
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Method* 9, 772. <http://dx.doi.org/10.1038/nmeth.2109>.
- Dayraud, C., Alié, A., Jager, M., Chang, P., Le Guyader, H., Manuel, M., Quéinnec, E., 2012. Independent specialisation of myosin II paralogs in muscle vs. non-muscle functions during early animal evolution: a ctenophore perspective. *BMC Evol. Biol.* 12, 107. <http://dx.doi.org/10.1186/1471-2148-12-107>.
- Dohrmann, M., Wörheide, G., 2013. Novel scenarios of early animal evolution – is it time to rewrite textbooks? *Integr. Comp. Biol.* 53, 503–511.
- Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sørensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452, 745–749.
- Erpenbeck, D., Nichols, S.A., Voigt, O., Dohrmann, M., Degnan, B.M., Hooper, J.N.A., Wörheide, G., 2007. Phylogenetic analyses under secondary structure-specific substitution models outperform traditional approaches: case studies with diploblast LSU. *J. Mol. Evol.* 64, 543–557.
- Ghada, B., Ahmed, B.A., Messaoud, M., Amel, S.-H., 2013. Genetic diversity and molecular evolution of the internal transcribed spacer (ITS) of nuclear ribosomal DNA in the Tunisian fig cultivars (*Ficus carica* L.; Moraceae). *Biochem. Syst. Ecol.* 48, 20–33.
- Gillespie, J.J., Yoder, M.J., Wharton, R.A., 2005. Predicted secondary structure for 28S and 18S rRNA from Ichneumonidea (Insecta: Hymenoptera: Apocrita): impact on sequence alignment and phylogeny estimation. *J. Mol. Evol.* 61, 114–137.
- Glennier, H., Hansen, A.J., Sørensen, M.V., Ronquist, F., Huelsenbeck, J.P., Willerslev, E., 2004. Bayesian inference of the metazoan phylogeny: a combined molecular and morphological approach. *Curr. Biol.* 14, 1644–1649.
- Granhag, L., Majaneva, S., Møller, L., 2012. First recordings of the ctenophore *Euplokamis* sp. (Ctenophora, Cydippida) in Swedish coastal waters and molecular identification of this genus. *Aquat. Inv.* 7, 455–463.
- Halanych, K.M., 2004. The new view of animal phylogeny. *Ann. Rev. Ecol. Evol. Syst.* 35, 229–256.
- Harbison, G.R., 1985. On the classification and evolution of the Ctenophora. In: Morris, S.C., George, J.D., Gibson, R., Platt, H.M. (Eds.), *The Origins and Relationships of Lower Invertebrates*. Oxford University Press, Oxford, pp. 78–100.
- Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martínez, P., Baguna, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W.E.G., Seaver, E., Wheeler, W.C., Martindale, M.Q., Giribet, G., Dunn, C.W., 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. R. Soc. B* 276, 4261–4270.
- Hernandez-Nicaise, M.L., Franc, J.M., 1994. Embranchement des Cténaïres. In: Grassé, P.P. (Ed.), *Traité de Zoologie*. Masson et Cie, Paris, pp. 943–1075.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Hyman, L.H., 1940. Ctenophora. In: Hyman, L.H. (Ed.), *The Invertebrates: Protozoa Through Ctenophora*. McGraw Hill, New York, pp. 662–696.
- Jager, M., Quéinnec, E., Chiori, R., Le Guyader, H., Manuel, M., 2008. Insights into the early evolution of SOX genes from expression analyses in a ctenophore. *J. Exp. Zool. B: Mol. Dev. Evol.* 310B, 650–667.
- Jager, M., Chiori, R., Alié, A., Dayraud, C., Quéinnec, E., Manuel, M., 2011. New insights on ctenophore neural anatomy: immunofluorescence study in *Pleurobrachia pileus* (Müller, 1776). *J. Exp. Zool. B: Mol. Dev. Evol.* 316B, 171–187.
- Kjer, K.M., 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Mol. Phylog. Evol.* 4, 314–330.
- Kohn, A.B., Citarella, M.R., Kocot, K.M., Bobkova, Y.V., Halanych, K.M., Moroz, L.L., 2012. Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore, *Pleurobrachia bachei*. *Mol. Phylog. Evol.* 63, 203–207.
- Komai, T., 1922. Studies on two aberrant ctenophores. In: *Coeleplana and Gastrodes*. The author, Kyoto, Japan.
- Komai, T., 1934. On the structure of Ctenoplana. *Kyoto Univ. Col. Sci. Mem.* B 9, 245–256.
- Lang, A., 1884. Die Polycladen (Seeplanarien) des Golfes von Neapel und der angrenzenden Meeresabschnitte: Eine Monographie. W. Engelmann, Leipzig.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lartillot, N., Philippe, H., 2004. CAT: un modèle phylogénétique Bayésien permettant de prendre en compte l'hétérogénéité des processus de substitution entre sites dans les alignements protéiques. *Biosystema* 22, 97–104.
- Lucic, D., Pesticor, B., Malej, A., Lopez-Lopez, L., Drakulovic, D., Onofri, V., Miloslavic, M., Gangai, B., Onofri, I., Benovic, A., 2011. Mass occurrence of the ctenophore *Bolinopsis vitrea* (L. Agassiz, 1860) in the nearshore southern Adriatic Sea (Kotor Bay, Montenegro). *Environ. Monit. Assess.* 184, 4777–4785.
- Maddison, W.P., Maddison, D.R., 2010. Mesquite: A Modular System for Evolutionary Analysis. Version 2.73. <http://mesquiteproject.org>
- Madin, L.P., Harbison, G.R., 1978. *Thalassocalyce inconstans*, new genus and species, an enigmatic ctenophore representing a new family and order. *Bull. Mar. Sci.* 28, 680–687.
- Mallatt, J., Craig, C.W., Yoder, M.J., 2010. Nearly complete rRNA genes assembled from across the metazoan animals: effects of more taxa, a structure-based alignment, and paired-sites evolutionary models on phylogeny reconstruction. *Mol. Phylog. Evol.* 55, 1–17.
- Manuel, M., 2009. Early evolution of symmetry and polarity in metazoan body plans. *C. R. Biol.* 332, 184–209.
- Markova, S.V., Burakova, L.P., Golz, S., Malikova, N.P., Frank, L.A., Vysotski, E.S., 2012. The light-sensitive photoprotein berovin from the bioluminescent ctenophore *Beroë abyssicola*: a novel type of Ca²⁺ regulated photoprotein: recombinant calcium-regulated photoprotein berovin. *FEBS J.* 279, 856–870.
- Martindale, M.Q., Henry, J.Q., 1999. Intracellular fate mapping in a basal metazoan, the ctenophore *Mnemiopsis leidyi*, reveals the origins of mesoderm and the existence of indeterminate cell lineages. *Dev. Biol.* 214, 243–257.
- Matsumoto, G.I., 1999. *Coeloplana thomsoni* sp. nov., a new benthic ctenophore (Ctenophora: Patyctenida: Coeloplanidae) from Western Australia. In: Walker, D.J., Wells, F.E. (Eds.), *The Seagrass Flora and Fauna of Rottnest Island, Western Australia*. Western Australian Museum, Perth, pp. 385–393.
- Maxmen, A., 2013. Comb jelly genetics suggest a radical redrawing of the tree of life. *Sci. News* 183, 20.
- Medlin, L., Elwood, H.J., Stickle, S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* 71, 491–499.
- Mills, C.E., 1987. Revised classification of the genus *Euplokamis* Chun, 1880 (Ctenophora: Cydippida: Euplokamidae n. fam.) with a description of the new species *Euplokamis dunlapae*. *Can. J. Zool.* 65, 2661–2668.

Please cite this article in press as: Simion, P., et al., Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora. *Zoology* (2014), <http://dx.doi.org/10.1016/j.zool.2014.06.004>

- 853 Mills, C.E., Electronic internet document available at <http://faculty.washington.edu/cemills/Ctenolist.html>. Published by the author, web page established March 1998. Last updated 15.05.12 2012. Phylum Ctenophora: list of all valid species names.
- 854
855
856
857 Mortensen, T., 1912. The Danish ingolf-expedition. Ctenophora, vol. V. Bianco Luno, Copenhagen.
- 858
859 Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., Müller, W.E.G., Nickel, M., Schierwater, B., Vacelet, J., Wiens, M., Wörheide, G., 2013. Deep metazoan phylogeny: when different genes tell different stories. *Mol. Phyl. Evol.* 67, 223–233.
- 860
861
862
863 Oliveira, O.M.P., Mianzan, H., Migotto, A.E., Marques, A.C., 2007. Chave de identificação dos Ctenophora da costa brasileira. *Biota Neotrop.* 7, 341–350.
- 864
865 Ospovat, M.F., 1985. On phylogeny and classification of the type Ctenophora. *Zool. žurnal* 647, 965–974.
- 866
867 Pang, K., Martindale, M.Q., 2008a. Ctenophores. *Curr. Biol.* 18, R1119–R1120.
- 868
869 Pang, K., Martindale, M.Q., 2008b. Developmental expression of homeobox genes in the ctenophore *Mnemiopsis leidyi*. *Dev. Genes Evol.* 218, 307–319.
- 870
871 Pang, K., Ryan, J.F., Baxevasis, A.D., Martindale, M.Q., 2011. Evolution of the TGF- β signaling pathway and its potential role in the ctenophore, *Mnemiopsis leidyi*. *PLoS ONE* 6, e24152. <http://dx.doi.org/10.1371/journal.pone.0024152>.
- 872
873 Pennisi, E., 2013. Nervous system may have evolved twice. *Science* 339, 391.
- 874
875 Peterson, K.J., Eernisse, D.J., 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. *Evol. Dev.* 3, 170–205.
- 876
877
878 Pett, W., Ryan, J.F., Pang, K., Mullikin, J.C., Martindale, M.Q., Baxevasis, A.D., Lavrov, D.V., 2011. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: insight from mtDNA and the nuclear genome. *Mitochondrial DNA* 22, 130–142.
- 879
880
881
882 Philippe, H., Roure, B., 2011. Difficult phylogenetic questions: more data, maybe; better methods, certainly. *BMC Biol.* 9, 91. <http://dx.doi.org/10.1186/1741-7007-9-91>.
- 883
884
885
886 Philippe, H., Telford, M.J., 2006. Large-scale sequencing and the new animal phylogeny. *Trends Ecol. Evol.* 21, 614–620.
- 887
888
889
890
891
892
893
894 Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houlston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D.J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G., Manuel, M., 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* 19, 706–712.
- 895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
- Pick, K.S., Philippe, H., Schreiber, F., Erpenbeck, D., Jackson, D.J., Wrede, P., Wiens, M., Alie, A., Morgenstern, B., Manuel, M., Wörheide, G., 2010. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol. Biol. Evol.* 27, 1983–1987.
- Podar, M., Haddock, S.H., Sogin, M.L., Harbison, G.R., 2001. A molecular phylogenetic framework for the phylum Ctenophora using 18S rRNA genes. *Mol. Phyl. Evol.* 21, 218–230.
- Purcell, J.E., 2012. Jellyfish and ctenophore blooms coincide with human proliferations and environmental perturbations. *Ann. Rev. Mar. Sci.* 4, 209–235.
- Rokas, A., King, N., Finnerty, J., Carroll, S.B., 2003. Conflicting phylogenetic signals at the base of the metazoan tree. *Evol. Dev.* 5, 346–359.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.-D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., NISC Comparative Sequencing Program, Smith, S.A., Putnam, N.H., Haddock, S.H.D., Dunn, C.W., Wolfsberg, T.G., Mullikin, J.C., Martindale, M.Q., Baxevasis, A.D., 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* 342, 1242592. <http://dx.doi.org/10.1126/science.1242592>.
- Stamatakis, A., 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Song, J.-I., Hwang, S.-J., Lee, S.-H., Park, J.-K., 2011. New records of creeping ctenophores, genus *Coeloplana* (Tentaculata: Platyctenida: Coeloplanidae), from Korea. *Korean J. Syst. Zool.* 27, 47–52.
- Telford, M.J., Copley, R.R., 2011. Improving animal phylogenies with genomic data. *Trends Genet.* 27, 186–195.
- Trégouboff, G., Rose, M., 1957. *Manuel de Planctologie Méditerranéenne*. Editions du CNRS, Paris.
- Voigt, O., Erpenbeck, D., Wörheide, G., 2008. Molecular evolution of rDNA in early diverging Metazoa: first comparative analysis and phylogenetic application of complete SSU rDNA secondary structures in Porifera. *BMC Evol. Biol.* 8, 69. <http://dx.doi.org/10.1186/1471-2148-8-69>.
- Yamada, A., Martindale, M.Q., Fukui, A., Tochinai, S., 2010. Highly conserved functions of the Brachyury gene on morphogenetic movements: insight from the early-diverging phylum Ctenophora. *Dev. Biol.* 339, 212–222.
- Zhao, G.H., Li, J., Mo, X.H., Li, X.Y., Lin, R.Q., Zou, F.C., Weng, Y.B., Song, H.Q., Zhu, X.Q., 2012. The second transcribed spacer rDNA sequence: an effective genetic marker for inter-species phylogenetic analysis of trematodes in the order Strigeata. *Parasitol. Res.* 111, 1467–1472.
- Zrzavy, J., Mihulka, S., Kepka, P., Bezdek, A., Tietz, D., 1998. Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. *Cladistics* 14, 249–285.

Erratum

Les corrections suivantes seront appliquées aux épreuves précédentes, avant leur publication :

Q2: "Université" avec un "U" majuscule.

Q4: ajout de la référence suivante dans la bibliographie:

Schierwater, B., Eitel, M., Jakob, W., Osigus, H.-J., Hadrys, H., Dellaporta, S. L., Kolokotronis, S.-O. & DeSalle, R., 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. PLoS biology 7, e20.

Q6: La dernière phrase devient : "This study was funded by a grant from the Institut Universitaire de France to M.M."

line 155: "As important results" devient "Among important results"

line 245: "only" supprimé dans "we only selected"

table 1, column 2, row 6: "St Rose" devient "Ste Rose"

table 1, column 1, row 53: "Deoipea" devient "Deiopea"

Table 2, upper legend, line 2: ajout d'un "B" après "supplementary Appendix"

Table 2, lower legend: ajout d'un caractère de séparation ";" entre "inference" and "Symbols"

Table 2, lower legend: "portioning" devient "partitioning"

line 493: ">75%" devient ">80%"

line 676-677: "Cambojjida" devient "Cambojiida"

line 692: "advice" devient "advices"

line 800: suppression de "Published by"

3.3 Compléments

3.3.1 Tentative d'analyse cladistique de caractères morpho-anatomiques

La diversité morphologique des cténophores a été formalisée en une matrice de caractères, afin de tenter de résoudre leurs relations phylogénétiques.³ Les états de caractères ont été renseignés soit à partir d'observations personnelles, soit, pour une grande partie, à partir de la littérature (voir références dans la légende de la Figure 3.1). Ce travail a eu pour résultat la constitution d'une matrice de 26 caractères pour 15 taxons (voir Figure 3.1 page ci-contre). Le résultat de l'analyse cladistique de cette matrice en maximum de parcimonie avec PAUP 4.0 est présenté en Figure 3.2 page 116.

De manière générale, la topologie obtenue à partir des données morphologiques est largement incongruente avec les résultats des analyses moléculaires (article 2 et ci-après sections 3.3.2 et 3.3.3). Deux informations principales ressortent de cette analyse : les Cestida sont proches des Lobata, ce qui a été suggéré par la quasi-totalité des auteurs précédents sur des bases morphologiques, et confirmé dans l'article 2 ; les Platyctenida et *Beroe* formeraient un groupe monophylétique supporté par la présence partagée de papilles aborales, d'une bouche largement extensible, et de l'absence de canal aboral. Cependant, on peut se demander si l'ouverture de la bouche des *Beroe* pour engloutir de grosses proies est réellement équivalente à la capacité qu'ont les platycténides de pouvoir ramper sur leur pharynx. De la même manière, l'absence de canal aboral chez les *Beroe* semble causée par l'espace disproportionné qu'occupe le pharynx par rapport au reste du corps, tandis que chez les Platyctenida, cette absence semble causé par l'applatissage du corps dans l'axe oral-aboral.

Cette analyse est un échec relatif pour plusieurs raisons : 1) Il y a peu de caractères morpho-anatomiques facilement formalisables chez cténophores ; 2) La description de certains taxons est insuffisamment détaillée et empêche la détermination de nombreux états de caractères (d'où le grand nombre d'états indéterminés dans la matrice présentée Figure 3.1) ; 3) Certains caractères correspondent à des autapomorphies, d'autres sont des caractères continus et potentiellement subjectifs, laissant un nombre réduit de caractères phylogénétiquement informatifs (22 sur 26 dans la matrice analysée) ; 4) L'utilisation d'OTUs à l'échelle des genres, familles et ordres, empêche la représentation de la diversité au sein de ces OTU, qui pourrait aider à la détection des homoplasies. Il est à noter que dans la matrice analysée ici, le polymorphisme au sein des taxons terminaux a été géré par une approche de type *exemplar* (choix d'une espèce au sein du taxon terminal), généralement préférable quand la phylogénie au sein des taxons terminaux n'est pas connue.

Ce travail préliminaire est bien entendu améliorable, par la redéfinition de certains caractères,

3. La majeure partie de ce travail a été réalisée par un étudiant en stage de Master 2, Nicolas Bekkouche, ma contribution s'étant limitée à amender la matrice et à la ré-analyser.

	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
Mertensidae	1	0	0	0	0	0	1	0	0	1	2	1	1	1	0	0	0	0	?	1	?	0	0	1	2	0
Euplokamididae	1	0	0	0	0	0	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0
Platyctenida	1	0	1	-	1	1	0	-	0	1	1	1	1	1	0	0	0	1	?	-	1	0	1	?	2	1
Lampeidae	?	?	0	0	0	0	1	1	?	1	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	?
Horniphora	1	1	0	0	0	0	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0
Pleurobrachia	1	1	0	0	0	0	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0
Beroe	0	1	1	1	1	1	0	-	-	0	-	0	-	-	-	-	0	1	0	0	1	0	0	0	1	?
Haekeliidae	0	1	1	0	0	0	1	1	1	1	2	1	0	0	-	-	0	1	?	0	1	0	0	0	0	?
Cestida	0	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	0	0	1	2	0	0	0	1	1	1
Thalassocalycida	1	1	0	0	0	0	1	0	0	1	2	0	0	1	0	0	0	0	?	0	?	1	0	?	2	1
Ocyropsis	0	1	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	2	0	1	0	1	1	1
Mnemiopsis	0	1	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	2	0	1	0	1	1	1
Leucothea	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	2	0	1	0	1	0	1	1
Deiopea	0	1	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	2	0	1	0	1	1	1
Bolinopsis	0	1	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	2	0	1	0	1	1	1

Description des caractères:

- 1: Canaux perradiaires. 0=absent; 1=présent
- 2: Canaux interradiaires. 0=absent; 1=présent
- 3: Connexion d'au moins un canal adradiaire au niveau de l'extrémité aborale d'un canal méridien. 0=absent; 1=présent
- 4: Anastomoses entre les canaux paragastriques et les canaux méridiens. 0=absent; 1=présent
- 5: Complexe circumoral. 0=absent; 1=présent
- 6: Diverticules sur les canaux méridiens. 0=absent; 1=présent
- 7: Canal aboral. 0=absent; 1=présent
- 8: Allongement du canal aboral. 0=absent; 1=présent
- 9: Canal synradiaire. 0=absent; 1=présent
- 10: Paire de tentacules. 0=absent; 1=présent
- 11: Position de la racine tentaculaire. 0=centrale; 1=orale; 2=aborale
- 12: Gaine tentaculaire. 0=absent; 1=présent
- 13: Gaines tentaculaires débouchent du côté aboral. 0=non; 1=oui
- 14: Tentilles. 0=absent; 1=présent
- 15: Tentilles enroulées au repos. 0=absent; 1=présent
- 16: Tentacules secondaires disposées le long d'une marge orale. 0=absent; 1=présent
- 17: Organe aboral profondément enfoncé au niveau du pôle aboral. 0=absent; 1=présent
- 18: Papilles aborales. 0=absent; 1=présent
- 19: Sillon cilié parcourant toute la rangée de peignes. 0=absent; 1=présent
- 20: Taille relative des rangées de peignes. 0=identiques; 1=rangées subtentaculaires plus longues; 2=rangées subpharyngiennes plus longues
- 21: Bouche largement extensible. 0=absent; 1=présent
- 22: Lobes. 0=absent; 1=présent
- 23: Lèvres orales. 0=absent; 1=présent
- 24: Nourrissage par tournoiement (chez les juvéniles pour les Lobata). 0=absent; 1=présent
- 25: Sens de compression du corps. 0=pas de compression; 1=dans le plan tentaculaire; 2=dans le plan pharyngien
- 26: Distribution des gonades dans les canaux méridiens. 0=continue; 1=discontinue

FIGURE 3.1 – Matrice de 26 caractères morphologiques chez les cténophores pour 15 OTUs. Certaines OTUs sont des ordres (Platyctenida, Cestida, Thalassocalycida), d'autres des familles (Mertensidae, Euplokamididae, etc) et d'autres des genres (*Beroe*, *Pleurobrachia*, etc). Sources bibliographiques utilisées pour le codage des caractères : Chun 1880; Delage and Hérouard 1901; Mayer 1912; ?; Komai 1922; Mortensen 1932; Dawydoff 1946; Trégouboff and Rose 1957; Madin and Harbison 1978; Harbison 1985; Ospovat 1985b; Brusca and Brusca 2003.

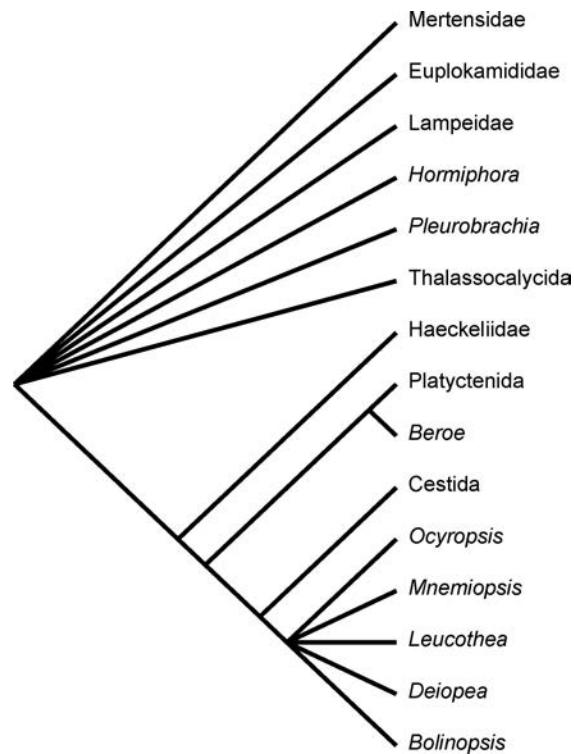


FIGURE 3.2 – Arbre phylogénétique des cténophores reconstruit en maximum de parcimonie à partir de la matrice morphologique présentée en Figure 3.1 page précédente. La topologie présentée est le consensus strict des 6 arbres minimaux (de 43 pas). Remarquez l’absence de résolution entre les familles de l’ordre des “Cydippida”.

l’ajout de quelques autres et l’utilisation d’OTUs à l’échelle des espèces, mais de toute manière, la morphologie générale des cténophores est clairement une source d’information phylogénétique très pauvre.

3.3.2 Apports des analyses phylogénomiques de l’article 1 pour la phylogénie des cténophores

Les analyses phylogénomiques réalisées au chapitre précédent (voir article 1) possèdent, grâce à nos efforts de séquençage, un échantillonnage taxonomique au sein des cténophores suffisant pour renseigner certaines de leur relations de parenté. Le premier résultat positif est que ces analyses sont toutes largement congruentes entre elles (Figure 3.3). Seules exceptions, les positions de *Lampea panzerina* et de l’espèce indéterminée “sp3” sont instables. *Lampea* serait soit le groupe-frère des Platyctenida (Figure 3.3 B et C), soit divergerait après ces derniers (Figure 3.3 A). Ces analyses sont particulièrement intéressantes, car elles résolvent les relations entre les Pleurobrachiidae, les Beroida et les Lobata + Cestida. Les Beroida sont ainsi plus proches des Lobata + Cestida que des Pleurobrachiidae, et au sein des Lobata, *Mnemiopsis* semble plus proche des Cestida que de *Bolinopsis*, rendant les Lobata paraphylétiques par rapport aux Cestida. Ces analyses suggèrent également que *Euplokamis dunlapae* représente le groupe-frère des autres cténophores échantillon-

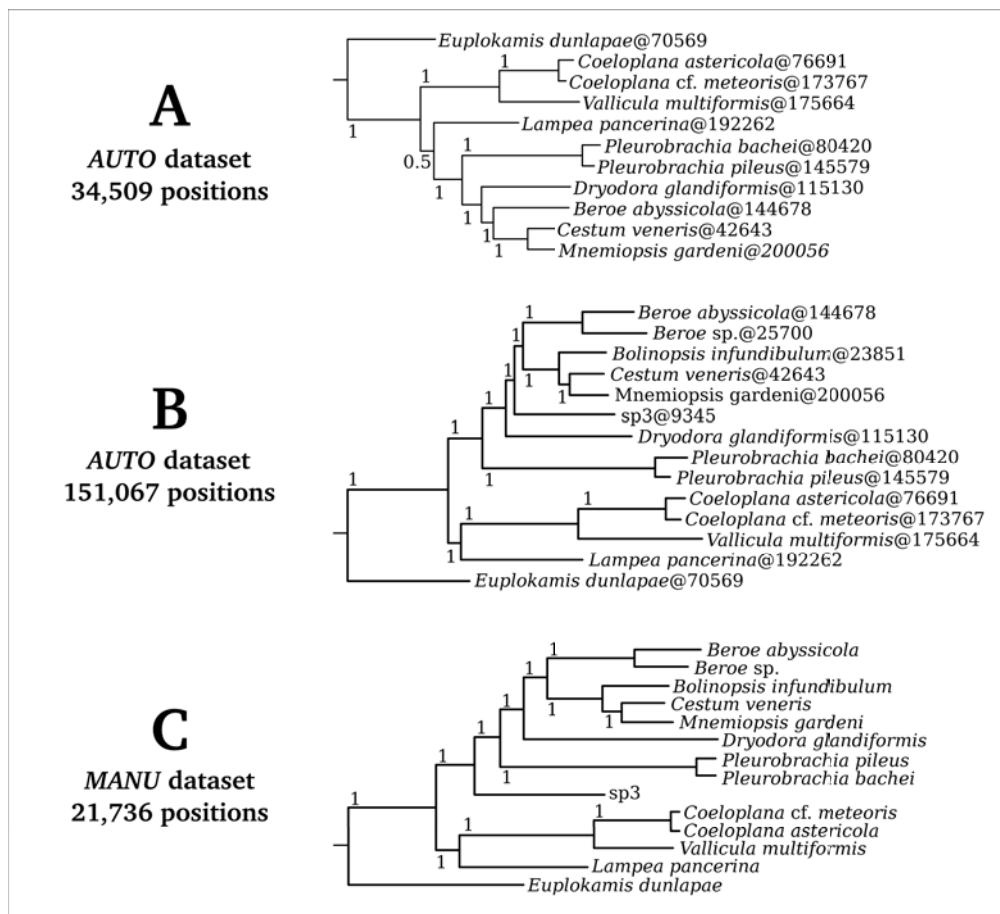


FIGURE 3.3 – Relations phylogénétiques au sein des cténophores telles que reconstruites lors des analyses phylogénomiques réalisées dans l'article 1. La position de *Lampea pancerina* (ainsi que d'un cténophore indéterminé "sp3") n'est pas résolue, mais toutes les analyses supportent l'existence d'un clade regroupant tous les cténophores sauf *Euplokamis dunlapae*. Pour les informations taxonomiques, voir article 2, Table 1.

nés. Enfin, l'existence de données transcriptomiques pour *Dryodora glandiformis* permet pour la première fois de déterminer sa place dans l'arbre des cténophores (aucune séquence de d'ADNr 18S n'est disponible pour cette espèce). Bien que basées sur un échantillonnage d'espèces limité, ces analyses sont parfaitement complémentaires de celles réalisées dans l'article 2, et apportent un éclairage nouveau sur les relations de parenté qui justement n'avaient pas pu être résolue précédemment.

3.3.3 Protocole dédié à la résolution de la position de la racine des cténophores

Le positionnement de la racine des cténophores est très instable lorsque l'on utilise le gène ribosomique ADNr 18S (voir article 2). Ce manque de résolution phylogénétique est dû au très fort éloignement des groupes externes utilisés pour enracer les cténophores, et au faible nombre, voir à l'absence, de caractères moléculaires informatifs à ce niveau de l'arbre (voir discussion article 2).

D'un autre côté, les analyses phylogénomiques résumées au paragraphe précédent suggèrent que le genre *Euplokamis* (membre de la famille des Mertensiidae d'après les résultats de l'article 2) est le groupe-frère du reste des cténophores. Cependant, il faut rester sceptique face à ce résultat : les supermatrices utilisées sont loin d'être insensibles aux biais systématiques (voir article 1), et l'échantillonnage utilisé ne contient que des gènes très conservés à l'échelle des métazoaires, et donc potentiellement faiblement informatifs pour résoudre les relations au sein des cténophores. Il est donc important de tester la position de cette racine par des approches alternatives.

Utilisation de duplications cténophores-spécifiques L'utilisation de groupes externes pose le problème de la très grande distance génétique existant entre les cténophores et les non-cténophores. Afin de contourner ce problème qui fragilise le positionnement de la racine des cténophores, il est possible d'utiliser un cas particulier qui nous permettrait d'enraciner les cténophores à partir d'un autre groupe de cténophores. Immédiatement après la duplication d'un gène chez un ancêtre, chacune des deux copies de ce gène va évoluer de manière indépendante chez les descendants. Si un gène a été dupliqué chez un ancêtre exclusif des cténophores actuels, avant la radiation évolutive de ces derniers, alors chaque cténophore possèdera deux versions de ce gène. La reconstruction de la phylogénie de ce gène dupliqué devrait faire apparaître deux groupes d'orthologie, chacun de ces groupes enracinant l'autre. Plus la duplication aura eu lieu récemment avant la radiation évolutive des cténophores actuels, plus la branche séparant ces deux groupes d'orthologie sera courte, et moins la position de la racine sera affectée par l'artéfact d'attraction des longues branches. Le corollaire négatif de ce protocole est qu'il implique la réalisation de phylogénies "simple-gène", qui pourraient ne pas contenir assez de signal phylogénétique pour résoudre les relations de parenté à l'échelle de l'embranchement des Ctenophora.

Duplication du gène *arp23* Parmi les gènes utilisés pour les analyses du chapitre précédent se trouve le gène *arp23*, dont l'examen de la phylogénie simple-gène a montré la caractéristique recherchée : toutes les espèces de cténophores possèdent deux versions du même gène, suggérant fortement l'existence d'une duplication cténophore-spécifique (les autres métazoaires n'en possèdent qu'une seule version)⁴. Seule l'espèce *Euplokamis dunlapae* ne possède qu'une seule version du gène.⁵ La reconstruction de la phylogénie à partir d'un alignement de séquences de la protéine *arp23*, longue de 384 acides aminés, a été réalisée avec le logiciel PhyML, sans groupe externe, et avec les paramètres suivants : utilisation du modèle d'évolution LG+Γ6, fréquences stationnaires

4. Lors de l'utilisation de ce gène dans le chapitre précédent, l'un des deux groupes d'orthologie avait bien entendu été supprimé, pour ne pas fausser les analyses phylogénétiques.

5. Il est théoriquement possible que *Euplokamis dunlapae* possède en réalité le second paralogue de ce gène, mais qu'il soit absent de son transcriptome. Néanmoins, cette possibilité est très improbable car le gène *arp23* est très fortement exprimé chez les métazoaires, le transcriptome d'*Euplokamis dunlapae* est de bonne qualité et présente une bonne couverture, et enfin, les deux paralogues de ce gène sont présents dans les transcriptomes de tous les autres cténophores, y compris dans ceux qui sont beaucoup moins complets.

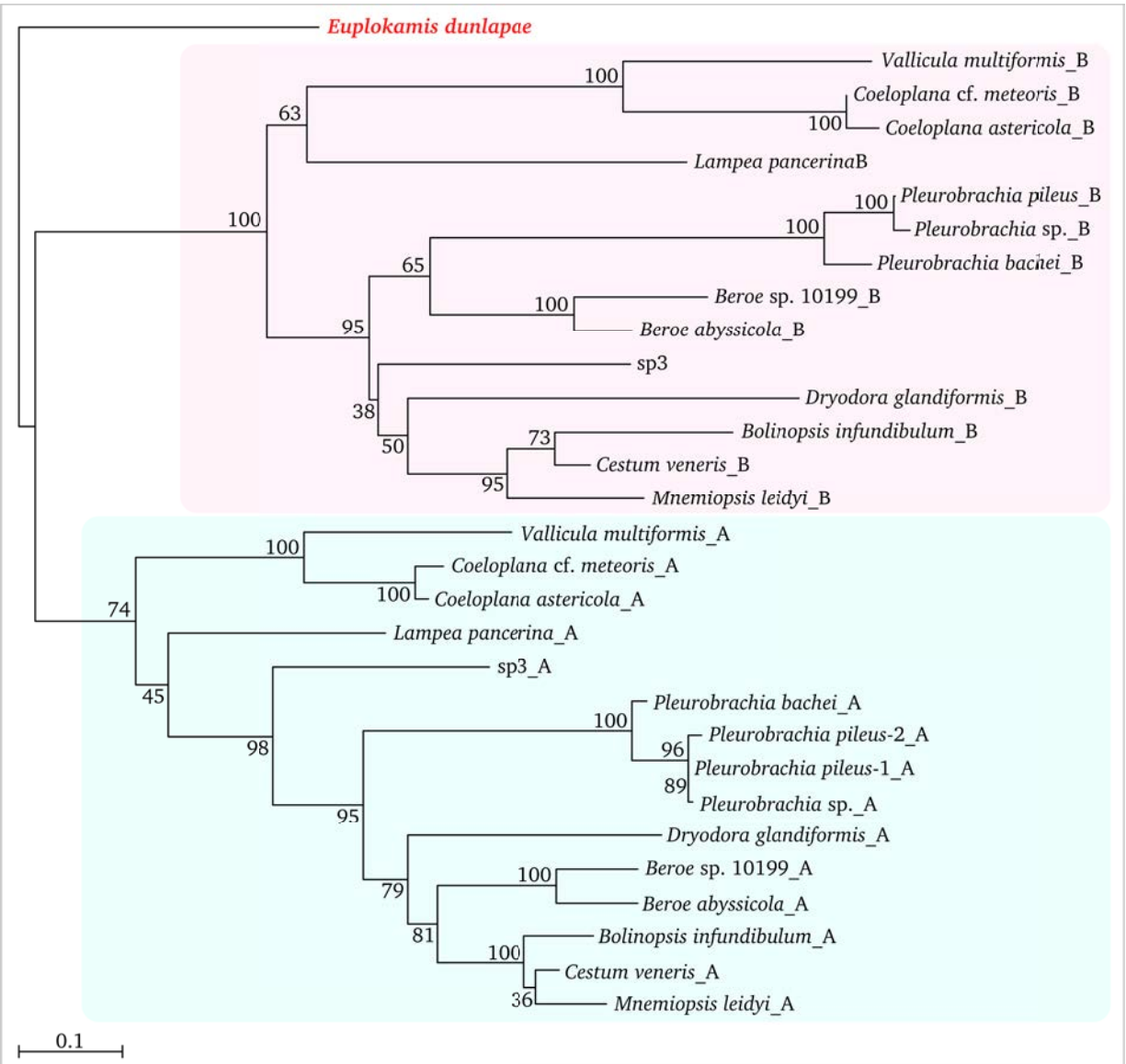


FIGURE 3.4 – Phylogénie du gène dupliqué *arp23*, reconstruite en Maximum de Vraisemblance (PhyML) avec le modèle LG+Γ6. Les valeurs de bootstraps sont représentées à chaque noeud. L'espèce *Euplokamis dunlapae*, représentée en rouge, ne possède qu'une seule version du gène. Les deux groupes d'orthologie sont encadrés en rouge et en vert.

déterminées empiriquement, estimation de la proportion de sites invariants, topologie de départ déterminée par BioNJ. Un total de 100 bootstraps ont ensuite été réalisés.

Les résultats de cette analyse sont extrêmement intéressants (Figure 3.4). Les deux groupes d'orthologie sont séparés par une branche courte, et l'espèce *Euplokamis dunlapae*, qui ne possède qu'un seul des deux paralogues, se positionne entre ces deux groupes. Cette topologie suggère très fortement que la duplication du gène a eu lieu chez les cténophores après la divergence du genre *Euplokamis*, ce qui implique que le genre *Euplokamis* serait le groupe-frère des autres cténophores représentés dans cette analyse. Ensuite, comme cela est visible dans chacun des deux groupes d'orthologie, les Platyctenida et *Lampea* divergent d'un groupe comprenant tous les cténophores

restants. Les deux groupes d'orthologie suggèrent des positions contradictoires du genre *Lampea*, mais aucune des deux n'est statistiquement soutenue (scores de bootstraps de 63 et 45). De manière attendue, les relations internes aux cténophores ne sont pas toutes résolues. En effet, tout comme les marqueurs ADNr 18S, ITS1, ADNr 5.8S et ITS2 n'avaient pas suffi à résoudre toutes les relations au sein des cténophores, il était prédictible que l'analyse du seul gène *arp23* ne comporte pas non plus assez de signal phylogénétique pour résoudre toutes les relations. Néanmoins, tous les noeuds soutenus (ç.-à-d. dont les supports de bootstraps sont supérieurs à 80) sont congruents à la fois avec les relations de parenté retrouvées avec les marqueurs ribosomiques (article 2) et avec celles retrouvées par les supermatrices phylogénomiques élaborées au chapitre précédent (voir article 1 et section 3.3.2).

Protocole de détection des duplications La confirmation de ces résultats par différents gènes est nécessaire, et potentiellement réalisable grâce au séquençage de données transcriptomiques pour un certain nombre de cténophores. J'ai donc mis en place un protocole afin de créer des clusters de séquences homologues, puis d'y détecter des duplications.

J'ai utilisé les transcriptomes de cténophores assemblés dans le chapitre précédent (après l'étape de dé-cross-contamination), pour lesquels les protéines ont été prédites à l'aide des logiciels GeneMarkS et Genemark (Besemer et al. 2001; Besemer and Borodovsky 2005). Le logiciel OrthoMCL a ensuite été utilisé (avec une granularité de 1.5) pour regrouper ces données protéiques en 19646 clusters de similarité. Des scripts Perl personnels ont été utilisés pour déterminer certaines métriques de chacun de ces clusters, comme le nombre d'espèces représentées et le nombre de séquences par espèce. Finalement, chacun des clusters où au moins 7 espèces de cténophores possédaient au moins 2 séquences ont été alignés avec l'algorithme Muscle avec les paramètres par défaut implémentés dans le logiciel Seaview. Ces alignements et leurs topologies reconstruites avec une méthode de distance ont ensuite été examinés afin de déterminer s'il s'agissait bien d'un cas de duplication où l'on pouvait identifier deux groupes d'orthologie. Cela a été le cas pour les deux clusters Cteno129 et Cteno320. Une dizaine d'autres alignements n'ont pas pu être analysés car ils correspondaient à des cas de paralogie multiple, qu'il était impossible de traiter avec notre échantillonnage taxonomique réduit.

Ce protocole possède une limite : il ne comporte pas d'étape permettant la vérification que les gènes dupliqués correspondent à des duplications cténophore-spécifiques. Si une duplication n'est en réalité pas cténophore-spécifique, mais plus ancienne, cela aura simplement pour effet de ne pas raccourcir suffisamment la longue branche des cténophores et d'empêcher une bonne résolution de la racine.

Les analyses phylogénétiques des gènes *Cteno129* et *Cteno320* ont été réalisées de la même manière que celle du gène *arp23*.

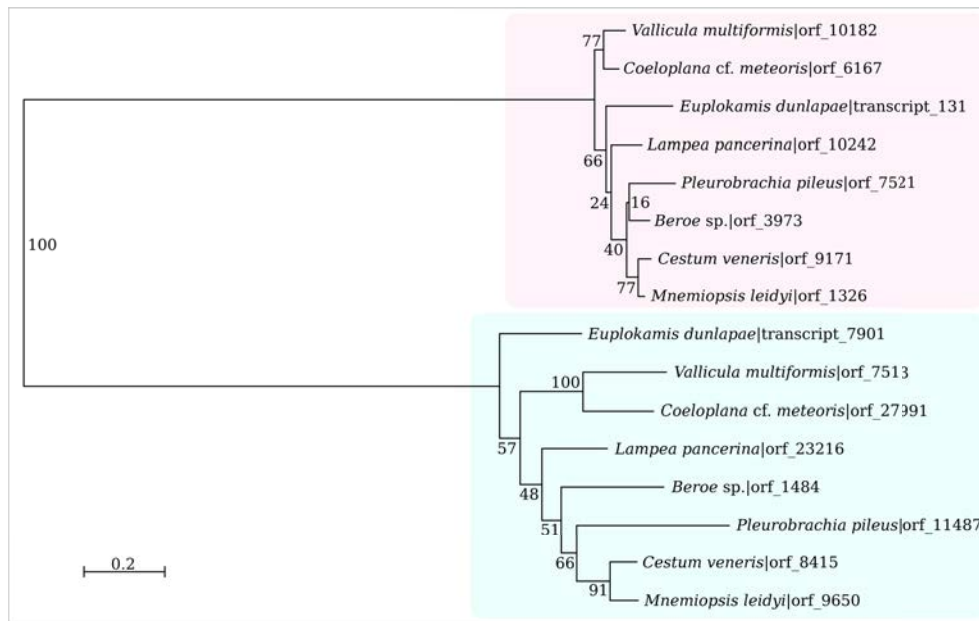


FIGURE 3.5 – Phylogénie du gène dupliqué *Cteno129*, reconstruite en Maximum de Vraisemblance (PhyML) avec le modèle LG+Γ6. Les valeurs de bootstraps sont représentées à chaque noeud. Les deux groupes d’orthologie sont encadrés en rouge et en vert. L’analyse phylogénétique du gène dupliqué *Cteno129* a été réalisée de la même manière que pour le gène *arp23*.

Duplication du gène Cteno129 (ASCC3) Les résultats de l’analyse des deux groupes d’orthologie du gène *Cteno129* suggèrent que soit *Euplokamis*, soit les Platyctenida sont le groupe-frère des autres cténophores (Figure 3.5). Le premier cas est donc en accord avec l’enracinement obtenu avec le gène *arp23*, bien que la longueur de la branche séparant les deux groupes d’orthologie soit bien plus longue que dans l’analyse précédente. Le second racinement obtenu est légèrement différent, et inverse les positions respectives d’*Euplokamis* et des Platyctenida. Aucun de ces deux racinements n’est fortement soutenu. Hormis cette différence, les topologies des deux groupes d’orthologie sont largement congruentes entre elles (mise à part la position de *Beroe* sp., qui n’est pas soutenue, ce qui est principalement dû à la faible longueur de ses séquences dans l’alignement) et sont également congruentes avec toutes les topologies intra-cténophores obtenues durant mes travaux.

Duplication du gène Cteno320 (SNRNP) Cette fois, bien que la topologie au sein des deux groupes d’orthologie du gène *Cteno320* soit identique (voir Figure 3.6 page suivante), leur enracinement respectif est contradictoire : le premier des deux groupes est enraciné sur *Euplokamis* de manière robuste (support de bootstrap de 86), en accord avec les analyses précédentes, mais le second groupe d’orthologie est enraciné sur *Mnemiopsis*. Ce second racinement, hautement contradictoire avec tous les résultats obtenus dans mon travail de thèse, est très certainement erroné. En effet, cette racine imposerait une interprétation relativement absurde de l’évolution morphologique des cténophores. La monophylie de l’ensemble Lobata (ici représenté par *Mnemiopsis*) + Cestida (*Cestum veneris*) est soutenue à la fois par la morphologie (section 3.3.1) et les données molé-

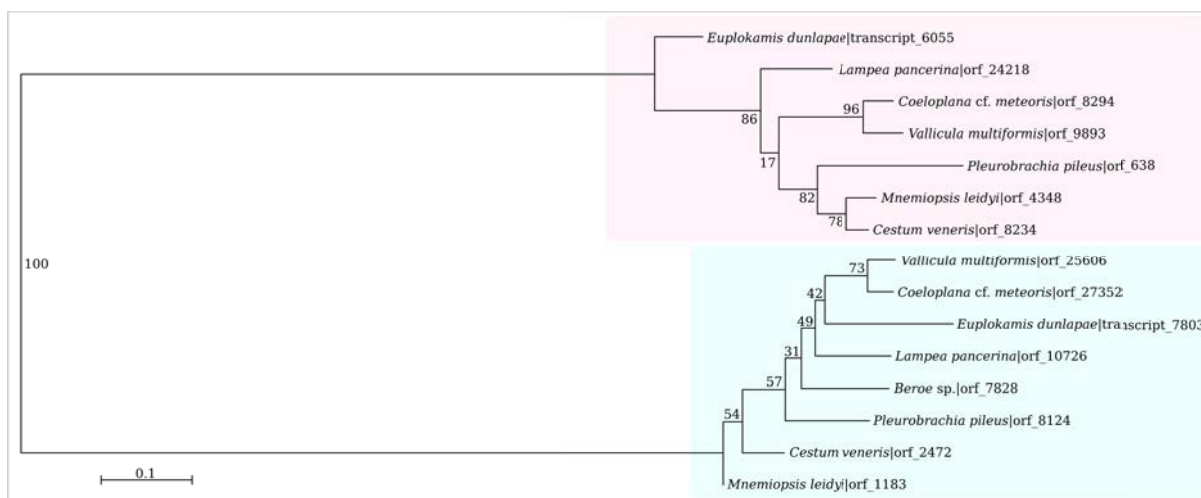


FIGURE 3.6 – Phylogénie du gène dupliqué *Cteno320*, reconstruite en Maximum de Vraisemblance (PhyML) avec le modèle LG+Γ6. Les valeurs de bootstraps sont représentées à chaque noeud. Les deux groupes d’orthologie sont encadrés en rouge et en vert. L’analyse phylogénétique du gène dupliqué *Cteno320* a été réalisée de la même manière que le gène *arp23*.

lares (article 2 et section 3.3.2). De plus, le soutien statistique de cette racine est très faible (ç.-à-d. seulement 54). L’enracinement de ce groupe d’orthologie sur *Mnemiopsis* est donc très probablement artificiel et issu d’un manque de résolution phylogénétique. Notez que, dans le premier groupe d’orthologie, le genre *Lampea* diverge cette fois avant les Platyctenida (mais de manière non soutenue), tandis que la plupart des autres analyses de cette thèse (articles 1 et 2) placent *Lampea* soit en groupe-frère des Platyctenida, soit divergeant après les Platyctenida.

La racine des cténophores Les résultats de ces trois analyses, comparées avec l’analyse des jeux de données 18S et ITS de ce chapitre (article 2) et les analyses des jeux de données phylogénomiques (article 1 et Figure 3.3), assemblent un faisceau d’arguments qui convergent tous vers la résolution de la position de la racine des cténophores : les Mertensiidae (au sens large, c’est-à-dire incluant les genres *Charistephane*, *Mertensia* et *Euplokamis*) sont très probablement le groupe-frère de l’ensemble des autres cténophores. L’ordre des Platyctenida serait le second clade à diverger des cténophores restants. La position du genre *Lampea* est, elle, toujours irrésolue : soit *Lampea* est le groupe-frère des Platyctenida, soit le troisième clade à diverger chez les cténophores.

3.3.4 Bilan de la phylogénie moléculaire et retour sur les caractères anatomiques

La combinaison de l’analyse des marqueurs 18S et ITS1 et 2 sur un large échantillonnage de cténophores, avec l’analyse de matrices phylogénomiques pour un nombre plus restreint d’espèces et avec trois analyses d’enracinement basées sur des gènes dupliqués, nous permet enfin d’appréhender en profondeur l’évolution des cténophores actuels. En effet, tous ces résultats sont compatibles

entre eux, et supportent une nouvelle phylogénie des cténophores, présentée en figure 3.7. La distribution des états de caractères morphologiques semble largement congruente avec cette phylogénie moléculaire. Des hypothèses de synapomorphies ont été reportées sur la figure 3.7 et dessinent un scénario évolutif nouveau pour les cténophores.⁶

Les analyses phylogénomiques présentées dans l'article 1 et les analyses phylogénétiques de gènes dupliqués chez les cténophores permettent en effet de résoudre les zones de l'arbre pour lesquelles les marqueurs ADNr 18S et ITS (article 2) ne possédaient pas de signal phylogénétique. Notamment, les relations de parenté entre les Pleurobrachiidae, les Beroidea + Haeckeliidae et les Lobata + Cestida + Thalassocalycida sont désormais résolues. De plus, le séquençage récent du transcriptome de *Dryodora glandiformis* nous permet désormais de déterminer sa position dans l'arbre. Enfin, la position de la racine des cténophores est désormais déterminée.

Un nouveau scénario évolutif Au sein de ce scénario, on peut noter l'existence d'un clade regroupant tous les cténophores possédant un complexe circumoral (clade 9 sur Figure 3.7), comme l'avait compris Ospovat (1985a) qui a nommé ce groupe les Cyclocoela. Également, plusieurs caractères anatomiques en lien avec la nutrition et chers à Harbison, comme la réduction de l'appareil tentaculaire ou la capacité d'extension de la bouche semblent également représenter des synapomorphies (clades 9, 11 et 12 sur Figure 3.7). Enfin, la distribution d'autres caractères morphologiques est en accord avec cette phylogénie, comme la compression du corps dans le plan pharyngien ou tentaculaire, même si l'orientation de leur évolution est pour l'instant impossible.⁷ La structure des canaux gastriques est également parfaitement congruente avec cette phylogénie moléculaire : l'existence des canaux perradiaires et adradiaires apparaît comme ancestrale aux cténophores actuels, tandis que l'acquisition des canaux interradiaires est une innovation d'un clade regroupant les Pleurobrachiidae et les Cyclocoela (clade 7 sur Figure 3.7). Plus tard, les canaux perradiaires ont été perdus par différentes espèces, mais il est pour l'instant impossible de mieux comprendre ces pertes en raison de l'irrésolution persistante des relations phylogénétique au sein du groupe unissant les Lobata, les Cestida et les Thalassocalycida (Lobata *sensu lato*, ou Panlobata).

Il est probable que l'évolution des espèces du groupe Lobata *sensu lato* (Panlobata) puisse être en partie comprise par le mode de connexion des canaux adradiaires aux canaux méridiens. Cela a été suspecté par Fanny Moser (Moser 1908) qui a remarqué que chez les Bolinopsidae et les Ocyropsidae, cette connexion se fait à l'extrémité aborale de tous les canaux méridiens, tandis que chez les Leucotheidae, *Eurhamphaea* et *Deiopea*, cette connexion se fait à l'extrémité

6. Dans cet arbre est représenté le genre *Ctenoplana*, bien qu'aucune donnée moléculaire ne soit disponible à ce jour pour ce taxon. Néanmoins, sa position phylogénétique au sein des Platyctenida est connue sur la base de ses caractéristiques morphologiques et est discutée en profondeur dans le travail de Rankin (1956). Il n'en est pas de même pour les genres *Tjalfiella*, *Lyrocteis* et *Savangia*, trois autres genres de Platyctenida, qui n'apparaissent donc pas dans cet arbre.

7. Déterminer l'état ancestral de ce caractère nécessiterait un groupe externe aux cténophores qui posséderait également ce caractère. Notre arbre est raciné, mais les états de caractères des groupes externes sont inapplicables : il est donc impossible d'orienter l'évolution des caractères qui diffèrent entre les clades 2 et 3 de la figure 3.7.

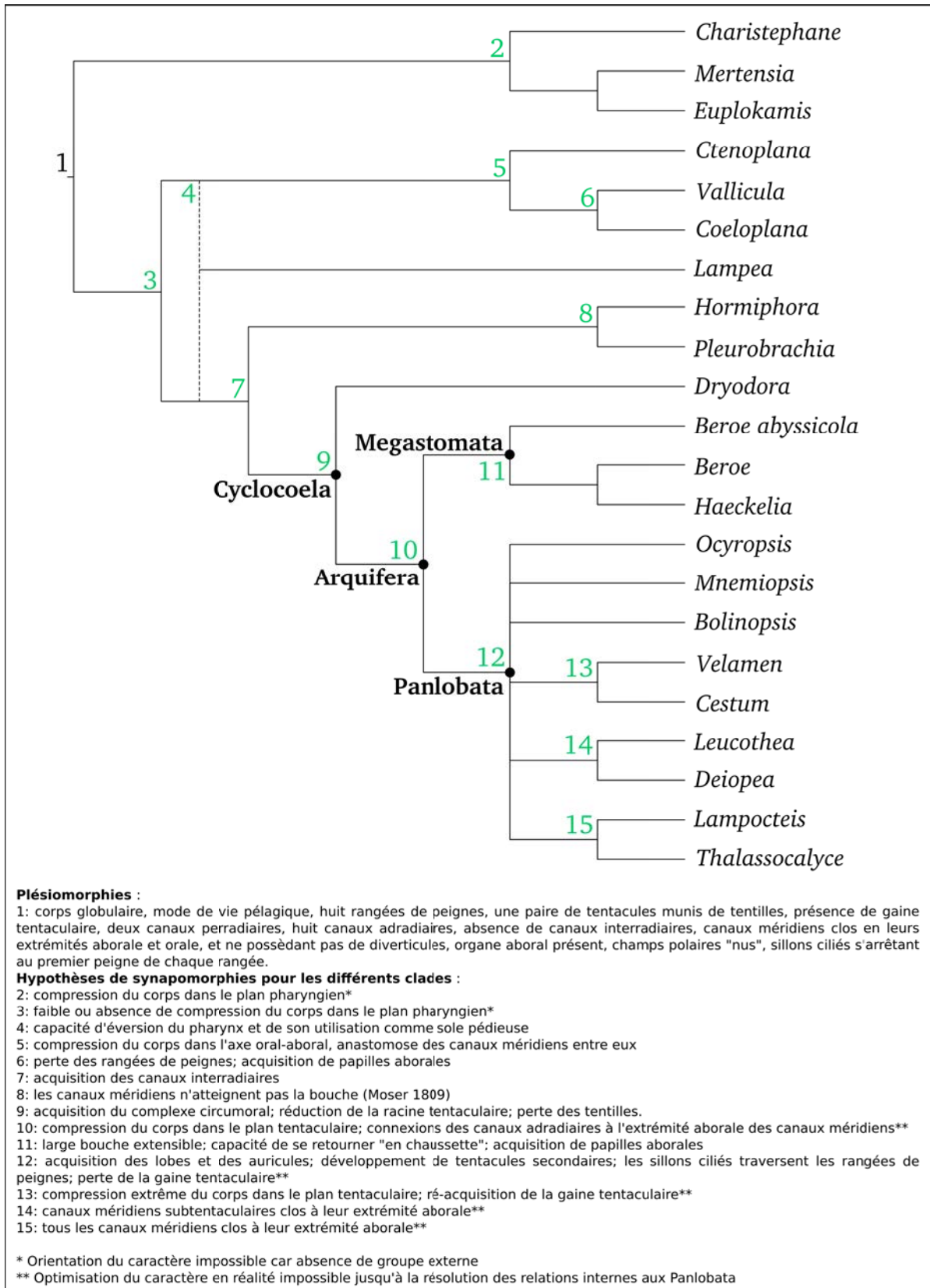


FIGURE 3.7 – Arbre bilan des analyses moléculaires et hypothèses d'évolution de caractères morpho-anatomiques. Les chiffres correspondent à une numérotation des clades (de 1 à 15).

aborale seulement pour les canaux méridiens subpharyngiens, mais pas pour les canaux méridiens subtentaculaires. Depuis, les genres *Thalassocalyce* et *Lampocteis* ont été découverts et décrits (Madin and Harbison 1978; Harbison et al. 2001), et leurs canaux adradiaires ne connectent jamais les canaux méridiens en leur extrémité aborale. Or, les analyses moléculaires de ce chapitre (article 2) indique d'une part une proche parenté entre *Thalassocalyce* et *Lampocteis* (clade 15 sur Figure 3.7), et d'autre part une proche parenté entre *Leucothea* et *Deiopea* (clade 14 sur Figure 3.7).

Certains caractères semblent en revanche de claires convergences évolutives chez les cténophores. La fusion des canaux méridiens en un réseau se retrouve ainsi à la fois chez certains Platyctenida et chez certains Beroida, tandis que des papilles aborales entourant les champs polaires existent chez les Coeloplanidae (Platyctenida) et dans le clade regroupant les Beroida et les Haeckeliidae.

Malgré ces avancées, trois problèmes restent à résoudre pour atteindre une compréhension encore plus complète de la systématique des cténophores actuels. Les voici par ordre de complexité croissante :

1. La position phylogénétique du genre *Lampea* doit être déterminée. La capacité des *Lampea* à utiliser leur pharynx pour se fixer à leur proie pourrait représenter une synapomorphie les unissant aux Platyctenida (voir discussion article 2), mais cette hypothèse reste à être confirmée par des données moléculaires.
2. Les relations phylogénétiques internes au Lobata *s.l.* (Panlobata) doivent être déterminées. Cette irrésolution est principalement due à une vitesse d'évolution très lente de ces groupes, et l'utilisation de marqueurs moléculaires plus variables sera nécessaire pour réunir suffisamment de signal phylogénétique. Compte tenu des longueurs de branches obtenues avec les matrices phylogénomiques du chapitre précédent entre *Mnemiopsis*, *Bolinopsis* et *Cestum*, il est clair qu'une approche phylogénomique devrait également permettre de résoudre ces relations de parenté.
3. Enfin, l'échantillonnage taxonomique des cténophores doit être augmenté, car de nombreux taxons sont encore absents des jeux de données moléculaires. Trois ordres sont manquants (il s'agit des Cambojiida, des Ganeshida et des Cryptolobiferida, dont l'étude devrait permettre d'éclaircir, entre autre, l'évolution des Lobata avec lesquels ces ordres ont de nombreux points communs ; voir section 1.1.2), ainsi que 10 familles. La récolte de nouvelles espèces est la plus difficile des tâches car elle ne peut avoir lieu que sur un temps long, avec beaucoup de moyens et de collaborations scientifiques.

Implications taxonomiques de ces résultats Cette nouvelle phylogénie des cténophores doit amener à des modifications de leur classification. Certaines de ces modifications ont déjà été suggérées par le passé, comme la suppression de l'ordre des Cydippida (Ospovat 1985a; Podar et al. 2001), et d'autres sont nouvelles car directement issues des résultats de cette thèse.

L'ordre des Cydippida est paraphylétique (voir article 2), et son utilisation doit être abolie.⁸ Les relations phylogénétiques unissant *Beroe abyssicola*, les *Haeckelia* et le reste du genre *Beroe* ont de nombreuses implications taxonomiques. Les deux classes Tentaculata et Nuda sont invalides, et d'autre part, l'ordre des Beroida, la famille des Beroidae et le genre *Beroe* doivent être redéfinis.

L'existence d'un clade unissant tous les cténophores possédant un complexe circumoral est en accord avec le travail de Ospovat (1985a), qui avait alors créé la classe des Cyclocoela. Ce groupe s'oppose aux Typhlocoela qui est quant à lui paraphylétique.

L'ordre des Lobata (tel que délimité classiquement) est paraphylétique car les Thalassocalycida sont le groupe-frère du genre *Lampocteis* et qu'il est probable que l'ordre des Cestida soit aussi inclus dans les Lobata. L'analyse du parcours des canaux méridiens autour de la "cloche" que forme le corps de *Thalassocalyce inconstans* montre que cette espèce possède en réalité deux lobes et quatre auricules, qui sont fusionnés entre eux (Madin and Harbison 1978). Le genre *Thalassocalyce* doit donc être intégré à l'ordre des Lobata. Les analyses phylogénomiques ainsi que l'analyse du gène dupliqué *arp23* suggèrent fortement que les Cestida sont eux-aussi des Lobata dérivés (voir article 1 et Figures 3.3 et 3.4).

Sur la figure 3.7 ont été reportés un certain nombre de noms de clades, dont trois sont nouveaux :

- Panlobata : regroupe les Lobata au sens strict avec les Cestida et les Thalassocalycida.⁹
- Megastomata : du grec *megas*, "grand, large" et *stoma*, "bouche". Ce nom fait référence à la capacité des Beroida et des Haeckeliidae d'étendre très largement leur bouche afin d'engloutir de larges proies.
- Arquifera : du latin *arquus*, "arc, arche" et *ferre*, "qui porte sur soi". Ce nom fait référence à la connexion des canaux adradiaires au niveau de l'extrémité aborale des canaux méridiens, formant ainsi un canal "continu", en forme d'arche.

8. De la même manière, l'utilisation de l'expression larve "cydippide" est incorrecte. De l'aveu même d'Harbison (1985), "the term 'cydippid larva' has little meaning."

9. Il est possible qu'à l'avenir, les ordres des Ganeshida et des Cryptolobiferida soient également inclus dans le groupe des Panlobata.

Chapitre 4

Recherche de caractères à l'échelle “micro-anatomique” par réalisation de marquages immunohistochimiques

4.1 Introduction

La morphologie des cténophores a été observée, décrite, et représentée par de nombreux auteurs. La première observation scientifique publiée de cténophores est due au docteur Martens, qui publia en 1675 les résultats de son voyage au Spitzberg et au Gröenland en 1671. Ses dessins ont été interprétés par Hyman (1940) comme représentant *Mertensia* et *Bolinopsis*. L'étude à l'échelle macroscopique de la diversité des cténophores et de leur systématique fut ensuite continuée, et l'on peut citer parmi les plus importants contributeurs Gronovius (1760), Fabricius (1780), Mertens (1834), Agassiz (1865), Chun (1880), Moser (1909), Mayer (1912), Mortensen (1912), Hyman (1940) ou encore Dawydoff (1946). Le cas de Libbie Hyman (Figure 4.1 page suivante) est particulier, car son apport à la systématique des cténophores n'est pas dû à la description de nouveaux échantillons ou de nouvelles espèces, mais correspond à une clarification et une formidable synthèse des connaissances de son époque sur les cténophores (Hyman 1940).

L'anatomie des cténophores a également beaucoup été étudiée à un niveau histologique, ultrastructural et cellulaire. Parmi les premiers à s'être intéressés à ce niveau anatomique, on peut citer Samassa, Hertwig et Chun. Plus tard, Horridge et Hernandez-Nicaise décrivent, entre autre, la structure cellulaire des systèmes neuro-sensoriels, et Hernandez-Nicaise décrit notamment une structure particulière des synapses des cténophores, appelée triade post-synaptique (Hernandez-Nicaise 1968, 1973). De son côté, Franc décrit la structure et les types cellulaires trouvés dans la mésoglée de certaines espèces (Franc 1988), tandis que Tamm s'intéressa notamment à la physiologie des cténophores et au fonctionnement de leur rangées de peignes (Tamm 1973, 1982).

L'objectif de ce chapitre est d'amorcer une exploration de la diversité anatomique des cténophores



FIGURE 4.1 – Portrait de Libbie Henrietta Hyman, auteur du traité "Invertebrates", comprenant six tomes. Elle reçut la médaille Daniel Giraud Elliot en 1951, la médaille d'or de la société Linnéenne de Londres en 1960 et la médaille d'or du Muséum Américain d'Histoire naturelle en 1969.

à un niveau inédit, intermédiaire entre leur morphologie macroscopique et l'ultrastructure de leur tissus et cellules vue en coupe. L'utilisation de marquages immunohistochimiques rend possible l'observation dans l'espace des structures anatomiques tout en offrant une résolution jusqu'à une échelle cellulaire. Cette technique permet donc la description de l'anatomie "fine" des organismes, et a déjà été utilisée à plusieurs reprises sur le cténophore *Pleurobrachia pileus* (Jager et al. 2008; Alié et al. 2011; Jager et al. 2011; Dayraud et al. 2012; Jager et al. 2013), permettant une amélioration considérable de nos connaissances de son organisation. Ce type d'approche est par ailleurs utilisé depuis plusieurs années sur de nombreux groupes d'animaux (voir par exemple Nakanishi et al. 2008).

Pour les cténophores, des marquages immunohistochimiques n'avaient jusqu'à maintenant été réalisés que chez la seule espèce *Pleurobrachia pileus*, et leur réalisation chez d'autres taxons de cténophores apparaissait comme une suite logique des précédents travaux de l'équipe. L'exploration comparée de la micro-anatomie des cténophores représente une source prometteuse de nouveaux caractères morphologiques. Leur découverte et description formelle pourrait contribuer à un renouveau de la taxinomie des cténophores, considérée comme fragile et problématique par la communauté scientifique (Mills 1998-present ; Bennema and Van Moorsel 2011)

La prochaine section de ce chapitre (article 3) consistera en une comparaison approfondie de la micro-anatomie de deux espèces particulièrement proches d'un point de vue morphologique et moléculaire, *Pleurobrachia pileus* (O. F. MÜLLER 1776), et *Pleurobrachia brunnea* (MAYER, 1912). Cette comparaison permettra d'évaluer le potentiel des marquages immunohistochimiques à détecter de nouveaux caractères diagnostiques à l'échelle des espèces, et ainsi à améliorer la systématique des cténophores. Après ce manuscrit suivra une section complémentaire présentant des marquages immunohistochimiques réalisés chez une espèce morphologiquement très différente et appartenant à l'ordre des Beroida : *Beroe gracilis* (dont la morphologie générale a été décrite au chapitre 1, section 1.1.1).

4.2 Article 3 : “New species-level diagnostic characters identified in ctenophores using simple fluorescence microscopy techniques : a case study comparing two *Pleurobrachia* species (Ctenophora : Pleurobrachiidae)”

Simion, P. ; Jager, M. ; Manuel, M. ; article formaté pour soumission à la revue *Zootaxa*.

New species-level diagnostic characters identified in ctenophores using simple fluorescence microscopy techniques: a case study comparing two *Pleurobrachia* species (Ctenophora: Pleurobrachiidae)

PAUL SIMION, MURIEL JAGER, MICHAEL MANUEL

UPMC Univ Paris 06, Institut de Biologie Paris–Seine, UMR 7138 “Evolution Paris–Seine” CNRS UPMC, Case 05, 7 quai St Bernard, 75005 Paris, France.

Corresponding author michael.manuel@snv.jussieu.fr

Abstract

Two closely-related ctenophore species (genus *Pleurobrachia* Fleming, 1822) are compared in order to evaluate the potential of fluorescence microscopy (after immunolabelling against two structural proteins and Dapi staining of cell nuclei) vs. classical morphology for uncovering characters of taxonomic interest. These species are the North Atlantic *Pleurobrachia pileus* (O.F. Müller, 1776) and the poorly known West Atlantic species *P. brunnea* Mayer, 1912 (newly recorded from Guadeloupe in Lesser Antillae). Their taxonomy is reviewed, *P. pileus* (O.F. Müller 1776) being established as *nomen protectum* against the older synonym *P. bicaudatus* (Linnaeus, 1767) and *P. rhodopis* Chun, 1880 as a junior synonym of *P. pileus*, and their phylogeny is reconstructed using a 18S rRNA + ITS1 + 5.8S rRNA + ITS2 molecular data set. An appreciable number of clear-cut anatomical characters distinguishing both species were obtained using immunofluorescence and Dapi fluorescent nuclei labelling within the apical organ, polar fields, epidermal nerve plexus and comb rows. It can thus be expected that wide-scale use of these techniques in the future will help clarifying species-level taxonomy within Ctenophora.

Key words: Atlantic, Ctenophora, Fluorescence microscopy, Immunolabelling, Morphology, Phylogeny, *Pleurobrachia*

Running title: COMPARISON OF TWO *PLEUROBRACHIA* SPECIES

Introduction

Members of the phylum Ctenophora are exclusively marine animals with a planktonic lifestyle, except in the benthic order Platyctenida. Despite worldwide distribution of the phylum, local abundance of some of the species, key importance of the group in planktonic food webs and strong ecological impact of ctenophore invasive species (Costello et al. 2012), the taxonomy of Ctenophora remains nowadays in very poor state.

Not only have molecular phylogeny studies revealed that the high-level classification of the phylum is in serious need of revision (Podar et al. 2001, Simion et al. 2014), but also the taxonomy of ctenophores at species level remains poorly understood (and even though the group is not particularly species-rich, especially when compared to the other non-bilaterian phyla Porifera and Cnidaria). Thus, there is considerable uncertainty concerning the number of valid described species, which according to Mills (1998-present) stands somewhere in between 150 and 200. This means that certainly a great number of the currently recognised nominal species (list established and curated by Mills 1998-present) are junior synonyms of other species names. In the absence of type material for most old species names, the identity of many of them is indeed unclear (particularly striking examples include the taxonomic conundrum around “*Callianira hexagona*”, recently solved by Bennema & van Moorsel 2011, or the uncertain identity of “*Beroe ovata*” auct., discussed in Mills et al. 1996). On the other hand, several ctenophore species have supposedly sub-cosmopolite distribution (e.g. *Cestum veneris* Lesueur, 1813, *Lampea panzerina* (Chun, 1880), etc.) based on global similarity of specimens throughout the world, but may in fact represent species complexes. C. Mills (1998-present) estimates that we may know only about half of the ctenophores that live in the sea, so many species may await formal description. This task is unfortunately hampered by the existing confusion, because it would be senseless to describe new species, whilst the identity of many of the previously-described ones is unclear.

Many factors contribute to this situation, including the small number of taxonomists working on ctenophores, the difficulties associated with macro-plankton sampling in the open ocean, and the fact that many ctenophore species have so delicate tissues that they cannot be preserved using classical fixation techniques (Mills & Haddock 2007).

An additional factor which hampers correct recognition and delimitation of ctenophore species is a paucity of morphological characters available to taxonomists. Whereas ctenophore orders or families can generally be easily recognised using features visible at macroscopic scale, at lower taxonomic levels the task becomes more difficult because the morphology is often rather homogeneous within families or genera. Most species are thus distinguished by only few and often subtle characters, sometimes rather difficult to appreciate on living specimens, and generally impossible to determine on fixed specimens (even for those ctenophores that do not disintegrate in fixatives).

Recent work by our lab on the sea gooseberry *Pleurobrachia pileus* (Müller, 1776) have contributed to realise that the ctenophore body displays considerable structural complexity at an organismal scale that can be said “micro-anatomical”, i.e. between global anatomy (accessible by naked-eye or stereomicroscope examination), and smaller scale tissue or cell features that classical histology and electron microscopy describe. Easy access to these “micro-anatomy scale” structural features is possible using epi-fluorescence microscopy, by examining fixed whole-mount specimens (or fragments of specimens) stained for instance by antibodies against structural proteins, or by DNA-binding molecules allowing visualisation of cell nuclei (such as Dapi or Hoescht). Using this approach, we previously re-described the complex architecture of the neuro-sensory system in *P. pileus* (Jager et al. 2011) as well as some aspects of muscular organisation (Dayraud et al. 2012).

This anatomical level can be suspected to represent a potential reservoir of taxonomically useful characters, notably for the diagnosis of different species within the same genus. In this study, we use epi-fluorescence microscopy on whole-mount preparations stained by two structural antibodies and Dapi to compare the micro-anatomy of epidermal structures (apical organ, polar fields, epidermal nerve plexus and comb rows) between *P. pileus* (specimens from the French coasts), and another species of the same genus, *P. brunnea* Mayer, 1912 (samples collected in Guadeloupe, Lesser Antillae). It was at first necessary to clarify the taxonomy of both species and to characterise their morphological differences at macroscopic scale, particularly since *P. brunnea* is a rather poorly-known species. In this context, we also sequenced and analysed molecular markers (18S rRNA and the highly variable ITS regions). The discussion is mostly dedicated to evaluate the potential of fluorescence microscopy to contribute to a renewal of species-level taxonomy of the Ctenophora, in light of results obtained by comparing the two selected *Pleurobrachia* species.

Material and methods

Specimen collection. Specimens of *P. pileus* used for morpho-anatomical observations were collected in the Villefranche-sur-Mer bay (south-eastern France, Mediterranean Sea, GPS N43°41' E07°19') by dragging a Régent plankton net (mesh size 700 µm), and in Gravelines (northern France, North Sea, GPS N51°01'20" E02°08'56") in the canal that brings sea water for cooling of the Gravelines nuclear plant, by drifting a KC Denmark diam. 50 cm plankton net (mesh size 500 µm) from the bridge. Specimens of *P. brunnea* were collected off the north-western coast of Basse-Terre island in Guadeloupe (GPS N16°24'10" W61°40 to 46') outside from the barrier reef in the Caribbean current, by boat towing of a WP2 plankton net (mesh size 250 µm). In all cases, ctenophores were sorted out from plankton samples immediately after collect and kept alive for a few hours in filtered sea water for observations and photography in the lab prior to fixation. See Table 1 for collection data associated with sequences used for molecular phylogenetic analysis.

Morphological observations. Most “classical morphology” observations were done with an Olympus SZX12 trinocular stereomicroscope and images were taken with a Spot FLEX Color Pixel Shift 64 Mp camera (Diagnostic Instruments Inc., USA) assisted by Spot Basic software. Habitus pictures of Figs 2C, 2E, 3C, 3F were reconstructed from Z-series of about 10 photos stacked using Helicon Focus Software (Helicon Soft Ltd.).

Immunofluorescence. For fixation, living animals were concentrated in a tube in filtered sea water to which was added an equal volume of paraformaldehyde 8% in phosphate-buffered saline (PBS: 10 mM Na₂HPO₄, 150 mM NaCl, pH 7.5); then the tube was gently agitated for 30 min of fixation at room temperature. Samples were washed quickly three times in PBTw (PBS + Tween 0.1%), then washed for 15 min in PBTw at room temperature. They were dehydrated through a graded series of ethanol then stored in methanol at -20°C. After stepwise rehydration to PBTr (PBT + Triton-X 100 0.01%), samples were permeabilised for 10 min in PBS + Triton-X 100 0.2%, then rinsed in PBTr. They were treated with PBTr + bovine serum albumine (BSA) 1% during 60 min for blocking. Then they were incubated for 4h at room temperature in PBTr + BSA 1% with the following primary antibodies: rat monoclonal anti-tyrosylated α -tubulin antibody (1:1000 dilution, AbD Serotec, Oxford, UK) and mouse monoclonal anti- β -actin antibody (Sigma Aldrich, St Quentin Fallavier, France). They were then rinsed 6 times for 15 min in PBTr. Incubation with secondary antibodies (Alexa Fluor® 568 goat anti-rat IgG and Alexa Fluor® 488 goat anti-mouse IgG; Molecular Probes, Invitrogen, Cergy Pontoise, France) was done for 4h in PBTr and followed by washing 6 times for 15 min in PBTr. All samples were stained with Dapi (1 µg/ml) for 30 min then washed twice for 15 min in PBTr. Animals were finally micro-dissected and mounted in Citifluor (Oxford Instruments, Gometz-la-Ville, France). All fluorescence microscopy and DIC images were produced using an Olympus BX61 microscope with Q-imaging camera and Image Pro Plus software (Mediacybernetics, ScopPro, Itteville, France).

Consideration of inter-individual variation. Character states ascribed to *P. pileus* in Table 2 are based on our observations over the last years on dozens of specimens from several localities (detailed in section Taxonomy). Assessment of intra-specific variation for *P. brunnea* was more limited, as all studied specimens come from the same locality. About 10 of the 50 collected specimens were treated for immunofluorescence study, and from their examination only characters states for which we observed no deviation within this series were formalised in Table 2 (right column).

Molecular phylogeny. Information about taxon sampling, specimen localities and data sources is presented in Table 1. For the species *P. pileus*, the molecular data set analysed in this study comprises sequences from three distinct localities. Sequence data from Villefranche-sur-Mer (Mediterranean Sea; *Pleurobrachia pileus* 1 in Table 1 and Fig. 4) were produced from specimens collected and identified by authors of this study and were published previously (Simion et al. 2014). Sequences assigned to *P. pileus* from eastern North America (*Pleurobrachia pileus* 2 in Table 1 and Fig. 4) are from Podar et al. (2001). Sequences of *P. pileus* from the Channel Sea (Roscoff, *Pleurobrachia pileus* 3 in Table 1 and Fig. 4) were extracted from an unpublished transcriptome assembly derived from 454 next-generation sequencing of specimens collected and identified by authors of this study. The molecular data for *P. brunnea* are from Simion et al. (2014) (specimens collected and identified by authors of this study). Outgroup species have been selected in order to represent the other Pleurobrachiidae genus *Hormiphora* L. Agassiz, 1860 as well as ctenophore taxa closely related to the family Pleurobrachiidae according to previous ribosomal RNA phylogenies (i.e. Beroida, Haeckeliidae, Lobata and Cestida; Podar et al. 2001, Simion et al. 2014). Less closely related ctenophores (e.g. Platyctenida, Lampeidae...) were not included in order to preserve the quality of the alignment of the variable ITS region.

Sequences were aligned using ClustalW (with the following parameters: BLOSUM matrix; gap open cost = 2; gap extension cost = 0.05; gap distance = 1; no end gaps) (Larkin et al. 2007). A few obvious alignment mistakes were corrected by hand, and ambiguously-aligned positions were trimmed from the alignment using Gblocks (with following parameters: p1 = 6; p2 = 6; p3 = 4; p4 = 2; p5 = All) (Castresana 2000). The combined 18S rRNA+ITS1+5.8S rRNA+ITS2 alignment comprises 2674 nucleotidic positions, and Gblocks trimming resulted in a final alignment of 2475 positions (including 214 parsimony-informative positions). Maximum Likelihood (ML) analyses of the trimmed alignment were performed using RAxML (Stamatakis 2014) as follows. First, 100 distinct ML analyses were performed using the GTR+ Γ model, previously shown to be the best-fit for this molecular data set (Simion et al. 2014). Following recommendations in the RAxML user manual (version 8.0.0. available online at <http://sco.h-its.org/exelixis/web/software/raxml/>), we did not add a parameter representing proportion of invariant positions. Then 1000 bootstrap replicates were computed from the same alignment and using the same evolution model. Finally, bootstrap values have been plotted onto the best tree from the first step.

Results

Taxonomy

Family Pleurobrachiidae Chun, 1880

Diagnosis. Body shape subspherical to ovoid, only slightly compressed in pharyngeal plane. Presence of one pair of tentacles bearing tentillae; base of tentacle housed in tentacle sheath, its opening located in aboral hemisphere; tentillae devoid of striated muscle. Absence of lobes and auricles. Subtentacular and subpharyngeal comb rows of approximately equal lengths. Perradial and interradianal canals present. Meridional canals ending blindly at both extremities; connected to adradial canals at some distance from their aboral extremity. Preys brought to mouth by spinning.

Note. The family Pleurobrachiidae as understood here does not comprise the genus *Euplokamis* Chun, 1880, in agreement with Mills (1998-present), Granhag et al. (2012) and Simion et al. (2014).

Genus *Pleurobrachia* Fleming, 1822

Type species. *Pleurobrachia pileus* (O.F. Müller, 1776)

Note. We refrain to formulate a diagnosis for this genus. Diagnoses of *Pleurobrachia* found in old works (e.g. Chun 1880, Mayer 1912) are no longer relevant because several genera described more recently have been allocated to the family Pleurobrachiidae. A correct diagnosis of *Pleurobrachia* would require prior revision of the family, which is beyond the scope of this study. For a list of genera currently placed in Pleurobrachiidae, and a list of *Pleurobrachia* species, see Mills (1998-present).

***Pleurobrachia pileus* (O.F. Müller, 1776)**

Figs 1, 2, 5A, 5B, 5E, 5G, 6A, 6C, 6E, 7A, 7C, 7E, 8A, 8C

Volvox bicaudatus Linnaeus, 1767:1325 (*nomen oblitum* by present action, see below).

Beroe pileus O.F. Müller, 1776:233 (*nomen protectum* by present action, see below)

Pleurobrachia pileus Fleming, 1822:612.

Cydidippe pileus Eschscholtz, 1829:24.

Callianira slabberi De Haan, 1827:150 (syn.: Bennema & van Moorsel 2011).

Callianira hexagona Eschscholtz, 1829:28 (syn.: Bennema & van Moorsel 2011)

Cydidippe pomiformis Patterson, 1838:109 (syn.: Moser 1903)

Pleurobrachia rhododactyla L. Agassiz, 1849:317 (syn.: Moser 1903).

Pleurobrachia rhodopis Chun, 1880:282 (**n. syn.**).

Reversal of precedence. In the 12th edition of the *Systema Naturae*, Linnaeus created the species name *Volvox bicaudatus* Linnaeus, 1767, accompanied by a short diagnosis (“*Volvox octagonus, tentaculis duobus elongatis*”). He mentioned that this was the animal previously described by Gronovius (1760) from the Belgian coast under the name *Beroe corpore octagono, sphaerico, tentaculis binis plumosis longissimis*. From the detailed description of Gronovius, his beautiful drawings (some of which here reproduced in Fig. 1), and the geographical origin of specimens, there can be no doubt that this is our *Pleurobrachia pileus*. Since Gronovius did not use binominal nomenclature, his name is not available, therefore the oldest available epithet for this species is *bicaudatus*. The synonymy of *pileus* with *bicaudatus* was still mentioned by O.F. Müller (1776) and by Eschscholtz (1829), but the name given by Linnaeus was ultimately forgotten (or perhaps discarded) by subsequent authors. It is for example absent from the lists of synonyms of *P. pileus* compiled by Chun (1898) (this author nevertheless still mentioning its identity with the species described by Gronovius), Moser (1903) and Mayer (1912). A possible explanation could be that Zoologists rejected this name because *Volvox* Linnaeus, 1758 had turned out to be (and is still today) a genus of green algae. However, in fact this has no bearing on the availability of *V. bicaudatus* as a name for an animal species (ICZN Art. 10.5). It follows that the valid name for the species under consideration here should be *Pleurobrachia bicaudatus* (Linnaeus, 1767). However, to our knowledge this name has never been used as the valid name for that species after 1899, and we give in Annexe I a list of 25 publications in the past 50 years using instead the name *Pleurobrachia pileus* (O.F. Müller 1776), with in total more than 10 authors involved, and spanning over more than 10 years. The conditions of ICZN Art. 23.9.1. are thus fully met and reversal of precedence can apply to protect the junior synonym *P. pileus*.

Type locality. Danish and Norwegian coasts (by inference from the geographical area covered by the work of Müller 1776).

References to description. The morphology of this species, summarised in Fig. 2, has been described in many classical works (e.g. L. Agassiz 1849, Moser 1903, Mayer 1912). The embryonic development has been described by A. Agassiz (1865, 1874) and Freeman (1977) and the development of comb rows in the adult by Tamm (2012). The parasites of *P. pileus* have been studied by Yip (1984). There have been many works investigating tissues and cell types of this species using electron microscopy (e.g. Horridge 1964a, 1964b, Horridge & Mackay 1964, Bargmann et al. 1972, Franc 1978, Hernandez-Nicaise 1968, 1974). Other works have documented various aspects of the physiology of this ctenophore, particularly the mechanism of comb plate ciliary beating (e.g. Sleigh 1968, 1972, Tamm 1984, Moss & Tamm 1987, 1993) and feeding biology (Moss 1991). Several recent studies using immunofluorescent labelling and/or gene expression analyses have provided additional detailed information about this species notably concerning neuro-sensory structures (Jager et al. 2011, 2013), muscular organisation (Dayraud et al. 2012), adult stem cell populations, cell renewal and formation of new combs at both extremities of comb rows (Alié et al. 2011).

Notes on distribution and synonymy. *Pleurobrachia pileus* is the North Atlantic sea gooseberry. It is notably very common along the North Atlantic coasts of Europe, the Channel Sea and the North Sea, where it is the only recorded *Pleurobrachia* species. It is also present in Arctic seas (Chun 1898, Moser 1903, Mayer 1912) and in the West Atlantic Ocean, along the shores of eastern Canada and USA, south to at least Chesapeake Bay (Mayer 1912, Bigelow 1926, Johnson & Allen 2012). These western Atlantic populations were at some time considered as a separate species, *P. rhododactyla* L. Agassiz, 1849. Our molecular phylogeny shows extremely short genetic distance between specimens from Woods Hole, MA (USA) and Roscoff (Channel Sea) (Fig. 4), in agreement with *P. rhododactyla* being a junior synonym of *P. pileus* (Moser 1903, Mayer 1912, Bigelow 1912).

Sea gooseberries of the Mediterranean Sea have been hitherto treated as a distinct species, *P. rhodopis* Chun, 1880. This species was said to differ from *P. pileus* by the comb rows proportionally shorter (not extending to the oral third of the body) and by the junction between adradial and meridional canals lying at the same level than the infundibulum (a bit closer to the aboral pole in *P. pileus*) (Chun 1880, Moser 1903). Chun (1880) himself seems to have been somewhat suspicious about the validity of his new species, as he ended the description saying “I suppose that all individuals [of the type series] were still very young and belong to a larger species, perhaps coming from the African coasts and brought in the Spring to the Mediterranean Sea by storms” (translation from the original German text). In addition, the two characters supposedly supporting recognition of *P. rhodopis* as a species distinct from *P. pileus* concern features that are known to vary depending on the state of contraction of specimens (as described by Mortensen 1912). Nevertheless, subsequent authors persisted in calling the Mediterranean specimens *P. rhodopis* (e.g. Moser 1903, Trégoubouff & Rose 1957, Franc 1978, Hernandez-Nicaise 1968, 1974, Riedl 1986, Mills et al. 1996, Buecher & Gasser 1998, Molinero et al. 2008, see also note by C. Mills in page 73 of the online supplementary information of Moroz et al. 2014), in some cases more or less explicitly suggesting a possible synonymy with *P. pileus* (e.g. Riedl 1986,

Mills et al. 1996). All these authors seem to agree that a single *Pleurobrachia* species is present in the Western Mediterranean Sea.

We argue that *P. rhodopis* is indeed a junior synonym of *P. pileus* on the following grounds:

- Over the last ten years we have studied hundreds of *Pleurobrachia* specimens coming from four main French localities: Villefranche-sur-Mer (Mediterranean Sea, close to North-West Italy); La Rochelle (Atlantic Ocean); Roscoff (Channel Sea) and Gravelines (North Sea). We never observed any noticeable morphological difference between specimens of these four localities (apart from size, on average larger in the more northern locality of Gravelines, in agreement with observations made by Bigelow (1926) along the eastern coast of USA). Notably, adults from Villefranche-sur-Mer do not have on average shorter comb rows than those from the Atlantic, Channel Sea or North Sea (compare Figs 2A and 2B); same for position of the junction between adradial and meridional canals (not shown);
- When examining large series of living *Pleurobrachia* specimens collected in Villefranche-sur-Mer, we observed important inter-individual variability concerning the ratio between comb row length and the distance between oral extremity of comb rows and the mouth. This variability seems due to a large extent to the degree of contraction of the body musculature, since isolated specimens with “*rhodopis*-like” morphology (i.e. with rather long comb-free oral region) often turned to a “*pileus*-like” morphology after some minutes or hours, and reciprocally. In 2004, we sorted out from a series of living specimens collected in Villefranche-sur-Mer five “*pileus*-like” and five “*rhodopis*-like” individuals in order to determine their ITS1 and ITS2 sequences. We obtained 100% identical ITS sequences for all of these specimens.
- Detailed investigation of micro-anatomy, notably of the neuro-sensory system, performed in our lab during the last 10 years using various antibodies did not reveal any structural difference between specimens from Villefranche-sur-Mer and specimens from La Rochelle, Roscoff or Gravelines (by contrast, the same techniques unveiled an appreciable number of structural differences with the other species *P. brunnea*, as described later on in this paper);
- The molecular tree shown in Fig. 4 indicates that the 18SrRNA + ITS1 + 5.8S rRNA + ITS2 sequence of a specimen from Villefranche-sur-Mer (Mediterranean Sea) is almost identical to that of a specimen from Roscoff (Channel Sea). The Mediterranean specimen is closer in the tree to the Channel Sea specimen than is the *P. pileus* specimen from West Atlantic (even if relationships between the three *P. pileus* sequences are statistically unsupported).

In conclusion, the Mediterranean Sea must now be considered part of the distribution area of *P. pileus*. Of note is that this synonymy has been already implicitly admitted in all previous studies from our lab which made use of *Pleurobrachia* material from Villefranche-sur-Mer (Jager et al. 2006, 2008, Derelle & Manuel 2007, Derelle et al. 2010, Philippe et al. 2009, Alié et al. 2011, Dayraud et al. 2012).

It remains to be seen if records of *P. pileus* from regions very distant from the North Atlantic (Antarctic; South Atlantic including western South Africa; Chile; Hawaii; New Zealand) (Moser 1908, 1909, Hirota 1974, Palma & Meruane 1975, Mayer 1912, Ralph 1950, Gibbons et al. 2003, Mianzan 1999, Mianzan et al. 2009) correspond to this or different species.

***Pleurobrachia brunnea* Mayer, 1912**

Figs 3, 5C, 5D, 5F, 5H, 6B, 6D, 6F, 7B, 7D, 7F, 8B, 8D

Pleurobrachia brunnea Mayer, 1912:14.

Euplokamis brunnea Mills, 1987:2667 [but see note added in proof in this reference, putting back the species into the genus *Pleurobrachia*]

Type locality. Off the coast of New Jersey (from Sandy Hook to Barnegat Bay).

Diagnosis. According to Mayer (1912) (see his original drawing reproduced in Fig. 3A), differs from *P. pileus* by : more oblong, egg-shaped body; pharynx opaque, yellow-brown; end of each tentacle thickened into terminal knob; tentacle with fewer side branches (tentillae), these highly contractile and often coiling into helices; more southern distribution.

Notes. In November 2012, we collected about 50 specimens of *P. brunnea* about 5 km off the north-western coast of Basse-Terre island, Guadeloupe (Lesser Antillae). These specimens displayed all diagnostic features mentioned by Mayer (1912) for *P. brunnea*, including oblong shape (Fig. 3B, 3C; compare with *P. pileus* Fig. 2A-C), terminal end of tentacles thickened (white arrows in Fig. 3D and 3E), tentillae less numerous than in *P. pileus*, more widely spaced, and in living animals most often kept contracted and coiled into helices even when tentacles are extended (Fig. 3D; compare with Fig. 2D). This remarkable behavioural feature had led Mills (1987) to suggest that *P. brunnea* might in fact belong to the genus *Euplokamis* Chun, 1880. For the same reason we initially thought that our ctenophore from Guadeloupe was a species of *Euplokamis*. However, in a note added in proof, Mills (1987) indicated that she could study by transmission electron microscopy the tentillae of *P. brunnea* specimens, which she collected in the Gulf Stream off Virginia (USA). They happened to lack the striated muscles representing the main diagnostic feature of the genus *Euplokamis* (Mills 1987). We examined the musculature of tentillae in our specimens from Guadeloupe (using phalloidin staining; results not shown) and likewise observed absence of striated muscle. Correlatively, our lab observations in Guadeloupe indicate that unlike *Euplokamis*, living specimens of *P. brunnea* do not extend their tentillae at high velocity when capturing preys.

The morphological differences at macroscopic scale between *P. brunnea* and *P. pileus* have been summarised in Table 2 (characters 1 to 6). Two of the characters mentioned by Mayer (1912) have been retained in this table although we cannot ascertain their diagnostic value, hence they are at present considered doubtful (asterisks in Table 2). These are pharynx colour (character 2), likely to vary with respect to the nature and abundance of preys, and terminal thickening of tentacles (character 3), also occasionally seen in *P. pileus*, and certainly strongly dependent on the state of muscle contraction. One of the listed characters (character 6) had not been described previously. In *P. brunnea*, intense spots of red pigment are apparent at the base of polar

fields and on each side of the apical organ in the tentacular plane (Fig. 3F, 3G). These red spots, which remain present after fixation, are either absent or very faint in *P. pileus* (Fig. 2E, 2F).

We report below a few additional “classical morphology” observations that we consider too preliminary to be formalised in Table 2:

- Swimming behaviour: specimens of *P. brunnea* seem to swim more rapidly and to be more reactive than those of *P. pileus*. However, this is a rather subjective appreciation, since we never had the opportunity to compare directly living specimens of both species, and we made no video recording of their behaviour.
- The epidermis of *P. brunnea* is more delicate and more fragile. This remains true when specimens are fixed. When rehydrated, *P. brunnea* specimens unlike those of *P. pileus* do not expand back to initial size and shape but remain more or less wrinkled particularly in the oral region (compare Figs 2C and 3C), which might reflect different properties of the mesoglea and/or musculature.
- Within the gastrovascular system, the interradial canals of *P. brunnea* seem wider than those of *P. pileus* (compare Fig. 2G and Fig. 3H). However, the aspect of gastro-vascular canals on fixed specimens is much variable among specimens of the same species (see Discussion). Further study is needed to assess the significance of this character.

Analyses of 18S rRNA, ITS1, 5.8S rRNA and ITS2 (Fig. 4) show that *P. brunnea* diverges significantly from the three sequenced specimens of *P. pileus*, to about the same extent than the Pacific species *P. bachei* L. Agassiz, 1860. The relationships between the three *Pleurobrachia* species present in this data set (*P. pileus*, *P. brunnea*, and *P. bachei*) are unresolved.

Ecology. According to Johnson & Allen (2012), in North American waters this species seems to live preferentially in the open ocean, by contrast to *P. pileus* which is known to proliferate mainly in near-shore calm waters (Bigelow 1926). Likewise in Guadeloupe, we found *P. brunnea* only off the coral barrier reef (bottom depth 30-50 meters), North-West of Basse-Terre, which is where the Caribbean current (coming from Florida) reaches the Guadeloupe archipelago. Other ctenophores collected together with *P. brunnea* were a single specimen of *Ocyropsis crystallina* Rang, 1828, several *Bolinopsis* sp., and one damaged specimen of *Beroe* sp. By contrast, *P. brunnea* was absent from the calm and shallow coastal waters on the inner side of the barrier reef, where juvenile and adult specimens of *Mnemiopsis gardeni* L. Agassiz, 1860 were on the contrary rather abundant.

Distribution. Eastern coast of USA from New Jersey to Florida; Guadeloupe (new record).

Comparison between *Pleurobrachia pileus* and *P. brunnea* using whole-mount DIC and fluorescence microscopy

This section is focused exclusively on clear-cut structural differences between *P. pileus* and *P. brunnea* revealed by our observations on multiple specimens of both *Pleurobrachia* species (these differences are summarised in Table 2, characters 7 to 16). The techniques used here were DIC microscopy, Dapi staining of cell nuclei and immunofluorescent staining with anti-tyrosylated- α -tubulin and anti- β -actin antibodies. Anatomical characters useful to distinguish both species were identified in the apical organ, polar fields, epidermal nerve plexus and comb rows.

Within the apical organ floor, four groups of large refringent cells called the lamellate bodies are easily visualised using DIC microscopy (Fig. 5A-D). Lamellate bodies are thought to be involved in photoreception (as already suggested by Chun 1880), based on ultrastructural features of the cells (Hernandez-Nicaise 1991) and expression of an opsin gene (Schnitzler et al. 2012). In *P. pileus*, the four lamellate bodies are well-separated, spherical and distributed in two pairs on both sides of the tentacular plane (Fig. 5A, 5B). By contrast in *P. brunnea*, both lamellate bodies of each pair are closely tied up in the pharyngeal plane (Fig. 5C, 5D) (Table 2, character 7). There is no visible separation between both lamellate bodies of each pair (at least in whole-mount preparations observed with DIC microscopy). Thus each pair looks like a single elongate mass of large refringent cells, with a concavity oriented towards the apical organ centre.

In Dapi-stained preparations of the *P. pileus* apical organ, both lamellate bodies of each pair are separated by two regular and parallel rows of nuclei, along the pharyngeal plane on each side of the apical organ centre (Fig. 5E, G; note that the cell nuclei of lamellate bodies are not visible in this focal plane because they are located deeper). This structure of unknown function was previously described in Jager et al. (2011; see their Figs 5B' and 6D). The corresponding cell nuclei in *P. brunnea* are not arranged in two rows but instead distributed according to the following pattern, well-defined and constant across individuals: from centre to periphery (left to right on Fig. 5H and in the boxed area of Fig. 5F), a V-shaped group of nuclei (with concavity towards centre), then a compact strand of nuclei (its width varying from one to three cell nuclei), followed by a rather wide ovoid mass of nuclei that comes into contact with the outer wall of the apical organ. This different arrangement between both species of the groups of cell nuclei sandwiched between lamellate bodies corresponds to character 8 in Table 2. Of note is that this character to be properly observed requires good accessibility to the apical organ (specimens fixed with the apical organ retracted are unsuitable) and very superficial optical focus (same level as statolith nuclei).

Polar fields are two elongated structures mostly made of ciliated epidermal cells, located on each side of the apical organ in the pharyngeal plane. They comprise a marginal zone and a central zone (respectively, ma Z and ce Z in Fig. 6) (terminology as in Jager et al. 2011). In Dapi stained preparations, the marginal zone itself is

differentiated into a peripheral area with very high density of cell nuclei (ma Z1) and a more internal area where cell nuclei are much sparser (ma Z2) (Fig. 6A, 6B). Within ma Z2 are housed the regularly spaced neuronal condensations called Z-bodies (two of them well visible on Fig. 6F above the dotted line; in Fig. 6E, Z-bodies are out of focus; see detailed description of these structures in Jager et al. 2011, 2013). The aspect of ma Z1 in Dapi-stained preparations is different in *P. pileus* vs. *P. brunnea* (compare Fig. 6A and 6B). Whereas in *P. brunnea*, nuclei density more or less continuously decreases from the outer to the inner border of ma Z1 (Fig. 6B), in *P. pileus* ma Z1 is more heterogeneous in terms of nuclei density (Fig. 6A), with (from the periphery inwards) a rather wide dense zone, followed by a narrow zone with lower nuclei density (white arrows in Fig. 6A), then again a thin narrow zone with higher nuclei density, adjacent to MZ2 (Table 2, character 9).

An additional difference in polar field structure between *P. pileus* and *P. brunnea* revealed by immunolabelling concerns the extension of the polar field muscle, a group of long tangential fibres oriented along the pharyngeal plane within the epidermis of the central zone (see Figs 7B, 7D and 7I in Jager et al. 2011). In *P. pileus*, the polar field muscle occupies a rather narrow area corresponding approximately to the internal half of the central zone (double arrow in Fig. 6C; detail in Fig. 6E with arrows indicating muscle fibres). The more peripheral area of the central zone does not contain polar field muscle fibres in *P. pileus*. By contrast in *P. brunnea*, the polar field muscle generally extends to the whole central zone (double arrow in Fig. 6D; muscle fibres indicated by arrows in detailed view of Fig. 6F), with approximately homogenous density of fibres from its periphery to the polar field midline, i.e. the pharyngeal plane (Table 2, character 10). Detailed comparison of the Z-bodies did not reveal any noticeable difference between both species.

Examination of the epidermal nerve plexus stained by the anti-tyrosylated- α -tubulin antibody revealed higher density (meshes on average twice smaller) in *P. brunnea* than in *P. pileus* (compare Figs. 7A and 7B; 7C and 7D) (Table 2, character 11). There were in addition rather spectacular differences between both species concerning the juxta-tentacular nerve cord, which is a longitudinally-oriented condensation of the epidermal nerve plexus in the tentacular plane (jtt in Fig. 7A-C; see also Fig. 3A-D in Jager et al. 2011). Firstly, in the aboral region of the body, the juxta-tentacular nerve cord starts closer to the apical organ in *P. pileus* than in *P. brunnea* (compare Fig. 7A and 7B) (Table 2, character 12). Furthermore, in *P. pileus* the juxta-tentacular nerve cord is detectable not only within the aboral hemisphere above the tentacle sheath opening (Fig. 7A), but also on the oral side of this opening (Fig. 7C). On the contrary in *P. brunnea* there is no detectable juxta-tentacular nerve cord orally from the tentacular sheath opening (Fig. 7D) (Table 2, character 13). When looking in more details at the juxta-tentacular nerve cord itself in both species (within the aboral hemisphere), it appears distinctly much thicker in *P. pileus* than in *P. brunnea* (compare Fig. 7E and 7F; these two pictures at same scale) (Table 2, character 14). Whereas in *P. pileus* the juxta-tentacular nerve cord is much wider than are the adjacent strands of the epidermal plexus (Fig. 7E), in *P. brunnea* it is at most twice thicker (Fig. 7F). Finally, in *P. pileus*, meshes of the epidermal nerve plexus adjacent to the juxta-tentacular nerve cord (aborally to the

tentacle sheath opening) are particularly large and have a marked longitudinal (oral/aboral) orientation (Fig. 7A, 7E). In *P. brunnea*, the corresponding meshes are not enlarged and have a polygonal shape (Fig. 7B, 7F) (character 15 in Table 2). Of note if that all these observations were made on specimens of approximately equivalent size for both species.

In the comb rows, we could identify a remarkably different arrangement the inter-plate fibres (strongly anti- β -actin immunoreactive) that connect adjacent basal cushions (the cellular component of combs) (Fig. 8). These fibres remain rather enigmatic but there is evidence that they play a supporting role rather than being muscle fibres in the strict sense (see discussion in Dayraud et al. 2012 p. 14). In *P. pileus*, a dense and narrow bundle of inter-plate fibres connects the central regions of adjacent basal cushions (Fig. 8A, 8C), while many short lateral “incomplete” fibres anchor to the oral side of each basal cushion. The latter are said “incomplete” in that they end within the inter-plate epidermis well before reaching the next comb (Fig. 8A, 8C). In *P. brunnea*, the arrangement of inter-plate fibres is different. Complete fibres (running from one comb to the following one) form a much wider bundle, and short incomplete lateral fibres are either absent or very few (Fig. 8B, 8D). This distinctive architecture of inter-plate fibres between both *Pleurobrachia* species corresponds to character 16 in Table 2.

Discussion

The main goal of this study was to evaluate the potential of fluorescent microscopy techniques (Dapi staining of cell nuclei and immunolabelling using antibodies against conserved structural proteins) as a source of characters that would help discriminating closely-related ctenophore species. For this we chose two members of the genus *Pleurobrachia*, tightly related genetically and phylogenetically but nevertheless undoubtedly representing two distinct species, and built upon our established experience in exploring and imaging the micro-anatomy of one of these species (*P. pileus*) using fluorescent microscopy. Morphological differences between *P. pileus* and *P. brunnea* apparent at macroscopic scale (with naked eye or stereomicroscope), and those issued from our observations using DIC or fluorescence microscopy, are listed separately in Table 2.

The number of distinctive characters between both species uncovered at micro-anatomic scale using the latter techniques (characters 7 to 16 in Table 2) is quite remarkable. Differences were found in the structure of the apical organ (arrangement of the lamellate bodies and distribution of the cell nuclei located in between each lamellate body pair), in some features of polar fields (aspect of the marginal zone in Dapi-stained preparations; relative extension in the transverse direction of the polar field muscle), in the longitudinal extension and thickness of the juxta-tentacular nerve cord and density of the epidermal nerve plexus, and finally within comb rows in the arrangement of inter-plate fibres. Some of these differences are likely to have interesting biological significance since they affect structures of obvious functional importance (juxta-tentacular nerve cord, inter-plate fibres), but it is beyond the scope of this study to interpret structural differences in functional/behavioural terms. Of note is that only some parts of the body have been investigated here, namely, epidermal structures, as they were previously described in details in *P. pileus* using the same techniques (notably in Jager et al. 2011, 2013, Dayraud et al. 2012). Therefore, there is certainly a potential for uncovering additional characters by looking at other parts of the body (e.g. mesogleal musculature; tentacular apparatus; gastrovascular system; mouth and pharynx ciliary structures, the latter shown by Tamm & Tamm 1993 to be taxonomically informative in the genus *Beroe*).

In order to alleviate the current weaknesses of morphological approaches to the diagnosis and characterisation of ctenophore species, a potential new source of anatomical characters must fulfil several important criteria. A first obvious one is quantity of characters. Ctenophore taxonomy is fragile to a large extent because of the low number of morphological characters provided by macroscopic-scale observation, this being true at all scales from species delimitation to establishment of relationships between ctenophore orders (Simion et al. 2014). The number of anatomical differences found between *P. pileus* and *P. brunnea* using DIC/Dapi/immunolabelling microscopy (Table 2) clearly demonstrates the validity of these techniques in providing more species-level discriminating characters than examination of specimens at macroscopic scale.

However, number of characters is not so important *per se*. We need “good” characters. It is not easy to define what a “good” character is in the context of taxonomy, but certainly this implies in particular that the differences should be clear-cut (discontinuous characters). This is the case of most novel characters reported here. For example, the distribution of lamellate bodies in the apical organ, the extension of the polar field muscle, the aspect of the juxta-tentacular nerve cord, and the arrangement of inter-plate fibres are all unambiguously distinctive in *P. pileus* vs. in *P. brunnea*. By contrast, many of the classically-used morphological characters are a matter of appreciation (being fundamentally continuous characters) and/or can be strongly affected by various factors including physiological state and age of specimens. For instance, *P. pileus* and *P. brunnea* certainly differ in terms of body shape (narrower in the latter, see above under Taxonomy); however the difference is not so clear-cut (body shape varies in response to changes in muscle contraction; specimens of *P. pileus* can sometimes adopt an ovoid shape quite similar to that usual for *P. brunnea*).

The notion of a “good” diagnostic character also implies that the differences must be consistent over intra-specific (inter-individual) variation, but this holds true for any kind of character so that in this respect there is no specificity for those discussed here. Clearly, even if we took into account inter-individual variation by looking at several specimens for each species and retaining only characters that seemed constant (see Material and Methods), it would be useful to study additional material of both species (e.g., *P. pileus* specimens from American coasts; *P. brunnea* specimens from more northern localities...) to confirm the diagnostic utility of the characters listed in the lower part of Table 2.

Another important requirement for a “good” character in the context of species delimitation, particularly critical as far as ctenophore are concerned, is that the character should be determinable using fixed specimens. Many macroscopic-level morphological characters that are widely used in ctenophore taxonomy do not fulfil this criterion. For example, the epidermis and mesoglea of *P. pileus* and *P. brunnea* behave differently upon fixation, such that rehydrated fixed specimens of the former species usually fully reverse to their initial size and shape (compare Fig. 2A-B and 2C), but not those of *P. brunnea*, which for that reason appear narrower (particularly in the oral region) that they are in living condition (compare Fig. 3B and 3C). Thus, it is not reasonable to use body shape for species discrimination when comparing fixed ctenophore specimens. Likewise, in the Taxonomy section we mentioned that the interradial canals of *P. brunnea* seem wider than those of *P. pileus* (see Figs 3H vs. 2G). However, the diameter of gastrovascular canals can be dramatically affected by muscle contraction and by fixation (e.g. we observed considerable variation in adradial canal diameter among fixed specimens of the same series for both species). Since we compared the aspect of interradial canals using fixed specimens of both species, it is thus not possible to ascertain the relevance of this macroscopic character (not included in Table 2). Even the aspect of tentillae in living specimens with tentacles extended, clearly the main distinctive feature between *P. brunnea* and *P. pileus* at macroscopic scale (living specimens of the former usually keeping their tentillae contracted and coiled, at least for rapid swimming, Fig.

3D), is of limited value with fixed specimens, since after fixation tentillae are contracted in both species. Thus, an obvious advantage of studying ctenophore morphology through the microscopy techniques used here is that this can be done in the lab on fixed material long after collecting.

In order to be really useful in taxonomic practice, a technique must furthermore be reasonably easy, quick and reproducible. As a counter-example, investigations using electron microscopy undoubtedly provide species-level discriminating characters, but in practice they cannot be used for species definition and routine identification because they are much too time-consuming. By contrast, fluorescent nuclei labelling and immunolabelling as employed in this study (without embedding and sectioning) are easy techniques, which after only 1-2 days of relatively simple steps of incubations and washings followed by micro-dissection of the samples will lead to microscopic preparations directly amenable to anatomical exploration. Another advantage of this approach (for instance over classical histology and transmission electron microscopy, which necessitate sectioning of the samples) is that it allows easy three-dimensional interpretation of the organisation at all scales from individual cells to general body structure. The two antibodies used here (anti-tyrosylated α -tubulin and anti- β -actin) are both commercially available and provide high-quality staining of most ctenophore cell types (including neuro-sensory cells, muscle cells, gland cells and all kinds of mesogleal cells, Jager et al. 2011). Finally, we have performed tests of these two antibodies (and Dapi staining of cell nuclei) on several distantly-related ctenophore species (in addition to Pleurobrachiidae, members of Lampeidae, Lobata and Cestida) and in all cases obtained efficient and high-quality staining, such that potentially any ctenophore species might be studied using these techniques. Admittedly a severe limitation to this claim is that appropriate fixation of the tissues is required. Unfortunately many ctenophores cannot be fixed using the procedure described above (Material and Methods); for instance tissues of cestids and most lobates will entirely disintegrate when immersed into paraformaldehyde (see e.g. Mills & Haddock 2007). However, according to our experience the fixation step should not be a problem at least for most species of the former order “Cydippida” (including for example Pleurobrachiidae, Mertensidae, Lampeidae) and for the Beroida (with some adjustments in the details, particularly concerning the permeabilisation step). For more fragile species, we have established a special fixation procedure which gives good results notably with the very delicate venus belt (*Cestum veneris*), and which will be described elsewhere.

In conclusion, we suggest that anatomical exploration using simple fluorescence microscopy techniques have a potential to become the new gold standard for morphological characterisation of ctenophore species. Possibly the expected benefits could be of the same order of magnitude as when entomologists began to study male genitalia (particularly the aedeagus) instead of only external morphology of insect species, towards the beginning of the 20th Century. Significant progress in ctenophore taxonomy will of course require parallel efforts, notably in terms of sampling (a notoriously critical bottleneck with these delicate macro-planktonic organisms) and molecular taxonomy. A standard barcoding marker remains to be determined for ctenophores. Here we used ITS sequences because there is an existing (albeit limited) ctenophore data set for this marker

(Podar et al. 2001, Simion et al. 2014), but other variable genes should be tested (e.g. mitochondrial cytochrome oxidase D). However, simply generating sequence data from collected specimens is useless if not associated with thorough morphological characterisation. Thus, a possible integrative strategy could involve (i) high-quality macroscopic photography of living specimens, measurement and macroscopic scale characterisation, (ii) fixation of the aboral region and at least part of the comb rows, for later elaboration of Dapi and antibody stained microscopic preparations, (iii) ethanol preservation of the rest of the specimen for DNA studies. Coordination between specialists of different countries is also a critical requirement for future progress in ctenophore taxonomy expected from this approach, notably for efficient sampling and data curation. Since microscopic preparations of fixed ctenophore tissues can be conserved on the long term, they could be used as type material for newly-described species. It would even be conceivable to go back to the type localities of previously-described nominal species in order to select and preserve specimens serving as neotypes, using the above-described strategy. The genus *Pleurobrachia* would be a good starting point to experiment the suggested approach. First the well-known Pacific species *P. bachei* should be characterised and compared with the two species studied here, and then it could be possible to reconstruct step by step the entire taxonomy of the genus by applying the same approach to material collected from all around the world.

Acknowledgements

We thank Olivier Gros for welcoming us in his lab at the Université des Antilles et de la Guyane of Pointe-à-Pitre, Guadeloupe, and for providing lab facilities. We are grateful to fisherman Jean-Michel Landres for boat rides off Sainte-Rose (Guadeloupe) and for his help in collecting ctenophores with the plankton net. Sailors of the Observatoire Océanologique in Villefranche-sur-Mer are also acknowledged for their help. Alicia Coste, João Gonçalves and Alexandre Alié contributed to collect of *P. pileus* specimens, and Latifa Belkessa to their maintenance in the lab. We thank C. Mills for insightful feedback after we sent her some photos of our *P. brunnea* specimens. Most ancient references (18th and 19th Centuries, early 20th Century) were accessed through the Biodiversity Heritage Library website. This work was supported by funding associated with M.M. membership to the Institut Universitaire de France (IUF).

References

- Agassiz, A. (1865) *Illustrated catalogue of the Museum of Comparative Zoology, at Harvard college. No II. North American Acalephae*. Sever & Francis, Cambridge, 227 pp.
- Agassiz, A. (1874) Embryology of the Ctenophorae. *Memoirs of the American Academy of Arts and Sciences*, 10 (3), 357-398.
- Agassiz, L. (1849) Contributions to the natural history of the Acalephae of North America. Part II. *Memoirs of the American Academy of Art and Sciences*, 4(10), 313-374 + 8 plates.
- Alié, A., Leclère, L., Jager, M., Dayraud, C., Chang, P., Le Guyader, H., Quéinnec, E. & Manuel, M. (2011) Somatic stem cells express Piwi and Vasa genes in an adult ctenophore : ancient association of « germline genes » with stemness. *Developmental Biology*, 350, 183-197.
- Bargmann, W., Jacob, K. & Rast, A. (1972) Über Tentakel und Colloblasten der Ctenophore *Pleurobrachia pileus*. *Zeitschrift für Zellforschung und mikroskopische Anatomie* 123, 121-152.
- Bennema, F.P. & van Moorsel, G.W.N.M. (2011) The status of *Callianira hexagona* (Slabber, 1778) (Ctenophora). *Zoologische Mededelingen*, 85, 825-833.
- Bigelow, H.B. (1912) The ctenophores of the “Albatross” expedition. *Bulletin of the Museum of Comparative Zoology*, 54, 369-404.
- Bigelow, H.B. (1926) *Plankton of the offshore waters of the Gulf of Maine*. U.S. Bureau of Fisheries Doc. 968. Government Print Office, Washington, 486 pp.
- Buecher, E. & Gasser, B. (1998) Estimation of predatory impact of *Pleurobrachia rhodopis* (cydippid ctenophore) in the northwestern Mediterranean Sea: in situ observations and laboratory experiments. *Journal of Plankton Research*, 20(4), 631-651.
- Castresana, J. (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, 17, 540-552
- Chun, C. (1880) *Die Ctenophoren des Golfes von Neapel und der angrenzenden Meeres-Abschnitte: eine Monographie*. Verlag von W. Engelmann, Leipzig, 313 pp., 18 plates.
- Chun, C. (1898) *Die Ctenophoren der Plankton-Expedition*. Verlag von Lipsius & Tischer, Kiel and Leipzig, 37 pp.
- Costello, J.H., Bayha, K.M., Mianzan, H.W., Shiganova, T.A. & Purcell, J.E. (2012) Transitions of *Mnemiopsis leidyi* (Ctenophora: Lobata) from a native to an exotic species: a review. *Hydrobiologia* 690, 21–46.
- Dayraud, C., Alié, A., Jager, M., Chang, P., Le Guyader, H., Manuel, M. & Quéinnec, E. (2012) Independent specialisation of myosin II paralogues in muscle vs. non-muscle functions during early animal evolution: a ctenophore perspective. *BMC Evolutionary Biology* 12, 107.
- De Haan, W. (1827) Berigt omtrent de *Callianira triloptera* Lamarck. *Bijdragen tot de natuurkundige wetenschappen, verzameld door H.C. van Hall, W. Vrolik & G.J. Mulder*, 2, 150-152.
- Derelle, R & Manuel, M. (2007) Ancient connection between NKL genes and the mesoderm? Insights from Tlx expression in a ctenophore. *Development, Genes and Evolution*, 217, 253-261.

- Derelle, R., Momose, T., Manuel, M., Da Silva, C., Wincker, P. & Houliston, E. (2010) Convergent origins and rapid evolution of « spliced leader » trans-splicing in Metazoa: insights from the Ctenophora and Hydrozoa. *RNA*, 16, 696-707.
- Eschscholtz, F. (1829) *System der Acalephen. Eine ausführliche Beschreibung aller Medusenartigen Strahlthiere*. Ferdinand Dümmler, Berlin, 190 pp. & 16 plates.
- Fleming, J. (1822) *The philosophy of zoology, or, a general view of the structure, functions, and classification of animals*. Volume II. Archibald Constable & Co, Edinburgh, Hurst, Robinson & Co, London, 618 pp.
- Franc, J.-M. (1978) Organization and function of ctenophore colloblasts: an ultrastructural study. *Biological Bulletin*, 155, 527-541.
- Freeman, G. (1977) The establishment of the oral-aboral axis in the ctenophore embryo. *Journal of Embryology and Experimental Morphology*, 42, 237-260.
- Gibbons, M.J., Buecher, E. & Thibault-Botha, D. (2003) Observations on the ecology of *Pleurobrachia pileus* (Ctenophora) in the southern Benguela ecosystem. *African Journal of Marine Science*, 25, 253-261.
- Granhag, L., Majaneva, S. & Møller, L. (2012) First recordings of the ctenophore *Euplokamis* sp. (Ctenophora, Cydippida) in Swedish coastal waters and molecular identification of this genus. *Aquatic Invasions*, 7, 455-463.
- Gronovius, L.T. (1760) Ejusdem observationes de animalculis aliquot marinae aquae innatantibus atque in littoribus belgicis obviis. *Acta Helvetica, physico-mathematico-anatomico-botanico-medica*, 4, 35-40, + plate IV, figs. 1-10.
- Hernandez-Nicaise, M.-L. (1968) Distribution et ultrastructure des synapses symétriques dans le système nerveux des Cténares. *Comptes Rendus de l'Académie des Sciences de Paris. Série D*, 267, 1731-1734.
- Hernandez-Nicaise, M.-L. (1974) Ultrastructural evidence for a sensory-motor neuron in Ctenophora. *Tissue & Cell*, 6(1), 43-47.
- Hernandez-Nicaise, M.-L. (1991) Ctenophora. In: Harrison, F.W. & Westfall, J.A. (Eds.), *Microscopic Anatomy of Invertebrates. Volume II: Placozoa, Porifera, Cnidaria, and Ctenophora*. John Wiley, New York, pp. 359-418.
- Hirota, J. (1974) Quantitative natural history of *Pleurobrachia bachei* in La Jolla bight. *Fishery Bulletin*, 72, 295-335.
- Horrige, G.A. (1964a) Presumed photoreceptive cilia in a ctenophore. *Quarterly Journal of Microscopical Science*, 105(3), 311-317.
- Horrige, G.A. (1964b) The giant mitochondria of ctenophore comb-plates. *Quarterly Journal of Microscopical Science*, s3, 301-310.
- Horrige, G.A. & Mackay, B. (1964) Neurociliary synapses in *Pleurobrachia* (Ctenophora). *Quarterly Journal of Microscopical Science*, 105(2), 163-174.
- Jager, M., Quéinnec, E., Houliston, E. & Manuel, M. (2006) Expansion of the SOX gene family predated the emergence of the Bilateria. *Molecular Phylogenetics and Evolution*, 39 (2), 468-477.

- Jager, M., Quéinnec, E., Chiori, R., Le Guyader, H. & Manuel, M. (2008) Insights into the early evolution of SOX genes from expression analyses in a ctenophore. *Journal of Experimental Zoology part B Molecular and Developmental Evolution*, 310B (8), 650-667.
- Jager, M., Chiori, R., Alié, A., Dayraud, C., Quéinnec, E. & Manuel, M. (2011). New insights on ctenophore neural anatomy: immunofluorescence study in *Pleurobrachia pileus* (Müller, 1776). *Journal of Experimental Zoology part B Molecular and Developmental Evolution* 316B (3), 171-187.
- Jager, M., Dayraud, C., Mialot, A., Quéinnec, E., Le Guyader, H. & Manuel, M. (2013) Evidence for involvement of Wnt signalling in body polarities, cell proliferation, and the neurosensory system in an adult ctenophore. *PLOS ONE*, 8(12), e84363.
- Johnson, W.S. & Allen, D.M. (2012) *Zooplankton of the Atlantic and Gulf coasts. A guide to their identification and ecology*. 2nd Edition. Johns Hopkins Univ. Press, Baltimore, 452 pp.
- Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T. & Higgins, D. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.
- Linnaeus, C. (1767) *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Editio duodecima. Tomus primus. Pars secunda. Impensis Direct. Laurentii Salvii, Holmia (Stockholm), pp. 533-1364.
- Mayer, A.G. (1912) *Ctenophores of the Atlantic coast of North America*. Carnegie Institution of Washington, Washington D.C., 58 pp.
- Mianzan, H.W. (1999) Ctenophora. In: Boltovskoy, D. (Ed.), *South Atlantic Zooplankton*. Backhuys Publishers, Leiden, pp. 561–573.
- Mianzan, H., Dawson, E.W. & Mills, C. (2009) *Phylum Ctenophora: comb jellies*. In: Gordon, D.P. (Ed.), *New Zealand inventory of biodiversity: 1. Kingdom Animalia: Radiata, Lophotrochozoa, Deuterostomia*. Canterbury University Press, Christchurch, pp. 49-58.
- Mills, C.E. (1987) Revised classification of the genus *Euplokamis* Chun, 1880 (Ctenophora: Cydippida: Euplokamidae n. fam.) with a description of the new species *Euplokamis dunlapae*. *Canadian Journal of Zoology*, 65, 2661-2668.
- Mills, C.E. (1998-present) *Phylum Ctenophora: list of all valid species names*. Electronic internet document established March 1998, last updated 16 March 2014. Distributed by the author from the URL <http://faculty.washington.edu/cemills/Ctenolist.html>
- Mills, C.E., Pugh, P.R., Harbison, G.R. & Haddock, S.H.D. (1996) Medusae, siphonophores and ctenophores of the Alborán Sea, south western Mediterranean. *Scientia Marina*, 60 (1), 145-163.
- Mills, C.E. & Haddock, S.H.D. (2007) *Ctenophora*. In: Carlton, J.T. (Ed.), *Intertidal invertebrates of the Central California coast*. 4th edition. University of California Press, Berkeley, 1019 pp.
- Molinero, J.C., Casini, M. & Buecher, E. (2008) The influence of the Atlantic and regional climate variability on the long-term changes in gelatinous carnivore populations in the northwestern Mediterranean. *Limnology and Oceanography*, 53(4), 1456-1467.

- Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M., Ptitsyn, A., Reshetov, D., Mukherjee, K., Moroz, T.P., Bobkova, Y., Yu, F., Kapitonov, V.V., Jurka, J., Bobkov, Y.V., Swore, J.J., Girardo, D.O., Fodor, A., Gusev, F., Sanford, R., Bruders, R., Kittler, E., Mills, C.E., Rast, J.P., Derelle, R., Solovyev, V.V., Kondrashov, F.A., Swalla, B.J., Sweedler, J.V., Rogaev, E.I., Halanych, K.M. & Kohn, A.B. (2014) The ctenophore genome and the evolutionary origins of neural systems. *Nature*, 510(7503), 109-114.
- Mortensen, T. (1912) Ctenophora. *The Danish Ingolf-Expedition*, 5(2), 1-98.
- Moser, F. (1903) *Die Ctenophoren der Siboga-Expedition*. E.J. Brill, Leiden, 34 pp., 4 plates.
- Moser, F. (1908) Neues über Ctenophoren. Mitt. III. Die Ctenophoren der deutschen Sudpolar-Expedition. *Zoologischer Anzeiger Leipzig*, 33, 750-759.
- Moser, F. (1909) Die Ctenophoren der Deutschen Südpolar-Expedition. *Zoologie*, 11, 117-192, plates 120-122.
- Moss, A.G. (1991) The physiology of feeding in the ctenophore *Pleurobrachia pileus*. *Hydrobiologia*, 216/217, 19-25.
- Moss, A.G. & Tamm, S.L. (1987) A calcium regenerative potential controlling ciliary reversal is propagated along the length of ctenophore comb plates. *Proceedings of the National Academy of Sciences of the United States of America*, 84, 6476-6480.
- Moss, A.G. & Tamm, S.L. (1993) Patterns of electrical activity in comb plates of feeding *Pleurobrachia* (Ctenophora). *Philosophical Transactions of the Royal Society of London, Series B*, 339, 1-16.
- Müller, O.F. (1776). *Zoologiae Danicae prodromus seu animalium Daniae et Norvegiae indigenarum characteres, nomine, et synonyma imprimis popularium*. Typis Hallageriis, Havnia (Copenhagen), 282 pp.
- Palma, S.G. & Meruane, J.Z. (1975) Aspectos ecológicos y crecimiento de *Pleurobrachia pileus* (Ctenophora) en la región de Valparaíso. *Investigative Marine Valparaíso*, 6, 25-40.
- Patterson, R. (1838). Description of the *Cydippe Pomiformis* mihi, (*Beroë Ovatus*, Flem.) with notice of an apparently undescribed species of Bolina, also found on the coast of Ireland. *Transactions of the Royal Irish Academy*, 19, 91-109.
- Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Deniel, E., Houlston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D.J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G. & Manuel, M. (2009) Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, 19, 706-712.
- Podar, M., Haddock, S.H.D., Sogin, M.L. & Harbison, G.R. (2001) A molecular phylogenetic framework for the phylum Ctenophora using 18S rRNA genes. *Molecular Phylogenetics and Evolution*, 21 (2), 218-230.
- Ralph, P.M. (1950) Ctenophores of the waters of Cook Strait and Wellington Harbour. *Transactions of the Royal Society of New Zealand*, 78 (1), 70-82.
- Riedl, R. (1986) *Fauna y flora del mar Mediterraneo, con 3512 ilustraciones*. Ediciones Omega, Barcelona, 858 pp.

- Schnitzler, C.E., Pang, K., Powers, M.L., Reitzel, A.M., Ryan, J.F., Simmons, D., Tada, T., Park, M., Gupta, J., Brooks, S.Y., Blakesley, R.W., Yokoyama, S., Haddock, S.H.D., Martindale, M.Q. & Baxevanis, A.D. (2012) Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidyi*: a new view of ctenophore photocytes. *BMC Biology*, 10, 107.
- Simion, P., Bekkouche, N., Jager, M., Quéinnec, E. & Manuel, M (2014) Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora. *Zoology*, e-pub ahead of print.
- Sleigh, M.A. (1968) Metachronal co-ordination of the comb plates of the ctenophore *Pleurobrachia*. *Journal of Experimental Biology*, 48, 111-125.
- Sleigh, M.A. (1972) Features of ciliary movement of the ctenophores *Beroë*, *Pleurobrachia* and *Cestus*. In: Clark, R.B. & Wootton, R. (Eds.), *Essays in Hydrobiology*. Exeter University Press, Exeter, pp. 119–136.
- Stamatakis, A. (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, advance access published 21 Jan 2014.
- Tamm, S.L. (1984) Mechanical synchronization of ciliary beating within comb plates of ctenophores. *Journal of Experimental Biology*, 113, 401-408.
- Tamm, S.L. (2012) Patterns of comb row development in young and adult stages of the ctenophores *Mnemiopsis leidyi* and *Pleurobrachia pileus*. *Journal of Morphology*, 273(9), 1050-1063.
- Tamm, S. L. & Tamm, S. (1993) Diversity of macrociliary size, tooth patterns, and distribution in *Beroë* (Ctenophora). *Zoomorphology*, 113, 79-89.
- Trégouboff, G. & Rose, M. (1957) *Manuel de planctonologie Méditerranéenne. Vol. I & II*. Centre National de la Recherche Scientifique, Paris.
- Yip, S.Y. (1984) Parasites of *Pleurobrachia pileus* Müller, 1776 (Ctenophora), from Galway Bay, western Ireland. *Journal of Plankton Research*, 6 (1), 107-121.

Figure legends

FIGURE 1. Drawings from Gronovius (1760), accompanying his description of a ctenophore to which Linnaeus (1767) explicitly referred when he created the species name *Volvox bicaudatus*. This ctenophore is clearly the species we now call *Pleurobrachia pileus*.

FIGURE 2. Morphological features of *Pleurobrachia pileus*. Oral pole on top in (A-D). (A) Living adult specimen from Villefranche-sur-Mer, France (Mediterranean Sea). (B) Living adult specimen from Gravelines, France (North Sea). (C) Fixed specimen from Gravelines shown after rehydration, in longitudinal view. (D) Living specimen from Villefranche-sur-Mer showing the aspect of tentillae usual for this species when the tentacles are extended. (E) Fixed specimen from Gravelines shown after rehydration, in aboral view. (F) Detailed view of the aboral pole (apical organ and proximal part of polar fields) in the same specimen than (E), showing absence of pigmentation. (G) Pattern of canal branching in the gastro-vascular system (fixed and rehydrated specimen from Gravelines), equatorial view. The specimen was dissected as to remove part of the aboral and oral hemispheres. Labels: ad: adradial canal; cr: comb row; int: interradial canal; per: perradial canal; pf: polar field; st: statolith; ttsh: tentacle sheath. Scale bars: A, D: 5 mm; B, C, E: 1 mm; F, G: 200 μ m.

FIGURE 3. Morphological features of *Pleurobrachia brunnea*. Oral pole on top in (A-D). (A) Drawing of *P. brunnea* from Mayer (1912, plate I, Fig. 2). (B) Living adult specimen from Guadeloupe. (C) Fixed specimen from Guadeloupe shown after rehydration, in longitudinal view. (D) Living specimen from Guadeloupe showing the typical aspect of tentillae (contracted and coiled) when the tentacles are extended (white arrow: terminal thickening or knob of tentacle axis). (E) Detailed view of the extremity of a tentacle in a fixed and rehydrated specimen from Guadeloupe, showing terminal knob (arrow) and coiling of tentillae (arrowheads). (F) Fixed specimen from Guadeloupe shown after rehydration, in aboral view. (G) Detailed view of the aboral pole (apical organ and proximal part of polar fields) in the same specimen than (F), showing pigmented red spots at the base of polar fields and in the wall of the apical organ. (H) Pattern of canal branching in the gastro-vascular system (fixed and rehydrated specimen from Guadeloupe), equatorial view. The specimen was dissected as to remove part of the aboral and oral hemispheres. Labels: ad: adradial canal; cr: comb row; int: interradial canal; per: perradial canal; pf: polar field; st: statolith; ttsh: tentacle sheath. Scale bars: B, C, D: 1 mm; E, G, H: 100 μ m; F: 500 μ m.

FIGURE 4. Phylogenetic relationships of Pleurobrachiidae, reconstructed from 18S rRNA+ITS1+5.8SrRNA+ITS2 sequences under Maximum Likelihood (see Material and Methods). The genus *Pleurobrachia* is boxed in light blue and the species *Pleurobrachia pileus* in beige. Refer to Table 1 for specimen collection data. Numbers below the branches are bootstrap support values. The scale bar indicates the inferred number of substitutions per site.

FIGURE 5. Structural comparison of the apical organs of *P. pileus* (A, B, E, G) and *P. brunnea* (C, D, F, H) using DIC light microscopy (A-D) and Dapi fluorescent staining of nuclei (E-H). Focus is in the apical organ floor at relatively superficial depth. Tentacular plane vertical and pharyngeal plane horizontal in all pictures. (B) and (D) correspond, respectively, to the boxed areas in (A) and (C), but shown in different individuals. Similarly, (G) and (H) correspond respectively to the areas boxed with dotted line in (E) and (F), but in different individuals. Labels: ba: balancers; lb: lamellate bodies; lit: lithocytes. Scale bars: A, C, E, F: 50 μm ; B, D: 10 μm ; G, H: 25 μm .

FIGURE 6. Structural differences observed within the polar fields between *P. pileus* (A, C, E) and *P. brunnea* (B, D, F). (A, B) Dapi staining of cell nuclei. The pictures show a portion of polar field along its whole width, at approximately equal distance from apical organ and distal extremity of polar field. Dotted lines delineate the different regions: outer marginal zone (ma Z1), inner marginal zone (ma Z2), central zone (ce Z). The white arrows in (A) point to a narrow stripe with lower density of nuclei within ma Z1, consistently observed in *P. pileus* but not in *P. brunnea* (B). (C, D) Portion of polar field stained with the anti-tyrosylated α -tubulin antibody. The double arrows indicate the extension of the polar field muscle, narrower in *P. pileus* (C) than in *P. brunnea* (D). (E, F) Higher magnification views of the boxed areas in (C) and (D). The dotted line materialised the limit between the marginal zone (above) and the central zone (below); the lower border of these pictures corresponds approximately to the polar field midline. The arrows in (E) and (F) point to muscle fibres of the polar field muscle. Note that two Z-bodies are visible in (F) (above the dotted line) while Z-bodies are not on focus in (E). Scale bars: A-F: 50 μm ; G, H: 25 μm .

FIGURE 7. Differences observed within the epidermal nerve plexus by anti-tyrosylated- α -tubulin immunolabelling, between *P. pileus* (A, C, E) and *P. brunnea* (B, D, F). (A, B) Partial view of the aboral hemisphere (aboral pole on top), centred on the tentacular plane. (C, D) Partial view of the equatorial region, centred on the tentacular plane, with the tentacle sheath opening (op) on top. (E, F) Higher magnification views of the areas boxed with dotted line respectively in (A) and (B), but with orientation rotated by 90° (tentacular plane horizontal). These pictures show that the juxta-tentacular nerve cord is much thicker in *P. pileus* (E) than in *P. brunnea* (F), neighbouring meshes of the nerve plexus being furthermore much wider in the former than in the latter (E and F shown at same scale; individuals of similar sizes). Labels: ao: apical organ; cg: ciliated groove; cp: polar field; jtt: juxta-tentacular nerve cord; op: tentacle sheath opening. Scale bars: A-D: 100 μm ; E, F: 50 μm ; G, H: 25 μm .

FIGURE 8. Comparison of the arrangement of inter-plate fibres within comb rows in *P. pileus* (A, C) and *P. brunnea* (B, D). Interplate fibres are visualised in these pictures using the anti- β -actin antibody (for pictures A and B, in combination with Dapi staining of nuclei). (A, B): Aboral extremity of one comb row; the five aboral-most combs are visible. (C, D) Pictures showing two adjacent basal cushions and corresponding inter-plate spaces, along a comb row at approximately equal distance from its aboral and oral extremities. Note that the

diffuse green colour in basal cushions in (A) and (C) corresponds to β -actin present in polster cell bodies, and appears stronger than in (B) and (D) only due to higher exposure time. Scale bars: 100 μ m.

Table 1
Data concerning ctenophore sequences used in this study.

Species	Family, order	Locality	GB accession number	Reference
<i>Pleurobrachia brunnea</i>	Pleurobrachiidae, “Cydippida”	Sainte-Rose, Guadeloupe	KJ754154, KJ754163	Simion et al. 2014
<i>Pleurobrachia pileus</i> 1	Pleurobrachiidae, “Cydippida”	Villefranche-sur-Mer, France	KJ754153, KJ754162	Simion et al. 2014
<i>Pleurobrachia pileus</i> 2	Pleurobrachiidae, “Cydippida”	Woods Hole, MA, USA	AF293678	Podar et al. 2001
<i>Pleurobrachia pileus</i> 3	Pleurobrachiidae, “Cydippida”	Roscoff, France	XXXXXXX	This study
<i>Pleurobrachia bachei</i>	Pleurobrachiidae, “Cydippida”	Santa Barbara, CA, USA	AF293677	Podar et al. 2001
<i>Hormiphora plumosa</i>	Pleurobrachiidae, “Cydippida”	Tortugas, FLO, USA	AF293676	Podar et al. 2001
<i>Haeckelia beehleri</i>	Haeckeliidae, “Cydippida”	Santa Barbara, CA, USA	AF293673	Podar et al. 2001
<i>Beroe cucumis</i>	Beroidae, Beroida	Gulf Stream, FLO, USA	AF293695	Podar et al. 2001
<i>Leucothea multicornis</i>	Leucotheidae, Lobata	Villefranche-sur-Mer, France	KJ754159, KJ754166	Simion et al. 2014
<i>Cestum veneris</i>	Cestidae, Cestida	Villefranche-sur-Mer, France	KJ754161, KJ754165	Simion et al. 2014

Table 2

Summary of morphological differences between *P. pileus* and *P. brunnea*. Data source: characters 1-5: Mayer (1912) (confirmed by observations reported here for characters 1, 4, 5); characters 6-16: this study. Asterisks: morphological characters of doubtful diagnostic value (see text, section Taxonomy).

Character	<i>P. pileus</i>	<i>P. brunnea</i>
Classical morphology		
1. Body shape	Sub-sphaerical to widely ovoid	Ovoid
2. Pigmentation of pharynx*	Whitish	Yellowish
3. Extremity of tentacle axis*	Not conspicuously thickened	Thickened into terminal knob
4. Number and spacing of tentillae	Tentillae numerous and rather closely spaced	Tentillae fewer (about 12 in specimens of length ca 5 mm), somewhat more widely spaced
5. Aspect of tentillae, tentacle extended	Generally extended and relaxed	Generally contracted and coiled into helices (at least for rapid swimming)
6. Pigmentation around apical organ	None or very faint	Four intense red spots on each side of polar field bases and two smaller red spots in apical organ wall
DIC or fluorescence microscopy		
7. Distribution of lamellate bodies (apical organ floor)	In two pairs, each lamellate body well individualised, rounded	Closely tied up two by two in pharyngeal plane
8. Cell nuclei between both lamellate bodies of each pair (superficial view)	Two parallel and regular rows	Arranged as in Fig. 5H
9. Density of cell nuclei in outer region of polar field marginal zone (ma Z1)	Discontinuous, with narrow stripe of lower nucleus density near internal border of ma Z1	Nucleus density more or less continuously decreasing from outer to inner border of ma Z1
10. Extension of polar field muscle in transverse direction	Spanning only internal half of central zone	Generally spanning whole central zone
11. Size of meshes of epidermal nerve plexus	Rather large.	On average about twice smaller (epidermal nerve plexus denser)
12. Aboral end of juxta-tentacular nerve cord	Closer to apical organ	Farther away from apical organ
13. Juxta-tentacular nerve cord on oral side of tentacle sheath opening	Present	Absent
14. Thickness of juxta-tentacular nerve cord	Very thick, much thicker than adjacent strands of epidermal nerve plexus	Only about twice thicker than adjacent strands of nerve plexus
15. Shape of meshes of epidermal nerve plexus near juxta-tentacular nerve cord	With marked longitudinal (oral/aboral) orientation	Polygonal
16. Arrangement of inter-plate fibres (comb rows)	Narrow central bundle of complete fibres and numerous lateral incomplete fibres	Wide bundle of complete fibres; lateral incomplete fibres very few

Figure 1

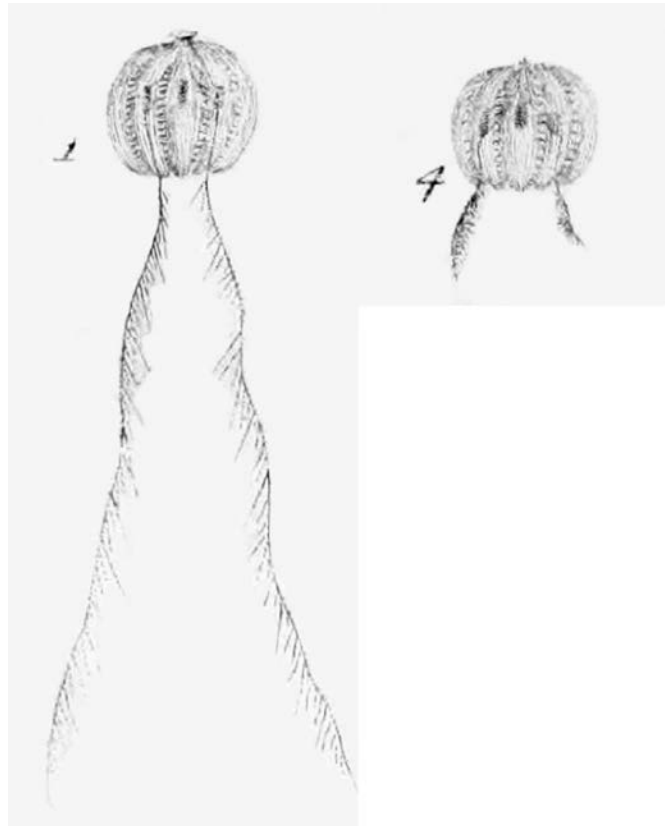


Figure 2

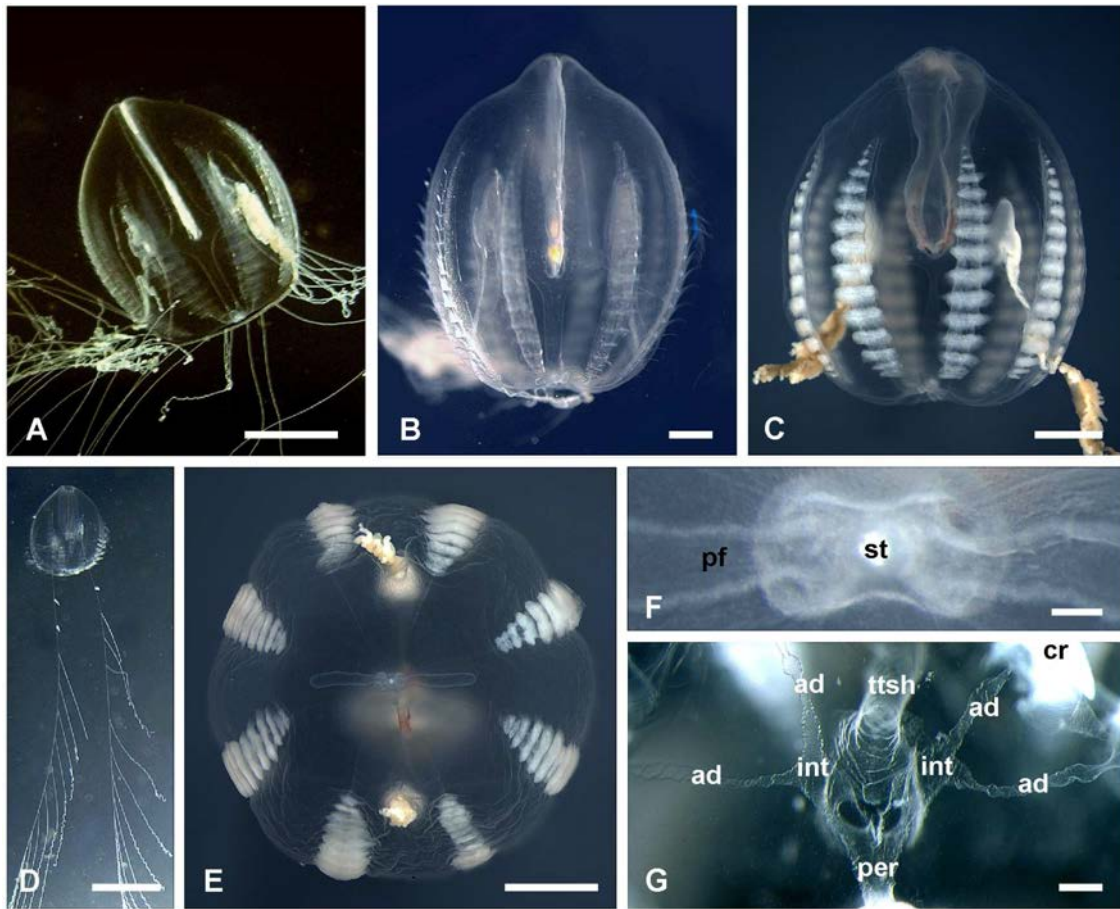


Figure 3

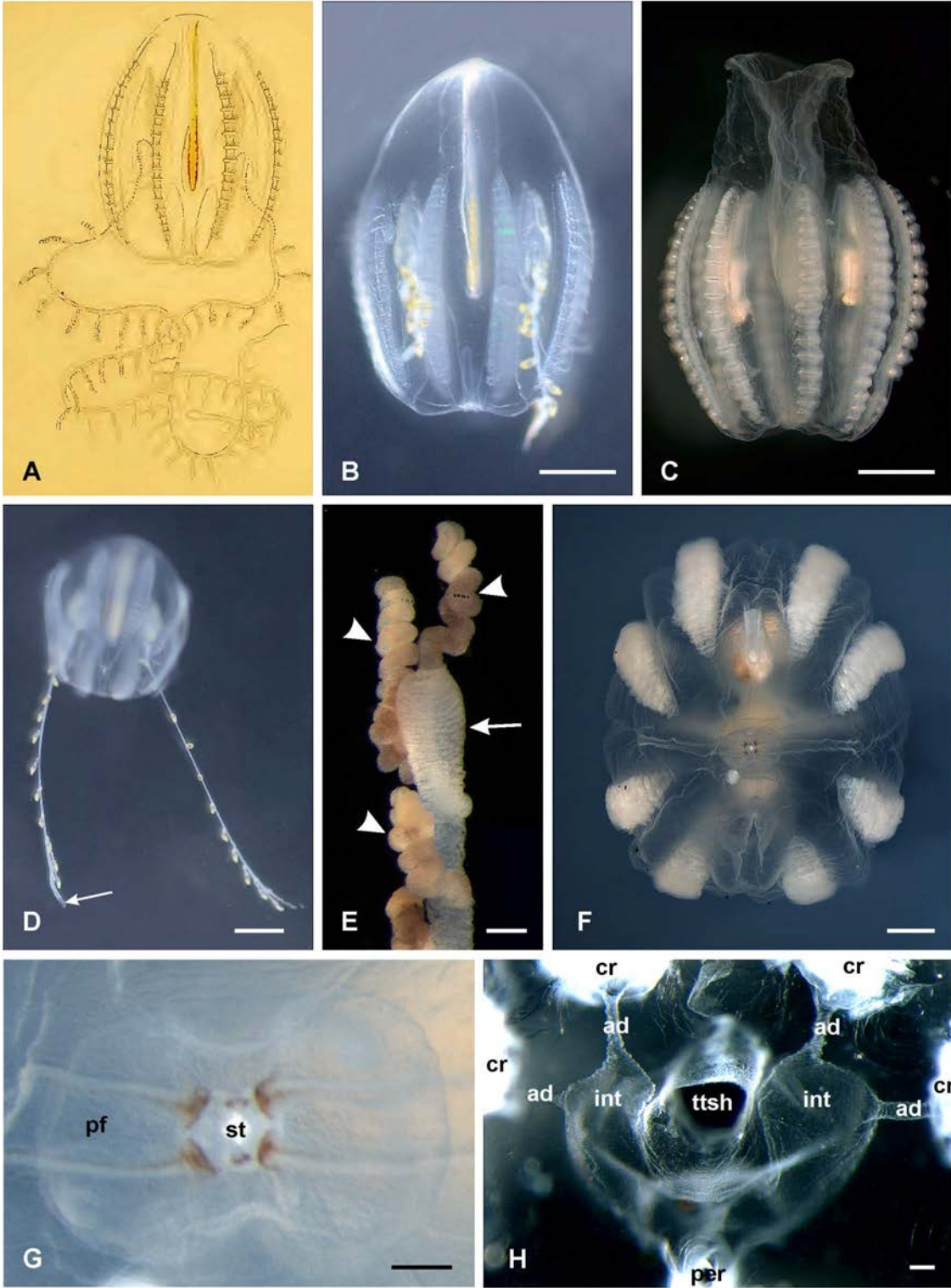


Figure 4

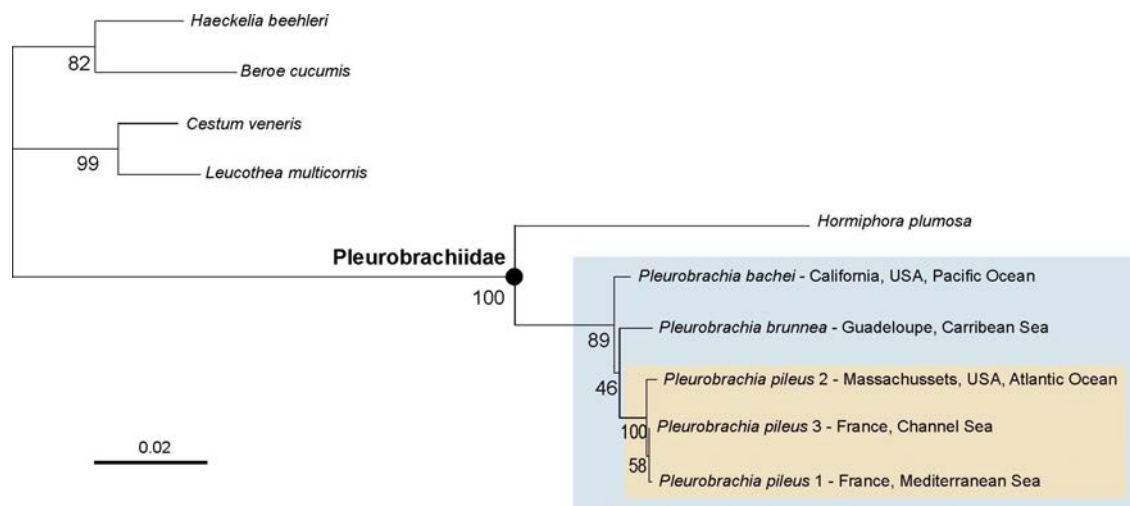


Figure 5

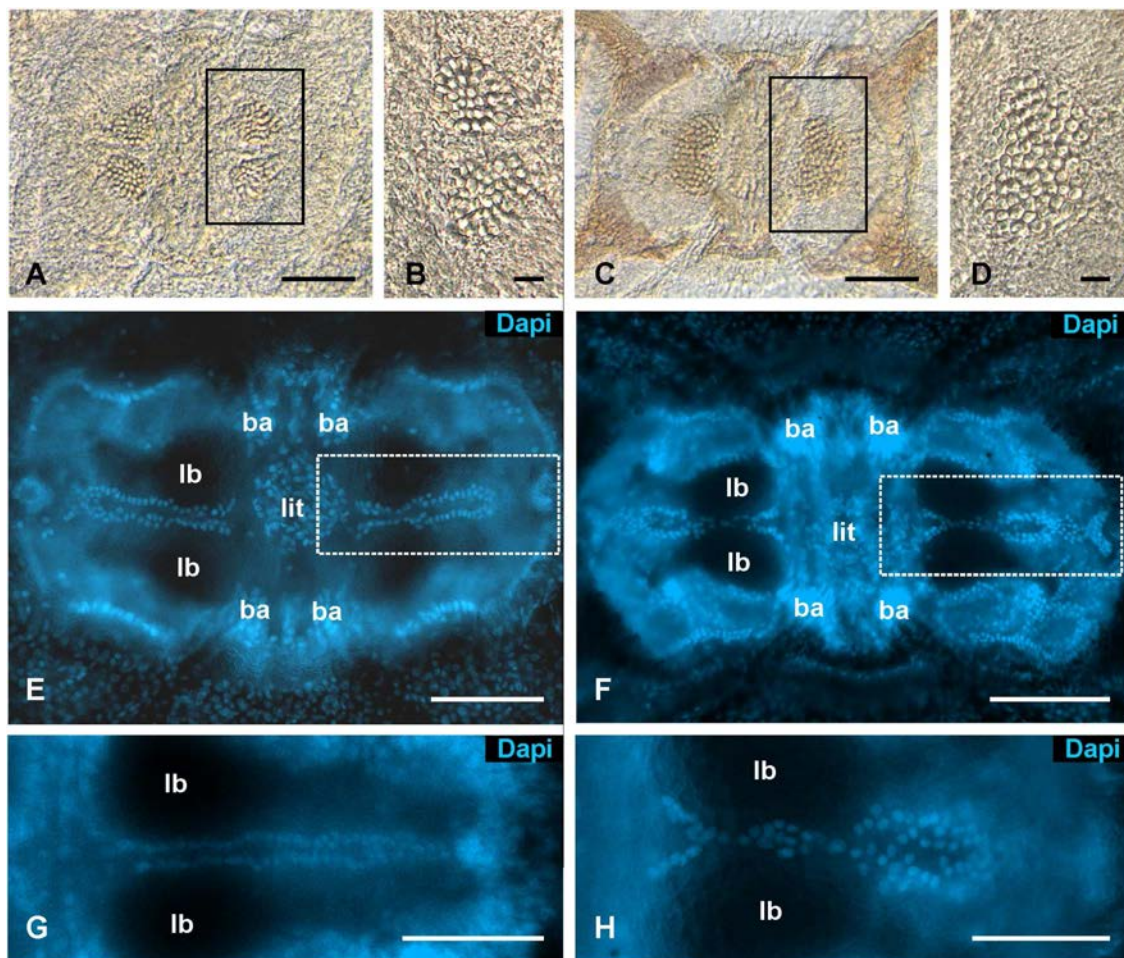


Figure 6

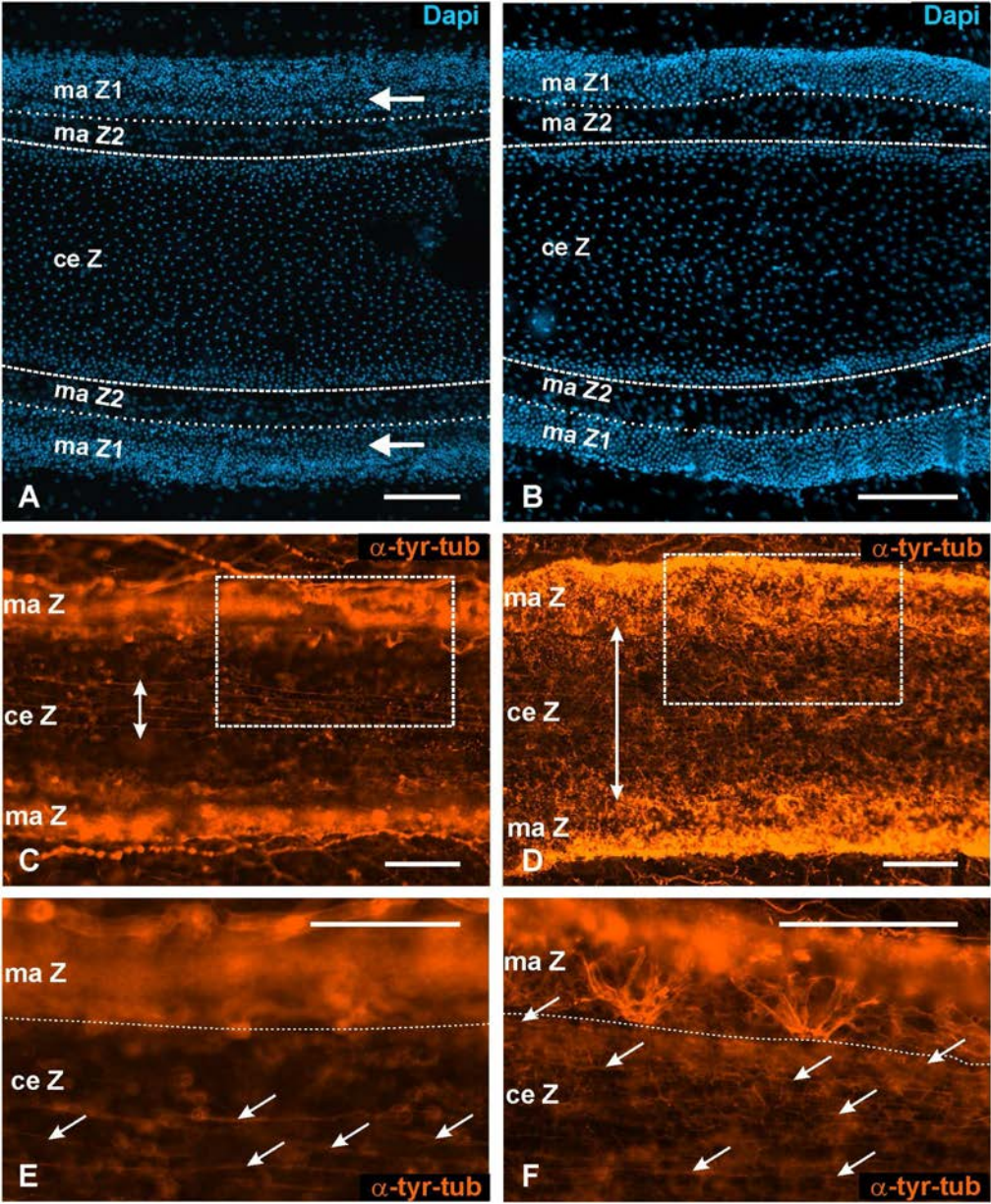


Figure 7

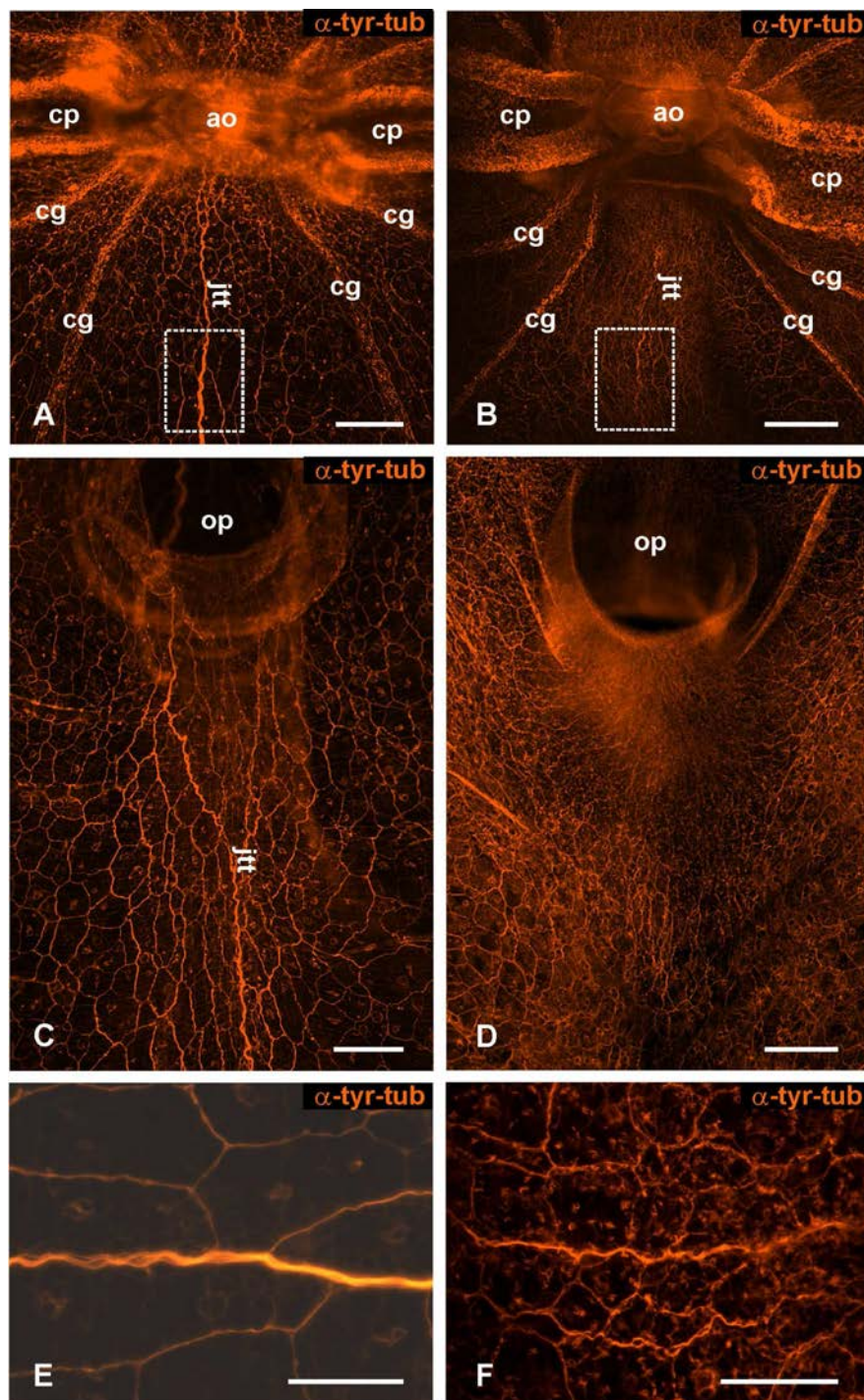
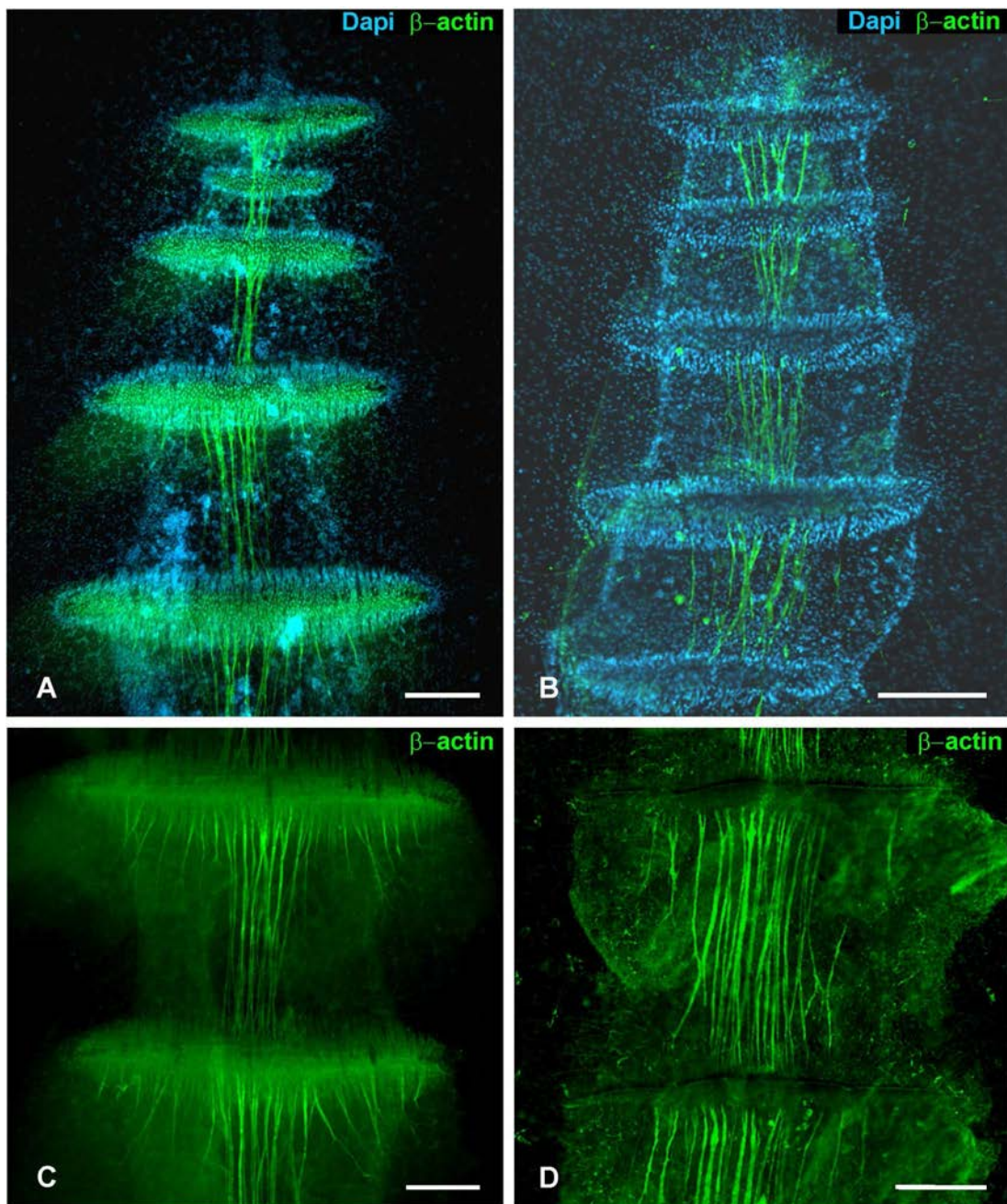


Figure 8



Annexe I

List of 25 publications in the past 50 years allowing protection of the name *Pleurobrachia pileus* (O.F. Müller 1776) in application of ICZN Art. 23.9.1.

- Alié, A., Leclère, L., Jager, M., Dayraud, C., Chang, P., Le Guyader, H., Quéinnec, E. & Manuel, M. (2011) Somatic stem cells express Piwi and Vasa genes in an adult ctenophore : ancient association of « germline genes » with stemness. *Developmental Biology*, 350, 183-197.
- Bargmann, W., Jacob, K. & Rast, A. (1972) Über Tentakel und Colloblasten der Ctenophore *Pleurobrachia pileus*. *Zeitschrift für Zellforschung und mikroskopische Anatomie* 123, 121-152.
- Dayraud, C., Alié, A., Jager, M., Chang, P., Le Guyader, H., Manuel, M. & Quéinnec, E. (2012) Independent specialisation of myosin II paralogues in muscle vs. non-muscle functions during early animal evolution: a ctenophore perspective. *BMC Evolutionary Biology* 12, 107.
- Derelle, R. & Manuel, M. (2007) Ancient connection between NKL genes and the mesoderm? Insights from Tlx expression in a ctenophore. *Development, Genes and Evolution*, 217, 253-261.
- Derelle, R., Momose, T., Manuel, M., Da Silva, C., Wincker, P. & Houliston, E. (2010) Convergent origins and rapid evolution of « spliced leader » trans-splicing in Metazoa: insights from the Ctenophora and Hydrozoa. *RNA*, 16, 696-707.
- Freeman, G. (1977) The establishment of the oral-aboral axis in the ctenophore embryo. *Journal of Embryology and Experimental Morphology*, 42, 237-260.
- Gibbons, M.J., Buecher, E. & Thibault-Botha, D. (2003) Observations on the ecology of *Pleurobrachia pileus* (Ctenophora) in the southern Benguela ecosystem. *African Journal of Marine Science*, 25, 253-261.
- Horridge, G.A. (1964a) Presumed photoreceptive cilia in a ctenophore. *Quarterly Journal of Microscopical Science*, 105(3), 311-317.
- Horridge, G.A. (1964b) The giant mitochondria of ctenophore comb-plates. *Quarterly Journal of Microscopical Science*, s3, 301-310.
- Horridge, G.A. & Mackay, B. (1964) Neurociliary synapses in *Pleurobrachia* (Ctenophora). *Quarterly Journal of Microscopical Science*, 105(2), 163-174.
- Jager, M., Quéinnec, E., Houliston, E. & Manuel, M. (2006) Expansion of the SOX gene family predated the emergence of the Bilateria. *Molecular Phylogenetics and Evolution*, 39 (2), 468-477.
- Jager, M., Quéinnec, E., Chiori, R., Le Guyader, H. & Manuel, M. (2008) Insights into the early evolution of SOX genes from expression analyses in a ctenophore. *Journal of Experimental Zoology part B Molecular and Developmental Evolution*, 310B (8), 650-667.
- Jager, M., Chiori, R., Alié, A., Dayraud, C., Quéinnec, E. & Manuel, M. (2011). New insights on ctenophore neural anatomy: immunofluorescence study in *Pleurobrachia pileus* (Müller, 1776). *Journal of Experimental Zoology part B Molecular and Developmental Evolution* 316B (3), 171-187.
- Jager, M., Dayraud, C., Mialot, A., Quéinnec, E., Le Guyader, H. & Manuel, M. (2013) Evidence for involvement of Wnt signalling in body polarities, cell proliferation, and the neurosensory system in an adult ctenophore. *PLOS ONE*, 8(12), e84363.
- Johnson, W.S. & Allen, D.M. (2012) *Zooplankton of the Atlantic and Gulf coasts. A guide to their identification and ecology*. 2nd Edition. Johns Hopkins Univ. Press, Baltimore, 452 pp.
- Mianzan, H., Dawson, E.W. & Mills, C. (2009) *Phylum Ctenophora: comb jellies*. In: Gordon, D.P. (Ed.), *New Zealand inventory of biodiversity: 1. Kingdom Animalia: Radiata, Lophotrochozoa, Deuterostomia*. Canterbury University Press, Christchurch, pp. 49-58.
- Moss, A.G. (1991) The physiology of feeding in the ctenophore *Pleurobrachia pileus*. *Hydrobiologia*, 216/217, 19-25.
- Moss, A.G. & Tamm, S.L. (1987) A calcium regenerative potential controlling ciliary reversal is propagated along the length of ctenophore comb plates. *Proceedings of the National Academy of Sciences of the United States of America*, 84, 6476-6480.
- Moss, A.G. & Tamm, S.L. (1993) Patterns of electrical activity in comb plates of feeding *Pleurobrachia* (Ctenophora). *Philosophical Transactions of the Royal Society of London, Series B*, 339, 1-16.
- Palma, S.G. & Meruane, J.Z. (1975) Aspectos ecológicos y crecimiento de *Pleurobrachia pileus* (Ctenophora) en la region de Valparaíso. *Investigative Marine Valparaíso*, 6, 25-40.
- Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Deniel, E., Houliston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D.J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G. & Manuel, M. (2009) Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, 19, 706-712.
- Podar, M., Haddock, S.H.D., Sogin, M.L. & Harbison, G.R. (2001) A molecular phylogenetic framework for the phylum Ctenophora using 18S rRNA genes. *Molecular Phylogenetics and Evolution*, 21 (2), 218-230.
- Tamm, S.L. (1984) Mechanical synchronization of ciliary beating within comb plates of ctenophores. *Journal of Experimental Biology*, 113, 401-408.
- Tamm, S.L. (2012) Patterns of comb row development in young and adult stages of the ctenophores *Mnemiopsis leidyi* and *Pleurobrachia pileus*. *Journal of Morphology*, 273(9), 1050-1063.
- Yip, S.Y. (1984) Parasites of *Pleurobrachia pileus* Müller, 1776 (Ctenophora), from Galway Bay, western Ireland. *Journal of Plankton Research*, 6 (1), 107-121.

4.3 Compléments - Micro-anatomie comparée à une échelle inter-ordinale : étude de *Beroe gracilis* KÜNNE, 1939

Introduction

Les résultats du travail précédent apportent des perspectives prometteuses pour la taxonomie des cténophores, en montrant que même en comparant deux espèces très proches entre elles, leur similarité morphologique apparente à l'échelle macroscopique cache de nombreuses différences structurales à l'échelle micro-anatomique. Les marquages immunohistochimiques chez les cténophores sont voués à jouer un rôle important dans l'avenir des travaux sur leur systématique.

La micro-anatomie des cténophores est probablement informative également à des échelles taxonomiques plus grandes. J'ai réalisé des marquages immunohistochimiques sur *Beroe gracilis*, une espèce de l'ordre des Beroida, morphologiquement et moléculairement éloignée des deux espèces de *Pleurobrachia* décrites précédemment (voir section 1.1.1). Cette section en présente les résultats préliminaires.

Matériels et méthodes

Récolte et identification Les individus étudiés ont été capturés sur les plages de l'Île de Ré en Mai 2012 et à Gravelines en Mai 2013 à l'aide de filets troubleaux et de filets à plancton. Leur appartenance à l'espèce *Beroe gracilis* est sans ambiguïté : 1) absence de diverticules au niveau des canaux méridiens; 2) taille correspondant à la description de Künne (1939); 3) localisation géographique cohérente avec la répartition connue de l'espèce (Müller 2004).

Fixation et marquage immunohistochimiques Tous les individus ont été fixés selon le protocole utilisé pour la fixation des espèces de *Pleurobrachia* décrit dans l'article 3. De la même manière, les marquages immunohistochimiques ont été réalisés en suivant strictement le protocole décrit dans l'article 3. Les marquages en phalloïdine (marqueur de l'actine F) ont été réalisés sur des spécimens qui n'ont pas été déshydratés, en les incubant 45 mn en PBTr + Phalloïdine 1/100 (stockée à 1 mg/ml en méthanol).

Observations préliminaires et discussion de la micro-anatomie de *Beroe gracilis*

La morphologie générale de *Beroe gracilis* a été présentée en détail dans le chapitre 1, section 1.1.1. Les Beroida se caractérisent principalement par l'absence d'appareil tentaculaire. On insistera particulièrement ici sur les structures anatomiques décrites pour les deux *Pleurobrachia* dans l'Article 3.

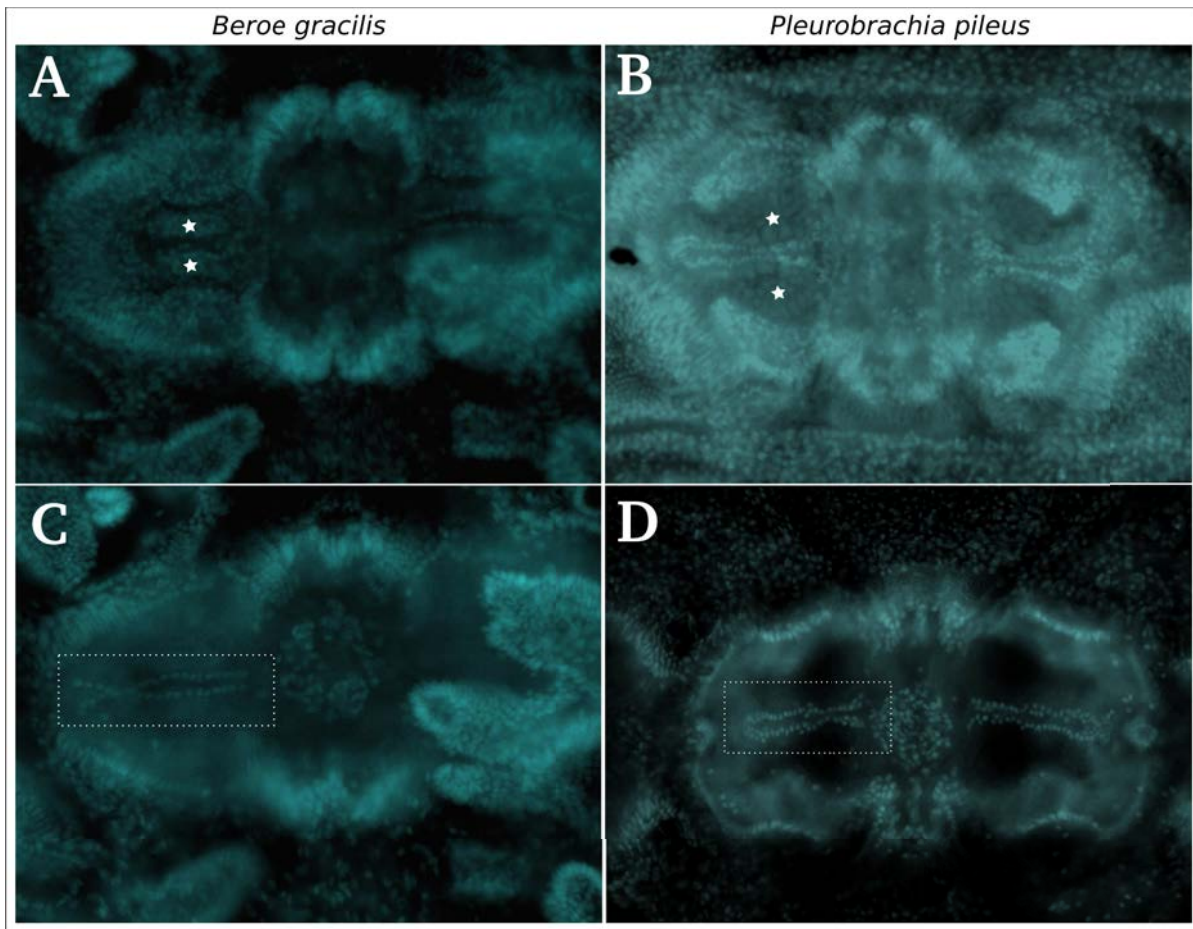


FIGURE 4.2 – Comparatif de structures de l’organe aboral vues en marquage DAPI chez *Beroe gracilis* et *Pleurobrachia pileus*. A et B) vues du plancher de l’organe aboral, les groupes de noyaux des corps réfringents sont marqués d’une étoile blanche. C et D) vues superficielles de l’organe aboral, où deux lignes de noyaux de formes différentes selon les espèces passent entre les corps réfringents.

Organe aboral La densité et la distribution des noyaux des corps réfringents, au niveau du plancher de l’organe aboral, sont différentes entre *Beroe gracilis* et *Pleurobrachia pileus* (Figure 4.2 A et B). Ces massifs de cellules sont plus denses et plus allongés (presque rectangulaire) chez *Beroe gracilis*. La disposition des noyaux des cellules passant entre ces deux corps réfringents est également différente entre les espèces (Figure 4.2 C et D). Chez *Beroe gracilis*, ces noyaux forment deux lignes parallèles dans la région proche du statolithe, comme chez *Pleurobrachia pileus*. En revanche, ces deux lignes se touchent ponctuellement dans une région éloignée du statolithe, tandis qu’au même endroit chez *Pleurobrachia pileus*, ces deux lignes ont tendance à présenter une plus forte densité de noyaux et à légèrement s’écarter l’une de l’autre (Figure 4.2 C et D). Il est nécessaire de noter ici que cette structure n’a pu être correctement observée que chez deux individus de *Beroe gracilis* seulement¹ et que la constance de la répartition de ces lignes de noyaux devra être confirmée dans de futures études.

1. Le faible nombre d’observations de ce caractère est dû à la difficile préparation de la région aborale disséquée chez *Beroe gracilis*, où l’organe aboral est très souvent profondément enfoncé dans l’organisme.

Fibres inter-peignes La répartition des fibres inter-peignes est très différente chez *Beroe gracilis* par rapport aux deux espèces de *Pleurobrachia*. Chez ces dernières, de chaque peigne partent des fibres centrales et latérales en direction du pôle oral et seules les fibres centrales atteignent le peigne suivant (voir article 3, Figure 8 et page suivante, Figure 4.3 F). Chez *Beroe gracilis*, de chaque peigne partent : 1) des fibres insérées au pôle oral qui ne rejoignent jamais le peigne suivant (que j'appellerai donc "fibres orales", voir Figure 4.3 A et B, points bleus); 2) des fibres très courtes insérées au pôle aboral, qui ne rejoignent jamais le peigne précédent (que j'appellerai donc "fibres aborales", voir Figure 4.3 A et B, points oranges). De plus, on trouve chez cette espèce un troisième type de fibres, qui forment une ligne centrale, ininterrompue et très fine, localisée sous la rangées de peigne (et que j'appellerai "fibres centrales", voir Figure 4.3 A et C, points magenta). Il est possible que les fibres centrales soient progressivement formées à partir des fibres aborales, pendant la croissance de la rangée de peignes (Figure 4.3 D et E, têtes de flèches) mais plus d'observations seront nécessaires pour éventuellement le démontrer.

Complexe sensoriel aboral et réseau nerveux épithélial En plus des observations rapportées ci-dessus (figure 4.2 et 4.3), d'autres structures semblent être des sources potentiels de caractères morphologiques. Mes tentatives de détection des fibres musculaires sous les champs polaires de *Beroe gracilis* ont échoué, il est donc possible que le "muscle" des champs polaires n'existe pas chez cette espèce. De manière générale, la morphologie des champs polaires de *Beroe gracilis* est très différente de ce qu'elle est chez *Pleurobrachia*, notamment car les champs polaires sont arrondis, plus courts et bordés par des papilles aborales (figure 4.4 A et B). Les réseaux nerveux épithéliaux chez ces deux espèces semblent relativement semblables (figure 4.4 C et D), mais je n'ai en revanche jamais observé de cordon nerveux juxta-tentaculaire chez *Beroe gracilis*. Cette structure semble ainsi absente, ce qui pourrait probablement être liée à la perte de l'ensemble de l'appareil tentaculaire chez les *Beroe*.

Anticipation d'un renouveau de la systématique des cténophores Les différences observées entre *Beroe gracilis* et *Pleurobrachia pileus* correspondent exactement à la définition de ce que sont des caractères en systématique car ce sont des structures clairement comparables (c'est-à-dire homologues) mais néanmoins différentes. C'est le cas des structures micro-anatomiques contenues dans l'organe aboral (comme les noyaux des corps réfringents) et des structures situées au niveau des rangées de peignes, où les types de fibres et leur distribution sont très variables. Nous anticipons que l'étude de la diversité de ces fibres inter-peignes à l'échelle des cténophores sera très informative d'un point de vue taxinomique.

Enfin, les marquages immunohistochimiques permettent également l'observation de la répartition et de l'aspect des dents macrociliaires présents dans la bouche des *Beroe* (Figure 4.5 A et B). La description de ces structures et la démonstration de leur utilité pour la classification des *Beroe* ont été fournies par Tamm and Tamm (1993) (les marquages immunohistochimiques ne sont pas

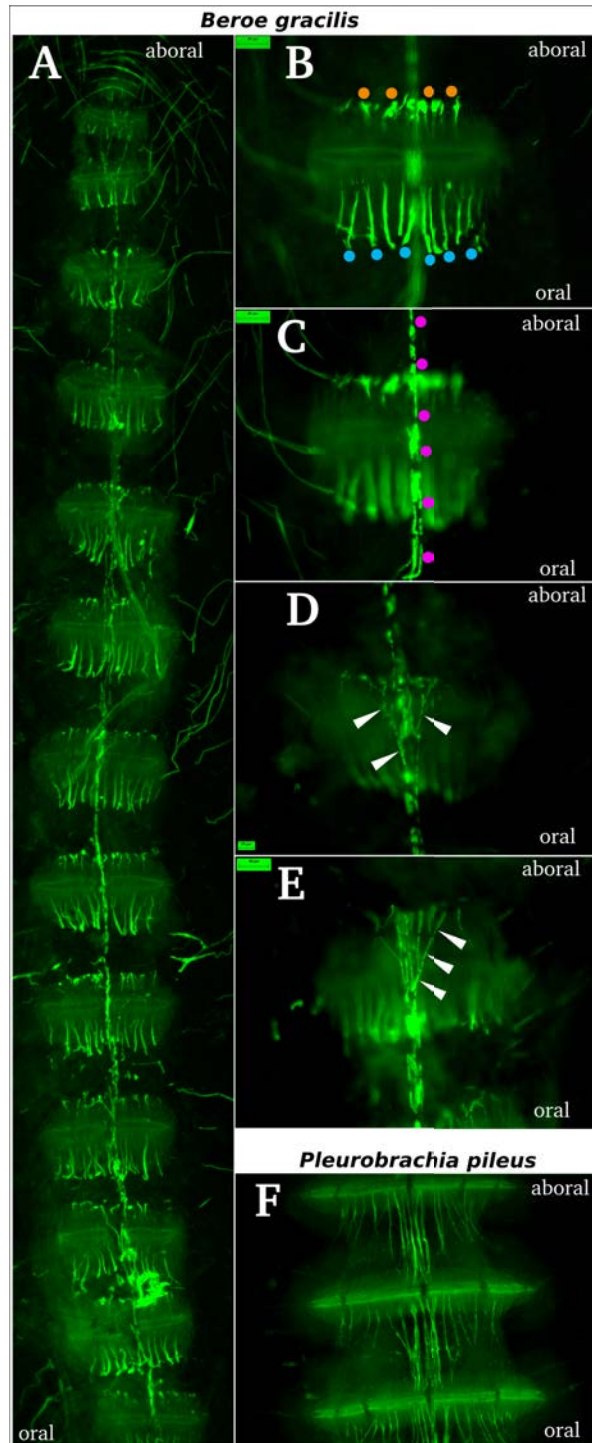


FIGURE 4.3 – Description des fibres inter-peignes chez *Beroë gracilis*, marquées en phalloïdine. A) vue générale d’une rangée de peignes. B et C) Détails de l’arrangement des fibres d’un peigne, avec une mise au point au niveau du coussinet basal (B) et à un niveau légèrement plus profond sous le peigne (C). D et E) Détails des prolongements des fibres aborales de certains peignes (têtes de flèches blanches) qui semblent rejoindre la ligne de fibres centrales. F) Organisation des fibres inter-peignes chez *Pleurobrachia pileus* pour rappel (voir aussi article 3, Figure 8 A et C). Les fibres orales, aborales et centrales sont localisées sur les images B et C par, respectivement, des pastilles bleues, oranges et magenta.

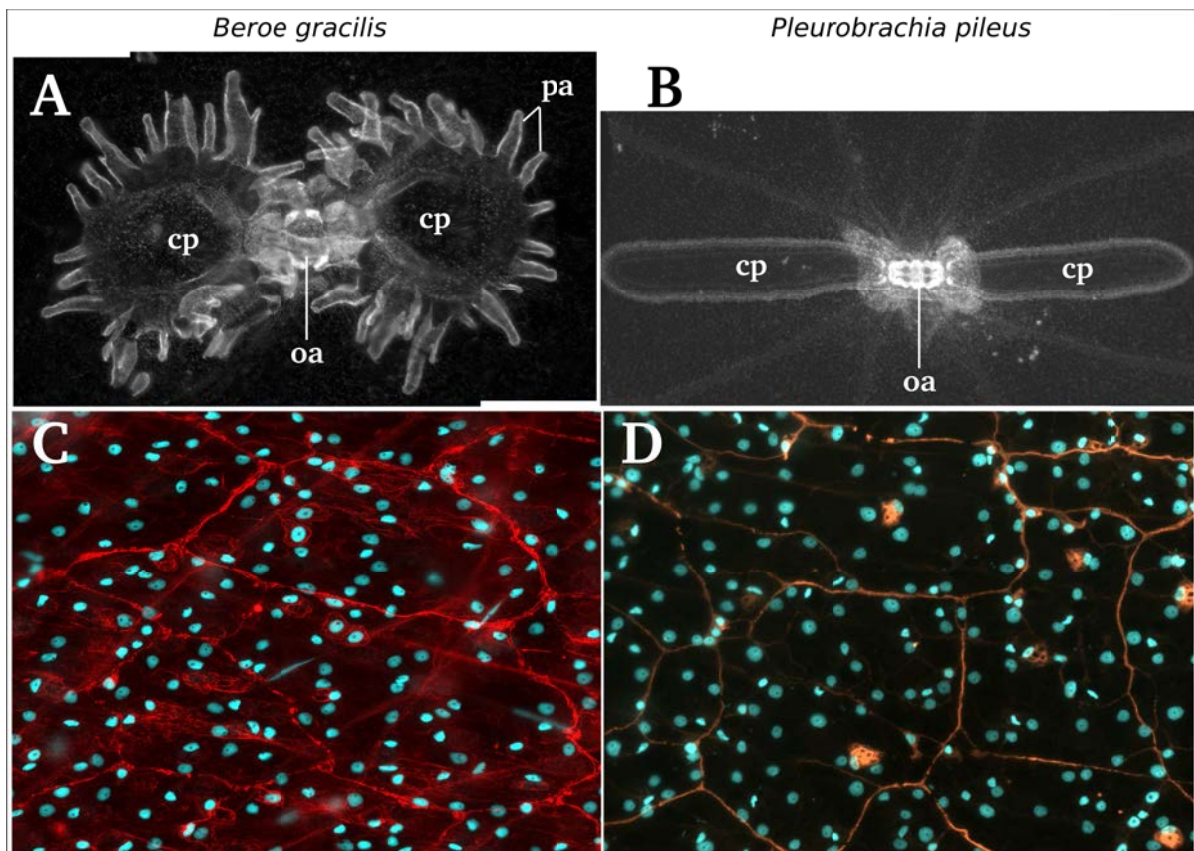


FIGURE 4.4 – Comparatif de l’aspect du complexe sensoriel aboral vu en marquage DAPI ainsi que du réseau nerveux épithélial vu par marquage de l’ α -tubuline tyrosylée et en DAPI, chez *Beroë gracilis* et *Pleurobrachia pileus*. A et B) Complexe sensoriel aboral vu en marquage DAPI chez *Beroë gracilis* (A) et *Pleurobrachia pileus* (B). C et D) Réseau nerveux épithélial vu en α -tubuline tyrosylée (rouge-orange) et DAPI (bleu) chez *Beroë gracilis* (C) et *Pleurobrachia pileus* (D). Légendes : oa : organe aboral ; cp : champs polaire ; pa : papilles aborales

nécessaires pour leur observation).

La comparaison de structures micro-anatomiques observées grâce à des marquages immunohisto-chimiques entre les espèces *Pleurobrachia pileus*, *Pleurobrachia brunnea* et *Beroë gracilis* (article 3 et observations préliminaires rapportées dans cette section) montre qu’elles représentent une source d’information importante qui devrait apporter beaucoup de nouveaux caractères informatifs aussi bien à l’échelle inter-spécifique qu’à l’échelle inter-ordinale.

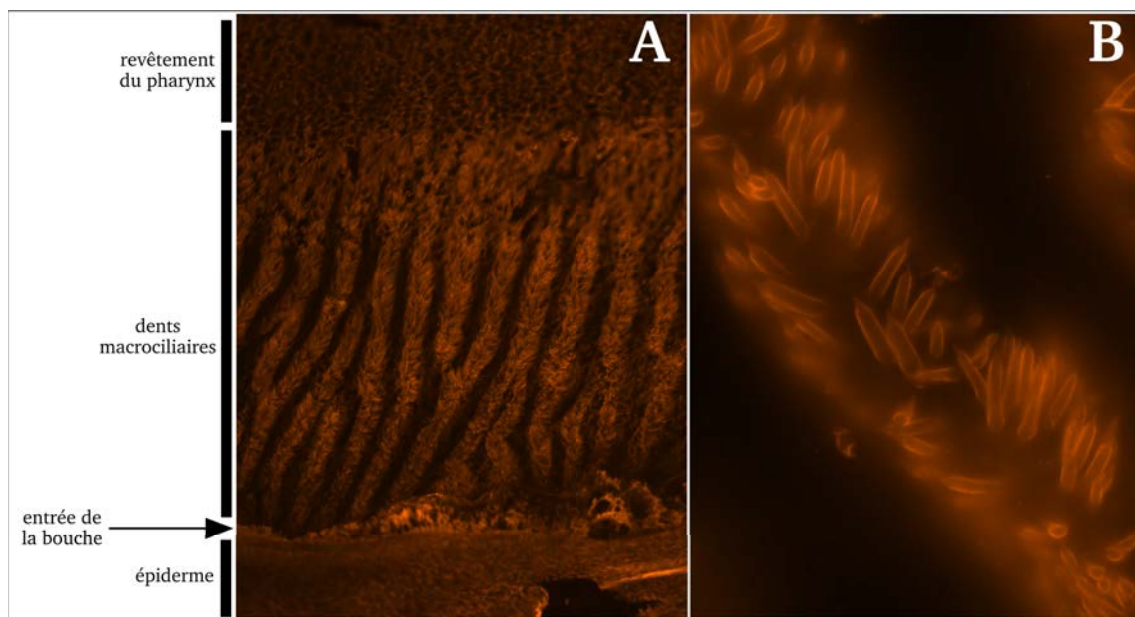


FIGURE 4.5 – Observation de la répartition et de la structure des dents macrociliaires de *Beroe gracilis*, vues en marquage anti α -tubuline tyrosylée. A) Répartition des dents macrociliaires au niveau de l'ouverture de la bouche. B) Détails de la morphologie des dents macrociliaires.

Chapitre 5

Conclusions générales et perspectives

5.1 Recherche de nouveaux caractères morpho-anatomiques utiles pour la taxonomie des cténophores (chapitre 4)

5.1.1 Synthèse des résultats

La taxonomie traditionnelle des cténophores, et particulièrement la définition des espèces au sein des différents genres, repose sur un faible nombre de caractères morpho-anatomiques, dont l'appréciation s'avère souvent délicate. Ceci explique en partie pourquoi l'identité et la validité de nombreuses espèces nominales décrites par le passé reste aujourd'hui très incertaine. L'insuffisance du nombre de caractères anatomiques disponibles chez les cténophores se fait aussi ressentir à des niveaux taxonomiques plus élevés (cf. Chapitre 3). M'appuyant sur l'expertise existante au sein de l'équipe concernant l'exploration détaillée de la "micro-anatomie" du cténophore *Pleurobrachia pileus* au moyen de la microscopie à épi-fluorescence, je me suis demandé si ces techniques (marquages fluorescents des noyaux cellulaires, marquage de certaines protéines structurales conservées au moyen d'anticorps) pourraient fournir de nouveaux caractères morpho-anatomiques potentiellement utiles pour la taxonomie des cténaires à différentes échelles. Les résultats présentés au chapitre 4 sont très encourageants de ce point de vue. En effet, la comparaison en immunofluorescence des structures épidermiques entre deux espèces proches du genre *Pleurobrachia* (*P. pileus* et *P. brunnea*) a permis d'identifier un plus grand nombre de caractères distinguant ces espèces que l'approche morpho-anatomique classique (Article 3). Des observations préliminaires focalisées sur les mêmes structures chez une espèce plus éloignée (un membre de l'ordre des Beroida, *Beroë gracilis*) suggèrent par ailleurs que cette approche peut aussi potentiellement apporter des caractères informatifs à plus grande échelle taxonomique.

5.1.2 Perspectives

Ces résultats très positifs invitent à la poursuite de l'observation de marquages immunohisto-chimiques chez les cténophores. A court terme, il sera possible d'exploiter cette nouvelle source d'information de trois manières différentes. D'abord, l'état des caractères "micro-anatomiques" décrits au chapitre 4 devrait être renseigné pour d'autres espèces de cténophores. Ensuite, seulement une partie de l'anatomie de *Pleurobrachia* a été utilisée (l'organe aboral, les rangées de peignes et le réseau nerveux épithélial) et il est fort probable que de nombreux autres caractères restent à découvrir dans des régions comme la racine tentaculaire, la mésoglée, la bouche et le pharynx. A moyen terme, l'utilisation d'anticorps différents (contre la tubuline acétylée, par exemple) permettrait d'augmenter encore le nombre de caractères diagnostiques observables. Ces pistes concernent la poursuite de nos efforts pour la mise en place d'un contexte taxonomique stable à partir de caractères morphologiques, mais l'utilisation de données moléculaires serait également souhaitable, et permettrait d'établir, progressivement bien sûr, les bases d'une taxonomie intégrative solide pour les cténophores. Cependant, les taux d'évolution entre les groupes de cténophores semblent très hétérogènes, et pourraient rendre difficile la détermination d'un marqueur moléculaire unique (Barcode) pour ces organismes. Enfin, l'étude de la morphologie des cténophores permettra, dans un contexte phylogénétique résolu, de reconstituer l'évolution de leurs structures anatomiques.

5.2 La reconstruction des relations phylogénétiques au sein des cténophores (chapitre 3)

5.2.1 Synthèse des résultats

Avant le début de cette thèse, les relations phylogénétiques à l'échelle des cténophores restaient extrêmement mal comprises. Le faible nombre de caractères anatomiques exploitables avait conduit les morphologistes à formuler des schémas évolutifs radicalement contradictoires, et une seule étude moléculaire, basée sur l'ADNr 18S (Podar et al. 2001) avait été publiée. Celle-ci établissait la non-monophylie de l'ordre des "Cydippida" et confirmait quelques relations de parenté suspectées par les morphologistes (comme l'existence d'un clade Haeckeliidae + Beroida, ou l'affinité des Cestida avec les Lobata), mais pour le reste, l'essentiel de la phylogénie des cténophores restait irrésolu. Dans ce travail, je me suis tout d'abord attaché à compléter le jeu de données de cette précédente étude, par l'intégration de taxons supplémentaires, et par l'ajout des séquences plus variables des deux ITS (*internal transcribed spacers*). Les analyses de ce nouveau jeu de données ont conduit à l'article 2 (sous presse dans *Zoology*). La paraphylie des Cydippida y est confirmée, et différents secteurs de l'arbre des cténophores sont précisés (p. ex. position des Mertensiidae s.l., des Platyctenida et des Lampeidae à l'extérieur d'un clade groupant tous les autres cténaïres ; paraphylie probable du genre *Beroe* ; aspects des relations entre genres et familles de Lobata). Cet

article laisse toutefois d'importants points d'ombres, l'un des plus critique étant la position de la racine de l'arbre des cténophores (= quelle lignée est groupe-frère de toutes les autres?). Une partie de ces problèmes a été éclaircie grâce aux analyses complémentaires présentées à la fin du chapitre 3 (section 3.3) : analyses de gènes dupliqués, et exploitation des arbres phylogénomiques (cf. chapitre 2 pour la construction du jeu de données phylogénomique). Ainsi, il apparaît désormais bien établi que les Mertensiidae (*sensu lato*, soit incluant le genre *Euplokamis*) sont groupe-frères des autres cténophores. Les relations entre les groupes principaux de cténaires sont désormais beaucoup mieux résolues (voir l'arbre bilan, chapitre 3, Figure 3.7 page 124,), même s'il demeure encore quelques irrésolutions, et que différentes familles ne sont pas encore représentées dans les arbres moléculaires.

5.2.2 Perspectives

Les relations phylogénétiques au sein du groupe unissant les Lobata, Cestida et Thalassocalycida ne sont pas résolues, et pourraient l'être à relativement court terme. A l'exception de *Thalassocalyce inconstans*, les espèces principales des Lobata et des Cestida sont fréquemment rencontrées, et le séquençage des ITS 1 et 2 (et éventuellement d'autres marqueurs variables) pour la plupart de ces espèces devrait facilement résoudre les relations de parenté, nous permettant du même coup de comprendre l'évolution d'un certain nombre de traits morphologiques, comme la présence ou l'absence de canal perradiaire ou le mode de connexion des canaux adradiaires aux canaux méridiens. Ce travail de séquençage pourra également s'inscrire dans un contexte de taxonomie moléculaire (voir les perspectives 5.1.2). En revanche, aller plus loin dans nos connaissances de l'évolution des cténophores à l'échelle des ordres et des familles ne sera faisable qu'à moyen ou long terme. En effet, le facteur limitant est désormais l'échantillonnage taxonomique, et la récolte d'espèces appartenant aux 13 familles encore manquantes risque d'être relativement longue. Si certains taxons pourront être récoltés assez rapidement, certains des groupes manquants n'ont pas été observés depuis des dizaines d'années, y compris les trois ordres des Cambojiida, Ganeshida et Cryptolobiferida. De ce point de vue, la priorité pour la communauté scientifique devra être l'organisation concertée de récoltes adaptées au macroplancton fragile.

5.3 Recherche de la position des cténophores dans l'arbre des métazoaires (chapitre 2)

5.3.1 Synthèse des résultats

La position des cténophores (et plus largement, la topologie de la base de l'arbre des métazoaires) reste très débattue. Notamment, plusieurs études phylogénomiques récentes ont soutenu un résultat fort surprenant, la position des cténophores comme groupe-frère de l'ensemble des autres

métazoaires. Afin d’aborder ce problème, j’ai travaillé à l’élaboration d’un jeu de données phylogénomique original de très grande ampleur, en collaboration avec principalement Hervé Philippe et Denis Baurain. Des transcriptomes de 22 espèces de spongiaires, cténophores et cnidaires ont été séquencés afin de représenter certains taxons clés qui manquaient précédemment dans les supermatrices analysées en phylogénomique. Un protocole bioinformatique original a été mis au point, impliquant l’élaboration de nombreux scripts inédits, afin de construire de manière semi-automatisée un ensemble d’alignements de plusieurs milliers de gènes, tout en traitant de la manière la plus rigoureuse possible les principales sources d’erreurs susceptibles d’affecter la reconstruction de la phylogénie (contaminations, paralogies, données manquantes) (jeu de données *AUTO*). En parallèle, nous avons actualisé l’échantillonnage taxonomique d’un jeu de données existant de 128 gènes (jeu de données *MANU*), ayant précédemment conduit à établir la monophylie des spongiaires et à suggérer une relation de groupe-frères entre cténophores et cnidaires (Philippe et al. 2009). Les deux jeux de données ont fait l’objet de différentes analyses explorant l’impact de l’échantillonnage taxonomique, du pourcentage de données manquantes, des modèles d’évolution, du biais de composition en acides aminés, du retrait de sites hétéropécilles ou de sites à taux d’évolution rapide. L’examen critique de l’ensemble des résultats conduit à rejeter l’hypothèse de la position “basale” des cténophores (comme étant due à un artéfact d’attraction de longues branches), le groupe-frère de l’ensemble des autres métazoaires étant plutôt les spongiaires (dont la monophylie est confirmée), en accord avec la vision traditionnelle. La position des cténophores reste toutefois incertaine ; en effet la plupart des analyses du jeu de données *AUTO* les positionnent comme groupe-frère d’un ensemble (placozoaires + cnidaires + bilatériens), alors que le jeu de données *MANU* contient un signal en faveur de la monophylie des coelentérés (cténophores + cnidaires). Une troisième hypothèse est soutenue par le jeu de données *AUTO* après retrait de beaucoup de sites à évolution rapide, ou retrait des taxons présentant le plus fort biais de composition en acides aminés : les cténophores groupe-frère des Bilateria. L’extrême longueur de la branche des cténophores rend donc difficile leur positionnement, même avec des supermatrices de très grande ampleur dont la construction a mobilisé tous les moyens disponibles pour minimiser les sources d’erreur potentielles.

5.3.2 Perspectives immédiates

L’observation de signaux phylogénétiques contradictoires entre, et au sein de nos deux jeux de données *AUTO* et *MANU* soulève un problème immédiat : parmi ces signaux, un seul (voir aucun !) correspond à l’histoire évolutive des métazoaires, les autres sont artéfactuels. Nos deux jeux de données doivent donc désormais être explorés très en profondeur, et il existe de nombreuses possibilités d’analyses pour ce faire. Des tests de congruence entre gènes pourraient permettre d’éliminer les moins informatifs d’entre eux, tandis que l’utilisation du modèle d’évolution CAT-BP pourrait permettre de réduire l’impact de l’hétérogénéité de composition en acides aminés

(au prix néanmoins d'un très long temps de calcul) sans réduire l'échantillonnage taxonomique de l'analyse. Le protocole bioinformatique réalisé au cours du chapitre 2 pourrait lui aussi être relativement rapidement amélioré. Par exemple, une version modifiée du logiciel PhyloTreePruner pourrait être créée, afin d'autoriser au cours de la procédure le ré-enracinement de l'arbre qui permettrait d'en extraire réellement la plus grande partition. L'application d'une telle procédure à l'ensemble des 4412 clusters originaux permettrait la réalisation d'une concaténation significativement plus longue et présentant moins de données manquantes. De plus, un certain nombre de nouveaux transcriptomes d'éponges ont été publiés récemment (Riesgo et al. 2014), et leur décontamination permettrait de les ajouter à nos jeux de données, améliorant ainsi notre échantillonnage taxonomique. Enfin, une conséquence à court terme de mon travail de phylogénomique sera la mise à disposition publique, lors de la publication de ces travaux, du script "DCC", conçu par Hervé Philippe, Michaël Manuel et moi-même, que j'ai écrit et validé sur nos propres données. Cet outil permettra à l'ensemble de la communauté scientifique travaillant en phylogénomique de supprimer les cross-contaminations des transcriptomes nouvellement séquencés (voir article 1 et explications ci-après).

5.3.3 Autres réflexions, implications et perspectives suscitées par ce travail

L'importance des contaminations dans les données de séquençage massif

L'immense majorité des données de séquençage de nouvelle génération (NGS) sont contaminées à des niveaux élevés, rendant leur utilisation dangereuse pour la reconstruction de relations de parenté. Ces contaminations sont issues de sources variables (comme la présence de symbiotes ou d'épibiontes sur l'échantillon récolté) dont la plus systématique est la cross-contamination. Il s'agit de la contamination réciproque d'échantillons qui sont regroupés (*pooling*, en anglais) pour être séquencés en même temps (procédure indispensable pour limiter le coût du séquençage). Le développement d'une méthode de De-Cross-Contamination (DCC) m'a permis de quantifier l'importance de ce phénomène. Ce dernier touche, sans exception, tous les transcriptomes étudiés ici. Son ampleur est très variable et semble difficilement prédictible. Par exemple, parmi les données séquencées par Moroz et al. (2014), le transcriptome de *Euplokamis dunlapae* comprend 0.87% de transcrits issus de cross-contaminations (soit 1474 séquences), tandis que le transcriptome de l'espèce indéterminée "sp3" en comprend 70.64% (soit 121398)! Un taux aussi extrême pourrait s'expliquer par une dégradation poussée des ARNs utilisés pour le séquençage, augmentant du même coup l'abondance relative des ARNs cross-contaminants. Ces observations ont deux implications majeures :

1. La plupart des jeux de données moléculaire NGS sont probablement touchés par ce phénomène.
2. L'utilisation d'un protocole similaire à celui conçu au chapitre 2 de cette thèse permet de dé-

tecter et d'éliminer ces contaminations. Cependant, l'utilisation d'un tel protocole requiert la connaissance du nombre de lectures participant à l'assemblage de chaque contigs, information souvent absente des bases de données publiques.

Si l'utilisation de ces jeux de données contaminés ne devrait pas avoir un impact majeur sur les résultats d'études quantitatives, comme les analyses d'expression différentielle (la présence et la nature des lectures contaminantes est probablement aléatoire et ne devrait pas fausser significativement la quantification de l'expression relative des transcrits à l'échelle d'une expérience avec réplicats), son impact peut être dramatique sur les résultats d'analyses "qualitatives", comme la phylogénomique ou la recherche de présence/absence de gènes dans un transcriptome. Pour s'en convaincre, il suffira de regarder la Figure 5.1 page suivante, résultat d'une analyse phylogénomique préliminaire (basée sur 800 gènes) que j'ai réalisé sur des données transcriptomiques sans De-Cross-Contamination. Dans cet arbre, l'ensemble des cténophores est artificiellement placé au sein des Octocorallia, simplement parce que les transcriptomes des espèces *Alcyonium*, *Coeloplana* et *Vallricula* ont été séquencés en Illumina dans la même expérience. Pour les mêmes raisons, on peut également constater dans cet arbre que *Liriope*, une trachyméduse, est positionnée au sein des éponges (Figure 5.1 page ci-contre ; *Pleraplysilla* faisait partie du même lot d'espèces séquencées). Un tel impact impose de prendre d'importantes précautions lors de l'utilisation de données moléculaires obtenues par des technologies de séquençage de seconde génération.¹

Dissection des signaux contradictoires des supermatrices et analyse des positions

Les analyses phylogénomiques de cette thèse sont en accord avec de nombreux autres travaux récents qui montrent la complexité et la multiplicité des signaux contradictoires contenus dans les supermatrices phylogénomiques (Jeffroy et al. 2006; Leigh et al. 2008; Nesnidal et al. 2010; Salichos and Rokas 2013; Nesnidal et al. 2013; Telford et al. 2014; Struck et al. 2014; Straub et al. 2014; Sharma et al. 2014; Meiklejohn et al. 2014). L'approche dite de "total evidence" présuppose que l'additivité du signal phylogénétique l'emportera sur le bruit de la matrice à mesure que les dimensions de la matrice augmenteront. En réalité, l'augmentation de la quantité de données moléculaires va de pair avec l'augmentation de leur hétérogénéité dans la matrice. Cette dernière est complexe à représenter pour les modèles d'évolutions, qui peinent alors à détecter le signal phylogénétique. De plus, l'augmentation de la quantité des données renforcent également les effets des biais systématiques (comme l'hétérogénéité des compositions en acide-aminés ou des taux de substitutions), car ils sont eux aussi additifs. Ces biais systématiques induisent de forts signaux non-phylogénétiques qui peuvent amener à la reconstruction d'une topologie fautive, mais robuste.

1. Il est probable que les prestataires de service de séquençage utilisent aussi le "pooling" de projets de laboratoires différents pour maximiser leurs capacités de séquençage. Une étape importante vers la dé-cross-contamination systématique des données NGS serait donc la divulgation par ces prestataires, le cas échéant, des regroupements de projets différents au sein d'un même "run".

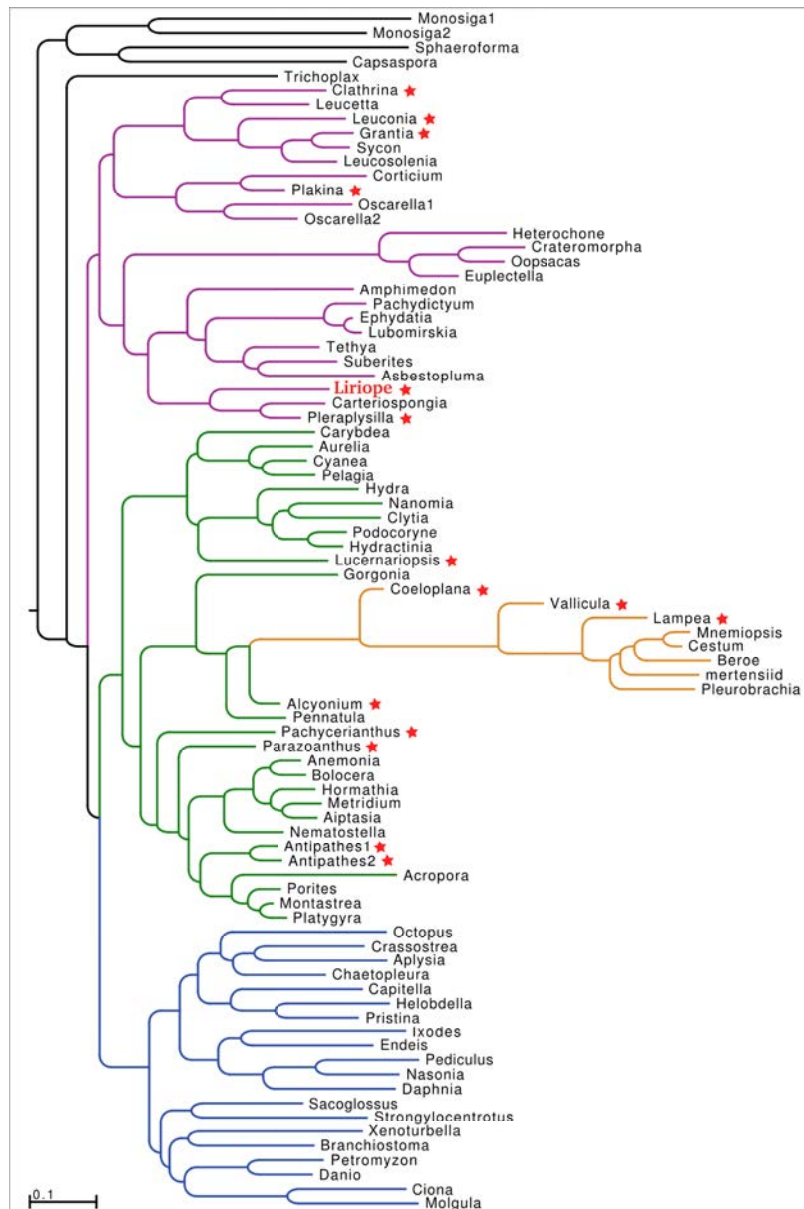


FIGURE 5.1 – Illustration du problème de l'utilisation de transcriptomes présentant des cross-contaminations pour reconstituer des relations phylogénétiques. Une supermatrice basée sur 800 gènes (10,547 positions) a été analysée avec le modèle CAT+ Γ sous Phylobayes. Remarquez l'étroite relation de parenté, clairement aberrante, entre les cténophores et *Alcyonium*, ainsi que la position de *Liriope*, une trachyméduse, proche de *Pleraplysilla* au sein des démosponges. *Coeloplana*, *Vallicula*, *Alcyonium*, *Liriope* et *Pleraplysilla* et les autres taxons marqués d'une étoile rouge avaient été regroupées pour être séquencées sur la même ligne en Illumina (étoiles rouges).

La détection de ces signaux contradictoires présents dans les supermatrices moléculaires est complexe et impose la réalisation de nombreuses analyses. Pour réduire l'impact des biais systématiques, il est par exemple possible de supprimer des taxons ou des gènes qui présenteraient des compositions en acides aminés trop différentes des autres, ou des taux d'évolution trop élevés. Il est également possible d'étudier directement les positions des supermatrices, et d'en évaluer diverses caractéristiques, dont le taux d'évolution. De nombreuses études ont déjà réalisé la suppression de catégories de positions, comme, dans le cas de séquences nucléotidiques codantes, les troisièmes positions de codons qui saturent plus rapidement que les autres à cause de la redondance du code génétique (Xia et al. 2003; Struck et al. 2007), ou encore les positions qui ont été mesurées empiriquement comme évoluant plus rapidement (chapitre 2 de cette thèse; Brinkmann and Philippe 1999; Pisani 2004; Pisani et al. 2012; Telford et al. 2014).

Une telle approche analytique menée au niveau des positions des supermatrices est cohérente avec des observations récentes qui montrent que le signal phylogénétique y est réparti de manière très hétérogène (certaines positions pouvant être très informatives et d'autres non), mais qu'il n'est pas réparti différentiellement selon les gènes (Evans et al. 2010; Smith et al. 2011; Struck 2013). Cela implique que la suppression de marqueurs entiers n'est peut être pas la solution optimale pour améliorer la qualité des supermatrices. Différentes méthodes existent pour estimer le taux d'évolution d'une position, comme celle implémentée dans TIGER (Cummins and McInerney 2011), qui calcule un score à chaque position, reflétant ainsi sa compatibilité avec les autres sites de la matrice, ou encore la méthode "slow-fast" (Brinkmann and Philippe 1999; Kostka et al. 2008), qui somme le nombre de changements d'états de chaque caractère au sein de groupes taxonomiques définis *a priori*. Evans et al. (2010) ont eu l'idée de catégoriser des positions selon leur vraisemblance pour l'une ou l'autre de deux topologies contradictoires (il s'agit donc d'une catégorisation *a posteriori*). Ils ont ensuite supprimé certaines positions qui favorisaient "trop" une topologie plutôt qu'une autre. Cette approche me suggère qu'il serait possible d'aller plus loin : au lieu de supprimer les positions "trop" favorables à l'une des deux topologie contradictoires, il serait intéressant de les rassembler en deux sous-matrices. L'analyse comparée (comme la mesure de degré de saturation ou de la composition en acides aminés) de ces deux sous-matrices pourrait permettre de mieux comprendre les sources du conflit topologique initial en apportant des arguments objectifs sur la qualité des données qui en sont responsables.

Une mise à jour souhaitable des datations moléculaires de l'origine des métazoaires

Les jeux de données phylogénomiques construits dans cette thèse, ainsi que les résultats de leur analyse, semblent représenter deux apports importants dans la perspective de la datation moléculaire de l'origine et de la diversification des métazoaires. En effet, les dernières datations publiées (Sperling et al. 2007; Peterson et al. 2008; Erwin et al. 2011) sont basées sur des données que l'on peut aujourd'hui regarder comme incomplètes :

1. La matrice moléculaire utilisée pour reconstituer les relations phylogénétiques à l'échelle des métazoaires et pour dater l'apparition des différents clades n'est composée que de 7 gènes.
2. Les cténophores sont absents de cette matrice.
3. Les cnidaires n'y sont représentés que par 7 espèces seulement.
4. Un seul point de calibration est utilisé en dehors des bilatériens.

De nombreuses études ont produit des données moléculaires dont l'utilisation pourrait grandement améliorer la qualité de ces datations (Philippe et al. 2009; Pick et al. 2010; Philippe et al. 2011; Nosenko et al. 2013; Ryan et al. 2013) et plus compréhensif encore est le jeu de données AUTO élaboré dans l'article 1 de cette thèse. Le coeur du processus de datation moléculaire repose sur la supermatrice analysée, et il est bien possible que l'ajout de plusieurs centaines de gènes, ainsi qu'une meilleure représentation taxonomique des non-bilatériens puisse fortement influencer sur les résultats de ces analyses.

D'autres problèmes affectent également la qualité de ces datations moléculaires. D'abord, il reste des incertitudes sur les noeuds profonds de l'arbre des métazoaires, alors que la topologie utilisée est un paramètre capital des datations moléculaires. Ensuite, la répartition et le nombre des points de calibrations utilisés lors de la datation sont cruciaux (Cartwright and Collins 2007; Duchêne et al. 2014) mais sont basés sur des interprétations de données paléontologiques qui peuvent être remises en cause. C'est le cas du point de calibration d'âge maximal des démosponges à 713 Ma, qui représente justement dans les analyses les plus récentes (Sperling et al. 2007; Peterson et al. 2008; Erwin et al. 2011) le seul point de calibration chez les non-bilatériens. Dans une revue publiée cette année, Antcliffé et al. (2014) redéterminent l'appartenance (ou non) de fossiles connus à l'embranchement des éponges. Les auteurs y rejettent vigoureusement l'utilisation de l'existence d'un biomarqueur (le 24-isopropyl cholestane) comme preuve de l'existence d'éponges à cette période. L'utilisation d'un seul point de calibration d'âge maximum (710 Ma pour l'origine des démosponges) conduit à une date d'apparition des métazoaires il y a 12,4 milliards d'années, et l'ajout d'un autre point de calibration d'âge maximum (560 Ma pour l'origine des bilatériens) ramène cette date à 1,6 milliards d'années. L'impact anticipé de la suppression du point de calibration des démosponges, associé à la faible taille de la matrice moléculaire utilisée précédemment, laisse penser que les datations moléculaires des origines des grandes lignées animales mériteraient d'être entièrement revues, notamment à l'aide de supermatrices moléculaires plus complètes.

Bibliographie

- Agassiz, A. (1865). *North American Acalephae*. Number XIV. Cambridge University press : Sever and Francis.
- Agassiz, L. (1860). *Contributions to the natural history of the United States of America*, volume VOL III. Boston : Little, Brown and Company.
- Alié, A., Leclère, L., Jager, M., Dayraud, C., Chang, P., Le Guyader, H., Quéinnec, E., and Manuel, M. (2011). Somatic stem cells express piwi and vasa genes in an adult ctenophore : Ancient association of “germline genes” with stemness. *Developmental Biology*, 350(1) :183–197.
- Antcliffe, J. B., Callow, R. H. T., and Brasier, M. D. (2014). Giving the early fossil record of sponges a squeeze. *Biological Reviews*, pages n/a–n/a.
- Bennema, F. and Van Moorsel, G. (2011). The status of *Callianira hexagona* (slabber, 1778) (ctenophora). *Zool. Med. Leiden*, 85.
- Besemer, J. and Borodovsky, M. (2005). Genemark : web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res*, 33(Web Server issue) :W451–W454.
- Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS : a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, 29(12) :2607–2618.
- Blanquart, S. and Lartillot, N. (2006). A bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Molecular Biology and Evolution*, 23(11) :2058–2071.
- Blanquart, S. and Lartillot, N. (2008). A site- and time-heterogeneous model of amino acid replacement. *Molecular Biology and Evolution*, 25(5) :842–858.
- Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J., and Le Parco, Y. (2001). Sponge paraphyly and the origin of Metazoa. *Journal of Evolutionary Biology*, 14(1) :171–179.
- Brinkmann, H. and Philippe, H. (1999). Archaea sister group of bacteria? indications from tree reconstruction artifacts in ancient phylogenies. *Molecular Biology and Evolution*, 16(6) :817–825.

- Brusca, R. and Brusca, G. (2003). *Invertebrates. 2nd Edition*. Sinauer, Sunderland, Massachusetts.
- Carré, D., Carré, C., and Mills, C. E. (1989). Novel cnidocysts of narcomedusae and a medusivorous ctenophore, and confirmation of kleptocnidism. *Tissue and Cell*, 21(5) :723–734.
- Cartwright, P. and Collins, A. (2007). Fossils and phylogenies : integrating multiple lines of evidence to investigate the origin of early major metazoan lineages. *Integrative and Comparative Biology*, 47(5) :744–751.
- Cartwright, P., Halgedahl, S. L., Hendricks, J. R., Jarrard, R. D., Marques, A. C., Collins, A. G., and Lieberman, B. S. (2007). Exceptionally preserved jellyfishes from the middle cambrian. *PLoS ONE*, 2(10) :e1121.
- Chen, J. and Zhou, G. (1997). Biology of the Chengjiang fauna. *Bull. Natl Mus. Nat. Sci. Taiwan*, 10 :11–105.
- Chen, J.-Y., Schopf, J. W., Bottjer, D. J., Zhang, C.-Y., Kudryavtsev, A. B., Tripathi, A. B., Wang, X.-Q., Yang, Y.-H., Gao, X., and Yang, Y. (2007). Raman spectra of a lower cambrian ctenophore embryo from southwestern shaanxi, china. *Proceedings of the National Academy of Sciences*, 104(15) :6289–6292.
- Chun, C. (1880). Die Ctenophoren des Golfes von Neapel und der angrenzenden Meeres-Abschnitte : eine Monographie. Fauna und Flora des Golfes von Neapel und der angrenzenden Meeres-Abschnitte. *Leipzig : Verlag von W. Engelmann*.
- Collins, D. (1996). The "evolution" of *Anomalocaris* and its classification in the arthropod class Dinocarida (nov) and order Radiodonta (nov). *Journal of Paleontology*, 70(2) :280–293.
- Conway Morris, S. and Collins, D. H. (1996). Middle cambrian ctenophores from the stephen formation, british columbia, canada. *Philosophical Transactions : Biological Sciences*, pages 279–308.
- Cummins, C. A. and McInerney, J. O. (2011). A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Systematic Biology*, 60(6) :833–844.
- Dawydoff, C. (1903). *Hydroctena salenskii* (etude morphologique sur un nouveau coelentère pélagique). *Zap. imp. Akad. Nauk. (Classe phys.-math.) (8eme sér.)*, 14 :1–16.
- Dawydoff, C. (1946). Contribution a la connaissance des ctenophores pelagiques des eaux de l'Indochine. *Bulletin Biologique France Belgique*, 80 :116–170.

- Dayraud, C. (2011). *Origine et évolution des muscles et du système neuro-musculaire : approches morpho-anatomique et moléculaire chez le cténaire Pleurobrachia pileus*. PhD thesis, Université Pierre et Marie Curie.
- Dayraud, C., Alié, A., Jager, M., Chang, P., Le Guyader, H., Manuel, M., and Quéinnec, E. (2012). Independent specialisation of myosin II paralogues in muscle vs. non-muscle functions during early animal evolution : a ctenophore perspective. *BMC Evolutionary Biology*, 12(1) :107.
- Delage, Y. and Hérouard, E. (1901). *Traité de Zoologie Concrète, 2ème partie*, chapter Les Coelentérés, pages 707–804. H. Le Soudier, Paris.
- Duchêne, S., Lanfear, R., and Ho, S. Y. W. (2014). The impact of calibration and clock-model choice on molecular estimates of divergence times. *Mol Phylogenet Evol*, 78 :277–289.
- Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sørensen, M. V., Haddock, S. H. D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M., Wheeler, W. C., Martindale, M. Q., and Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452(7188) :745–749.
- Dutheil, J. and Boussau, B. (2008). Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evolutionary Biology*, 8(1) :255.
- Dzik, J. (2002). Possible ctenophoran affinities of the precambrian "sea-pen" *Rangea*. *Journal of Morphology*, 252(3) :315–334.
- Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D., and Peterson, K. J. (2011). The cambrian conundrum : Early divergence and later ecological success in the early history of animals. *Science*, 334(6059) :1091–1097.
- Evans, N. M., Holder, M. T., Barbeitos, M. S., Okamura, B., and Cartwright, P. (2010). The phylogenetic position of myxozoa : Exploring conflicting signals in phylogenomic and ribosomal data sets. *Molecular Biology and Evolution*, 27(12) :2733–2746.
- Fabricius, O. (1780). *Fauna Groenlandica, systematice sistens, Animalia Groenlandiae occidentalis hactenus indagata, quoad nomen specificum, triviale, vernaculumque synonyma auctorum plurium, descriptionem, locum, victum, generationem, mores, usum, capturamque singuli prout detegendi occasio fuit, maximaque parte secundum proprias observationes*. Hafniae [Copenhagen] Ioannis Gottlob Rothe.
- Foster, P. G. and Hickey, D. A. (1999). Compositional Bias May Affect Both DNA-Based and Protein-Based Phylogenetic Reconstructions. *Journal of Molecular Evolution*, 48(3) :284–290.

- Foster, P. G., Jermini, L. S., and Hickey, D. A. (1997). Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. *Journal of Molecular Evolution*, 44(3) :282–288.
- Franc, J. (1988). The mesoglea of ctenophores. *Bulletins de la Société Zoologique de France*, 113 :347–351.
- Gerschwin, L., Zeidler, W., and Davie, P. J. F. (2010). Ctenophora of australia. *Memoirs of the Queensland Museum*, 54(3).
- Ghabooli, S., Shiganova, T. A., Briski, E., Piraino, S., Fuentes, V., Thibault-Botha, D., Angel, D. L., Cristescu, M. E., and Macisaac, H. J. (2013). Invasion pathway of the ctenophore *Mnemiopsis leidyi* in the mediterranean sea. *PLoS One*, 8(11) :e81067.
- Glennner, H., Hansen, A. J., Sørensen, M. V., Ronquist, F., Huelsenbeck, J. P., and Willerslev, E. (2004). Bayesian inference of the metazoan phylogeny : a combined molecular and morphological approach. *Current Biology*, 14(18) :1644–1649.
- Greve, W. (1970). Cultivation experiments on North Sea ctenophores. *Helgoländer wissenschaftliche Meeresuntersuchungen*, 20(1-4) :304–317.
- Gronovius, L. T. (1760). Observationes de animalculis aliquot marinae aquae innatantibus atque in littoribus Belgicis obviis. *Acta Helvetica, physico-mathematico-anatomico-botanico-medica*, 4 :35–40, Tab. IV.
- Groussin, M., Boussau, B., and Gouy, M. (2013). A branch-heterogeneous model of protein evolution for efficient inference of ancestral sequences. *Systematic Biology*, 62(4) :523–538.
- Haddock, S. H. D. (2007). Comparative feeding behavior of planktonic ctenophores. *Integrative and Comparative Biology*, 47(6) :847–853.
- Harbison, G., Madin, L., and Swanberg, N. (1978). On the natural history and distribution of oceanic ctenophores. *Deep Sea Research*, 25(3) :233 – 256.
- Harbison, G. R. (1985). On the classification and evolution of the Ctenophora. In *The Origins and Relationships of lower Invertebrates*, pages 78–100. Oxford University Press, Oxford.
- Harbison, G. R., Matsumoto, G. I., and Robison, B. H. (2001). *Lampocteis cruentiventer* gen. nov., sp. nov. : a new mesopelagic lobate ctenophore, representing the type of a new family (class tentaculata, order lobata, family lampoctenidae, fam. nov.). *Bulletin of marine science*, 68(2) :299–311.

- Hausdorf, B., Helmkamp, M., Meyer, A., Witek, A., Herlyn, H., Bruchhaus, I., Hankeln, T., Struck, T. H., and Lieb, B. (2007). Spiralian Phylogenomics Supports the Resurrection of Bryozoa Comprising Ectoprocta and Entoprocta. *Molecular Biology and Evolution*, 24(12) :2723–2729.
- Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., Martinez, P., Baguna, J., Bailly, X., Jondelius, U., Wiens, M., Muller, W. E. G., Seaver, E., Wheeler, W. C., Martindale, M. Q., Giribet, G., and Dunn, C. W. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings of the Royal Society B : Biological Sciences*, 276(1677) :4261–4270.
- Hennig, W. (1966). *Phylogenetic Systematics*. University of Illinois Press.
- Hernandez-Nicaise, M. (1968). Distribution and ultrastructure of symmetrical synapses in the ctenaires nervous system. *Comptes rendus Hebdomadaires des Séances de l'Académie des Sciences, série D : Sciences Naturelles*, 267(21) :1731–1734.
- Hernandez-Nicaise, M. (1973). The nervous system of ctenophores. iii. ultrastructure of synapses. *Journal of Neurocytology*, 2(3) :249–263.
- Hernandez-Nicaise, M. (1991). *Microscopic anatomy of Invertebrates, Volume 2 : Placozoa, Porifera, Cnidaria and Ctenophora*, chapter Ctenophora, pages 359–418. Wiley-Liss, Inc.
- Hernandez-Nicaise, M. and Franc, J. (1994). Embranchement des cténaires. In *Traité de Zoologie*, volume Tome III, pages 943–1075. Masson et Cie, Paris, p.p. grassé edition.
- Horridge, G. (1974). *Coelenterate Biology*, chapter Recent studies on the Ctenophora, pages 439–468. Academic Press, New York.
- Hyman, L. (1940). *The Invertebrates*, volume I. Protozoa through Ctenophora, chapter VIII The Radiate Phyla – Phylum Ctenophora, pages 662–695. McGraw Hill.
- Ivantsov, A. (2009). New reconstruction of kimberella, problematic vendian metazoan. *Paleontological Journal*, 43(6) :601–611.
- Jager, M., Chiori, R., Alié, A., Dayraud, C., Quéinnec, E., and Manuel, M. (2011). New insights on ctenophore neural anatomy : Immunofluorescence study in *Pleurobrachia pileus* (müller, 1776). *Journal of Experimental Zoology Part B : Molecular and Developmental Evolution*, 316B(3) :171–187.
- Jager, M., Dayraud, C., Mialot, A., Quéinnec, E., le Guyader, H., and Manuel, M. (2013). Evidence for involvement of wnt signalling in body polarities, cell proliferation, and the neuro-sensory system in an adult ctenophore. *PLoS One*, 8(12) :e84363.

- Jager, M., Quéinnec, E., Chiori, R., Le Guyader, H., and Manuel, M. (2008). Insights into the early evolution of *SOX* genes from expression analyses in a ctenophore. *Journal of Experimental Zoology Part B : Molecular and Developmental Evolution*, 310B(8) :650–667.
- Jayaswal, V., Jermin, L. S., Poladian, L., and Robinson, J. (2011). Two stationary nonhomogeneous markov models of nucleotide sequence evolution. *Systematic Biology*, 60(1) :74–86.
- Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics : the beginning of incongruence? *Trends in Genetics*, 22(4) :225–231.
- Just, J., Kristensen, R., and Olesen, J. (2014). *Dendrogramma*, New Genus, with Two New Non-Bilaterian Species from the Marine Bathyal of Southeastern Australia (Animalia, Metazoa *incertae sedis*) - with Similarities to Some Medusoids from the Precambrian Ediacara. *PLoS ONE*, 9(9) :e102976.
- Kim, J., Kim, W., and Cunningham, C. W. (1999). A new perspective on lower metazoan relationships from 18S rDNA sequences. *Molecular Biology and Evolution*, 16(3) :423–427.
- Kircher, M., Heyn, P., and Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*, 12.
- Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the illumina platform. *Nucleic Acids Research*, 40(1) :e3.
- Kocot, K. M., Citarella, M. R., Moroz, L. L., and Halanych, K. M. (2013). PhyloTreePruner : A Phylogenetic Tree-Based Approach for Selection of Orthologous Sequences for Phylogenomics. *Evolutionary Bioinformatics*, 9 :429–435.
- Komai, T. (1922). *Studies on Two Aberrant Ctenophores, Coeloplana and Gastrodes*. Published by the author, Kyoto, Japan.
- Komai, T. (1963). *Lower Metazoa, Comparative Biology and Phylogeny*, chapter A Note on the Phylogeny of the Ctenophora, pages 181–188. Univ. California Press.
- Kostka, M., Uzlikova, M., Cepicka, I., and Flegr, J. (2008). Slowfaster, a user-friendly program for slow-fast analysis and its application on phylogeny of *Blastocystis*. *BMC Bioinformatics*, 9 :341.
- Künne (1939). Die Beroe (Ctenophora) der südlicher Nordsee, *Beroe gracilis* n. sp. *Zoologischer Anzeiger*, 127 :172–174.
- Lang, A. (1884). Die Polycladen (Seeplanarien) des Golfes von Neapel und der angrenzenden Meerabschnitte. *Fauna und Flora des Golfes von Neapel*, 11 :1–688.

- Lartillot, N. and Philippe, H. (2004). CAT : un modèle phylogénétique bayésien permettant de prendre en compte l'Hétérogénéité des processus de substitution entre sites dans les alignements protéiques. *Biosystema*, 22 :97–104.
- Laurin-Lemay, S., Brinkmann, H., and Philippe, H. (2012). Origin of land plants revisited in the light of sequence contamination and missing data. *Current Biology*, 22(15) :R593–R594.
- Leigh, J. W., Susko, E., Baumgartner, M., and Roger, A. J. (2008). Testing congruence in phylogenomic analysis. *Systematic Biology*, 57(1) :104–115.
- Lendenfeld, R. V. (1885). Notes on a beroid of port jackson. *Proceedings of the Linnean Society of New South Wales*, 9 :968–976.
- Lopez, P., Casane, D., and Philippe, H. (2002). Heterotachy, an important process of protein evolution. *Molecular Biology and Evolution*, 19(1) :1–7.
- Madin, L. and Harbison, G. (1978). *Thalassocalyce inconstans*, new genus and species, an enigmatic ctenophore representing a new family and order. *Bulletin of Marine Science*, 28(4) :680–687.
- Mallatt, J., Craig, C. W., and Yoder, M. J. (2012). Nearly complete rRNA genes from 371 animalia : updated structure-based alignment and detailed phylogenetic analysis. *Molecular phylogenetics and evolution*, 64(3) :603–617.
- Manuel, M., Borchiellini, C., Alivon, E., Le Parco, Y., Vacelet, J., and Boury-Esnault, N. (2003). Phylogeny and Evolution of Calcareous Sponges : Monophyly of Calcinea and Calcaronea, High Level of Morphological Homoplasy, and the Primitive Nature of Axial Symmetry. *Systematic Biology*, 52(3) :311–333.
- Martindale, M. and Henry, J. (1997a). Reassessing embryogenesis in the ctenophora : The inductive role of e(1) micromeres in organizing ctene row formation in the ‘mosaic’ embryo, *Mnemiopsis leidyi*. *Development*, 124(10) :1999–2006.
- Martindale, M. Q. and Henry, J. Q. (1997b). *Embryology : constructing the organism*, chapter Ctenophorans, the comb jellies, pages 87–111. Sinauer, Sunderland, MA.
- Martindale, M. Q. and Henry, J. Q. (1999). Intracellular fate mapping in a basal metazoan, the ctenophore *Mnemiopsis leidyi*, reveals the origins of mesoderm and the existence of indeterminate cell lineages. *Developmental Biology*, 214(2) :243 – 257.
- Mayer, A. (1912). *Ctenophores of the Atlantic coast of North America*. Number 162. Carnegie Institution of Washington.

- Medina, M., Collins, A. G., Silberman, J. D., and Sogin, M. L. (2001). Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proceedings of the National Academy of Sciences*, 98(17) :9707–9712.
- Meiklejohn, K. A., Danielson, M. J., Faircloth, B. C., Glenn, T. C., Braun, E. L., and Kimball, R. T. (2014). Incongruence among different mitochondrial regions : A case study using complete mitogenomes. *Molecular Phylogenetics and Evolution*, 78(0) :314 – 323.
- Mertens (1834). *fascic. I. Polypos, Acalephas Discophoras et Siphonophoras, nec non Echinodermata continens.*, chapter Prodromus descriptionis animalium ab H. Mertensio observatorum, pages 204–275. Bulletin de l'Academie imperiale des sciences de St. Petersburg.
- Mills, C. (1998-present). Phylum ctenophora : list of all valid species names. <http://faculty.washington.edu/cemills/Ctenolist.html>.
- Mills, C., Pugh, P., Harbison, G., and Haddock, S. (1996). Medusae, siphonophores and ctenophores of the Alboran Sea, south western Mediterranean. *Scientia Marina*, 60(1) :145–163.
- Moroz, L. L., Kocot, K. M., Citarella, M. R., Dosung, S., Norekian, T. P., Povolotskaya, I. S., Grigorenko, A. P., Dailey, C., Berezikov, E., Buckley, K. M., Ptitsyn, A., Reshetov, D., Mukherjee, K., Moroz, T. P., Bobkova, Y., Yu, F., Kapitonov, V. V., Jurka, J., Bobkov, Y. V., Swore, J. J., Girardo, D. O., Fodor, A., Gusev, F., Sanford, R., Bruders, R., Kittler, E., Mills, C. E., Rast, J. P., Derelle, R., Solovyev, V. V., Kondrashov, F. A., Swalla, B. J., Sweedler, J. V., Rogaev, E. I., Halanych, K. M., and Kohn, A. B. (2014). The ctenophore genome and the evolutionary origins of neural systems. *Nature*, 510(7503) :109–114.
- Mortensen, T. (1912). *The Danish Ingolf-Expedition*, volume Vol V. Zoological Museum.
- Mortensen, T. (1932). Ctenophora from the "michel sars" north atlantic deep-sea expedition 1910. *Reports of the "Michel Sars" North Atlantic deep-sea expedition 1910, Bergen Museum*, 3 :1–9.
- Moser, F. (1908). Japanische Ctenophoren. *Abt. Mat. Phys. Klasse K. Bayer. Acad. Wiss.*, 4, I suppl.
- Moser, F. (1909). Die Ctenophoren der deutsche Südpolar-Expedition 1901-1903. *Deutsche Südpolar-Expedition, XI, Zoologie III* :116–192, 3 pls.
- Müller, Y. (2004). *Faune et flore du littoral du Nord, du Pas-de-Calais et de la Belgique*. ICB Nord Pas de Calais.
- Nakanishi, N., Yuan, D., Jacobs, D., and Hartenstein, V. (2008). Early development, pattern, and reorganization of the planula nervous system in *Aurelia* (cnidaria, scyphozoa). *Development Genes and Evolution*, 218(10) :511–524.

- Nesnidal, M. P., Helmkampf, M., Bruchhaus, I., El-Matbouli, M., and Hausdorf, B. (2013). Agent of whirling disease meets orphan worm : Phylogenomic analyses firmly place myxozoa in Cnidaria. *PLoS ONE*, 8(1) :e54576.
- Nesnidal, M. P., Helmkampf, M., Bruchhaus, I., and Hausdorf, B. (2010). Compositional Heterogeneity and Phylogenomic Inference of Metazoan Relationships. *Molecular Biology and Evolution*, 27(9) :2095–2104.
- Nielsen, C., Scharff, N., and Eibye-Jacobsen, D. (1996). Cladistic analyses of the animal kingdom. *Biological Journal of the Linnean Society*, 57(4) :385–410.
- Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., Müller, W. E., Nickel, M., Schierwater, B., Vacelet, J., Wiens, M., and Wörheide, G. (2013). Deep metazoan phylogeny : When different genes tell different stories. *Molecular Phylogenetics and Evolution*, 67(1) :223–233.
- Oliveira, O. M. P., Feliú, G., and Palma, S. (2014). *Beroe gracilis* (ctenophora) from the humboldt current system : first occurrence of this species in the southern hemisphere. *Zootaxa*, 3827(1) :397–400.
- Ospovat, M. (1985a). Phylogeny and classification of the type Ctenophora. *Zoologichetskyi Zhurnal*, 64(7) :965–974.
- Ospovat, M. (1985b). Relationships between the orders Cryptolobiferida, Lobiferida and Thalas-socalycida (Ctenophora). *Zoologichetskyi Zhurnal*, 64(6) :805–812.
- Pagel, M. and Meade, A. (2008). Modelling heterotachy in phylogenetic inference by reversible-jump markov chain monte carlo. *Philosophical Transactions of the Royal Society of London, B : Biological Sciences*, 363(1512) :3955–3964.
- Peterson, K. J., Cotton, J. A., Gehling, J. G., and Pisani, D. (2008). The ediacaran emergence of bilaterians : congruence between the genetic and the geological fossil records. *Philos Trans R Soc Lond B Biol Sci*, 363(1496) :1435–1443.
- Peterson, K. J. and Eernisse, D. J. (2001). Animal phylogeny and the ancestry of bilaterians : inferences from morphology and 18S rDNA gene sequences. *Evolution & Development*, 3(3) :170–205.
- Peterson, K. J., Summons, R. E., and Donoghue, P. C. J. (2007). Molecular palaeobiology. *Palaeontology*, 50(4) :775–809.
- Pett, W., Ryan, J. F., Pang, K., Mullikin, J. C., Martindale, M. Q., Baxevanis, A. D., and Lavrov, D. V. (2011). Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi* : Insight from mtDNA and the nuclear genome. *Mitochondrial DNA*, 22(4) :130–142.

- Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., and Baurain, D. (2011). Resolving difficult phylogenetic questions : why more sequences are not enough. *PLoS biology*, 9(3) :e1000602.
- Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D. J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G., and Manuel, M. (2009). Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, 19(8) :706–712.
- Philippe, H. and Telford, M. J. (2006). Large-scale sequencing and the new animal phylogeny. *Trends in Ecology & Evolution*, 21(11) :614–620.
- Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., and Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. *BMC evolutionary biology*, 5(1) :50.
- Pick, K. S., Philippe, H., Schreiber, F., Erpenbeck, D., Jackson, D. J., Wrede, P., Wiens, M., Alie, A., Morgenstern, B., Manuel, M., and Wörheide, G. (2010). Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Molecular Biology and Evolution*, 27(9) :1983–1987.
- Pisani, D. (2004). Identifying and removing fast-evolving sites using compatibility analysis : An example from the arthropoda. *Systematic Biology*, 53(6) :978–989.
- Pisani, D., Feuda, R., Peterson, K. J., and Smith, A. B. (2012). Resolving phylogenetic signal from noise when divergence is rapid : a new look at the old problem of echinoderm class relationships. *Molecular Phylogenetic and Evolution*, 62(1) :27–34.
- Podar, M., Haddock, S. H., Sogin, M. L., and Harbison, G. (2001). A molecular phylogenetic framework for the phylum ctenophora using 18S rRNA genes. *Molecular Phylogenetics and Evolution*, 21(2) :218–230.
- Rankin, J. J. (1956). The structure and biology of *Vallicula multiformis*, gen. et sp. nov., a platyctenid ctenophore. *Journal of the Linnean Society of London, Zoology*, 43(289) :55–71.
- Riesgo, A., Farrar, N., Windsor, P. J., Giribet, G., and Leys, S. P. (2014). The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Molecular Biology and Evolution*, 31(5) :1102–1120.
- Rokas, A., King, N., Finnerty, J., and Carroll, S. B. (2003). Conflicting phylogenetic signals at the base of the metazoan tree. *Evolution and Development*, 5(4) :346–359.
- Roure, B., Baurain, D., and Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, 30(1) :197–214.

- Ryan, J. F., Pang, K., Schnitzler, C. E., Nguyen, A.-D., Moreland, R. T., Simmons, D. K., Koch, B. J., Francis, W. R., Havlak, P., NISC Comparative Sequencing Program, Smith, S. A., Putnam, N. H., Haddock, S. H. D., Dunn, C. W., Wolfsberg, T. G., Mullikin, J. C., Martindale, M. Q., and Baxevanis, A. D. (2013). The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science*, 342(6164) :1242592–1242592.
- Salichos, L. and Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, 497(7449) :327–331.
- Schierwater, B., Eitel, M., Jakob, W., Osigus, H.-J., Hadrys, H., Dellaporta, S. L., Kolokotronis, S.-O., and Desalle, R. (2009). Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biol*, 7(1) :e20.
- Schram, F. (1991). *The early evolution of Metazoa and the significance of problematic taxa*, chapter Cladistic analysis of metazoan phyla and the placement of fossil problematica, pages 35–46. Cambridge Univ. Press, WY.
- Seravin, L. (1995). Brief history of the study of jelly-fishes of the genus *Beroe* and differential characters *B. cucumis* (Ctenophora, Atentaculata, Beroida). *Zoologichetskyi Zhurnal*, 74(12) :3–15.
- Seravin, L. N. (1994). The systematic revision of the genus *Mnemiopsis* (ctenophora, lobata). *Zoologichetskyi Zhurnal*, 73 :9–18.
- Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C., and Giribet, G. (2014). Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution*.
- Shu, D., Morris, S., Han, J., Li, Y., Zhang, X., Hua, H., Zhang, Z., Liu, J., Guo, J., Yao, Y., and Yasui, K. (2006). Lower cambrian vendobionts from china and early diploblast evolution. *Science*, 312(5774) :731–734.
- Smith, S. A., Wilson, N. G., Goetz, F. E., Feehery, C., Andrade, S. C. S., Rouse, G. W., Giribet, G., and Dunn, C. W. (2011). Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, 480(7377) :364–367.
- Sorensen, M., Funch, P., Willerslev, E., Hansen, A., and Olesen, J. (2000). On the phylogeny of the metazoa in the light of Cyclophora and Micrognathozoa. *Zoologischer Anzeiger*, 239(3-4) :297–318.
- Sperling, E., Pisani, D., and Peterson, K. (2007). Poriferan paraphyly and its implications for precambrian palaeobiology. *Geological Society, London, Special Publications*, 286(1) :355–368.

- Stanley, G. D. and Sturmer, W. (1983). The first fossil ctenophore from the lower devonian of west germany. *Nature*, 303(5917) :518–520.
- Stanley, G. D. and Sturmer, W. (1987). A new fossil ctenophore discovered by x-rays. *Nature*, 328(6125) :61–63.
- Straub, S. C., Moore, M. J., Soltis, P. S., Soltis, D. E., Liston, A., and Livshultz, T. (2014). Phylogenetic signal detection from an ancient rapid radiation : Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Molecular Phylogenetics and Evolution*, 80(0) :169 – 185.
- Struck, T. H. (2013). The impact of paralogy on phylogenomic studies : A case study on annelid relationships. *PLoS ONE*, 8(5) :e62892.
- Struck, T. H. and Fisse, F. (2008). Phylogenetic position of nemertea derived from phylogenomic data. *Molecular Biology and Evolution*, 25(4) :728–736.
- Struck, T. H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., and Halanych, K. M. (2007). Annelid phylogeny and the status of Sipuncula and Echiura. *BMC Evolutionary Biology*, 7 :57.
- Struck, T. H., Wey-Fabrizius, A. R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., Kück, P., Herlyn, H., and Hankeln, T. (2014). Platyzoan paraphyly based on phylogenomic data supports a non-coelomate ancestry of Spiralia. *Molecular Biology and Evolution*.
- Tamm, S. (1973). Mechanisms of ciliary co-ordination in ctenophores. *Journal of Experimental Biology*, 59(1) :231–245.
- Tamm, S. (1982). *Electrical conduction and behavior in "simple" Invertebrates*, chapter Ctenophora, pages 266–358. Oxford : Clarendon Press.
- Tamm, S. and Tamm, S. (1987). Massive actin bundle couples macrocilia to muscles in the ctenophore *Beroë*. *Cell Motility and the Cytoskeleton*, 7(2) :116–128.
- Tamm, S. and Tamm, S. L. (1991a). Actin pegs and ultrastructure of presumed sensory receptors of *Beroë* (ctenophora). *Cell and Tissue Research*, 264(1) :151–159.
- Tamm, S. and Tamm, S. L. (1995). A giant nerve net with multi-effector synapses underlying epithelial adhesive strips in the mouth of *Beroë* (ctenophora). *Journal of neurocytology*, 24(9) :711–723.

- Tamm, S. L. (2012). Patterns of comb row development in young and adult stages of the ctenophores *Mnemiopsis leidyi* and *Pleurobrachia pileus*. *Journal of Morphology*, 273(9) :1050–1063.
- Tamm, S. L. (2014). Cilia and the life of ctenophores. *Invertebrate Biology*, 133(1) :1–46.
- Tamm, S. L. and Tamm, S. (1991b). Reversible epithelial adhesion closes the mouth of *Beroë*, a carnivorous marine jelly. *Biological Bulletin*, 181 :11.
- Tamm, S. L. and Tamm, S. (1993). Diversity of macrociliary size, tooth patterns, and distribution in *Beroë* (ctenophora). *Zoomorphology*, 113 :79–89.
- Tamura, K. and Kumar, S. (2002). Evolutionary distance estimation under heterogeneous substitution pattern among lineages. *Molecular Biology and Evolution*, 19(10) :1727–1736.
- Tang, F., Bengtson, S., Wang, Y., Wang, X.-l., and Yin, C.-y. (2011). *Eoandromeda* and the origin of ctenophora. *Evolution & Development*, 13(5) :408–414.
- Tang, F., Yin, C., Bengtson, S., Liu, P., Wang, Z., and Gao, L. (2008). Octoradiate spiral organisms in the ediacaran of south china. *Acta Geologica Sinica - English Edition*, 82(1) :27–34.
- Telford, M. J. (2013). Evolution. the animal tree of life. *Science*, 339(6121) :764–766.
- Telford, M. J. and Copley, R. R. (2011). Improving animal phylogenies with genomic data. *Trends in Genetics*, 27(5) :186–195.
- Telford, M. J., Lowe, C. J., Cameron, C. B., Ortega-Martinez, O., Aronowicz, J., Oliveri, P., and Copley, R. R. (2014). Phylogenomic analysis of echinoderm class relationships supports asterozoa. *Proceedings of Biological Science*, 281(1786).
- Trégouboff, G. and Rose, M. (1957). *Manuel de planctonologie méditerranéenne : Tome 1*, chapter Ctenophora (=Acnidaria), pages 400–419. CNRS, Paris.
- Waggett, R. and Costello, J. (1999). Capture mechanisms used by the lobate ctenophore, *Mnemiopsis leidyi*, preying on the copepod *Acartia tonsa*. *Journal of Plankton Research*, 21(11) :2037–2052.
- Welch, V., Vigneron, J. P., Lousse, V., and Parker, A. (2006). Optical properties of the iridescent organ of the comb-jellyfish *Beroë cucumis* (ctenophora). *Physiological Reviews E*, 73 :041916.
- Xia, X., Xie, Z., Salemi, M., Chen, L., and Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, 26(1) :1 – 7.
- Xiao, S., Yuan, X., and Knoll, A. H. (2000). Eumetazoan fossils in terminal proterozoic phosphorites? *Proceedings of the National Academy of Sciences*, 97(25) :13684–13689.

Zhang, X. and Reitner, J. (2006). A fresh look at *Dickinsonia* : removing it from vendobionta. *Acta Geologica Sinica-English Edition*, 80(5) :636–642.

Zrzavy, J., Mihulka, S., Kepka, P., Bezdek, A., and Tietz, D. (1998). Phylogeny of the Metazoa Based on Morphological and 18S Ribosomal DNA Evidence. *Cladistics*, 14(3) :249–285.

Annexe A

Appendice : article 1

Supplementary Data

<i>Chlamydomonas reinhardtii</i>	PRJ NA21061
<i>Arabidopsis lyrata lyrata</i>	PRJ NA49545
<i>Chondrus crispus</i>	PRJ NA193762
<i>Acanthamoeba castellanii</i>	PRJ NA193615
<i>Dictyostelium purpureum</i>	PRJ NA63531
<i>Bigeloviella natans</i>	PRJ NA27939
<i>Toxoplasma gondii</i> ME49	PRJ NA32719
<i>Oxytricha trifallax</i>	
<i>Tetrahymena thermophila</i> SB210	PRJ NA16792
<i>Perkinsus marinus</i> ATCC 50983	PRJ NA46451
<i>Symbiodinium</i>	
<i>Phytophthora infestans</i>	PRJ NA49677
<i>Emiliana huxleyi</i>	PRJ NA222302
<i>Trypanosoma cruzi</i>	PRJ NA15540
<i>Batrachomyxium dendrobatidis</i> JAM81	PRJ NA225502
<i>Rhizopus delemar</i>	PRJ NA13066
<i>Phanerochaete carnosae</i>	PRJ NA38425
<i>Aspergillus oryzae</i> RIB40	PRJ NA28175
<i>Debaryomyces hansenii</i> CBS767	PRJ NA12410
<i>Nostoc punctiforme</i> PCC 73102	PRJ NA57767
<i>Escherichia coli</i>	PRJ NA176127
<i>Rhizobium etli</i> bv mimosae	PRJ NA213896
<i>Simkania negevensis</i>	PRJ NA68451
<i>Bacillus firmus</i> DS1	PRJ NA193594
<i>Streptomyces aurantiacus</i> J A 4570	PRJ NA209656
<i>Prevotella bergensis</i> DSM 17361	PRJ NA55893
<i>Marinobacter adhaerens</i> HP15	PRJ NA162009
<i>Marinomonas mediterranea</i>	PRJ NA64753
<i>Idiomarina loihiensis</i> L2TR	PRJ NA205256
<i>Pyrococcus abyssi</i> GE5	PRJ NA62903
<i>Sulfolobus solfataricus</i> P2	PRJ NA57721
<i>Candidatus Nitrosopumilus</i>	PRJ NA176130
<i>Capsaspora owczarzaki</i>	PRJ NA193613, PRJ NA20341
<i>Monosiga brevicollis</i>	PRJ NA28133, PRJ NA19045
<i>Salpingoeca rosetta</i>	PRJ NA193541, PRJ NA37927
<i>Amphimedon queenslandica</i>	ENSEMBL
<i>Trichoplax adhaerens</i>	ENSEMBL
<i>Mnemiopsis leidyi</i>	NHGR
<i>Nematostella vectensis</i>	ENSEMBL
<i>Acropora digitifera</i>	COMPAGEN
<i>Hydra magnipapillata</i>	COMPAGEN
<i>Daphnia pulex</i>	ENSEMBL
<i>Ixodes scapularis</i>	ENSEMBL
<i>Tribolium castaneum</i>	ENSEMBL
<i>Crassostrea gigas</i>	ENSEMBL
<i>Aplysia californica</i>	PRJ NA209509
<i>Capitella teleta</i>	ENSEMBL
<i>Hellodella robusta</i>	ENSEMBL
<i>Strongylocentrotus purpur.</i>	ENSEMBL
<i>Saccoglossus kowalevskii</i>	JGI
<i>Branchiostoma floridae</i>	UNIPROT
<i>Homo sapiens</i>	UNIPROT

Table 1: List of species used in our database when looking for non-holozoan contaminations in transcriptomes (DC1). Species, data source.

Daphnia pulex	ENSEMBL
Ixodes scapularis	ENSEMBL
Tribolium castaneum	ENSEMBL
Crassostrea gigas	ENSEMBL
Aplysia californica	#PRJ NA209509
Capitella teleta	ENSEMBL
Hellobdella robusta	ENSEMBL
Strongylocentrotus purpur.	ENSEMBL
Saccoglossus kowalevskii	JGI
Branchiostoma floridae	UNIPROT
Homo sapiens	UNIPROT
Hymenolepis microstoma	# http://www.ncbi.nlm.nih.gov/protein/?term=txid85433[Organism%3Aexp]
Clonorchis sinensis	# http://www.ncbi.nlm.nih.gov/protein/?term=txid79923[Organism%3Aexp]
Ascaris suum	# http://www.ncbi.nlm.nih.gov/protein/?term=txid6253[Organism%3Aexp]
Necator americanus	# http://www.ncbi.nlm.nih.gov/protein/?term=txid51031[Organism%3Aexp]
Trichinella spiralis	# http://www.ncbi.nlm.nih.gov/protein/?term=txid6334[Organism%3Aexp]
Adineta vaga	#Wey-Fabrizius
Bugula neritina	#SRX015704

Table 2: List of species used in the database when looking for bilaterian contaminations in transcriptomes (DC2). Species, data source.

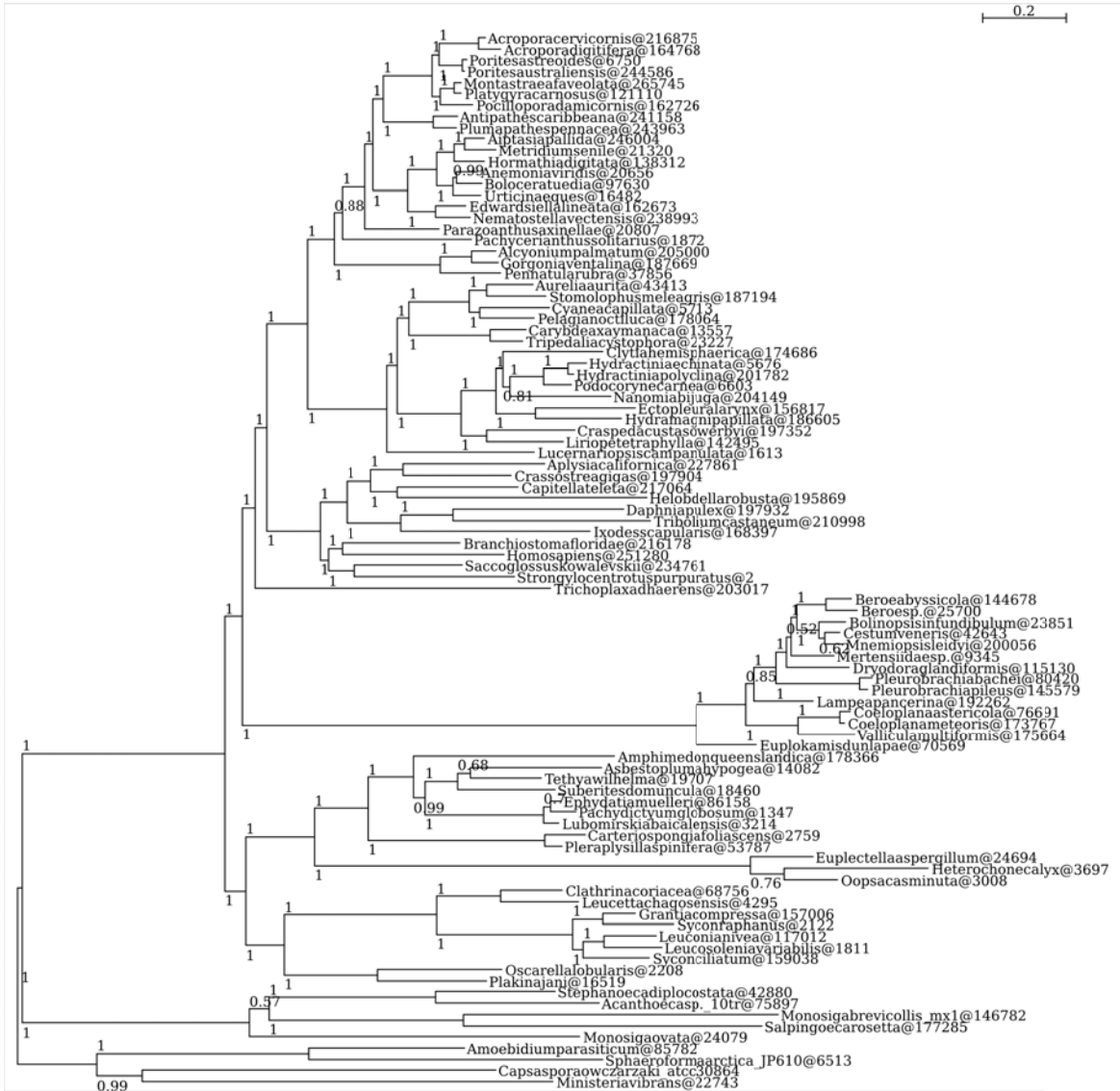


Figure 1: AUTO-1143.94.PB50

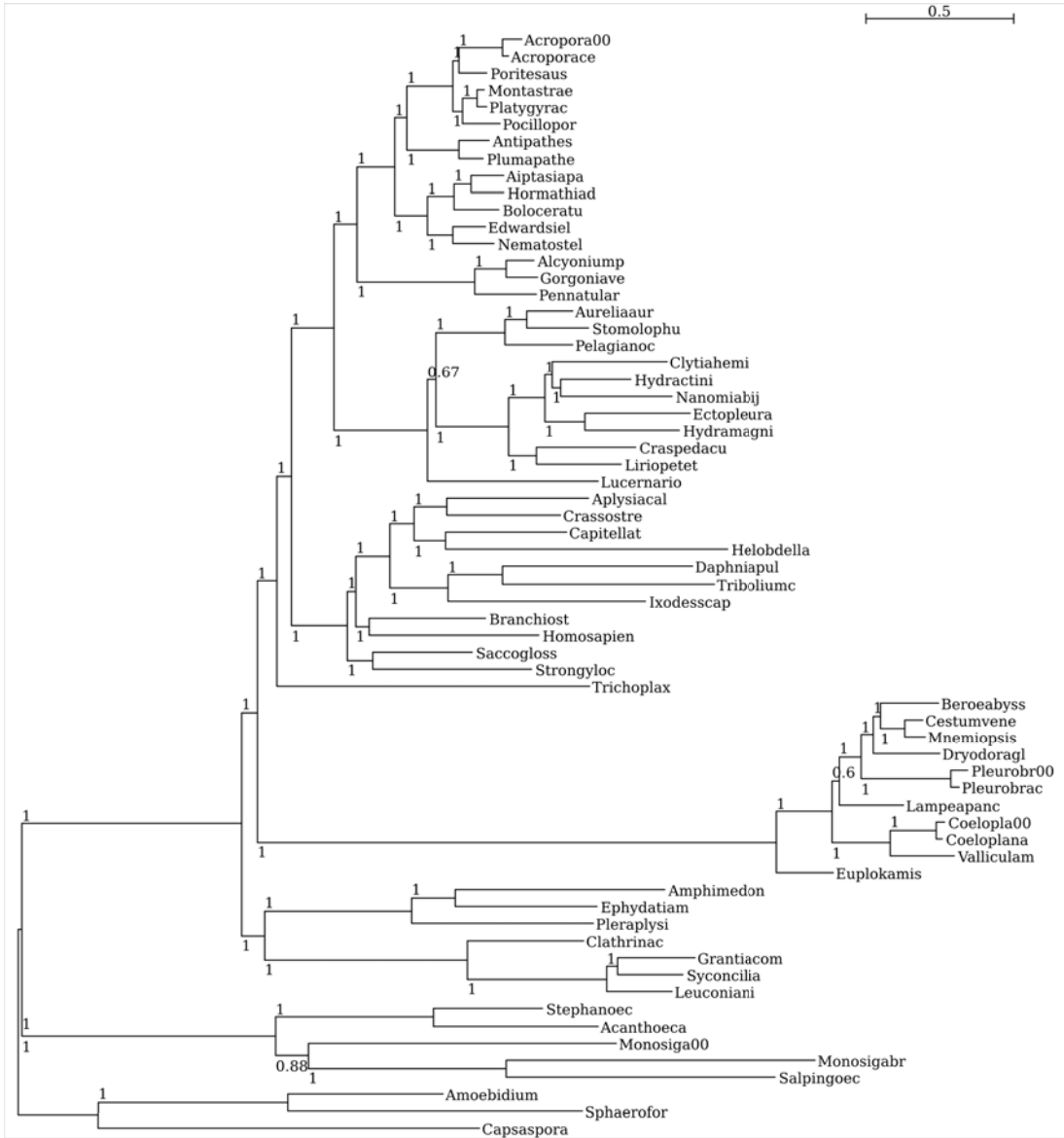


Figure 2: AUTO-1143.65_PB70

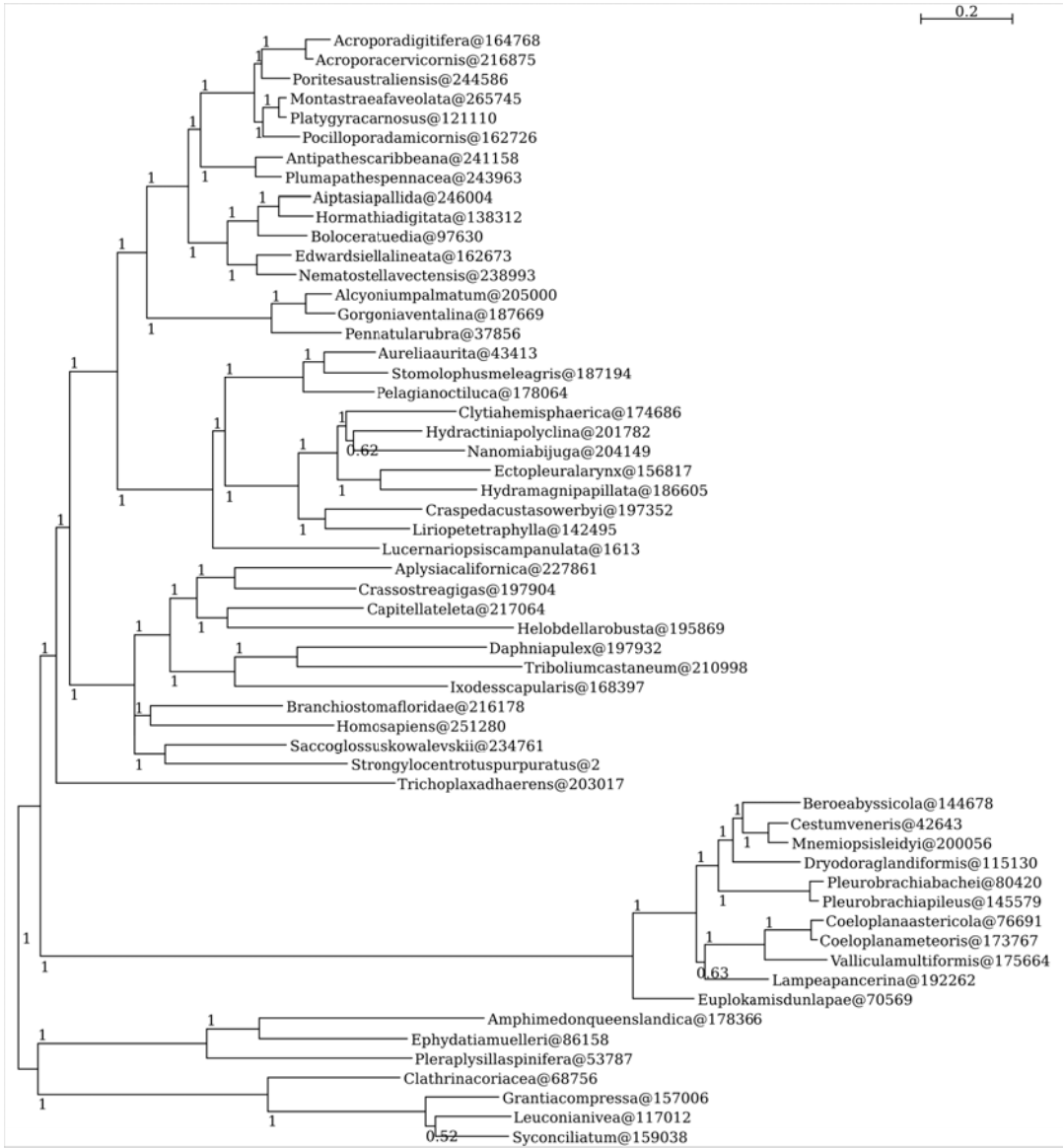


Figure 3: AUTO-1143_65_PB50 with no outgroups

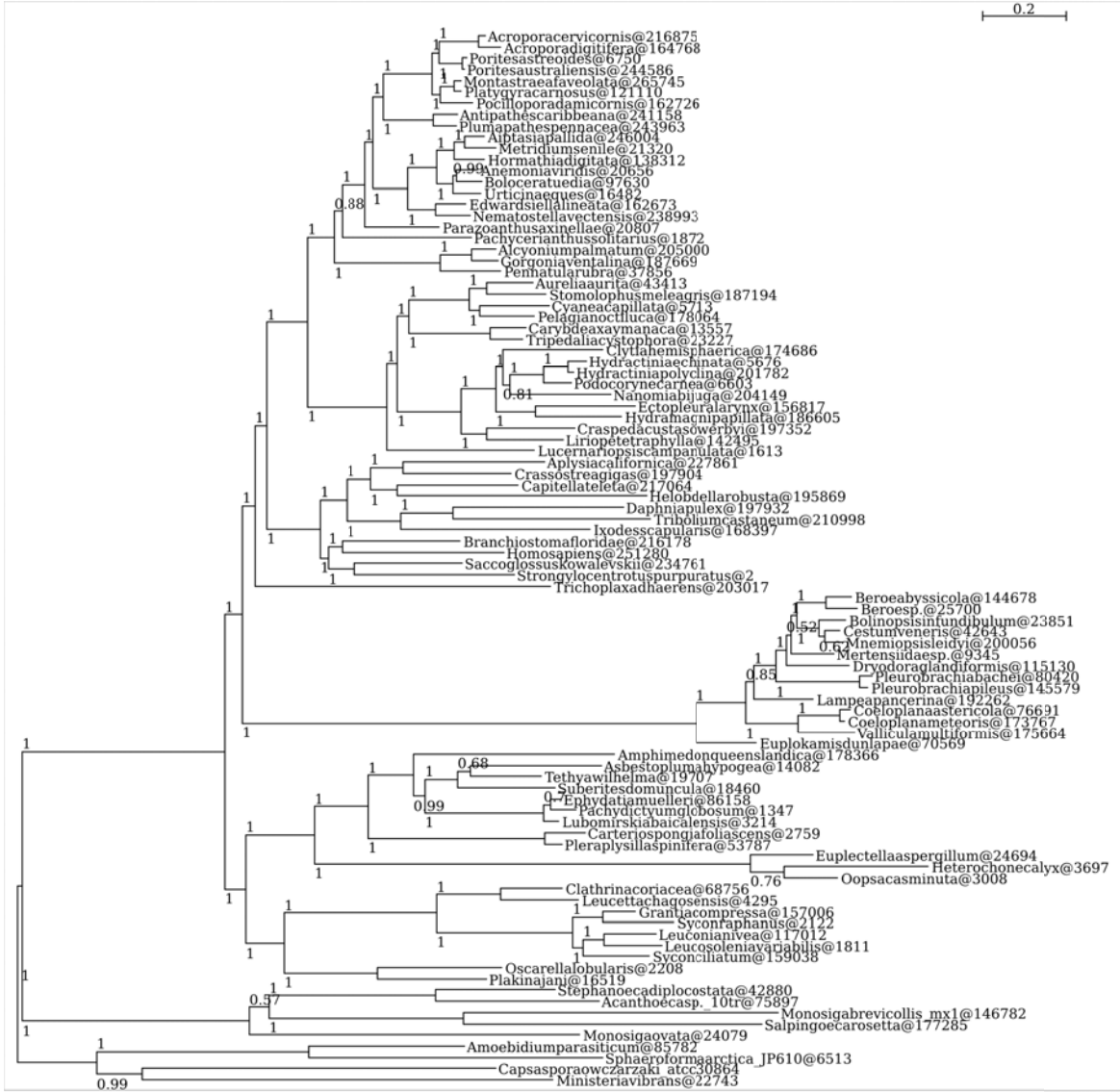


Figure 4: AUTO-1143.94_PB50 with no outgroups

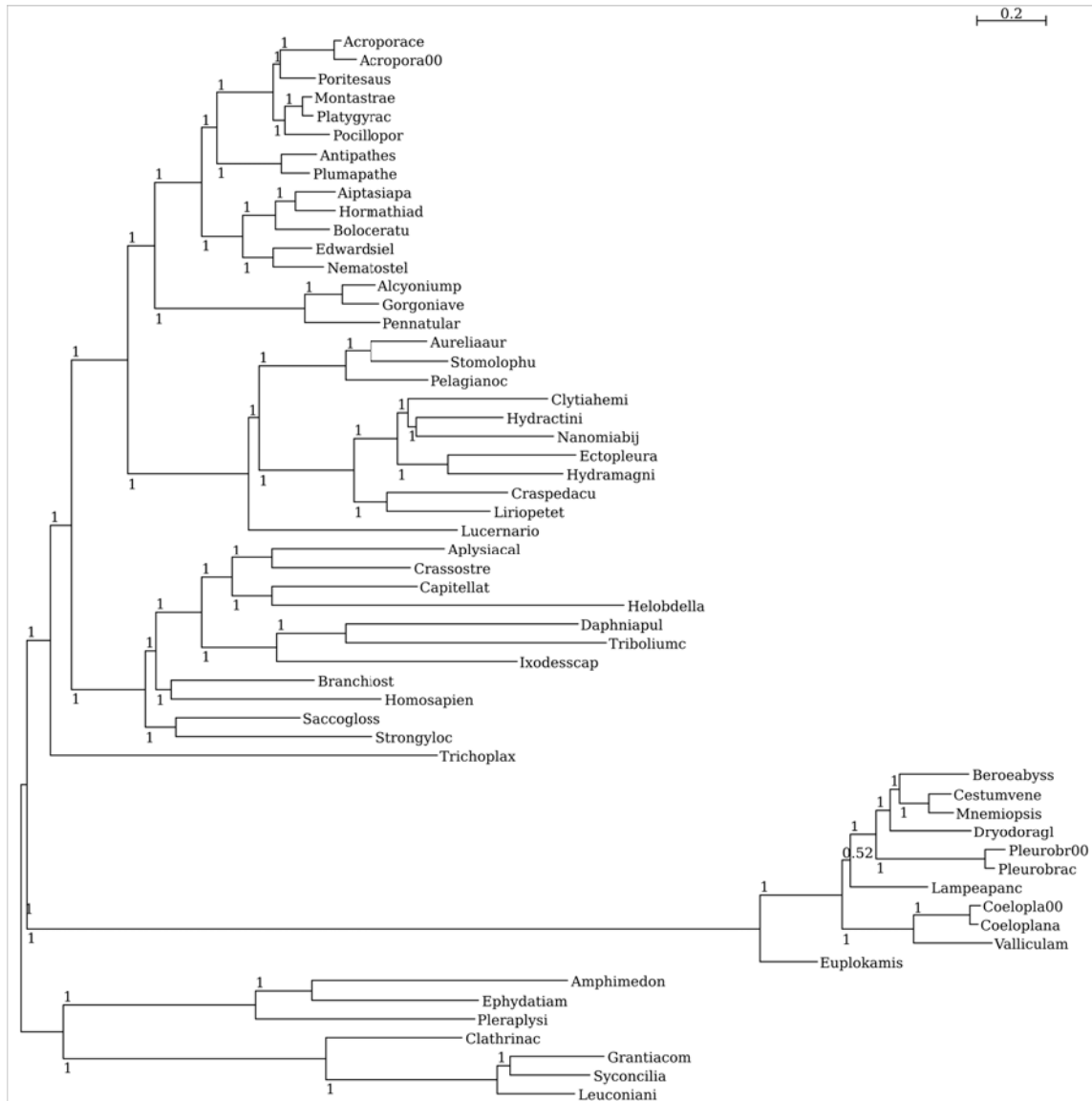


Figure 5: AUTO-1143.65_PB70 with no outgroups

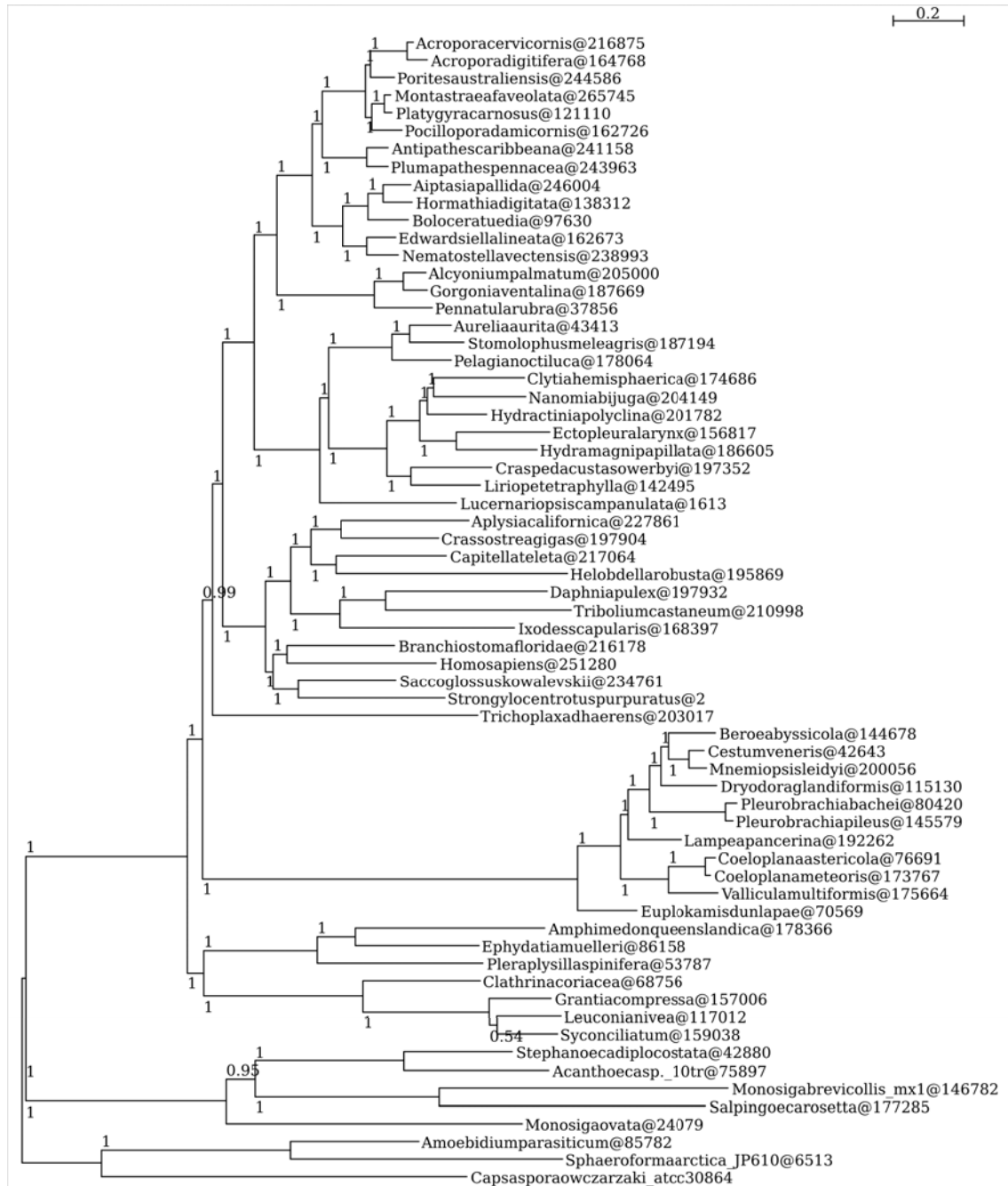


Figure 6: AUTO-1143_65_PB50_PIP10

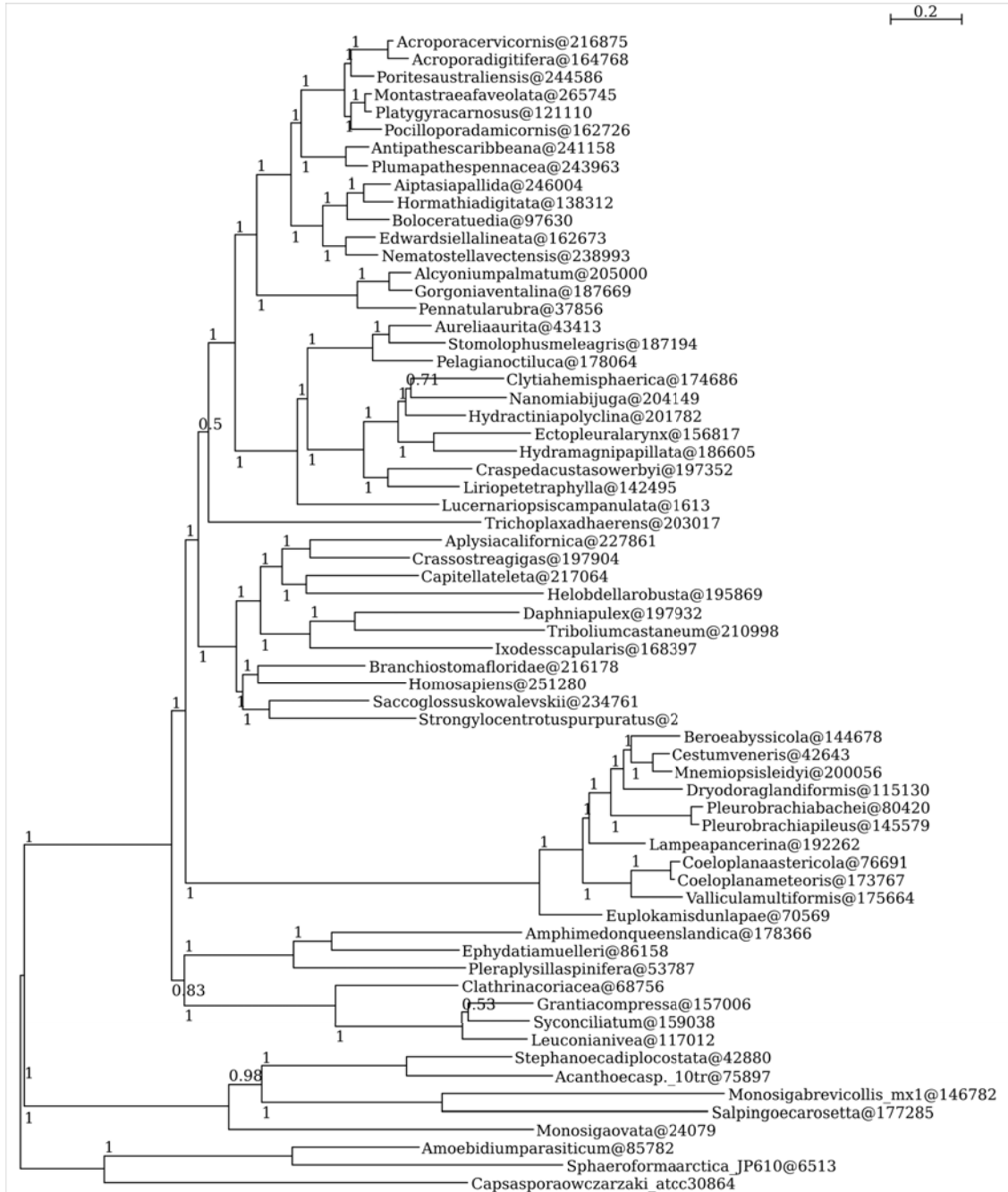


Figure 7: AUTO-1143_65_PB50_PIP20

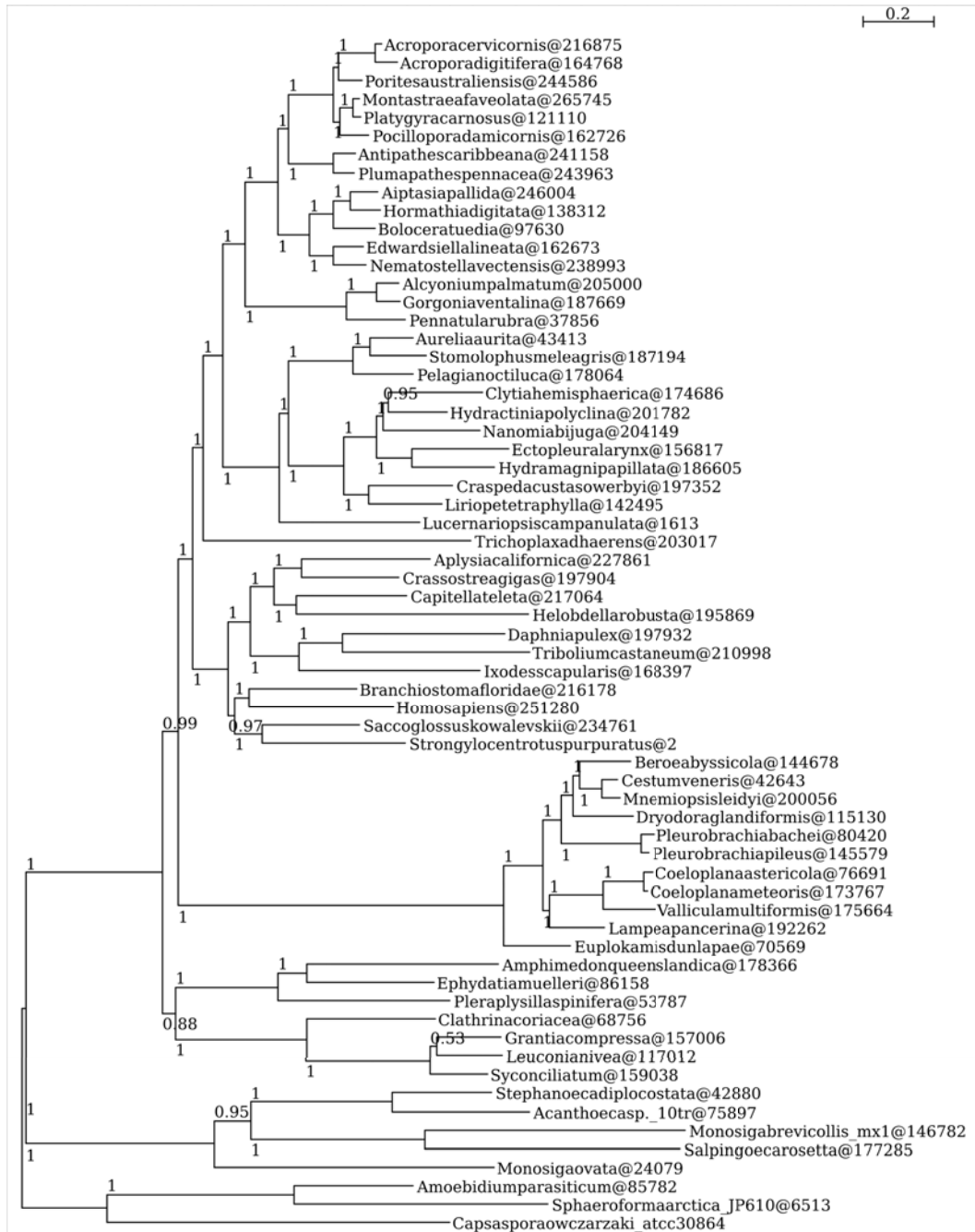


Figure 8: AUTO-1143_65_PB50_PIP30

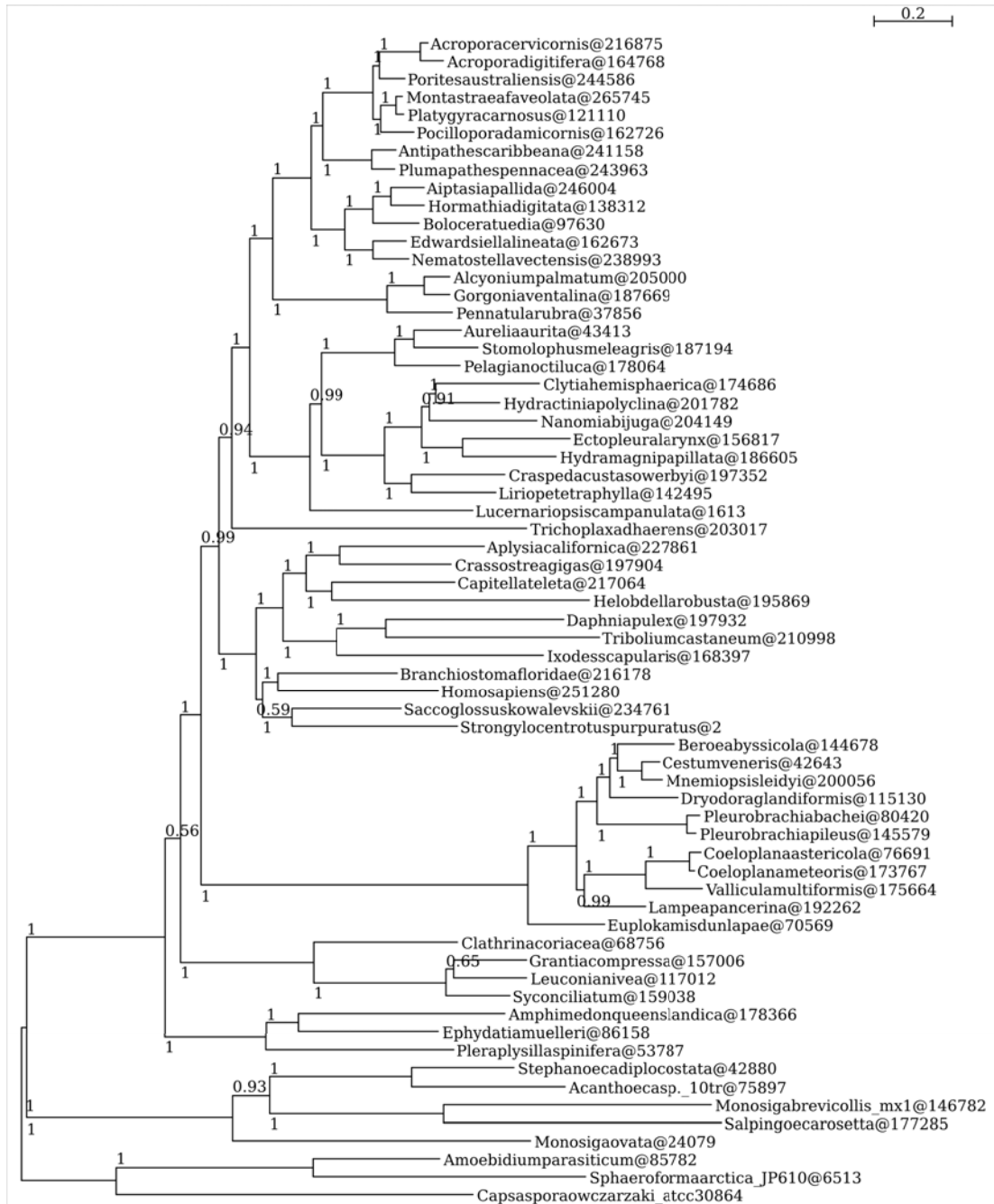


Figure 9: AUTO-1143_65_PB50_PIP40

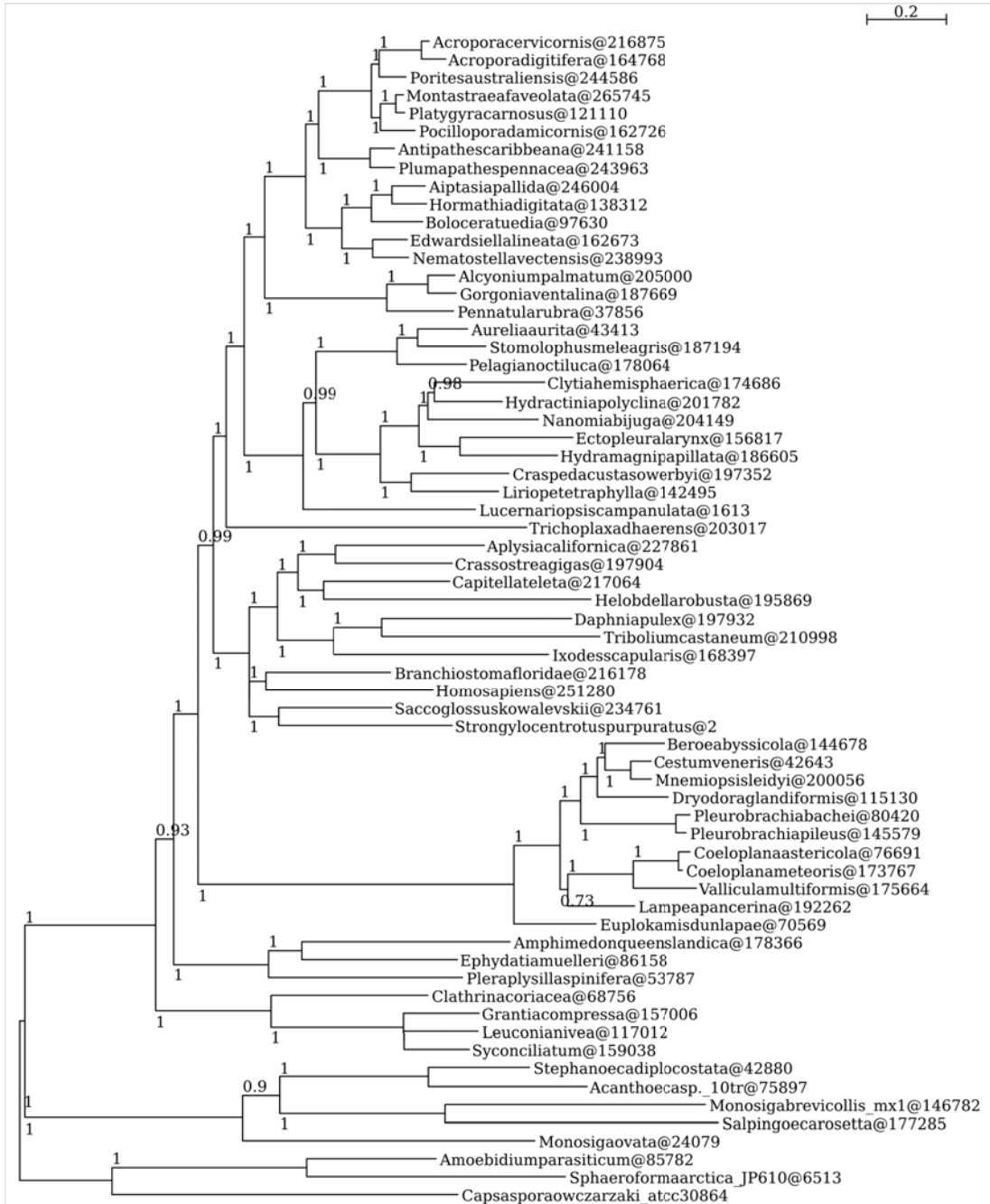


Figure 10: AUTO-1143_65_PB50_PIP50

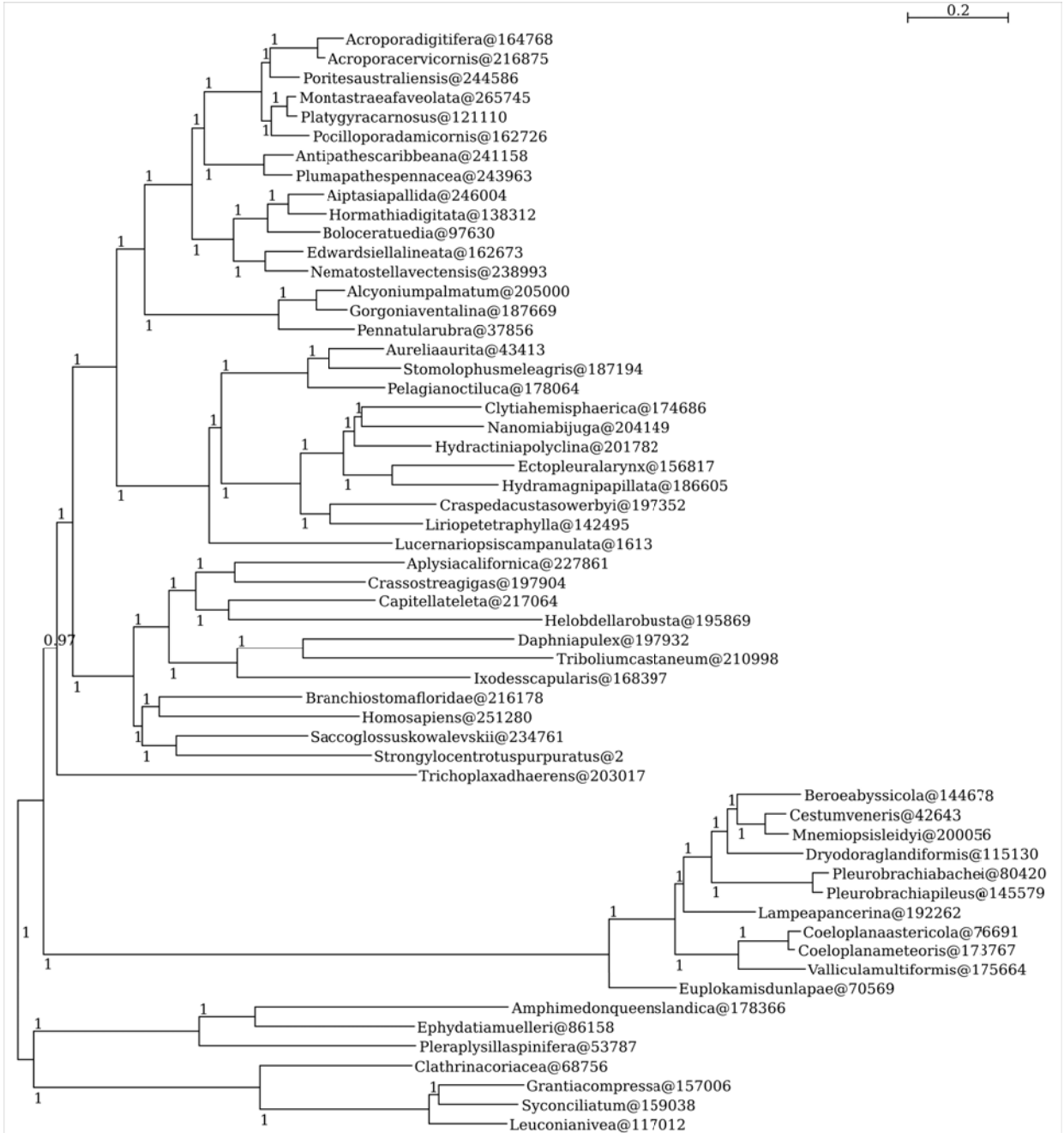


Figure 11: AUTO-1143_65_PB50_PIP10 with no outgroups

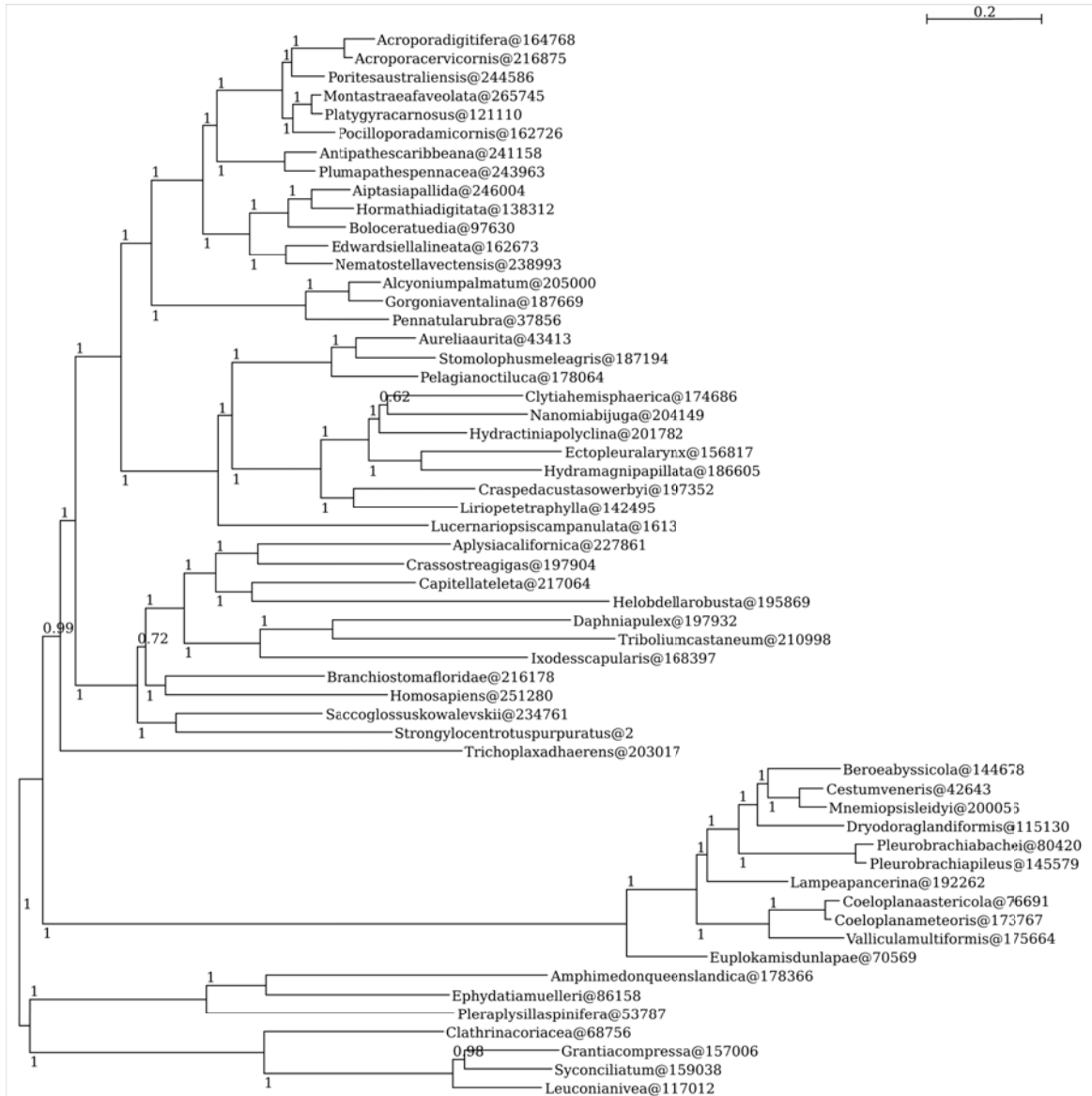


Figure 12: AUTO-1143_65_PB50_PIP20 with no outgroups

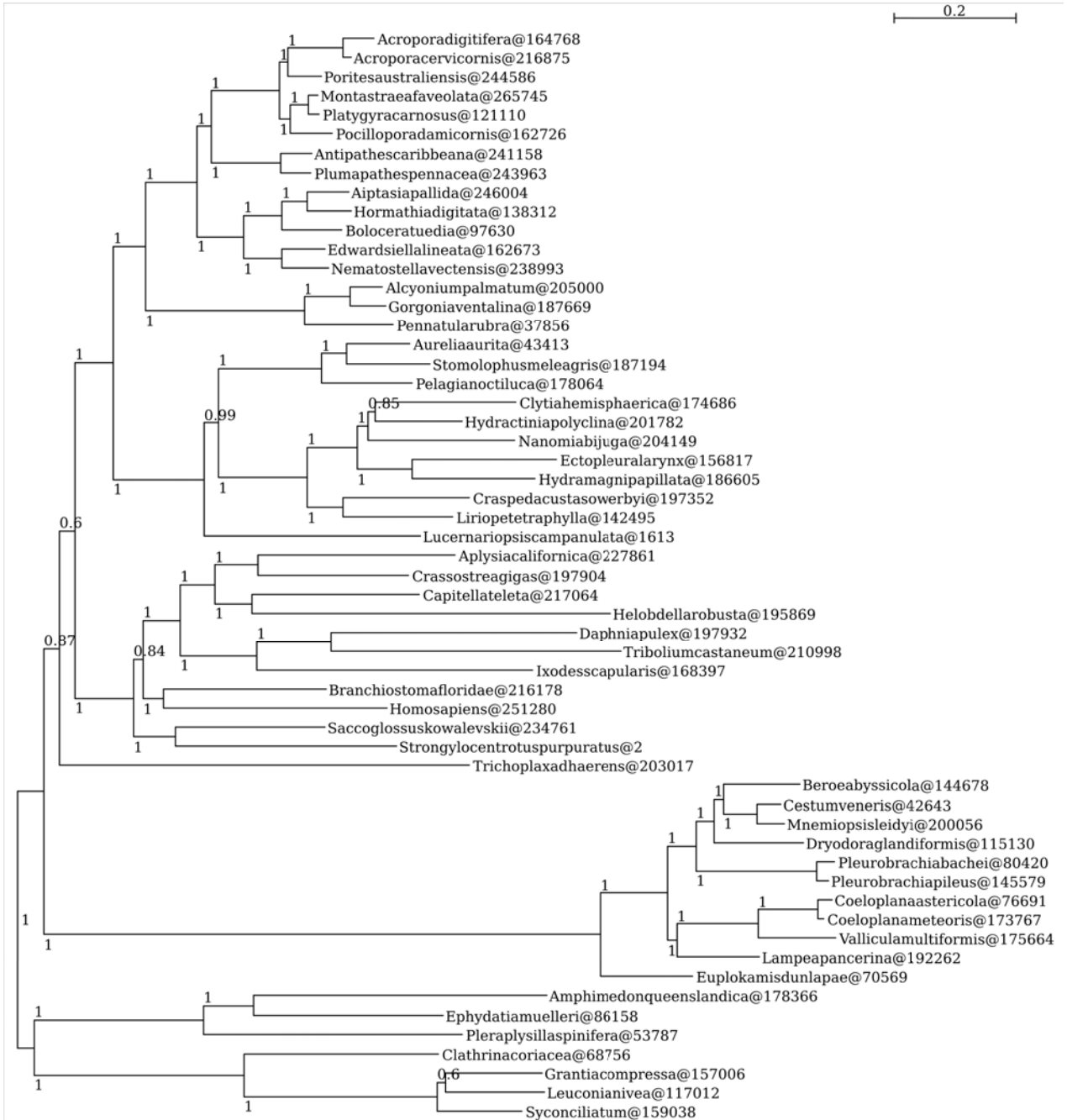


Figure 13: AUTO-1143_65_PB50_PIP30 with no outgroups

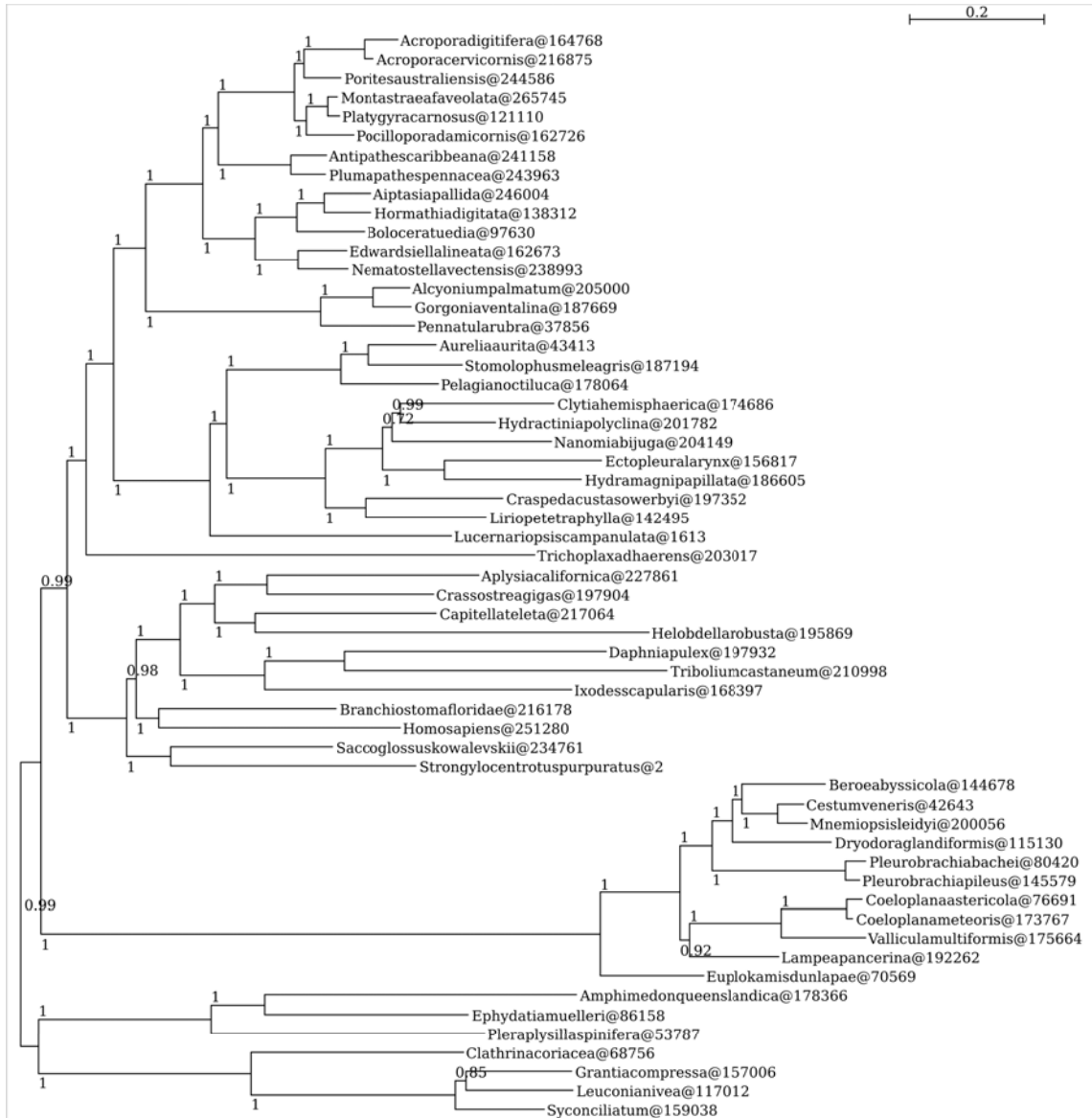


Figure 14: AUTO-1143_65_PB50_PIP40 with no outgroups

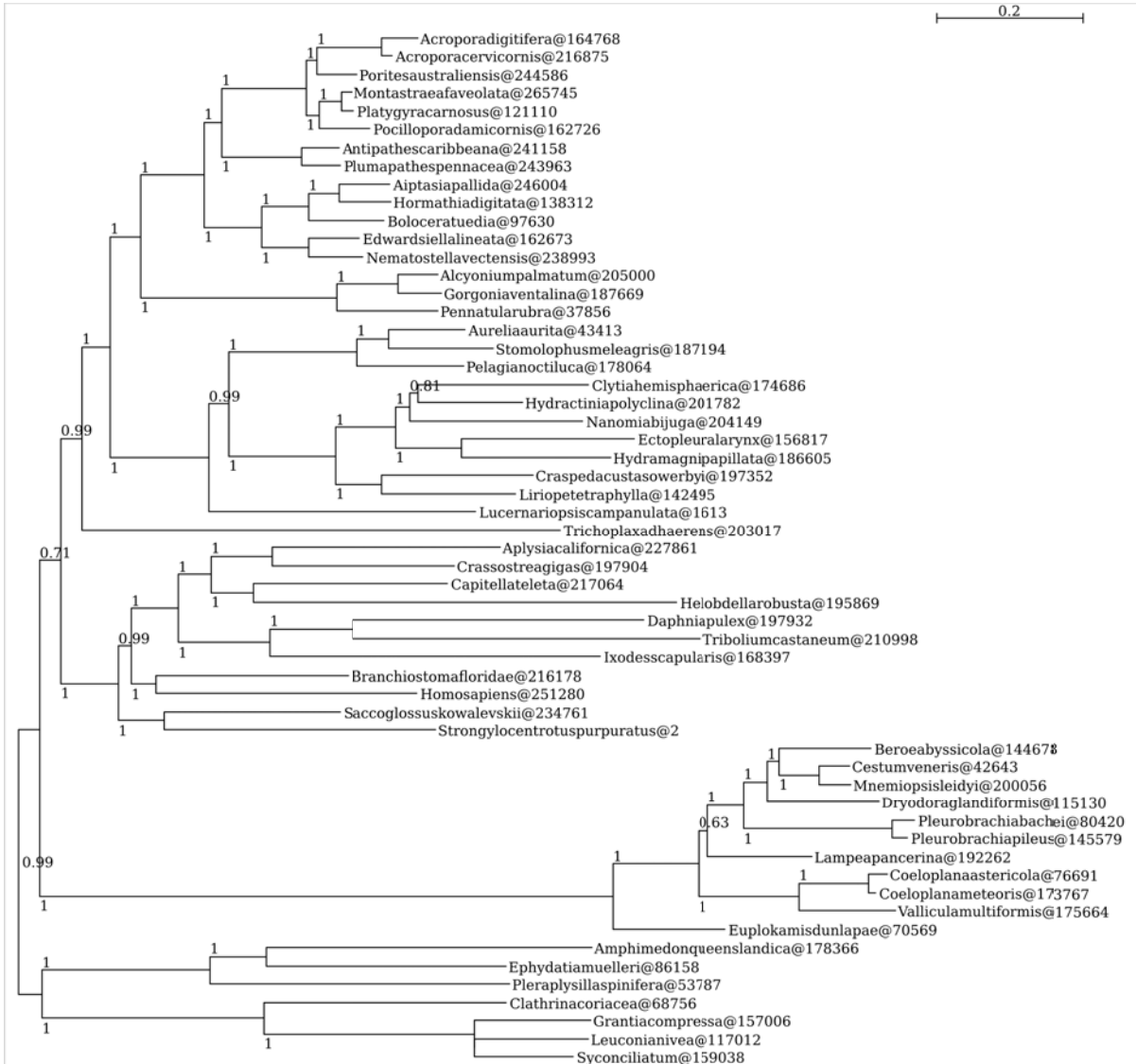


Figure 15: AUTO-1143_65_PB50_PIP50 with no outgroups

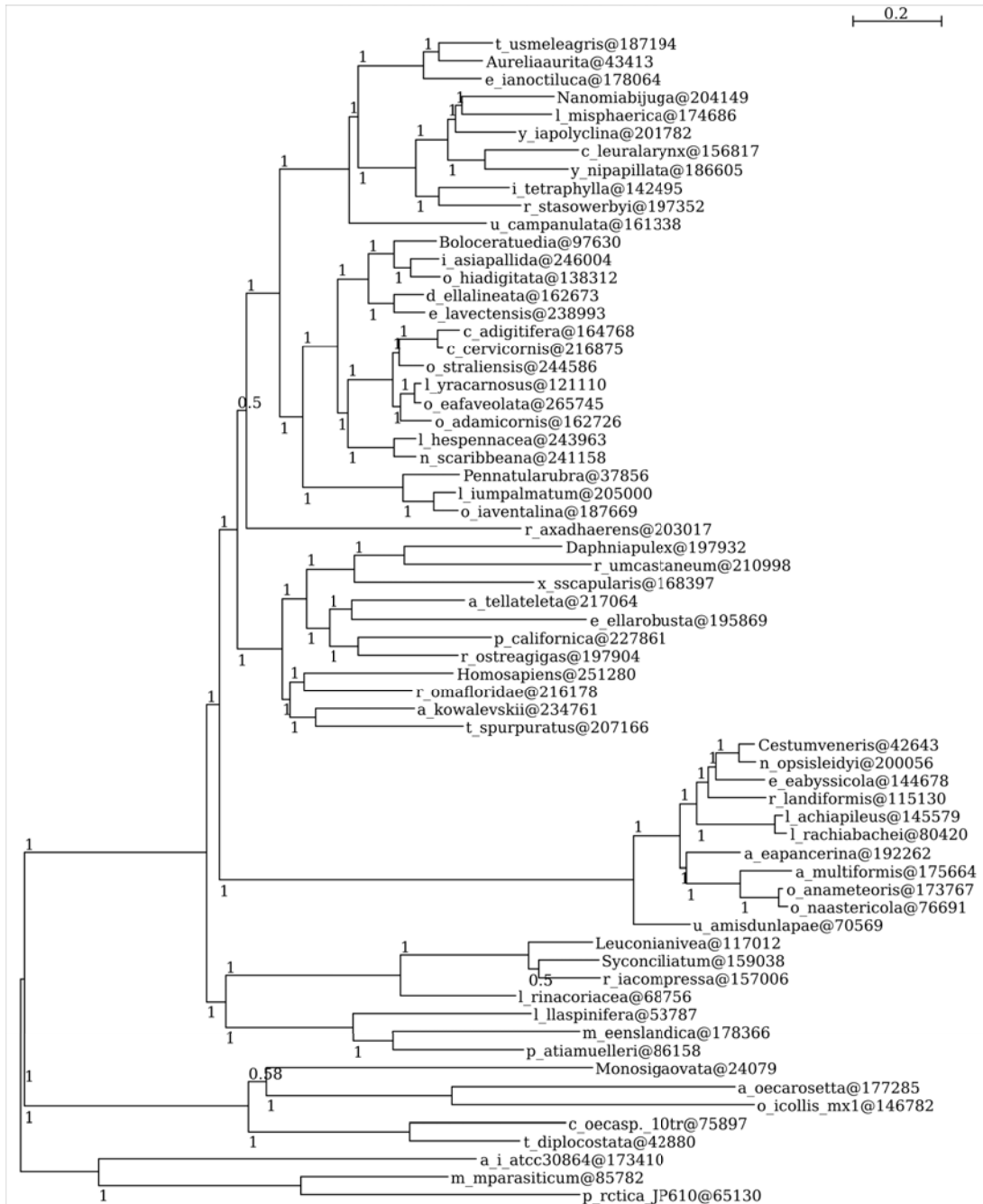


Figure 16: AUTO-1143_65_PB50_Bin19

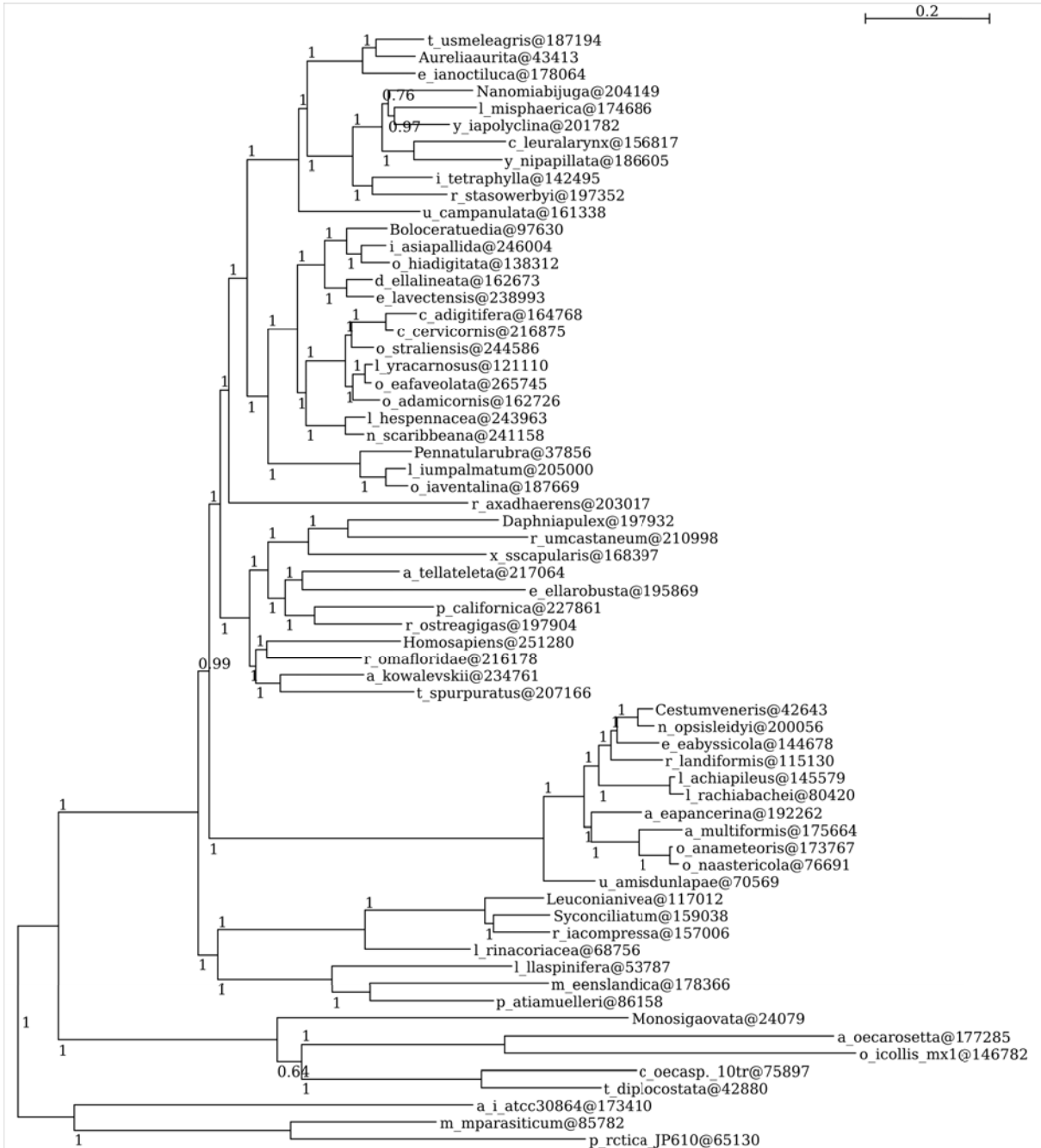


Figure 17: AUTO-1143.65_PB50_Bin18

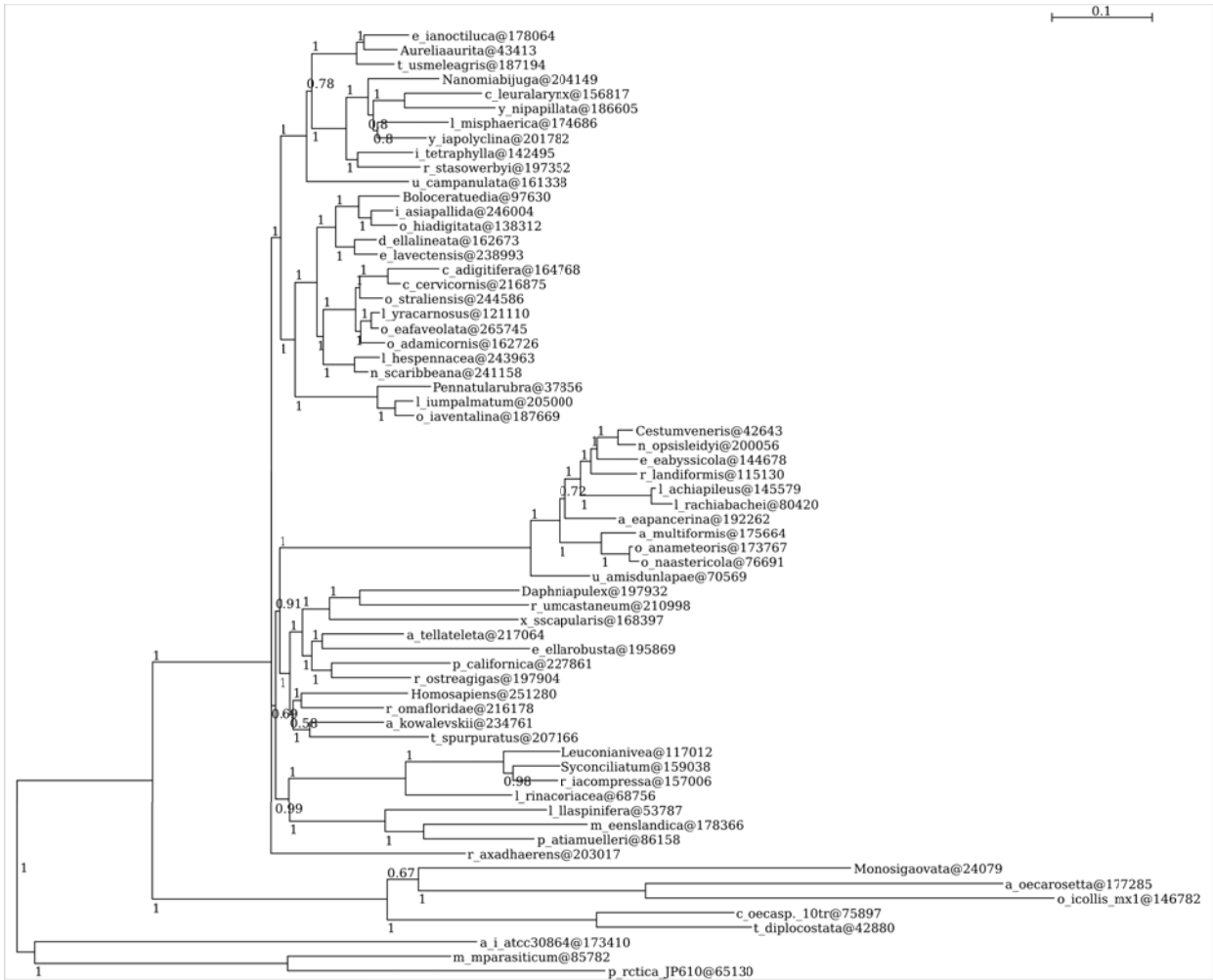


Figure 18: AUTO-1143.65_PB50_Bin17

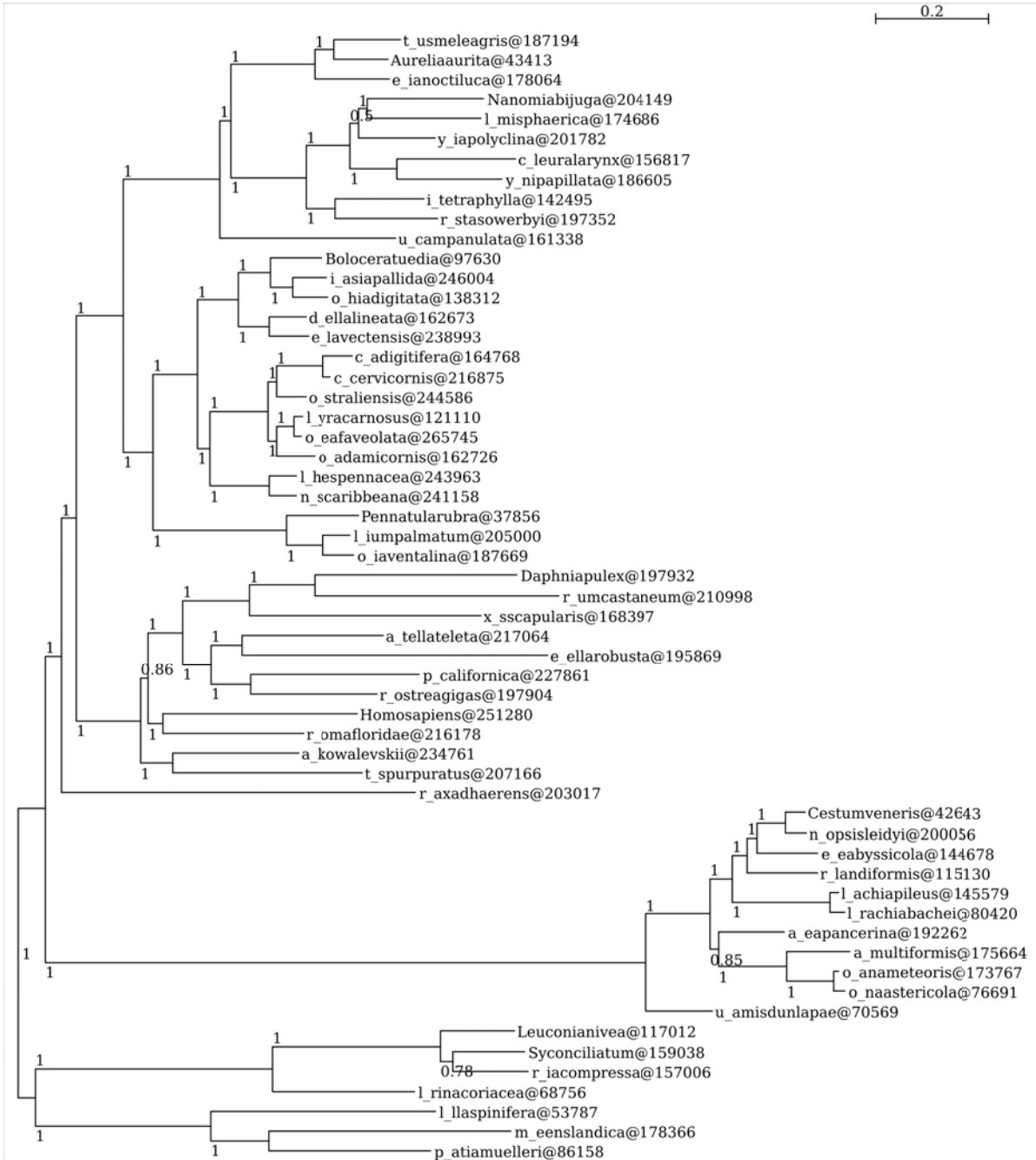


Figure 19: AUTO-1143_65_PB50_Bin19 with no outgroups

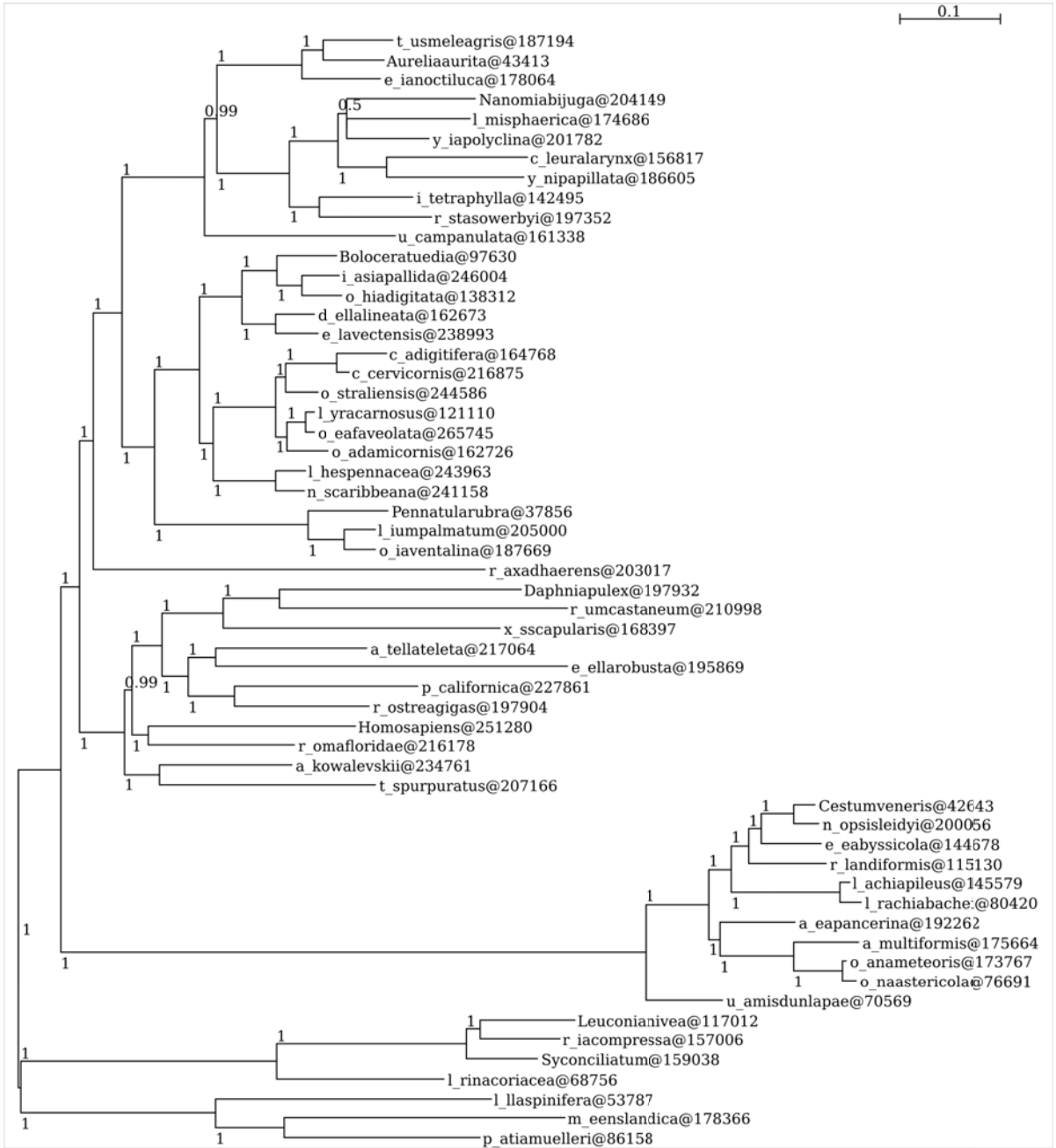


Figure 20: AUTO-1143_65_PB50_Bin18 with no outgroups

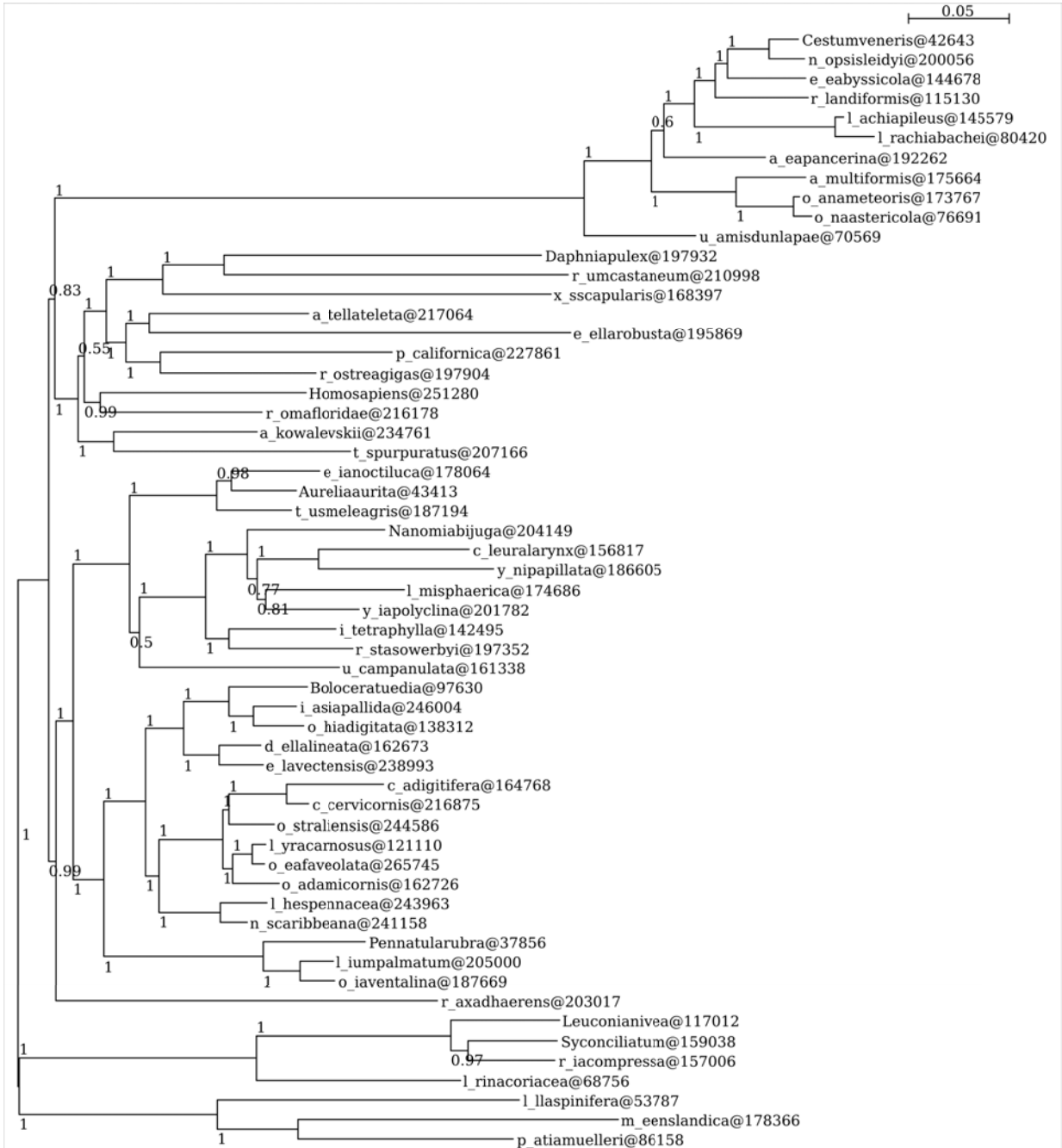


Figure 21: AUTO-1143_65_PB50_Bin17 with no outgroups

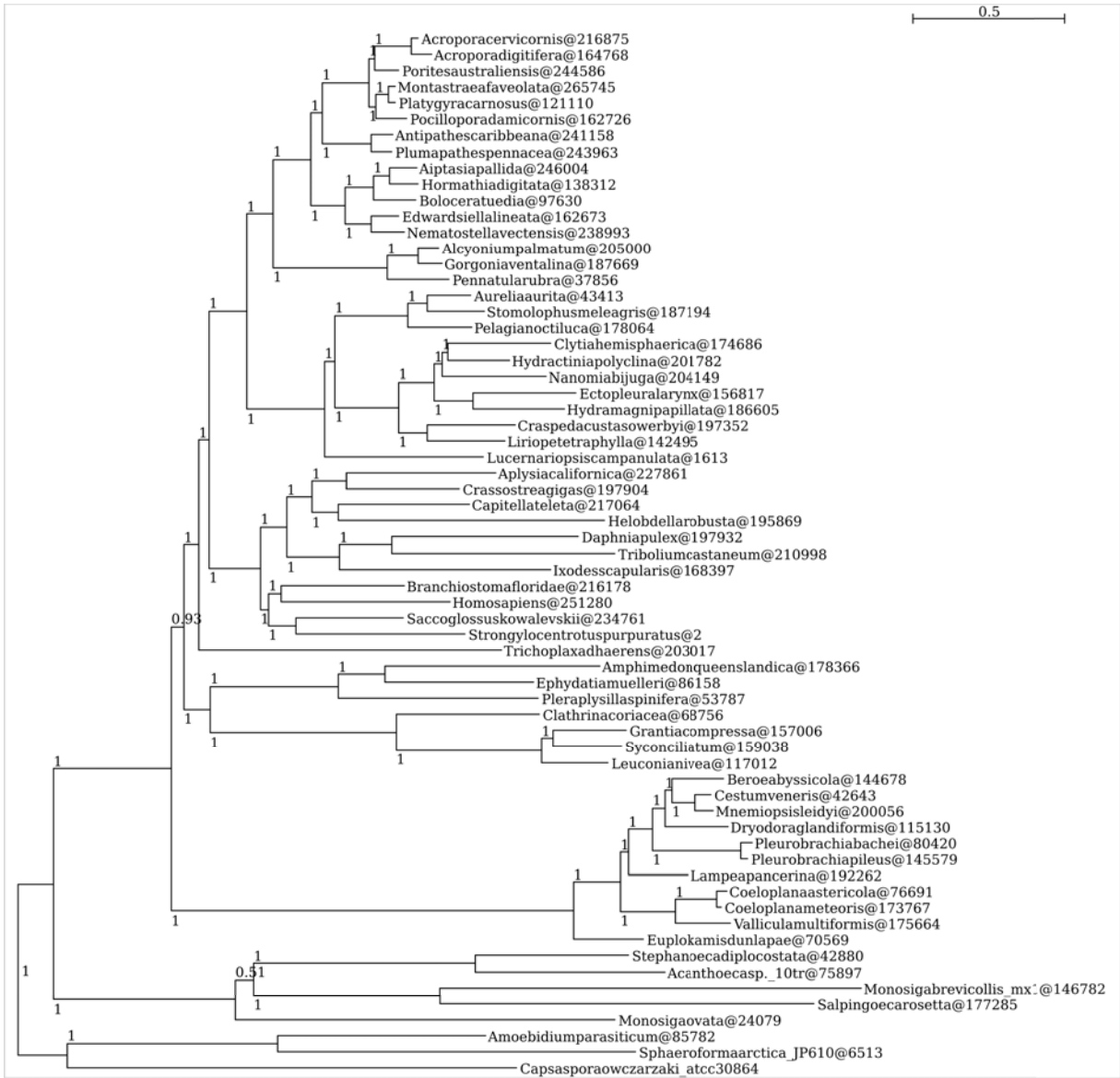


Figure 22: AUTO-1143.65_PB50 with CAT+GTR+ Γ evolution model

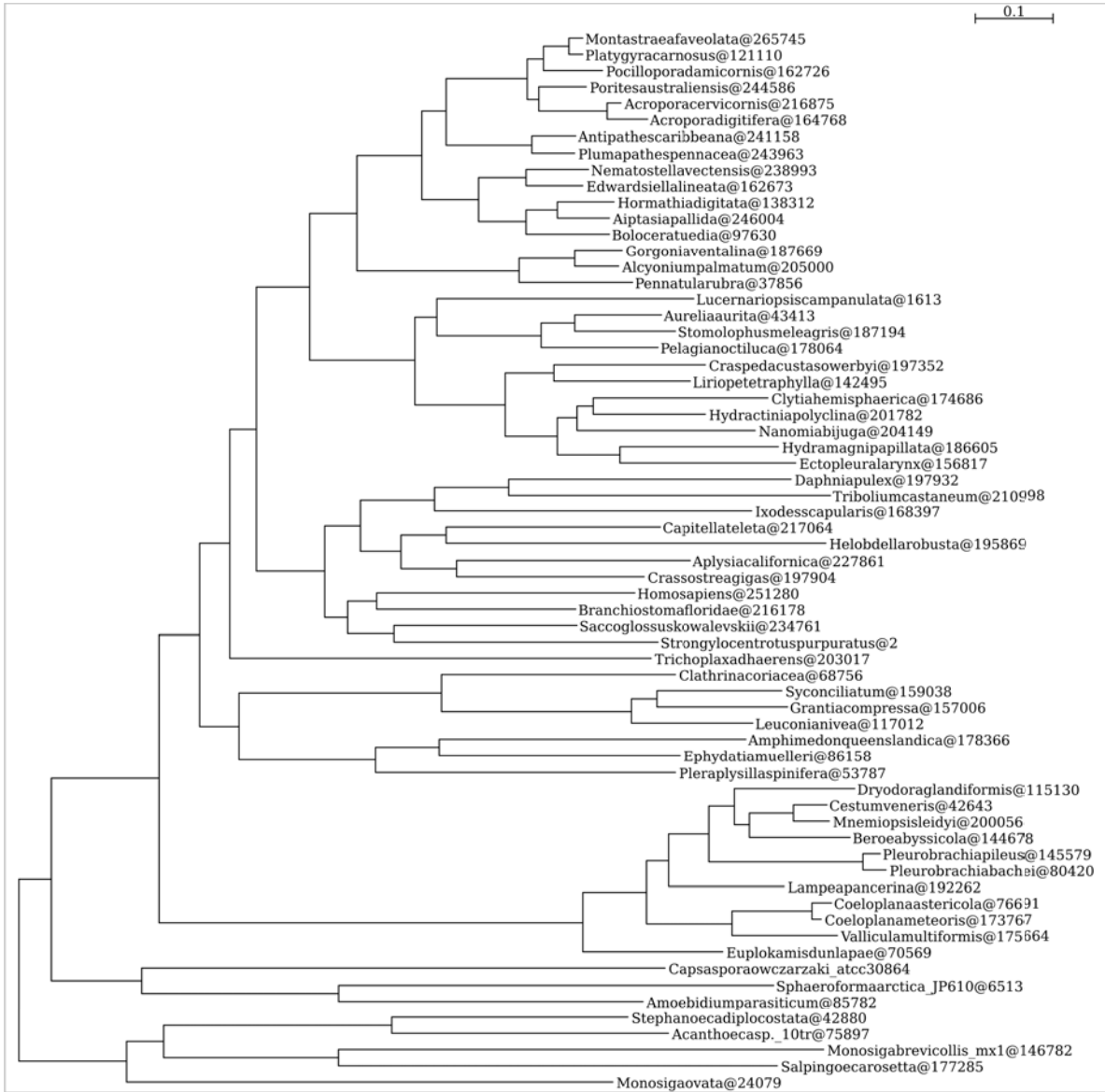


Figure 23: AUTO-1143_65_PB50 with LG+ Γ evolution model

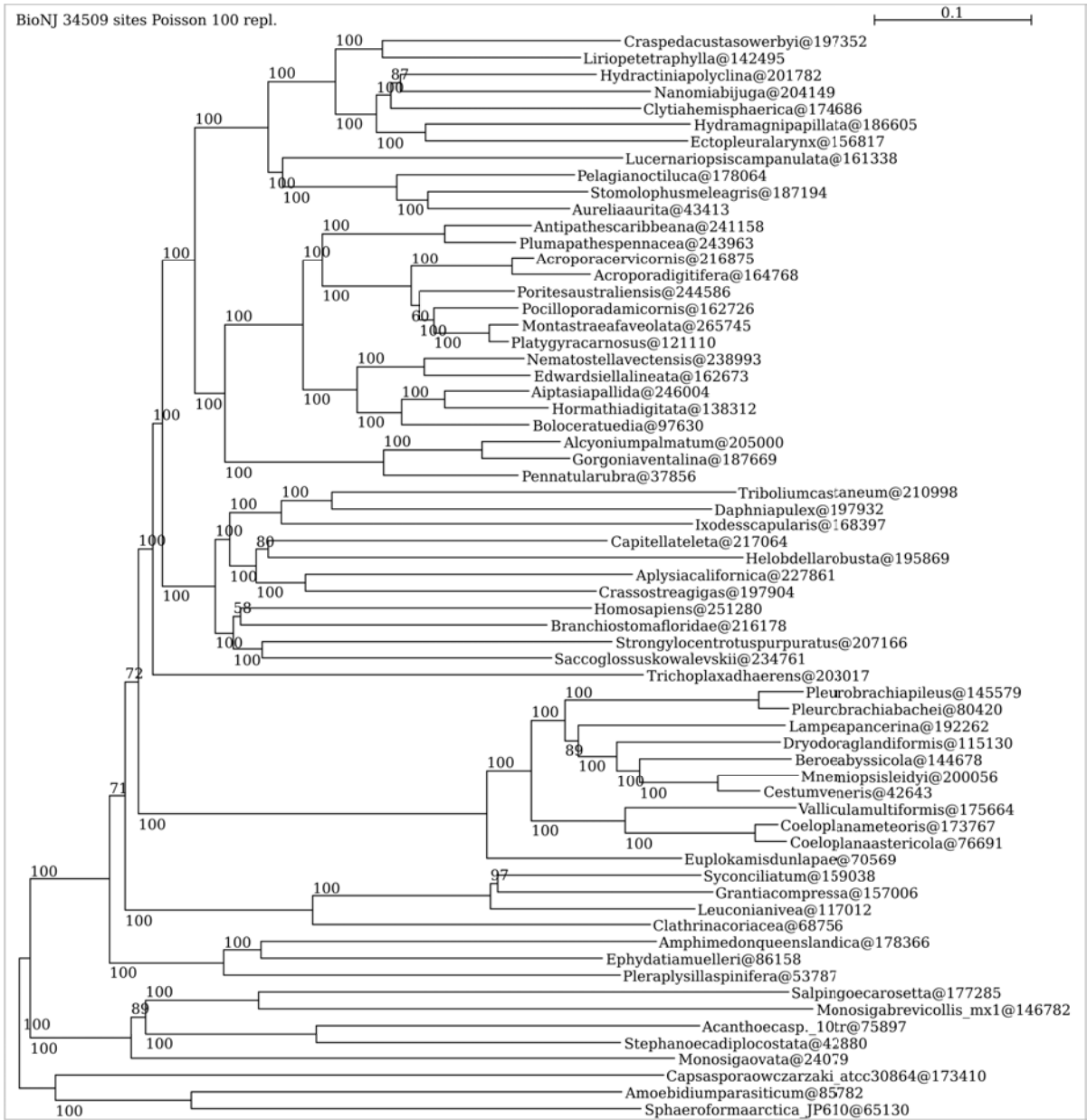


Figure 24: AUTO-1143.65_PB50 with BioNJ using a poisson distribution

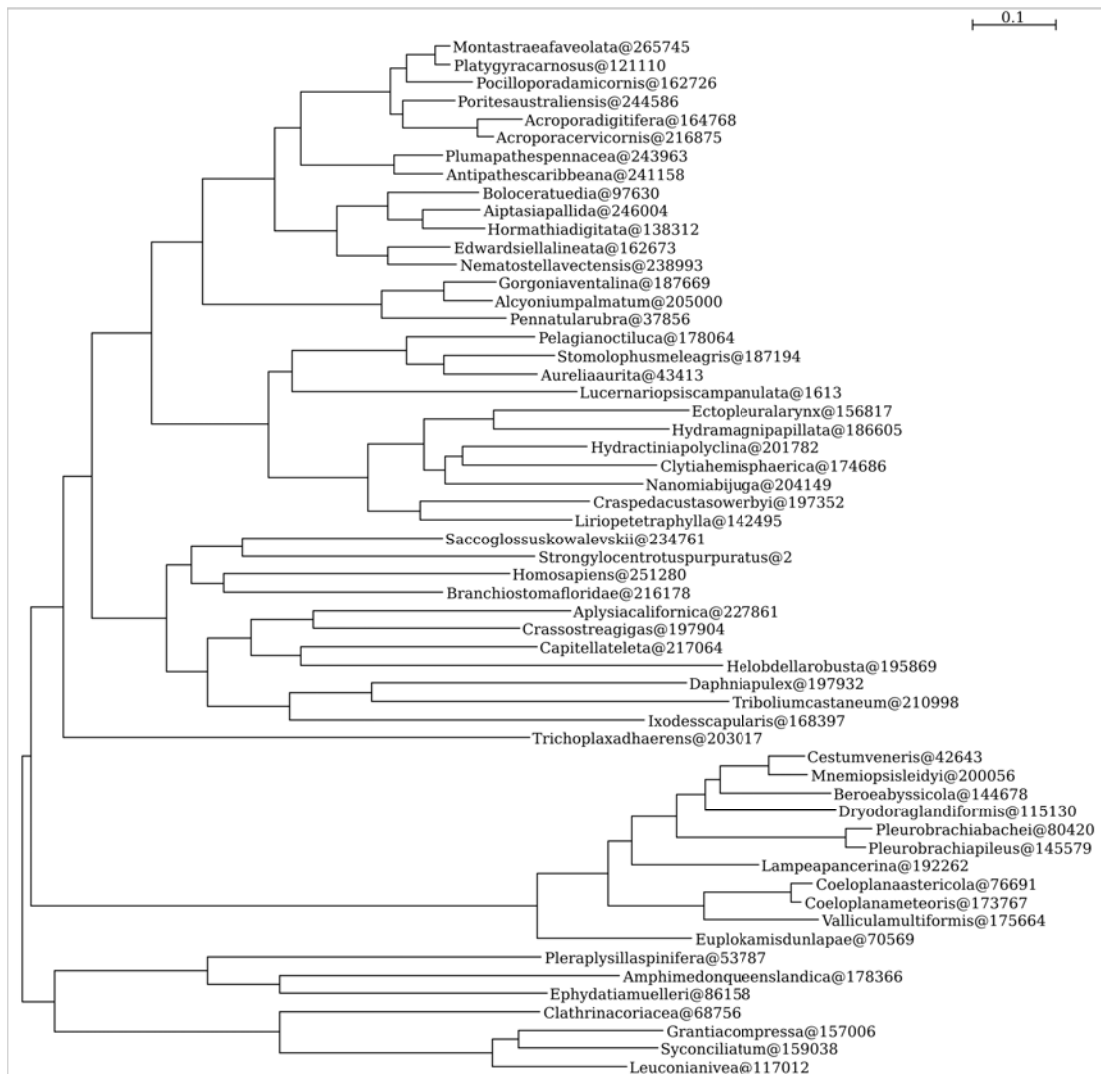


Figure 25: AUTO-1143.65_PB50 with LG+Γ evolution model and no outgroups

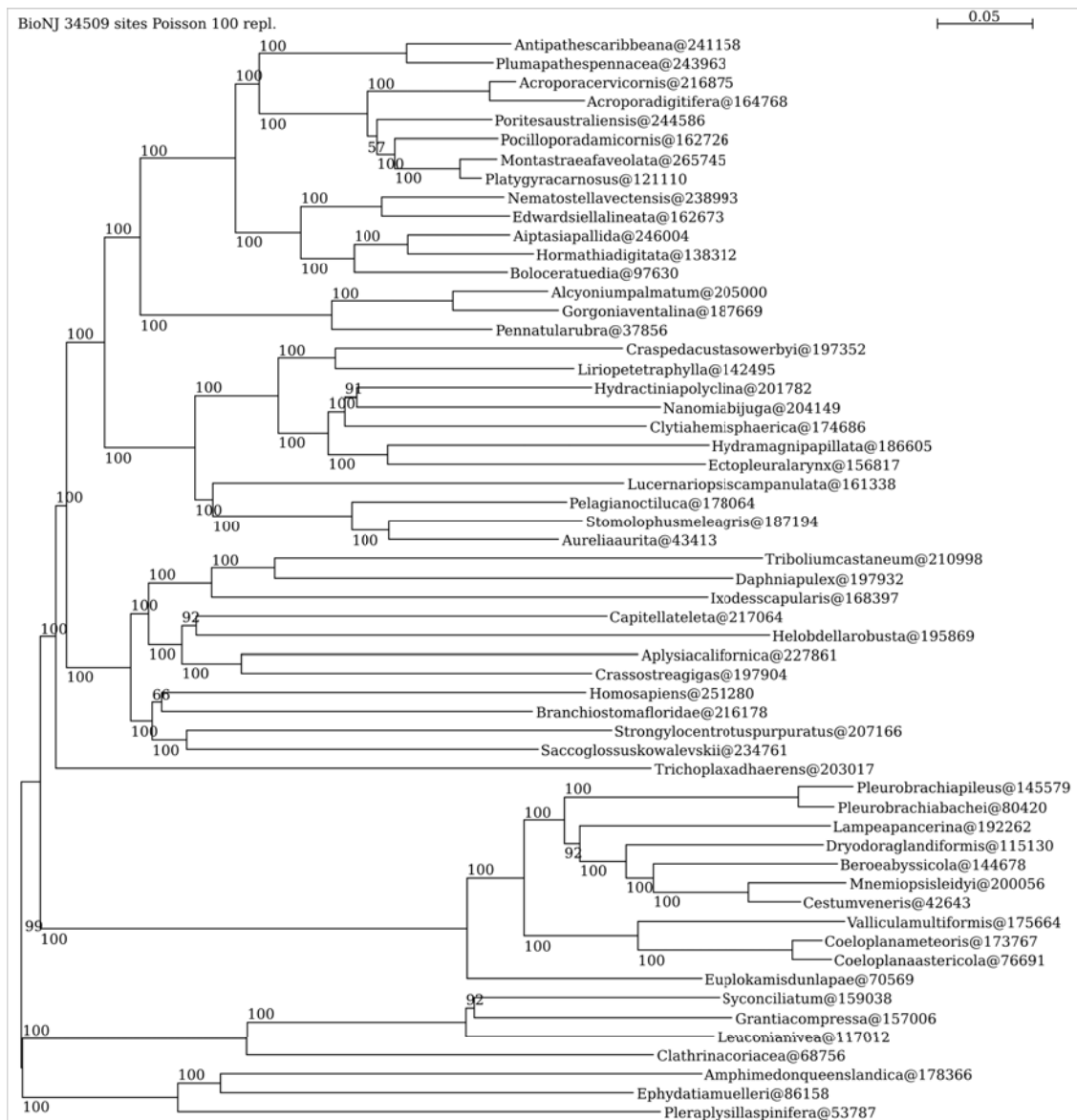


Figure 26: AUTO-1143_65_PB50 with BioNJ using a poisson distribution and no outgroups

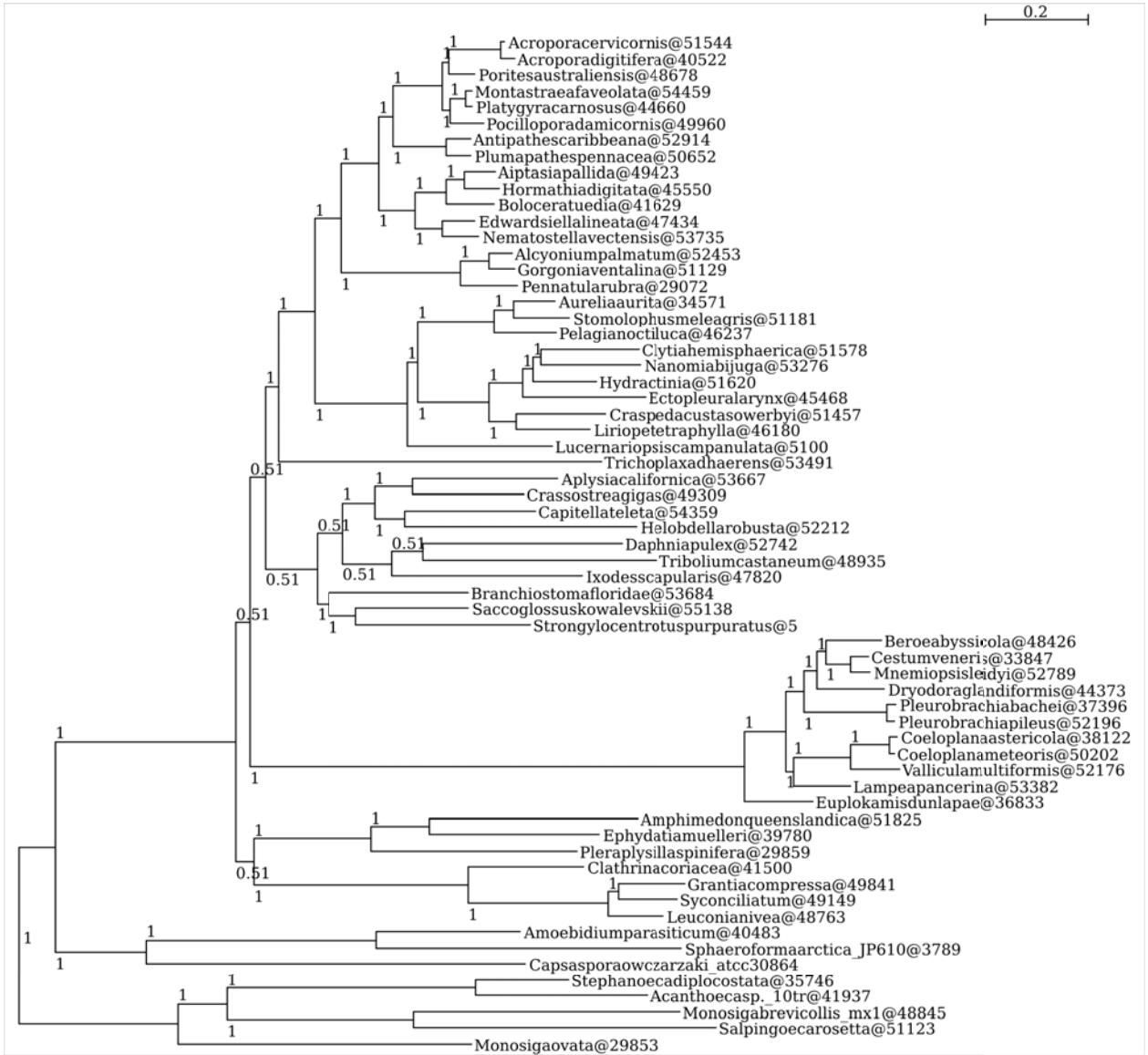


Figure 27: 65_PB50+UPCB_116

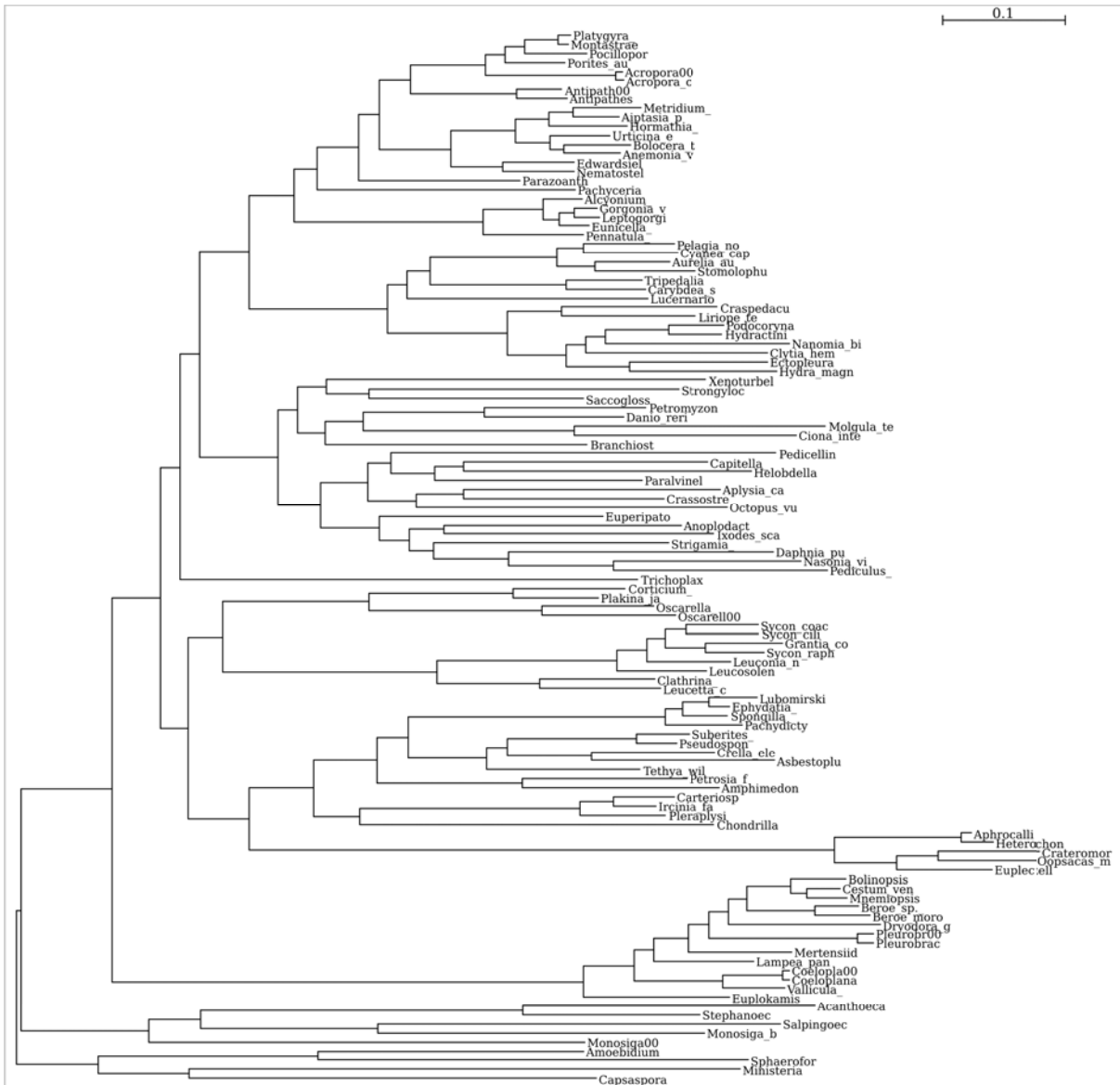


Figure 28: UPCB-116 with LG+Γ

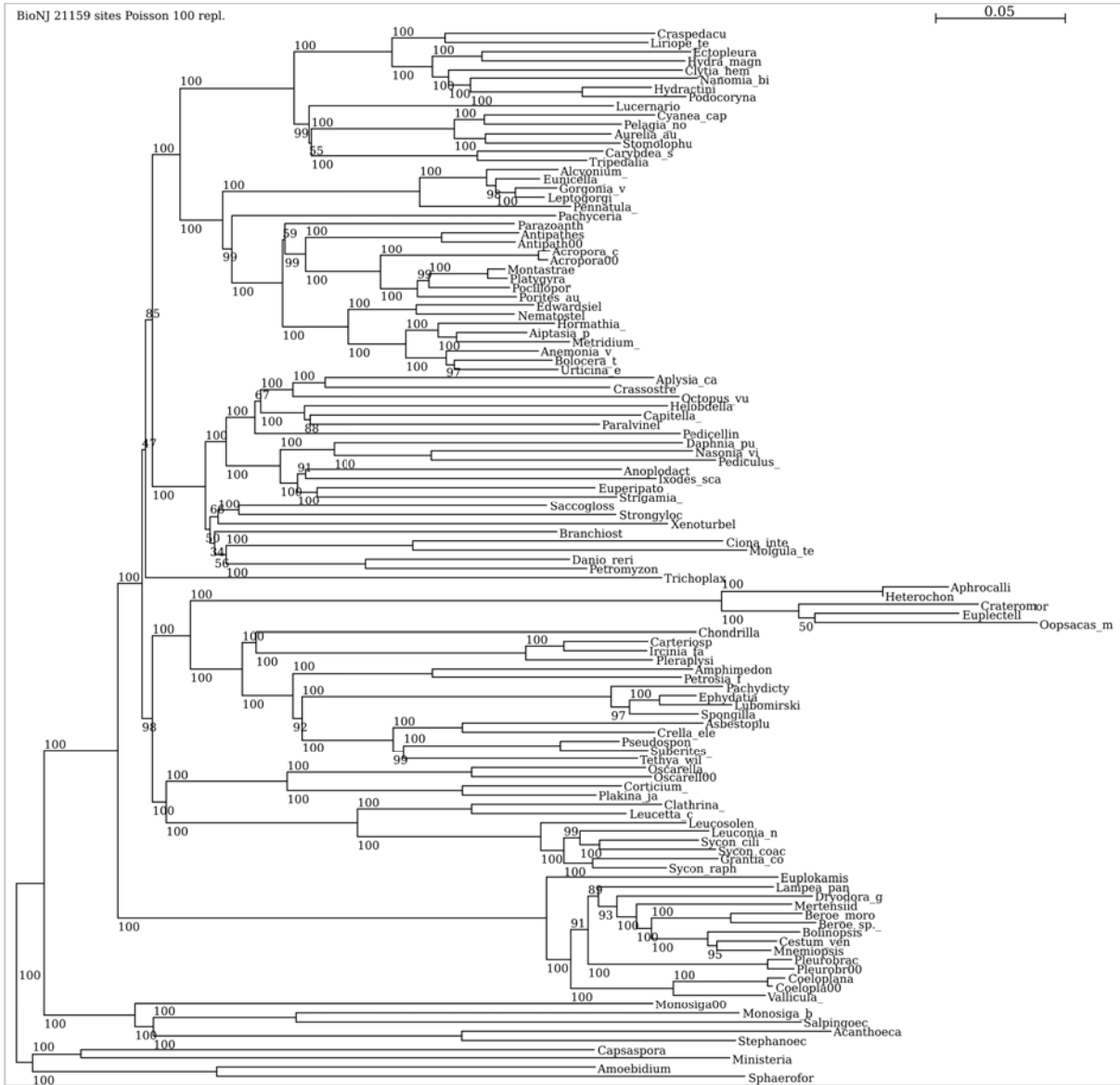


Figure 29: UPCB_116 with BioNJ using a poisson distribution

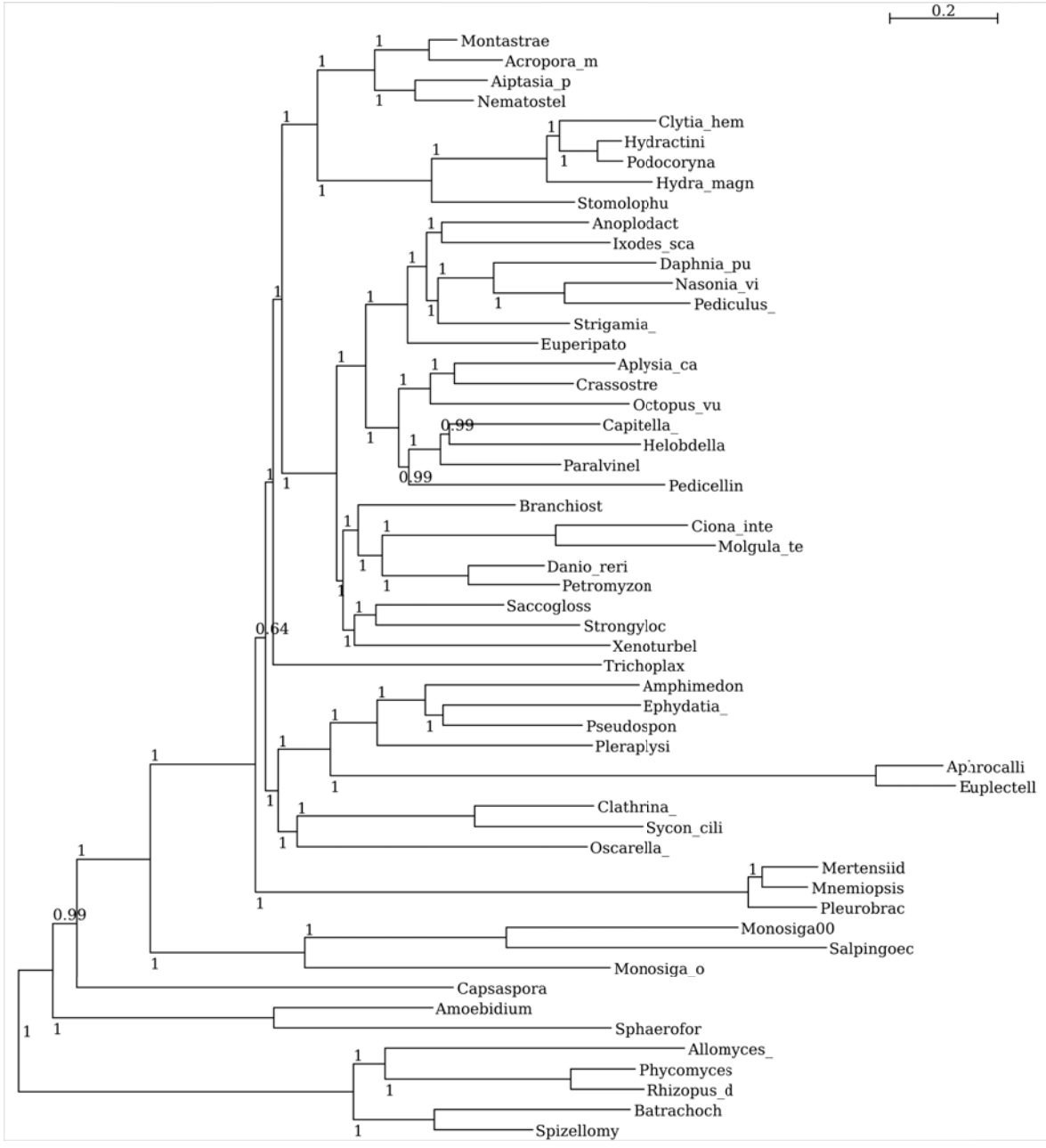


Figure 30: UPCB_55 with CAT+GTR+ Γ evolution model

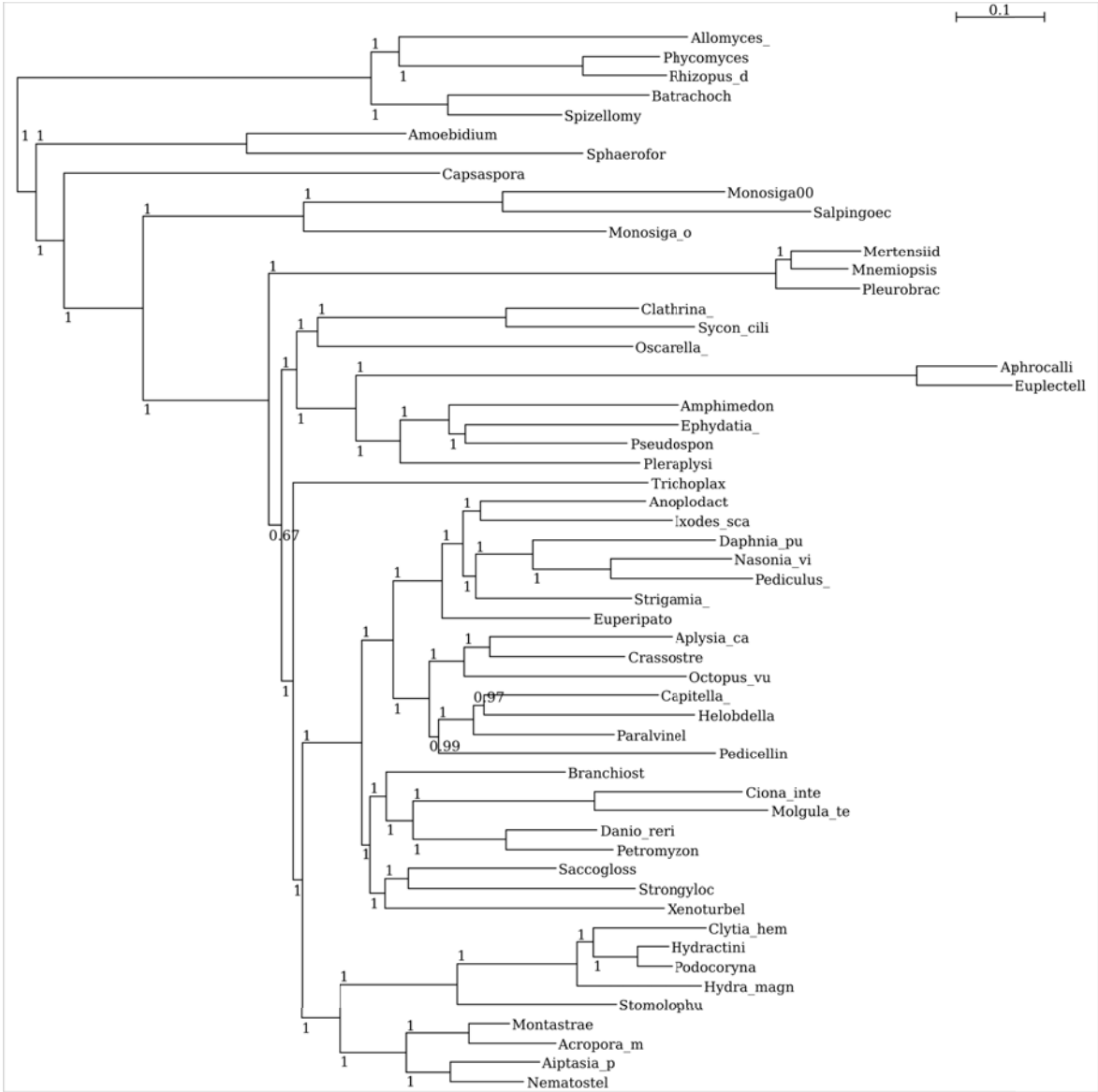


Figure 31: UPCB_55

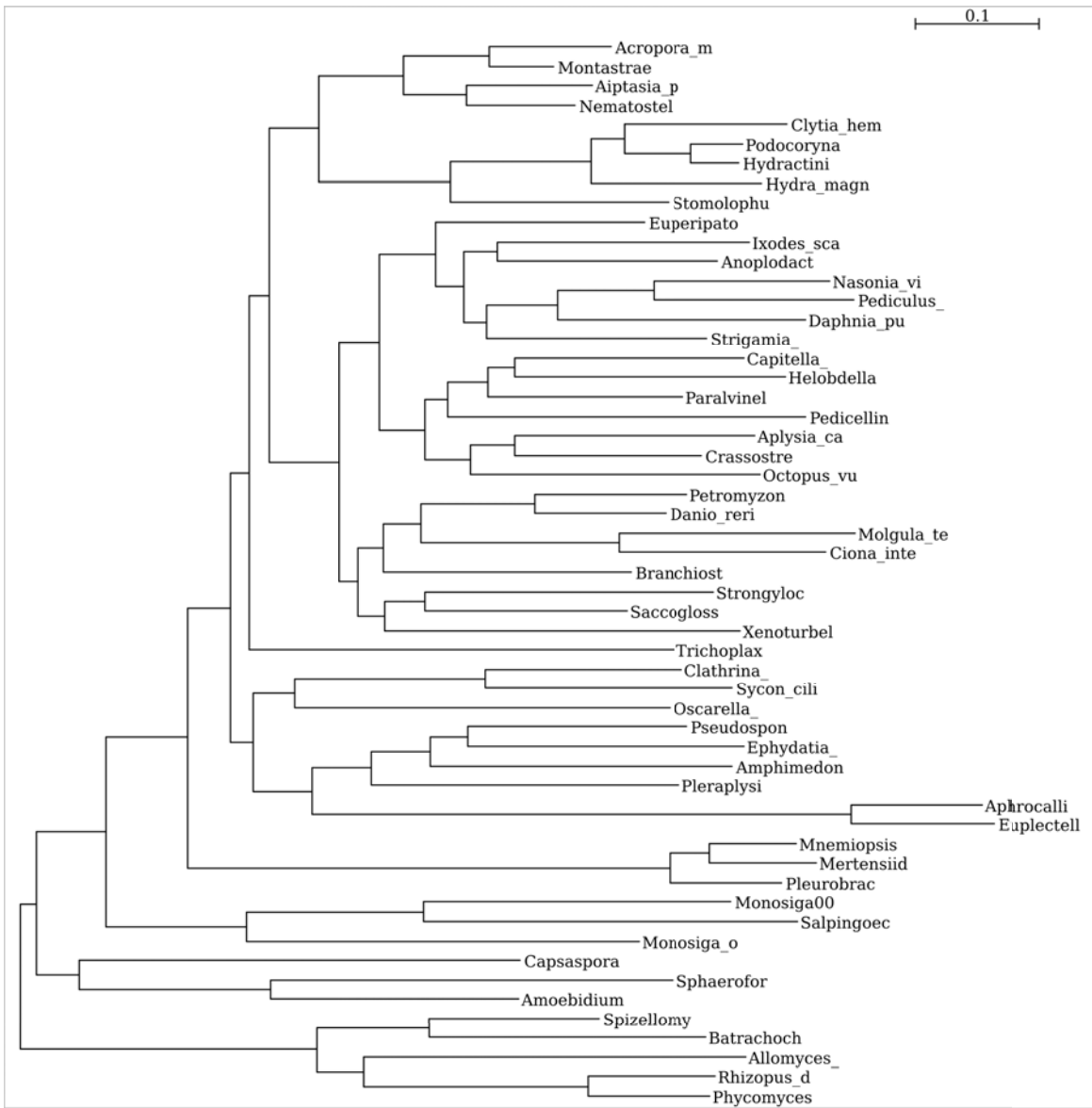


Figure 32: UPCB_55 with LG+Γ

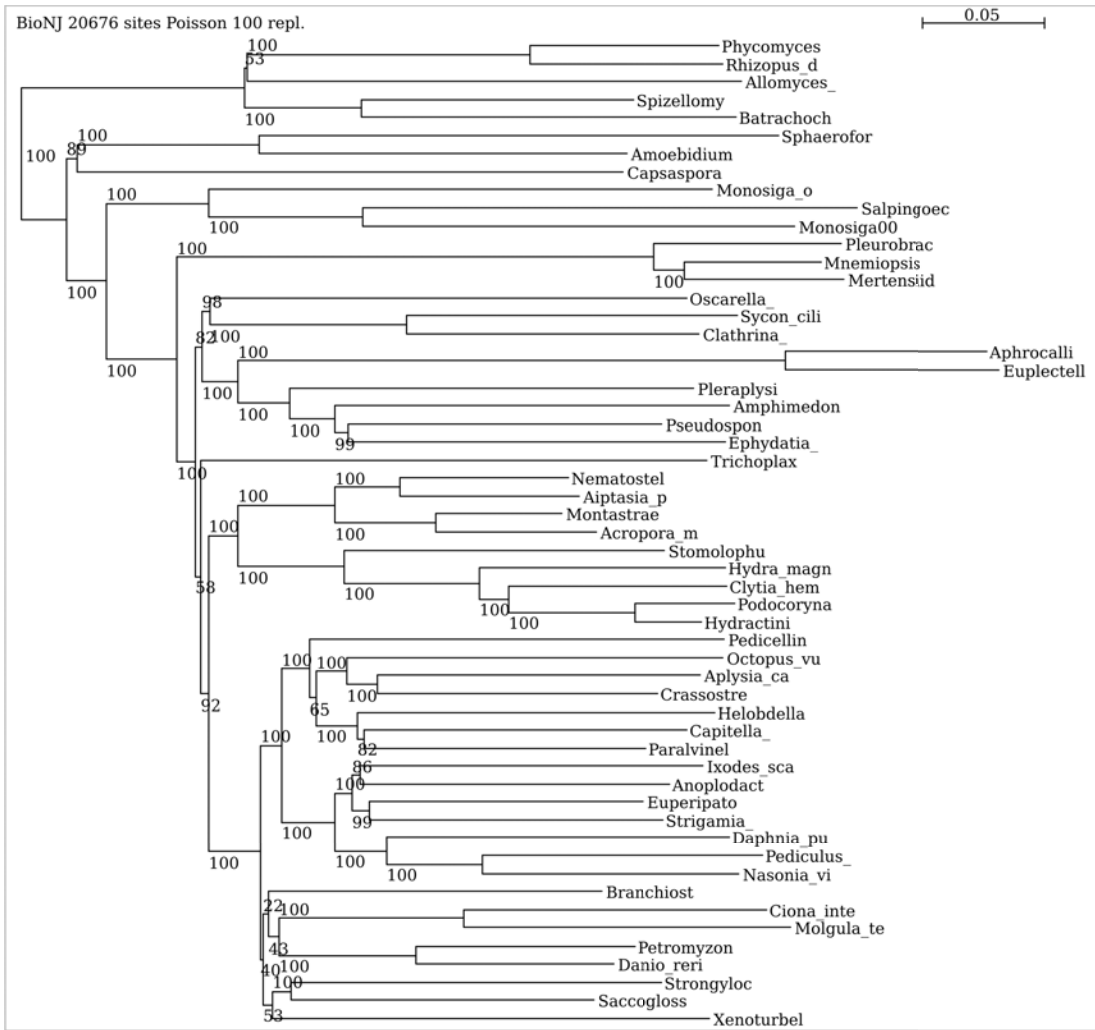


Figure 33: UPCB.55 with BioNJ using a poisson distribution

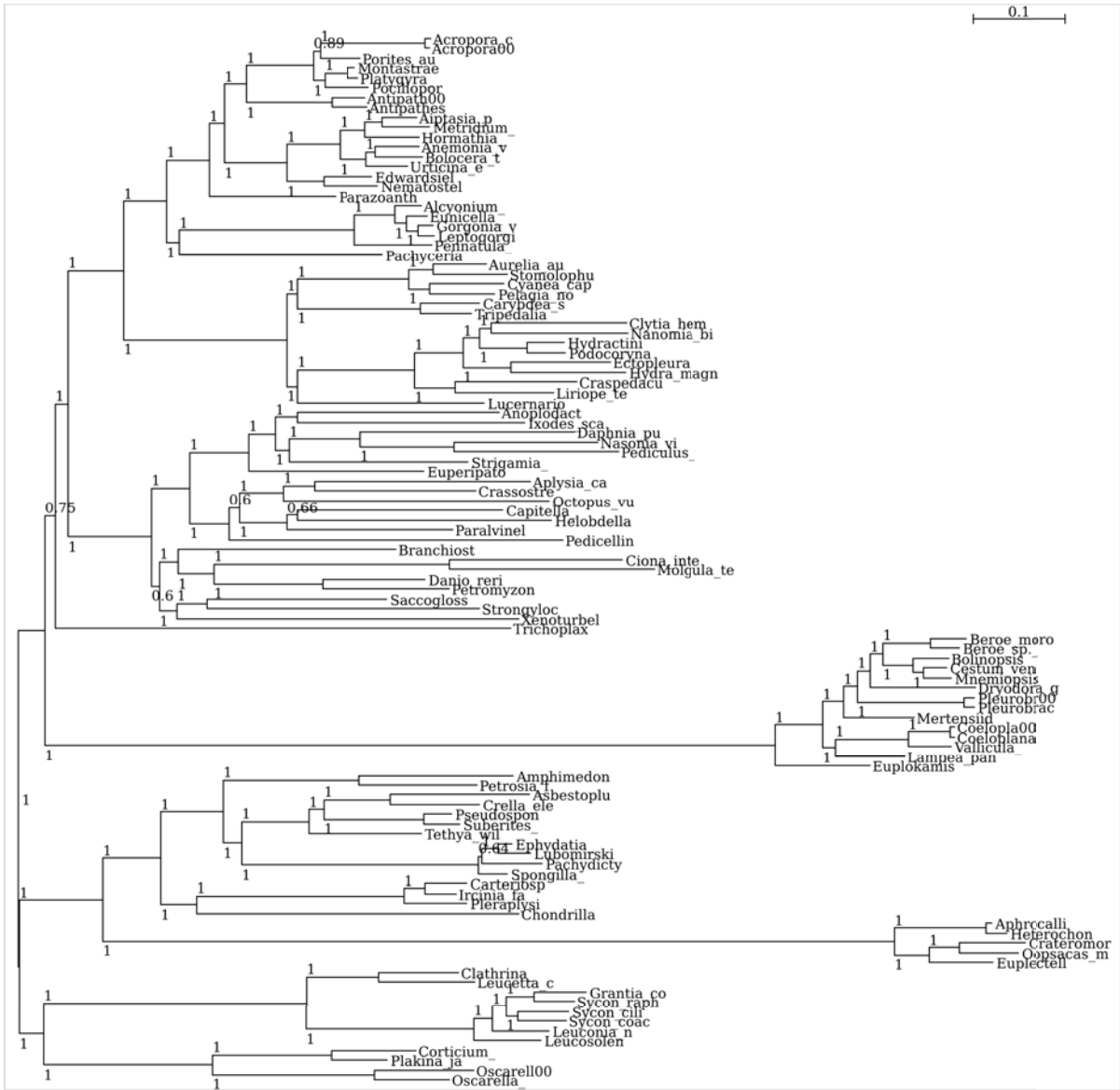


Figure 34: UPCB_116 with no outgroups

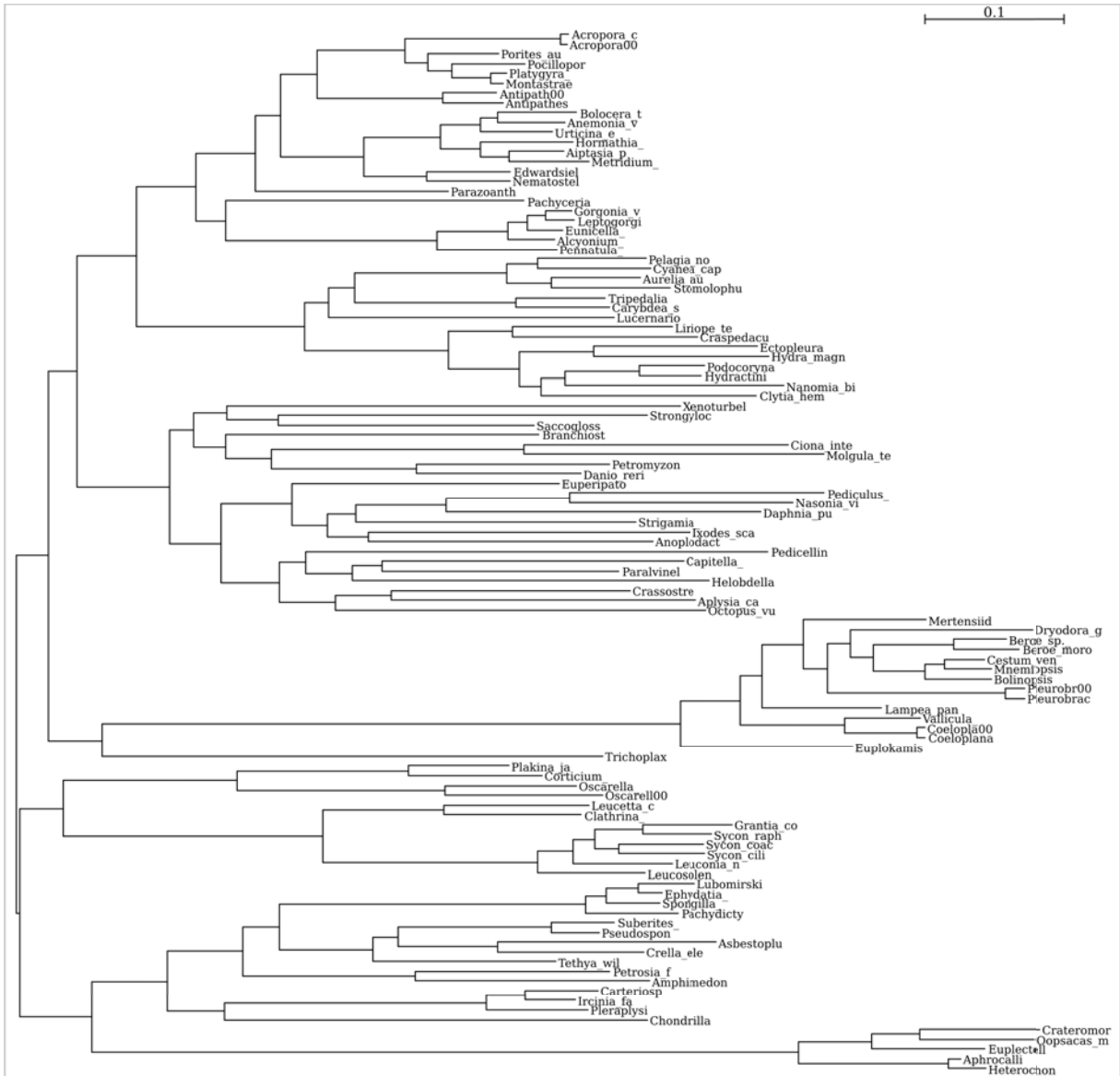


Figure 35: UPCB_116 with LG+Γ and no outgroups

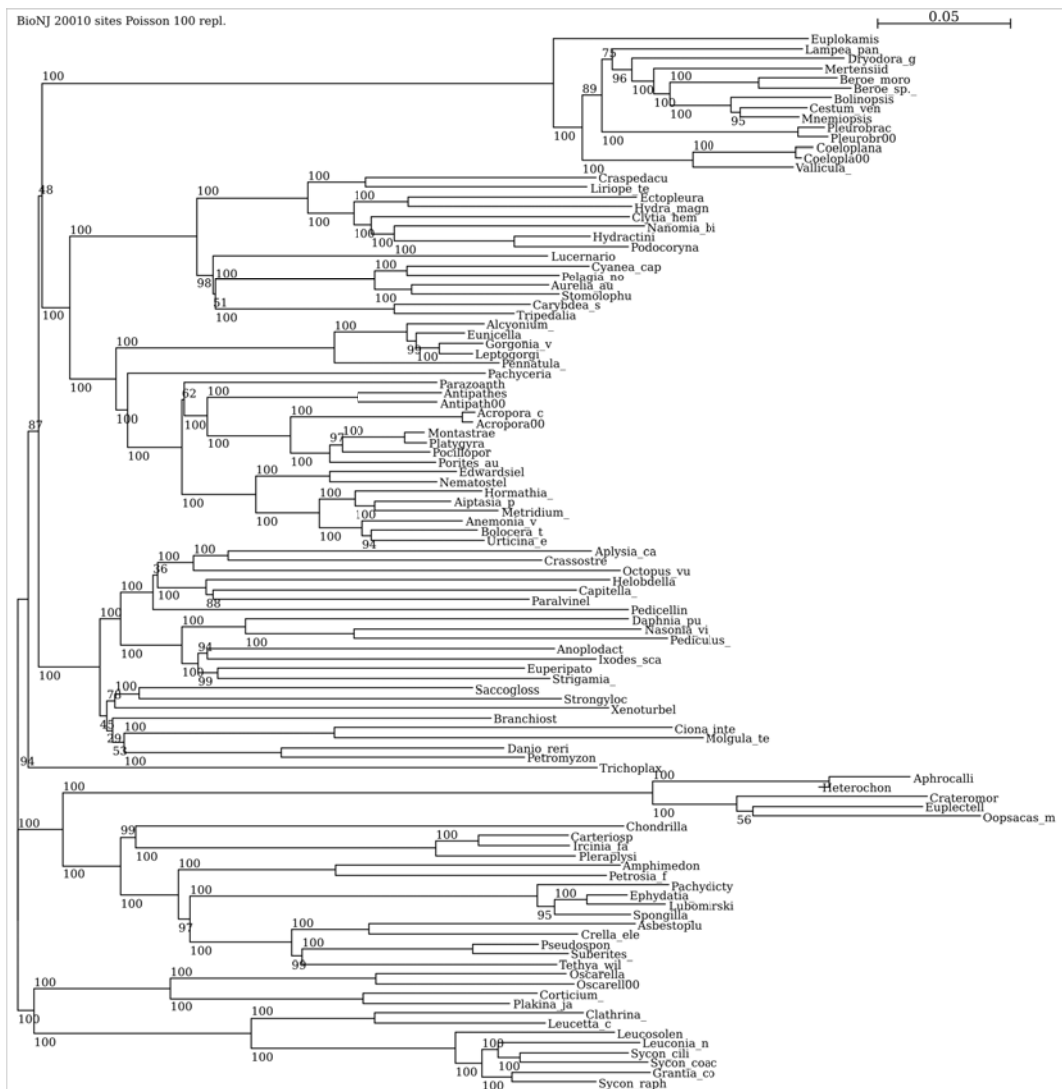


Figure 36: UPCB_116 with BioNJ using a poisson distribution and no outgroups

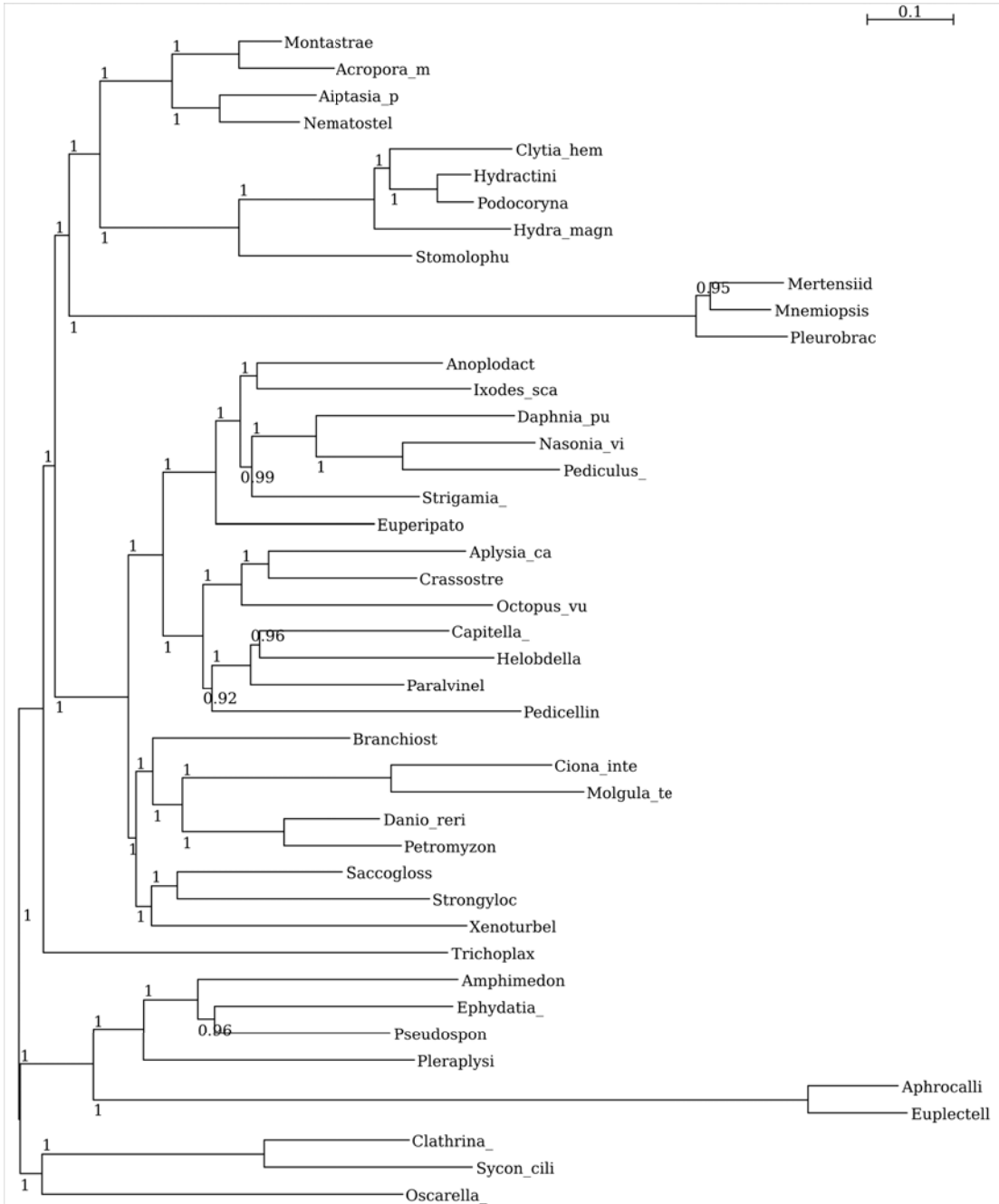


Figure 37: UPCB_55 with no outgroups

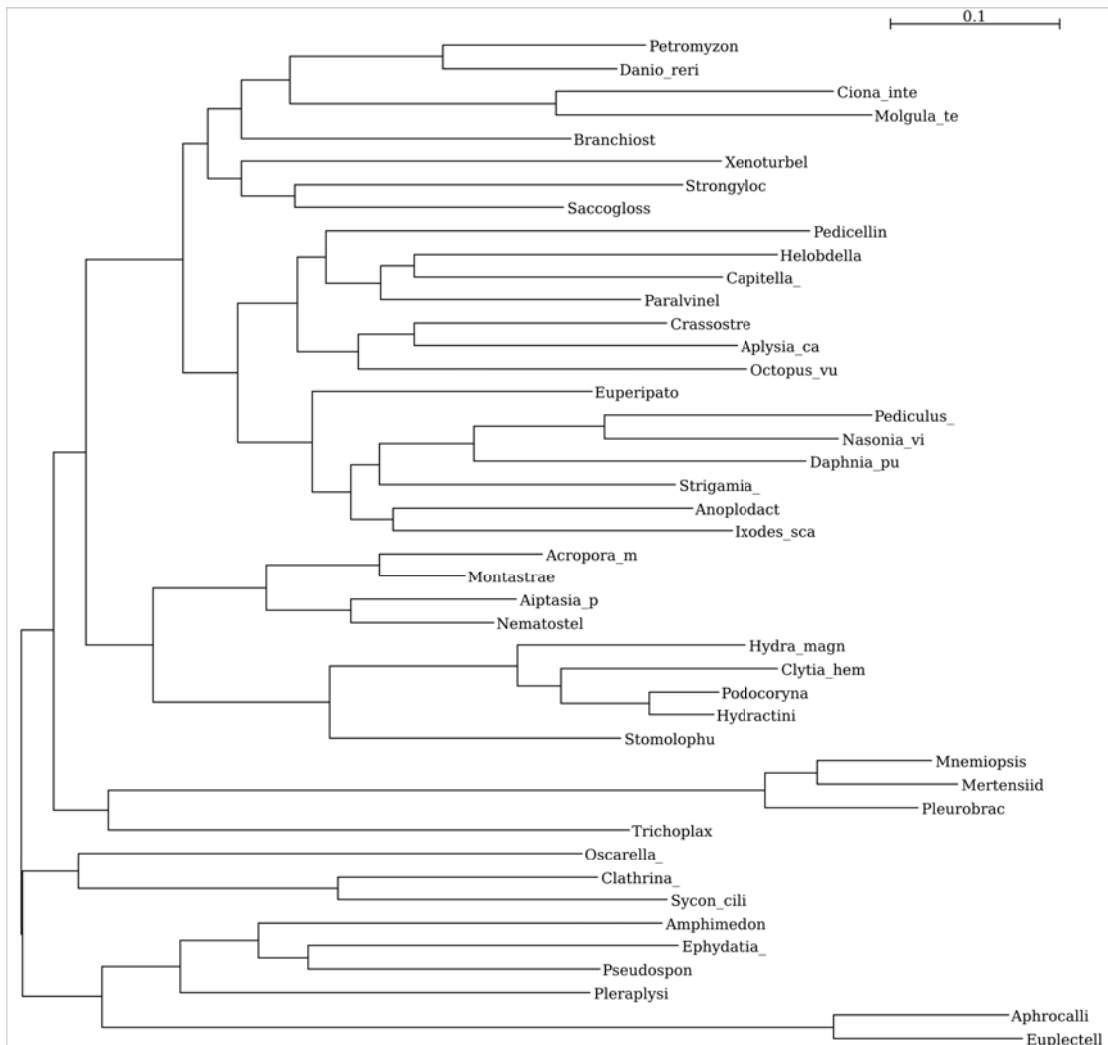


Figure 38: UPCB_55 with LG+Γ and no outgroups

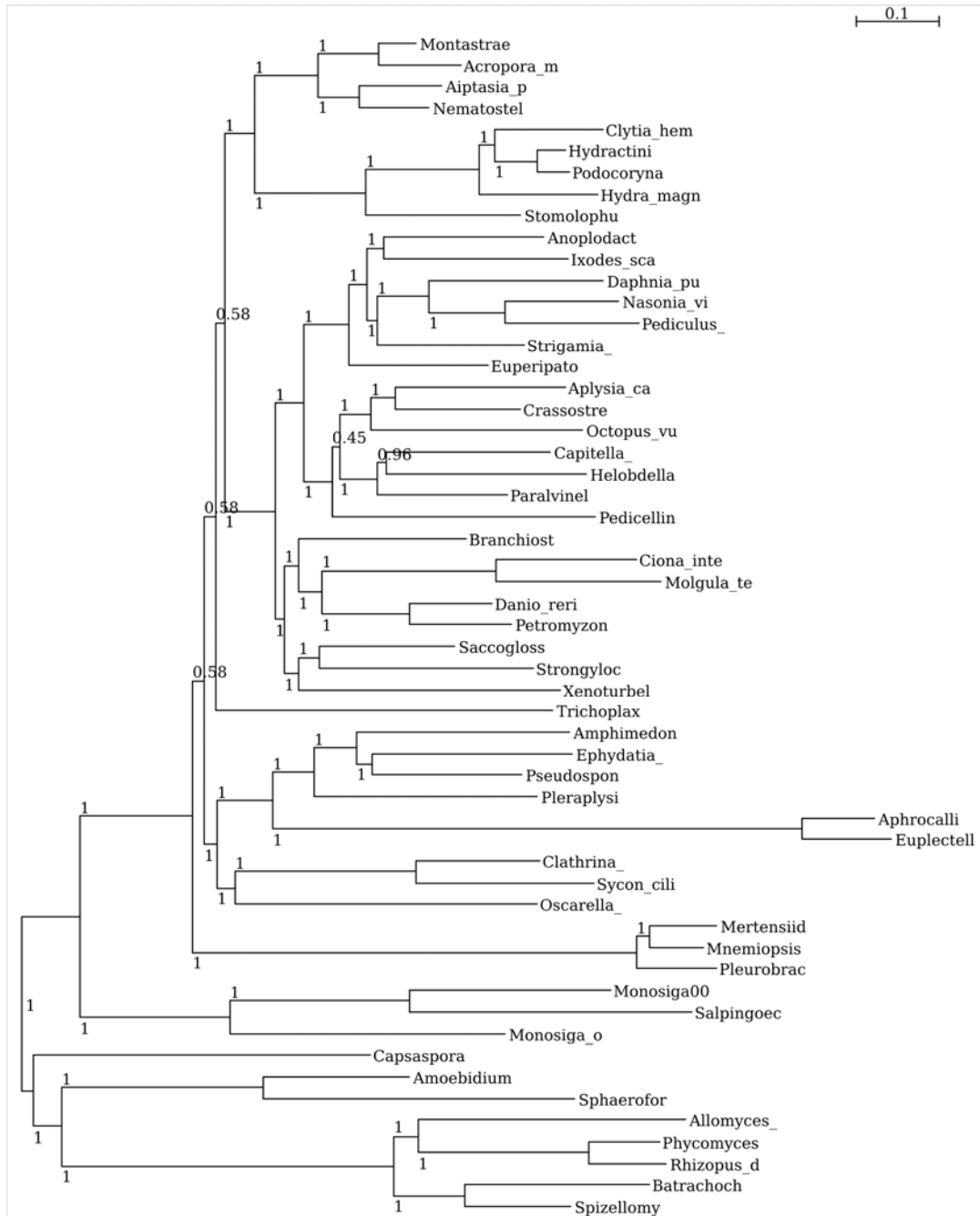


Figure 40: UPCB_55_PIP10

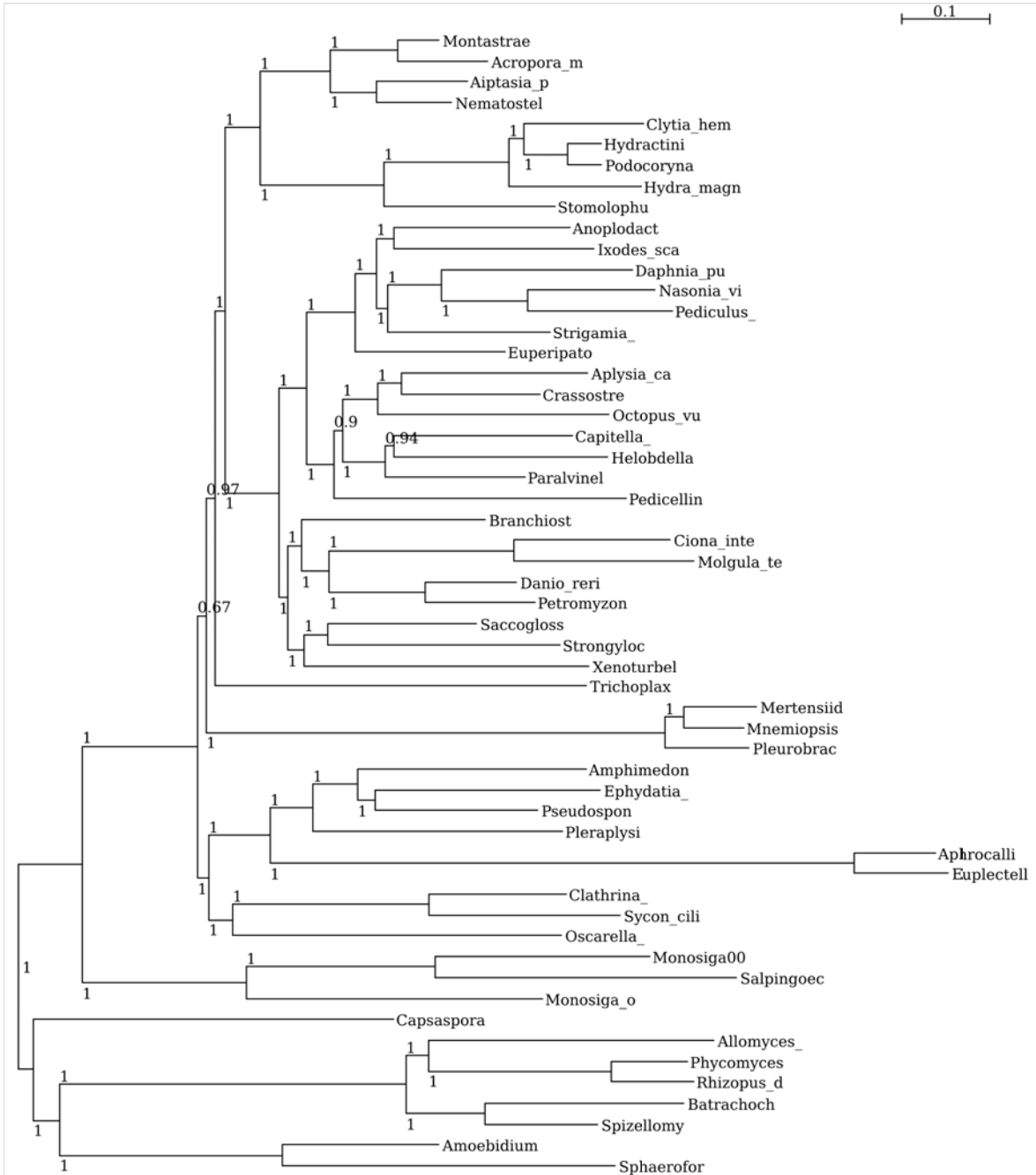


Figure 41: UPCB_55_PIP20

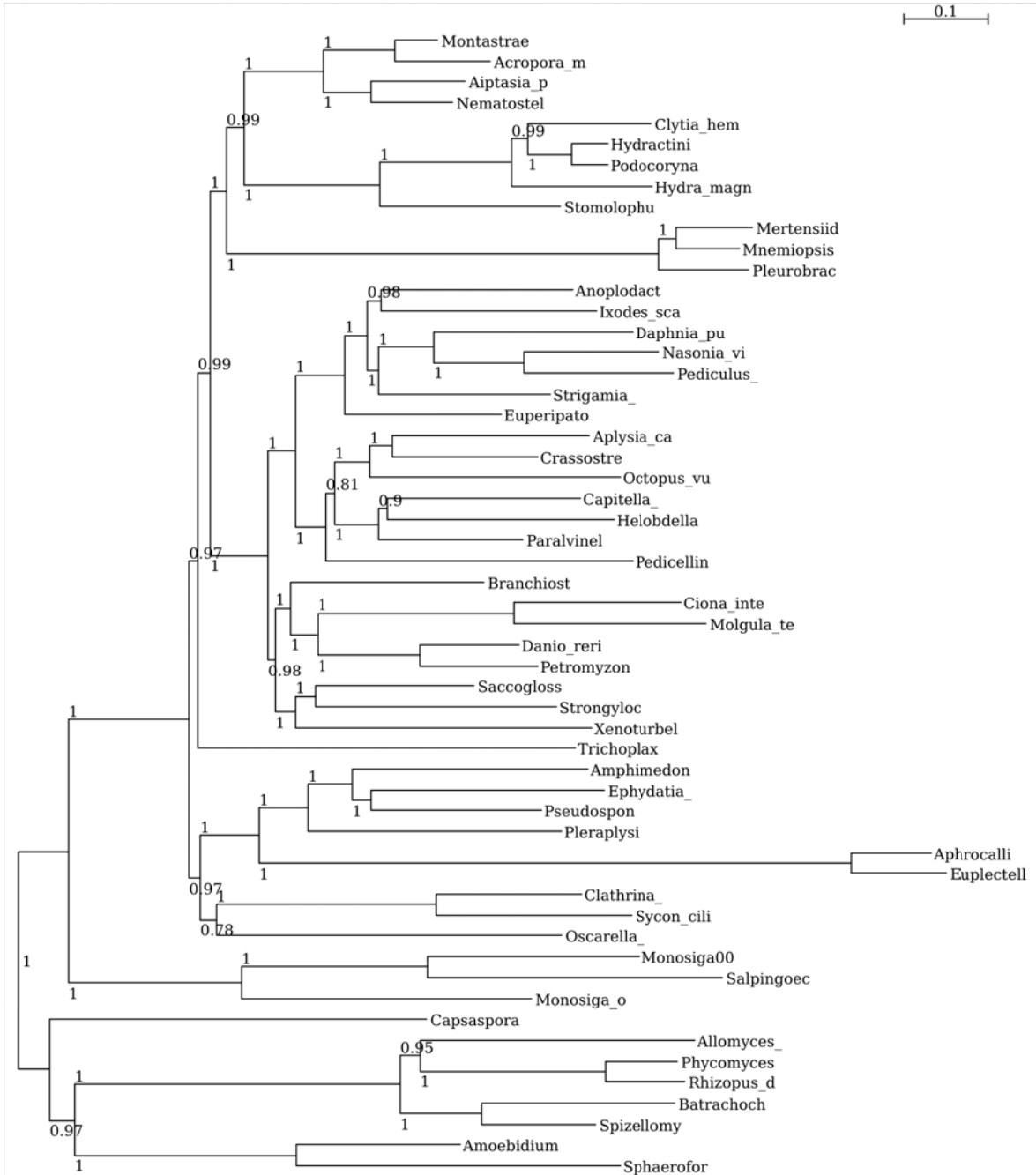


Figure 42: UPCB_55_PIP40

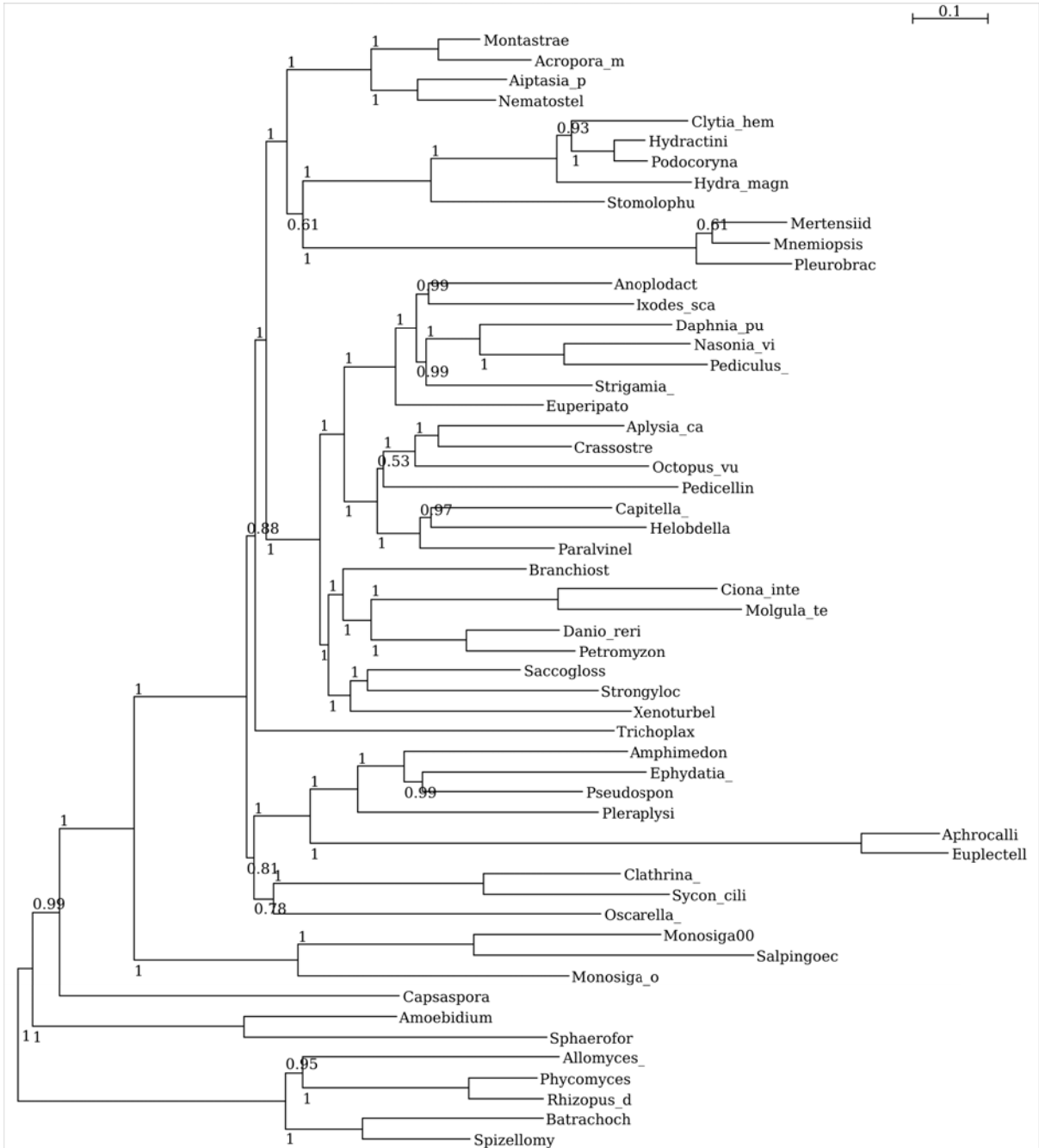


Figure 43: UPCB_55_PIP50

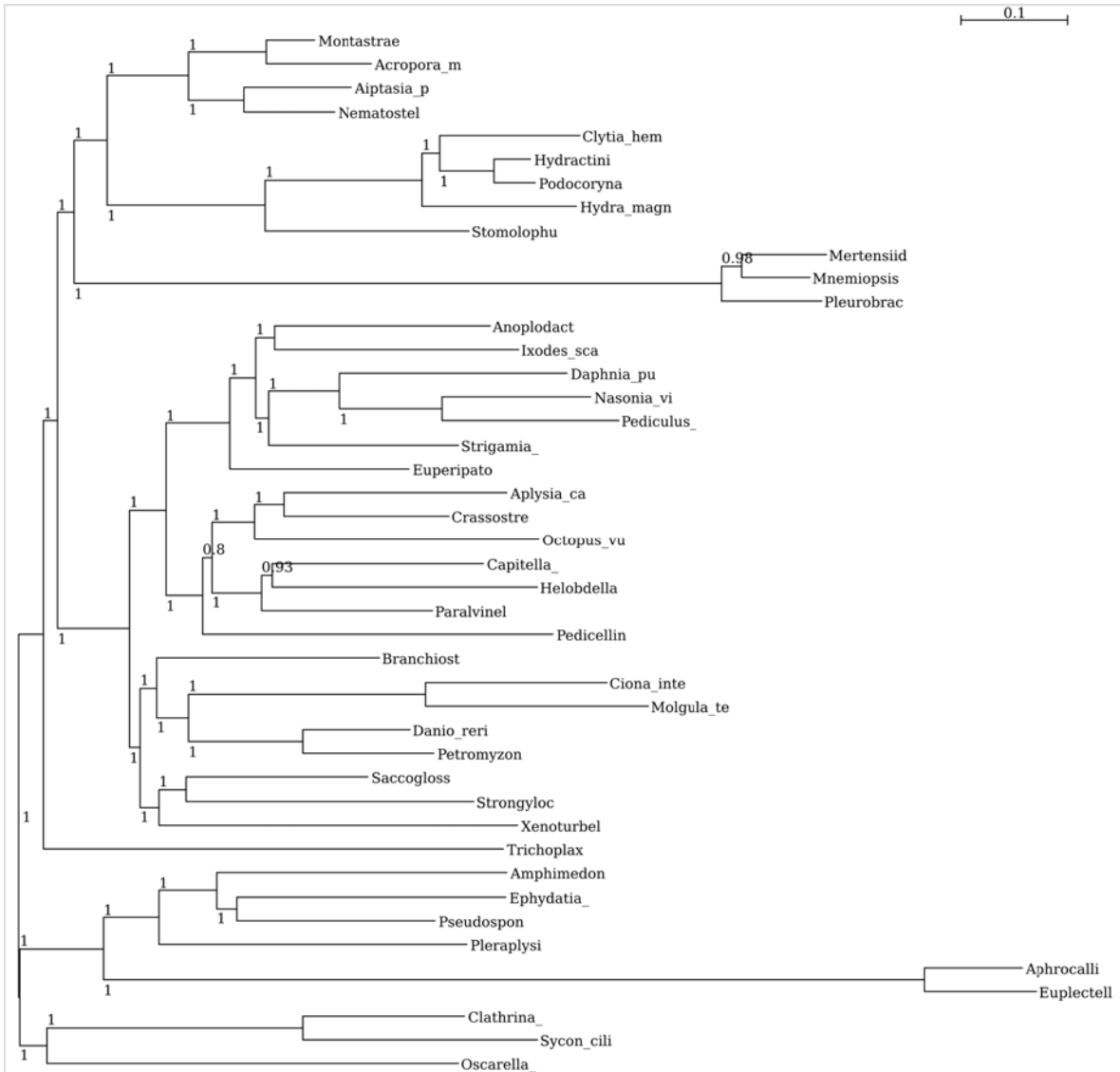


Figure 44: UPCB_55_PIP10 with no outgroups

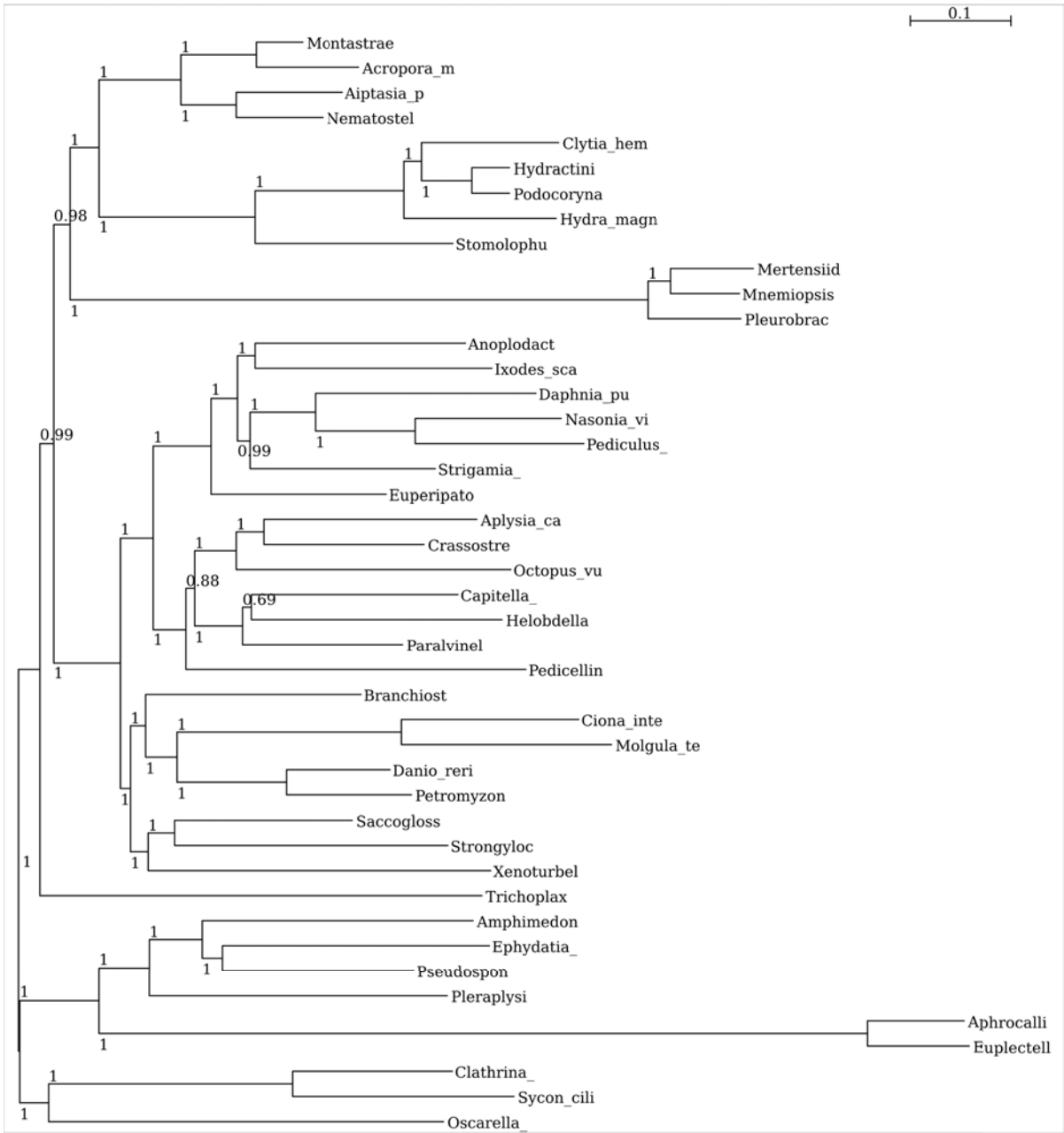


Figure 45: UPCB_55_PIP20 with no outgroups

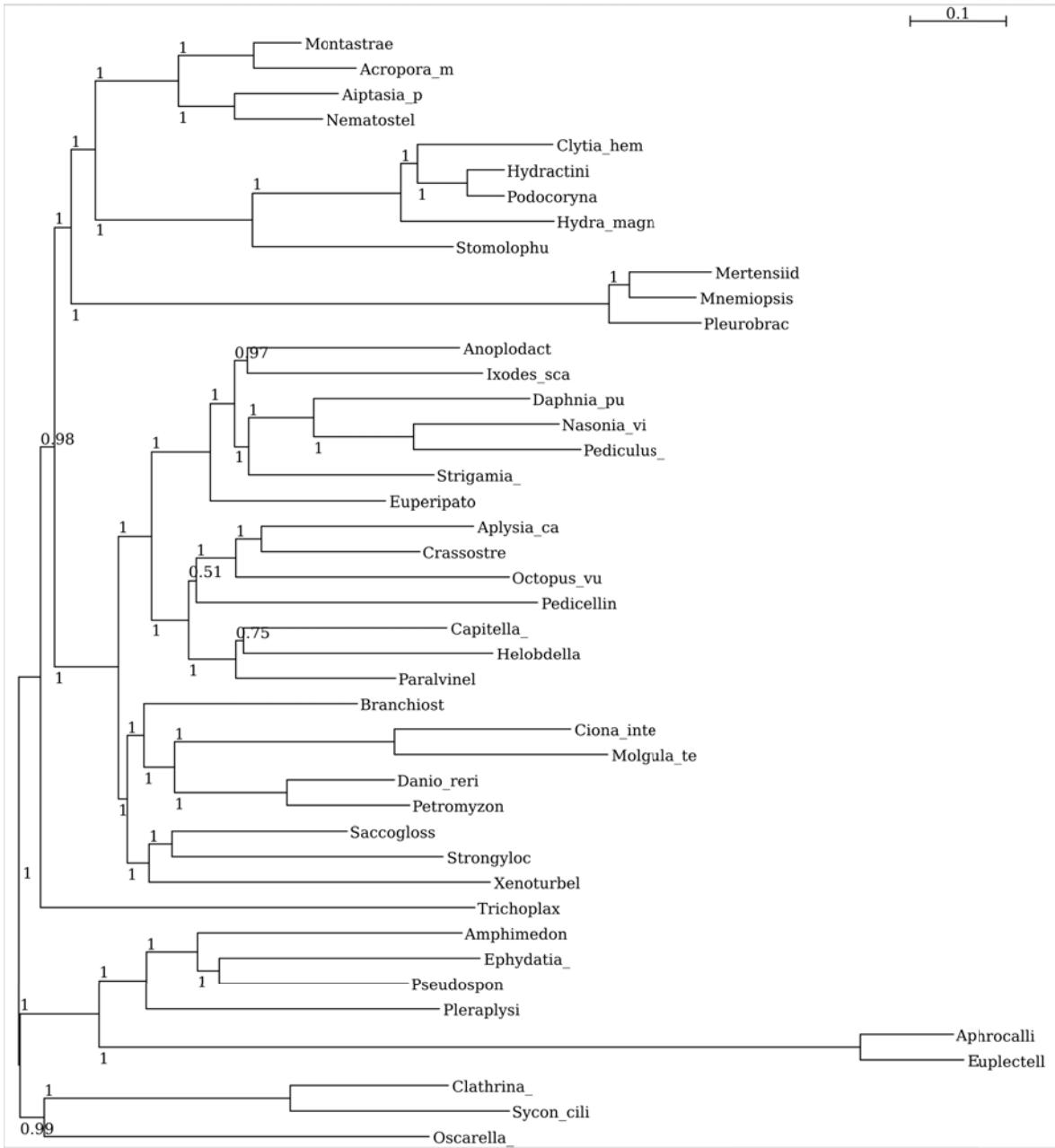


Figure 46: UPCB_55_PIP30 with no outgroups

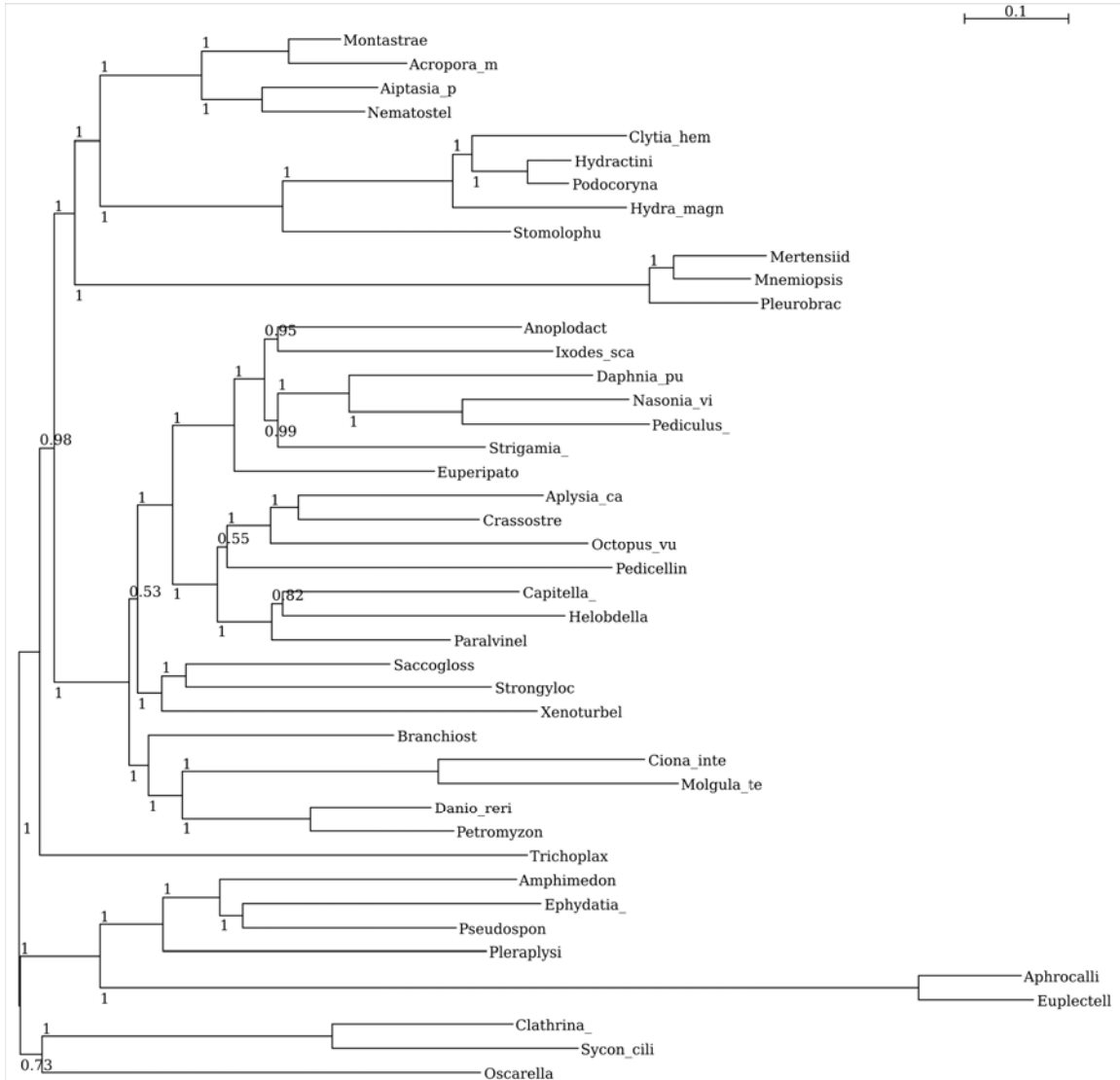


Figure 47: UPCB_55_PIP40 with no outgroups

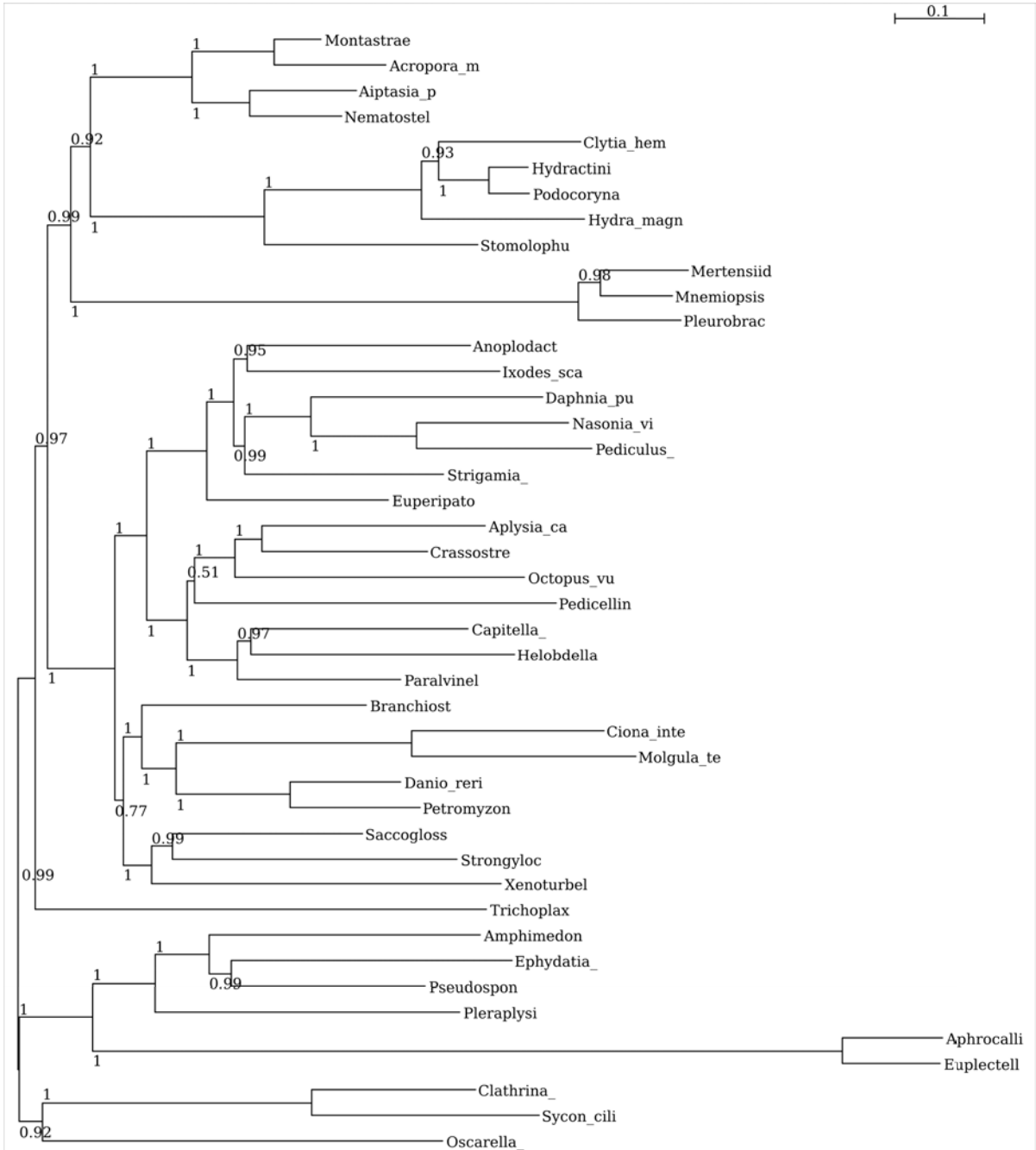


Figure 48: UPCB_55_PIP50 with no outgroups

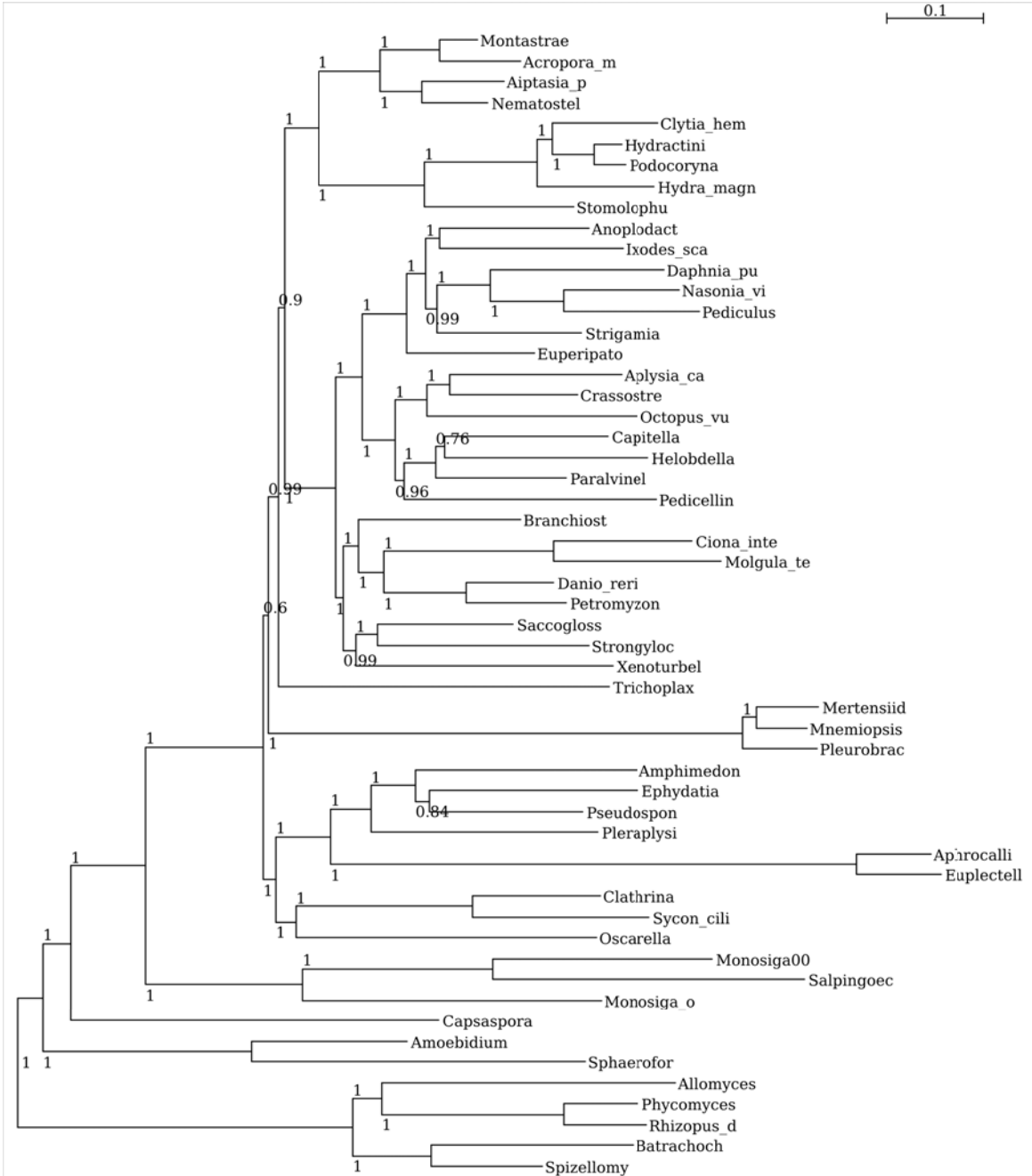


Figure 49: UPCB_55_Bin19

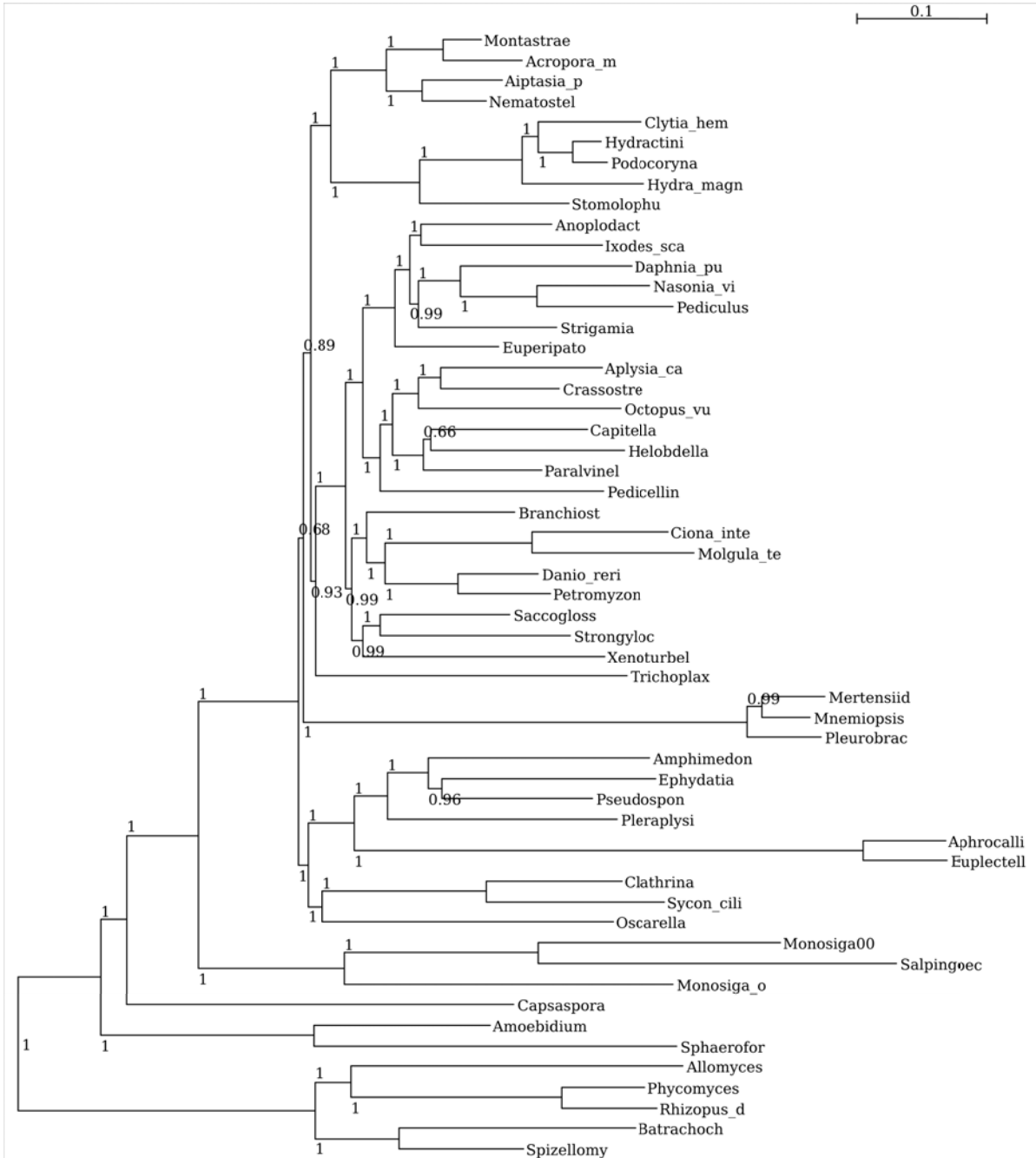


Figure 50: UPCB_55_Bin18

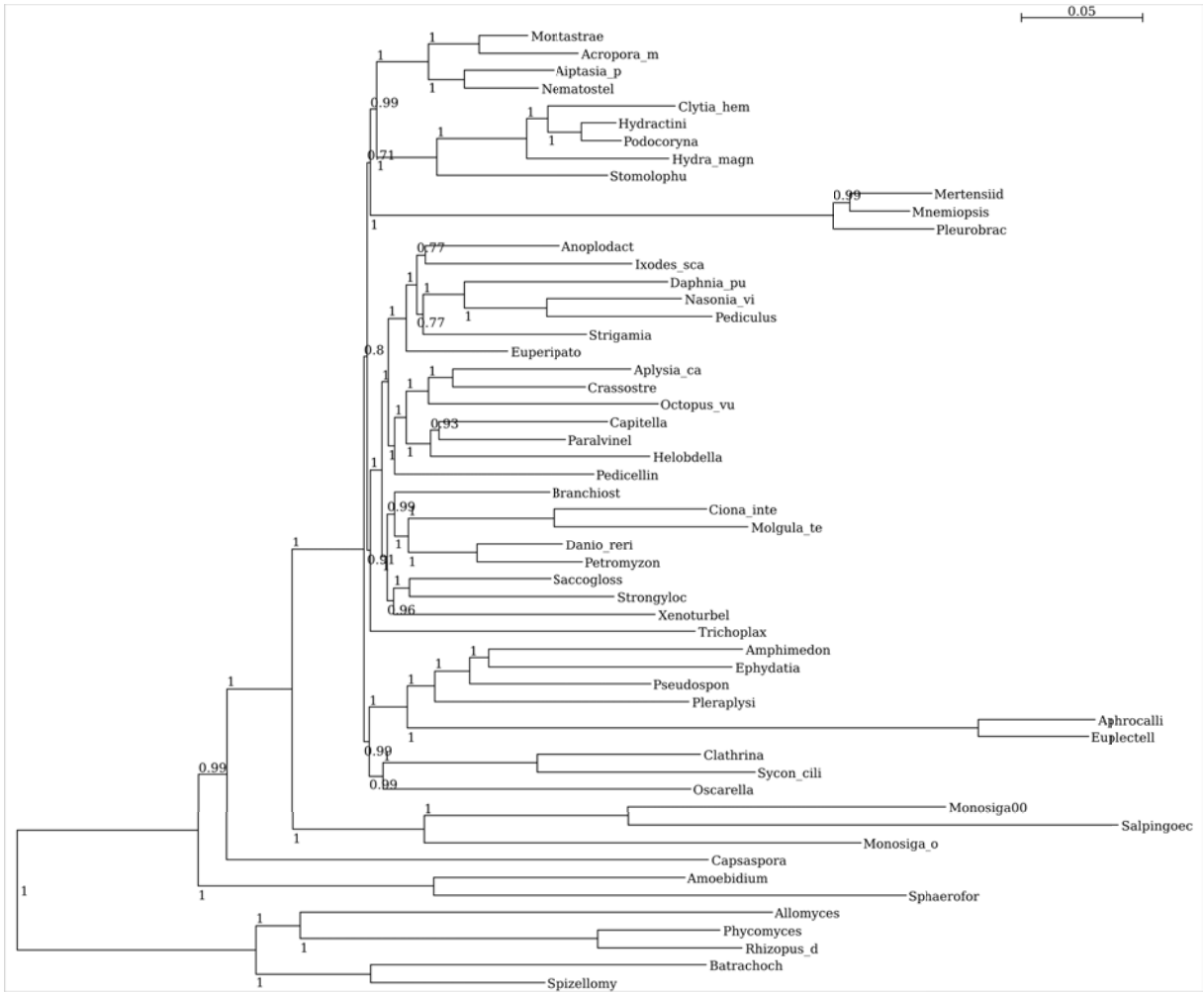


Figure 51: UPCB_55_Bin17

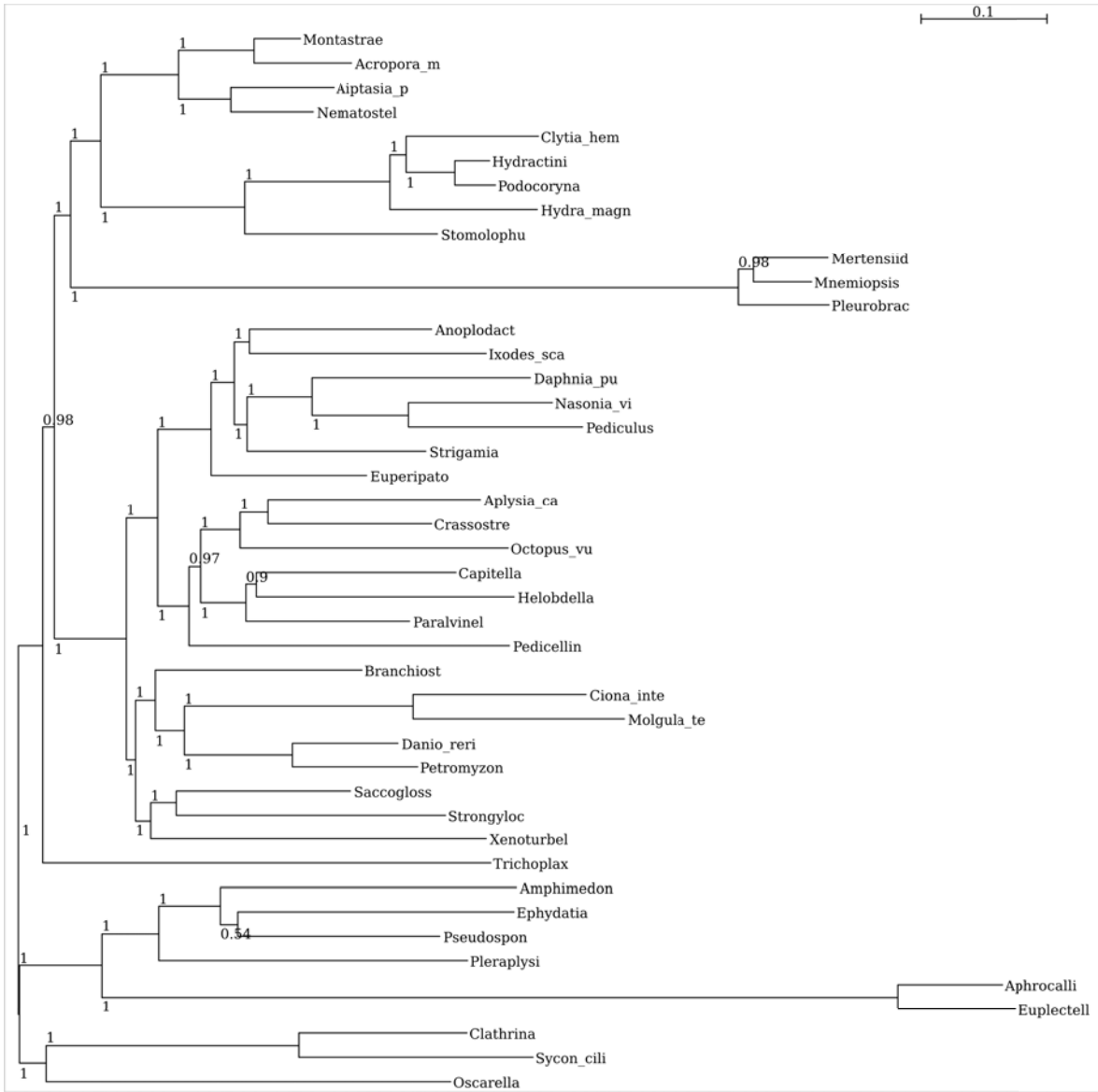


Figure 52: UPCB_55_Bin19 with no outgroups

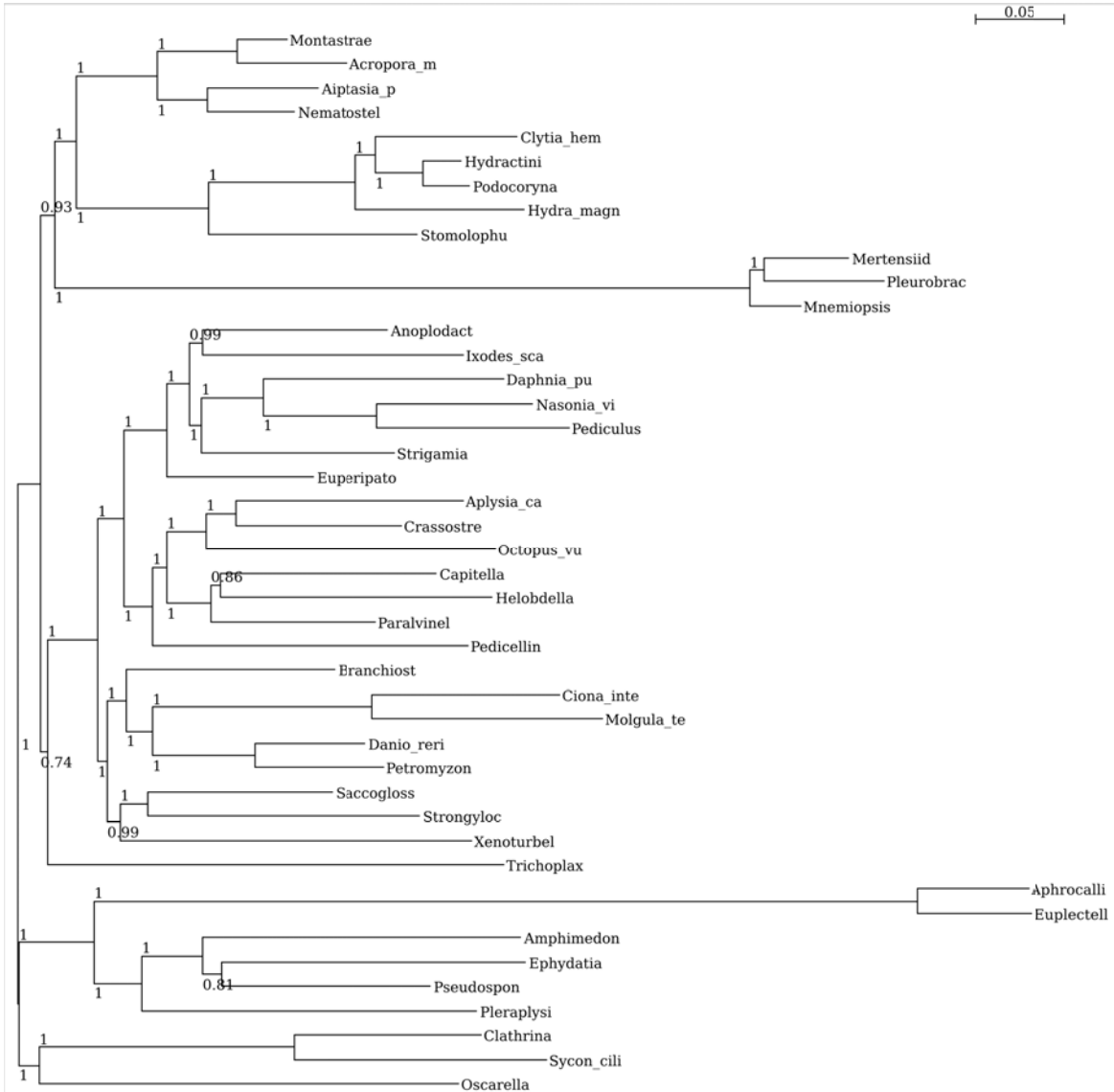


Figure 53: UPCB_55_Bin18 with no outgroups

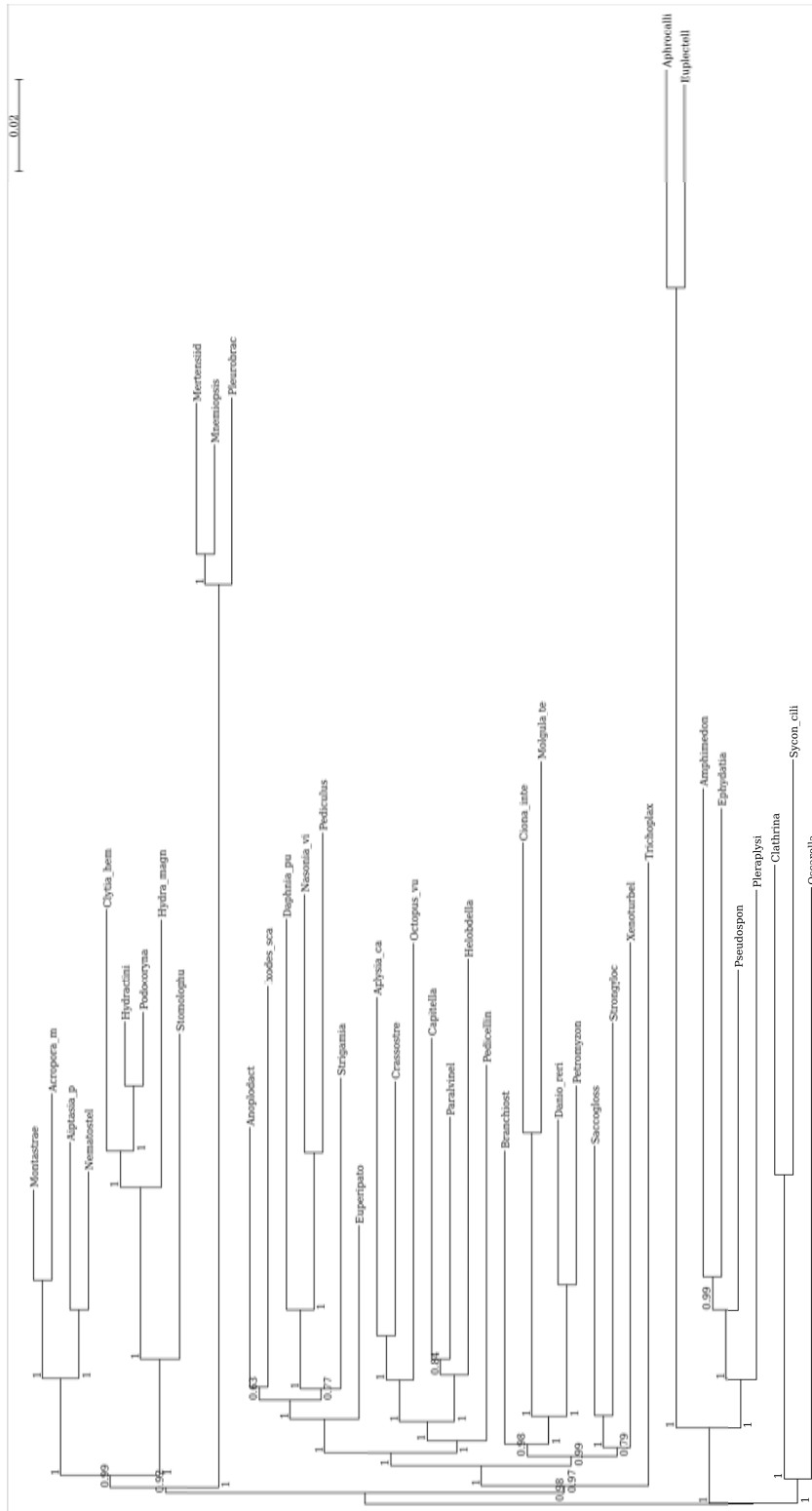


Figure 54: UPCB_55_Bin17 with no outgroups

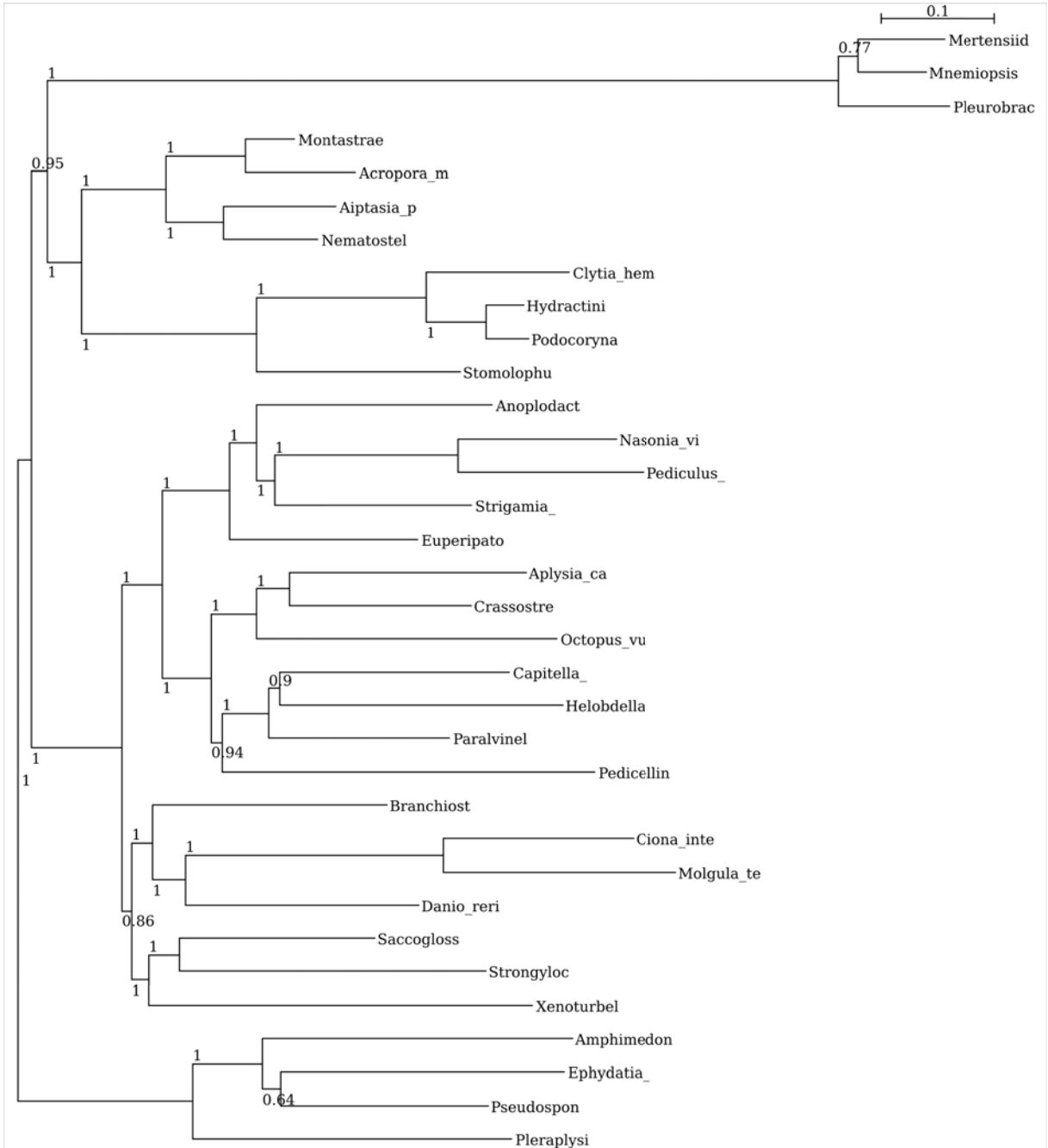


Figure 55: UPCB_55_NOCOMP

Annexe B

Appendice : article 2

SUPPLEMENTARY DATA APPENDIX A

Table S1: List of 73 taxa extracted from Mallatt et al. 2010
(as coded in their dataset « 01_rRNA371animalsMarch252011.docx »).

MONOSIGA
MONOSIGAOVATA
SACCHAROMYCES
MUCOR
SALPINGOECA
ICHTHYOPHONUS
PROTEROSPONGIA
GLYCERA
AMPHITRITE
ARION
CHAETOPLEURA
CEREBRATULUS
PHORONIS
CALLIPALLENE
MISUMENOPS
SCUTIGERA
HOMARUS
TRIOPS
PODURA
CALLIBAETIS
TENEBRIO
STRONGYLOCENTROTUS
SACCOGLOSSUS
PETROMYZON
RAJA
AMBYSTOMA
TRICHOPLAX
NAUSITHOE
ANTIPATHES
PODOCORYNE
AEGINA
AGLAUOPSIS
CARYBDEA
CHRYSAORA
HALICLYSTUS
CATOSTYLUS
PANTACHOGON
SCRIPPSIA
PORPITA
HYDRA
FABIENNA
MELICERTISSA
MONTASTREA
ATOLLA
NECTOPYRAMIS
HYDRACTINIA
CLYTIA
CYANEA
METRIDIVM
ANEMONIA
NEMATOSTELLA
PLEUROBRACHIA
BEROE
MNEMIOPSIS
MERTENSIA
LUBOMIRSKIA

EPHYDATIA
PACHYDICTYUM
AMPHIMEDON
HETEROCHONE
CRATEROMORPHA
OOPSACAS
SYCON
LEUCETTA
LEUCOSOLENIA
SUBERITES
RHABDOCALYPTUS
MYCALE
TRACHYCLADUS
TETHYA
OSCARELLA
RHIZAXINELLA
AXINELLA

Table S2: List of 7 porifera sequences imported from Genbank. Species, accession number.

<i>Sericolophus hawaiiicus</i>	190350254
<i>Hertwigia sp.</i>	222419928
<i>Leuconia nivea</i>	12711461
<i>Plakortis simplex</i>	38327059
<i>Chondrilla nucula</i>	187961570
<i>Igemella notabilis</i>	187961577
<i>Haliclona mediterranea</i>	38327054

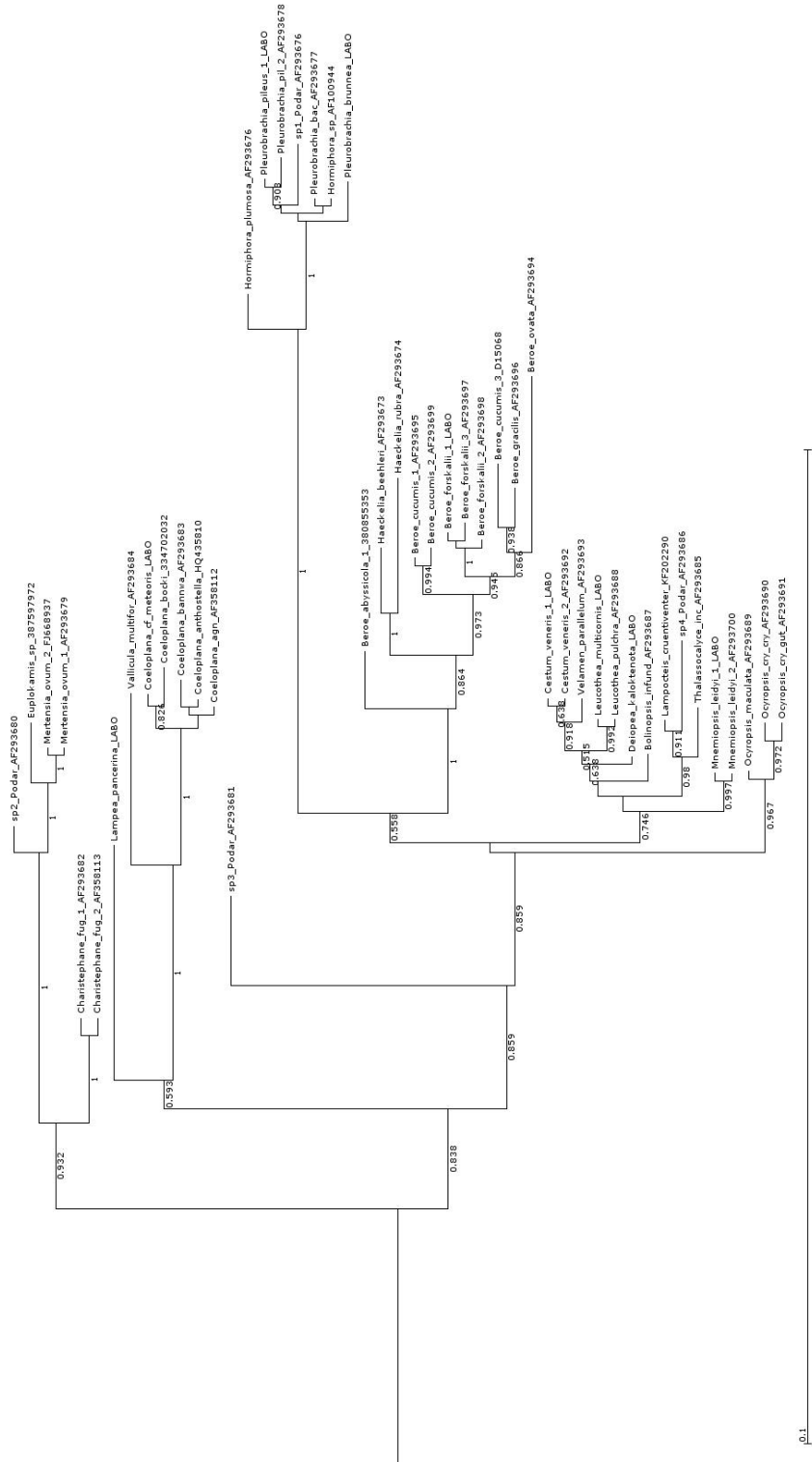


Figure S2 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, GTR+G+I model) from the 18S dataset using all the outgroups without taking into account the 18S rRNA secondary structure.

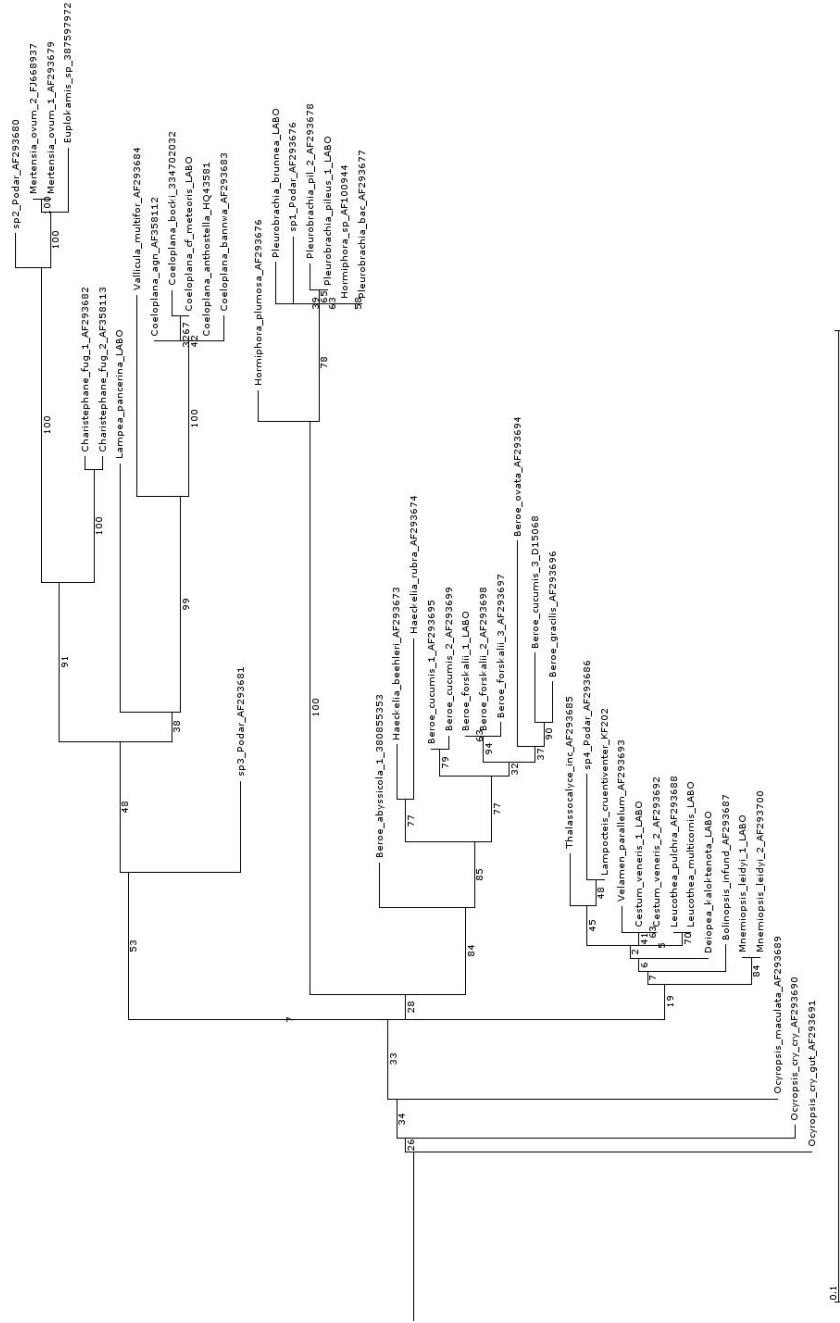


Figure S3 : Ctenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, S16 doublet model) from the 18S dataset using all the outgroups and taking into account the 18S rRNA secondary structure.

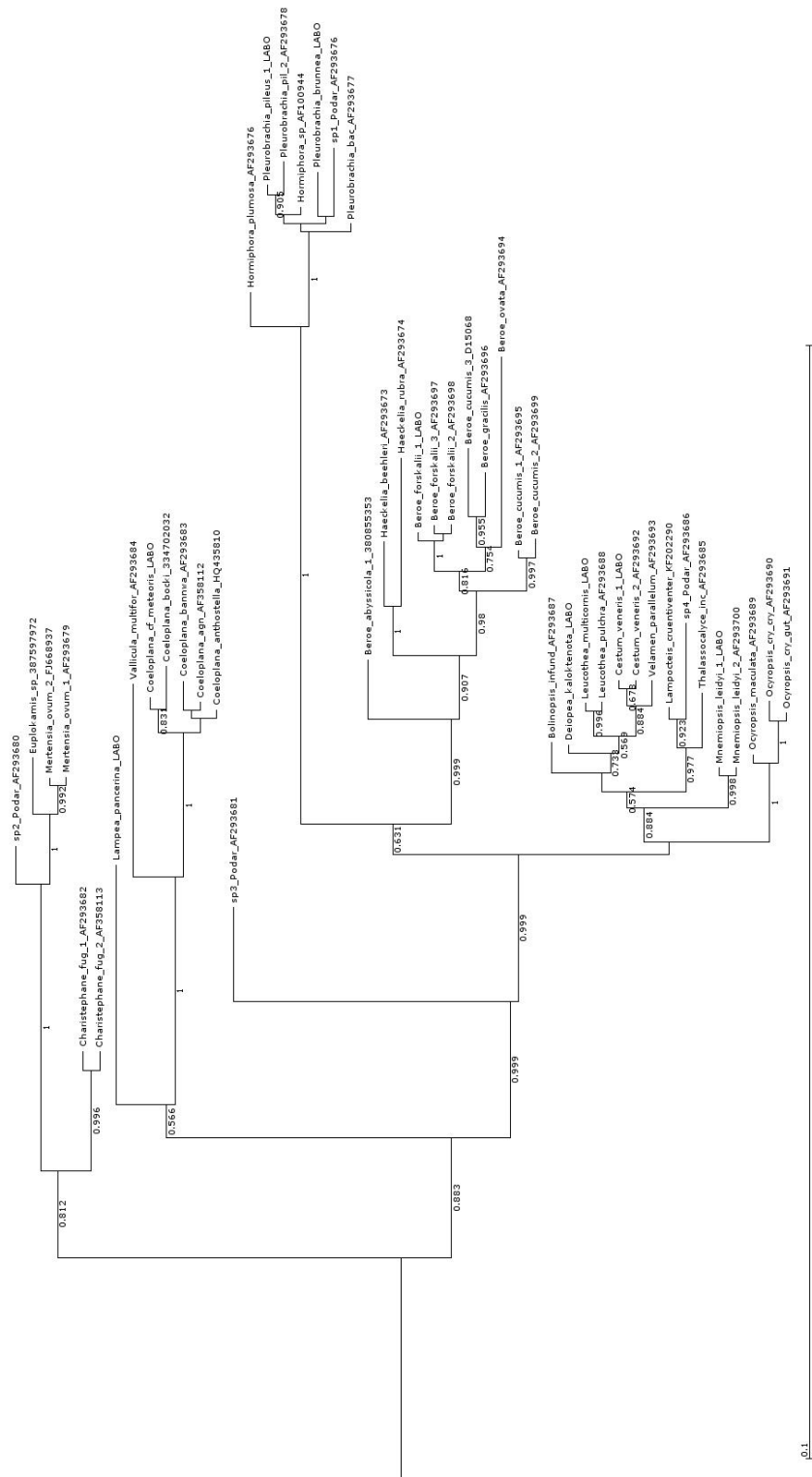


Figure S4 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, S16 doublet model) from the 18S dataset using all the outgroups and taking into account the 18S rDNA secondary structure.

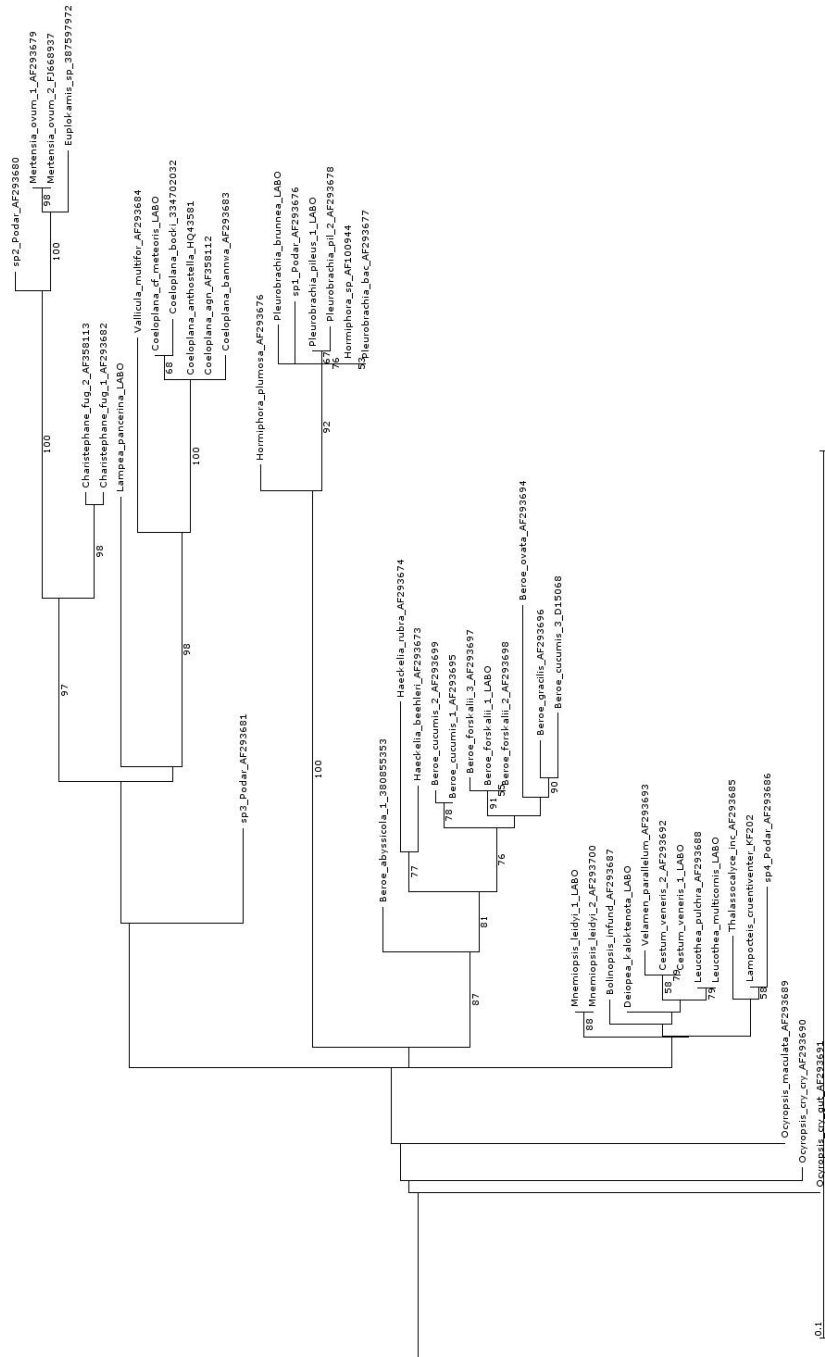


Figure S5 : Ctenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, GTR+G+I model) from the 18S dataset using cnidarians as outgroups without taking into account the 18S rRNA secondary structure.

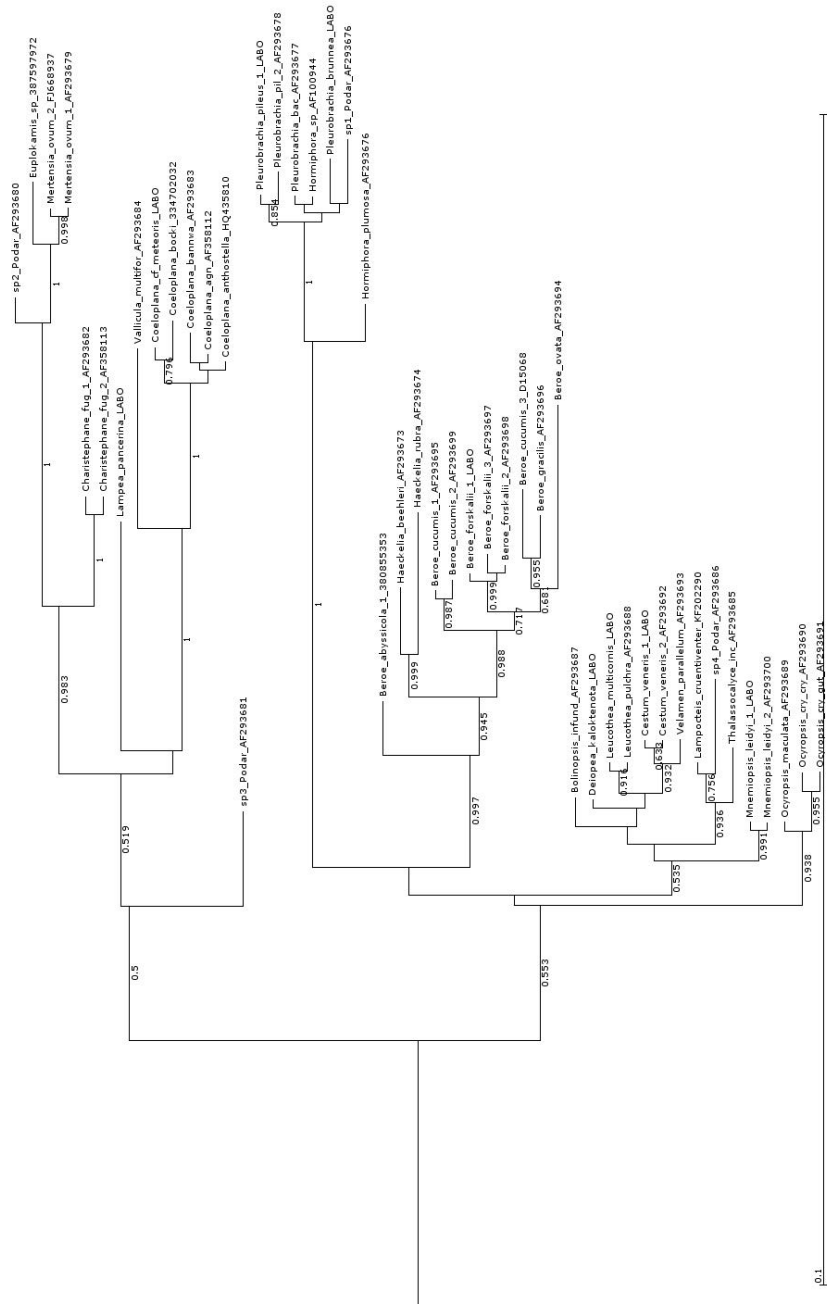


Figure S6 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, GTR+G+I model) from the 18S dataset using cnidarians as outgroups without taking into account the 18S rRNA secondary structure.

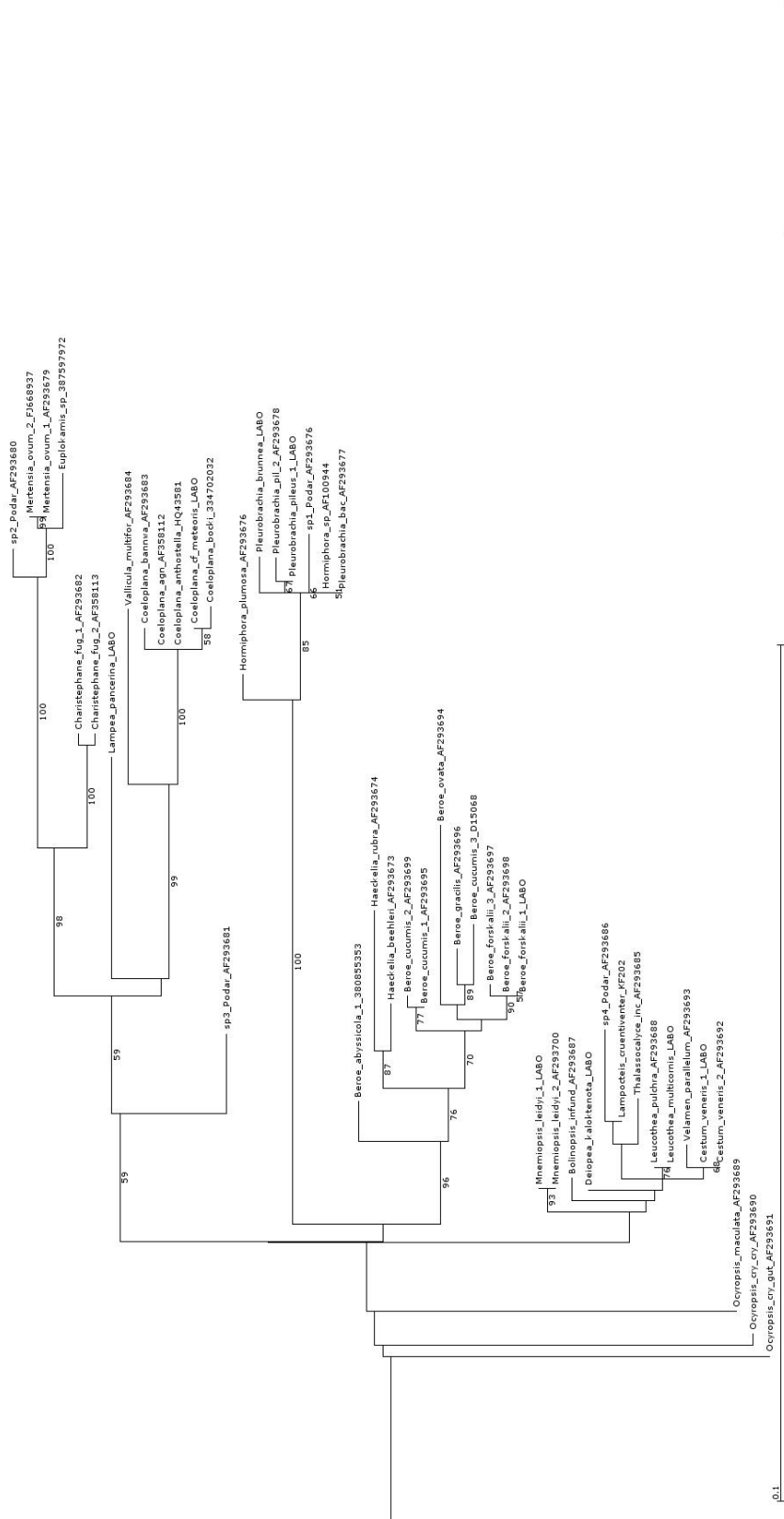


Figure S7 : Ctenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, S16 doublet model) from the 18S dataset using cnidarians as outgroups and taking into account the 18S rRNA secondary structure.

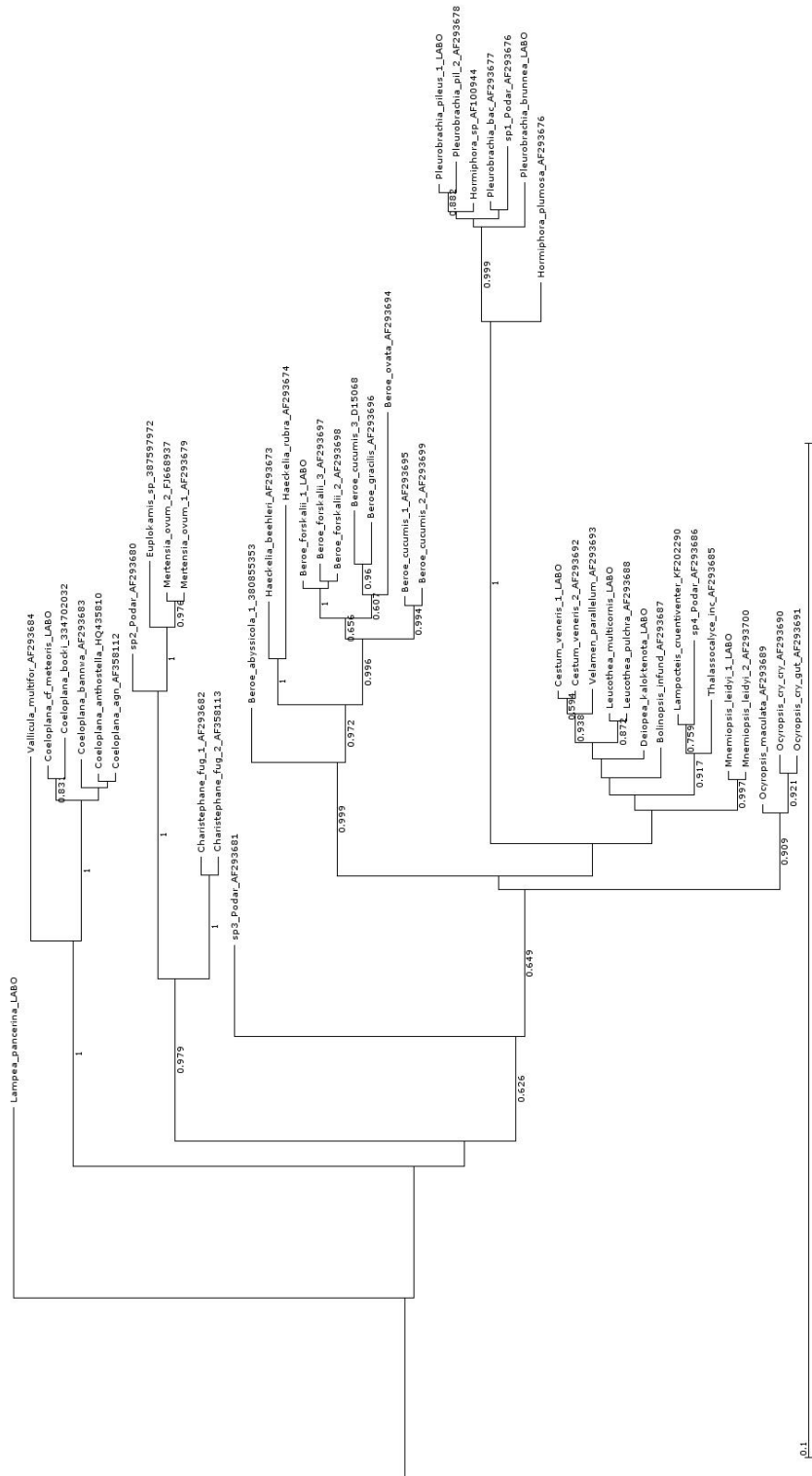


Figure S8 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, S16 doublet model) from the 18S dataset using cnidarians as outgroups and taking into account the 18S rRNA secondary structure.

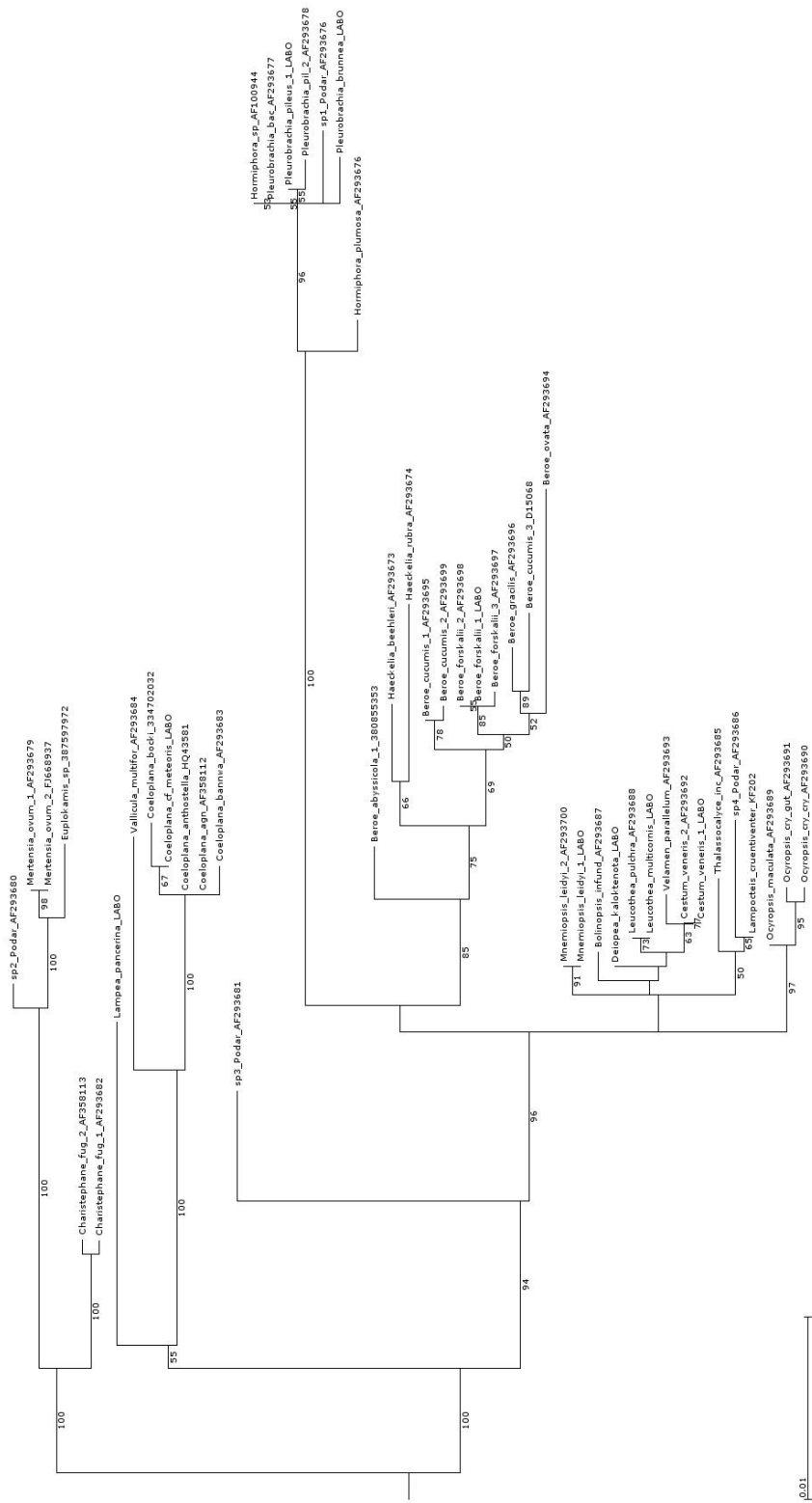


Figure S9 : Cenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, GTR+G+I model) from the 18S dataset using no outgroups and without taking into account the 18S rRNA secondary structure. The tree is unrooted.

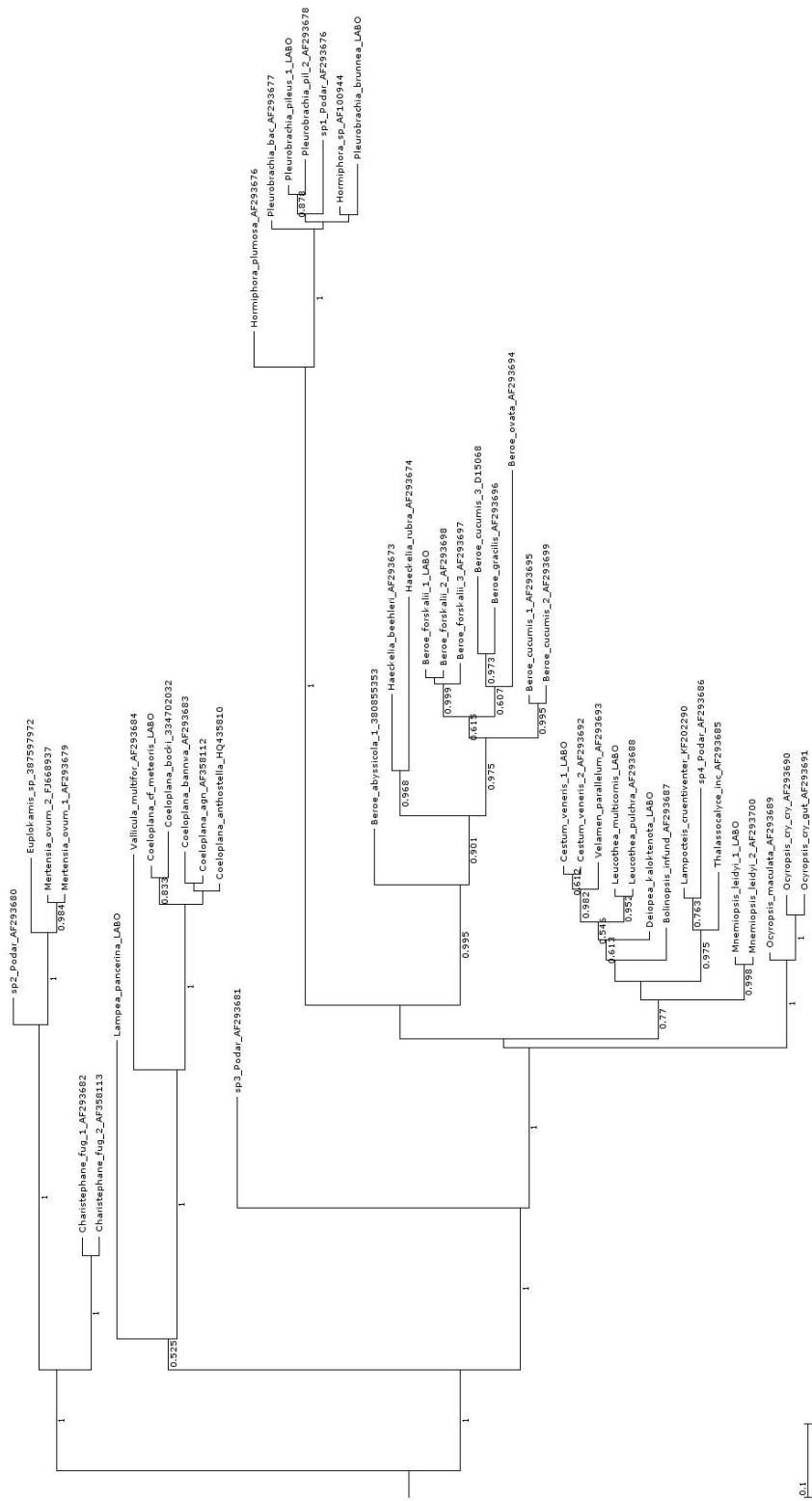


Figure S10 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, GTR+G+I model) from the 18S dataset using no outgroups and without taking into account the 18S rRNA secondary structure. The tree is unrooted.

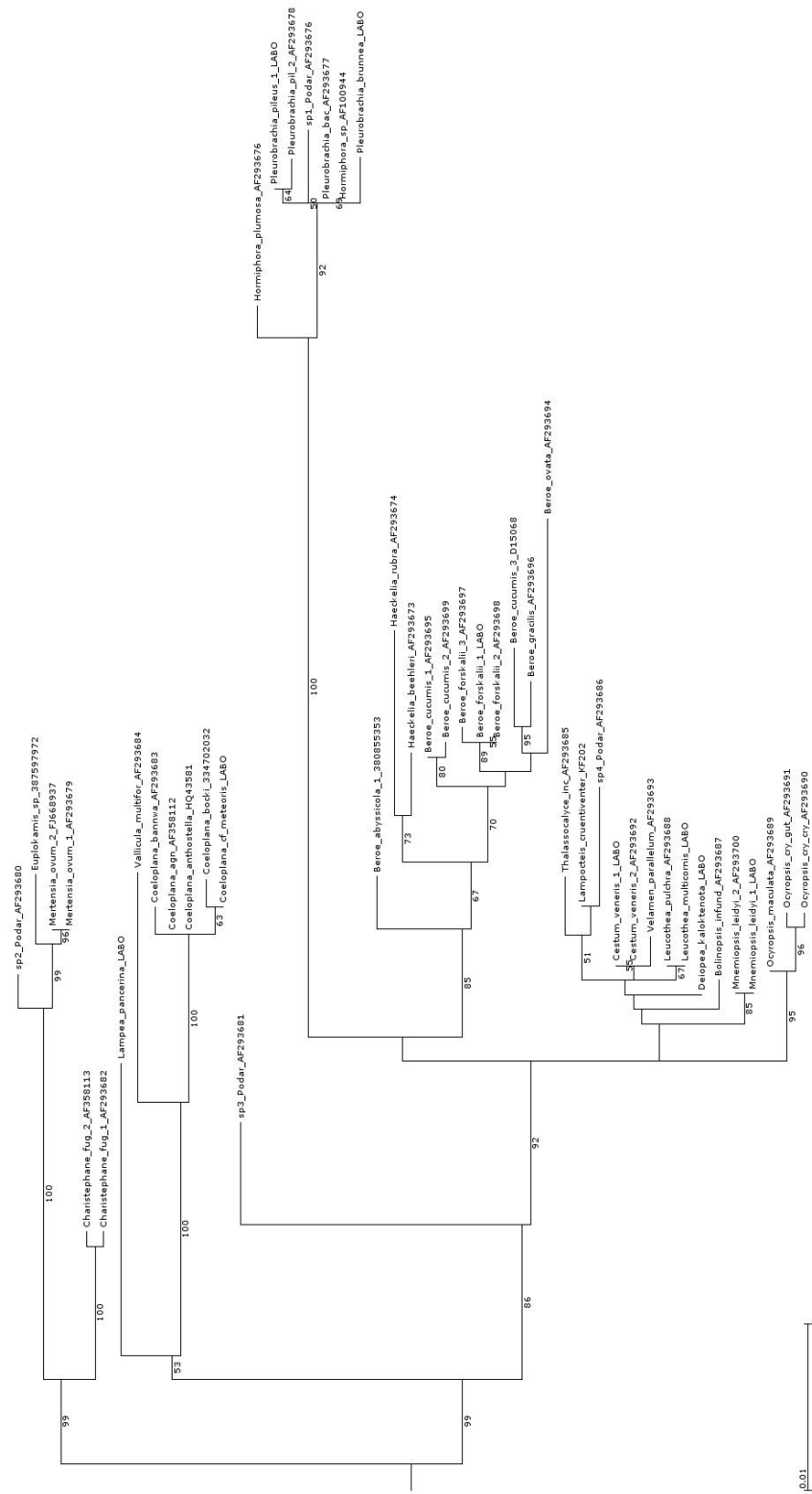


Figure S11 : Ctenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, S16 doublet model) from the 18S dataset using no outgroups and taking into account the 18S rRNA secondary structure. The tree is unrooted.

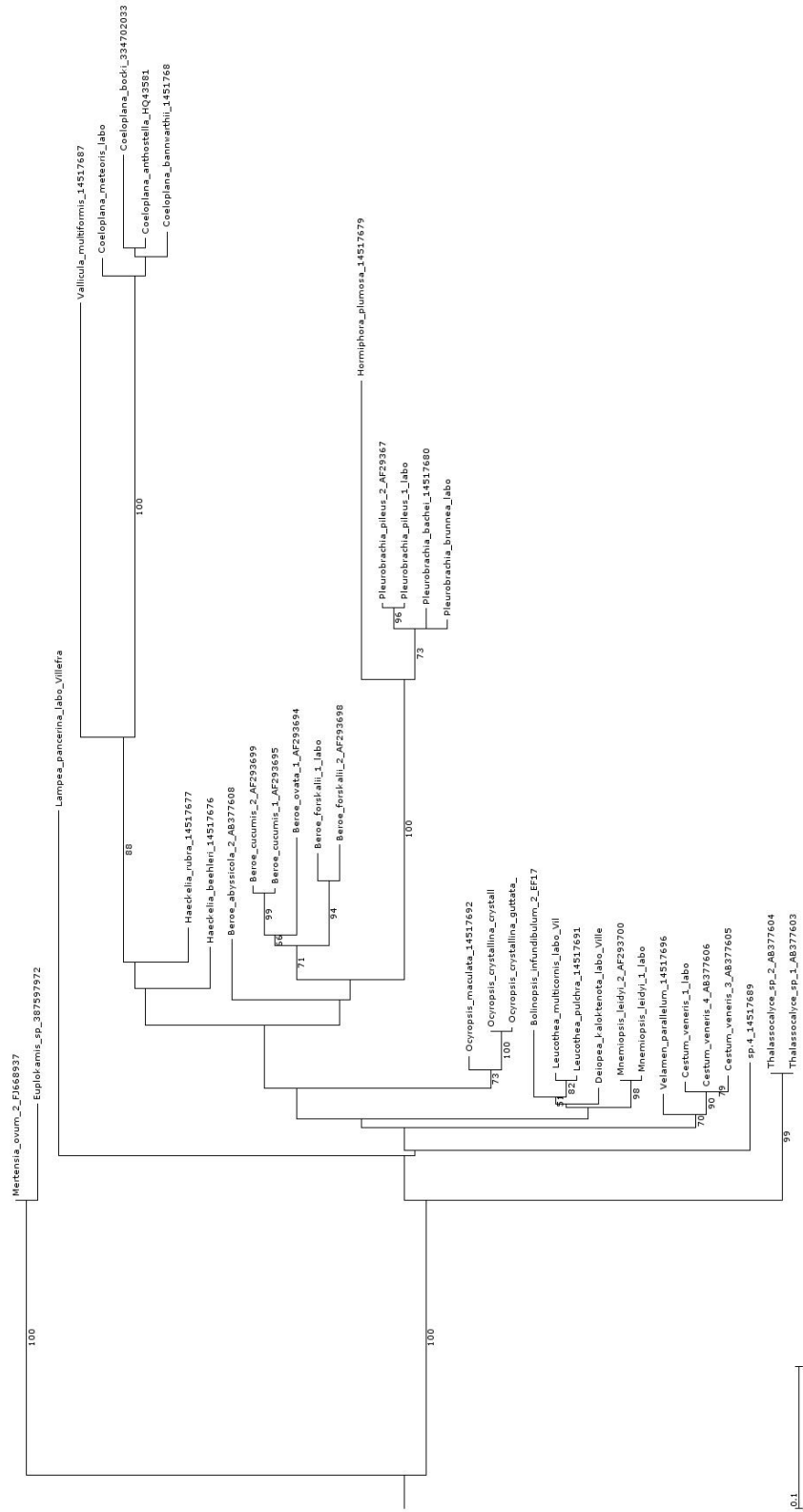


Figure S13 : Ctenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, GTR+G+I model) from the ITS dataset using no outgroups and without partitioning the ITS1, 5.8S rRNA and ITS2. The tree is unrooted.

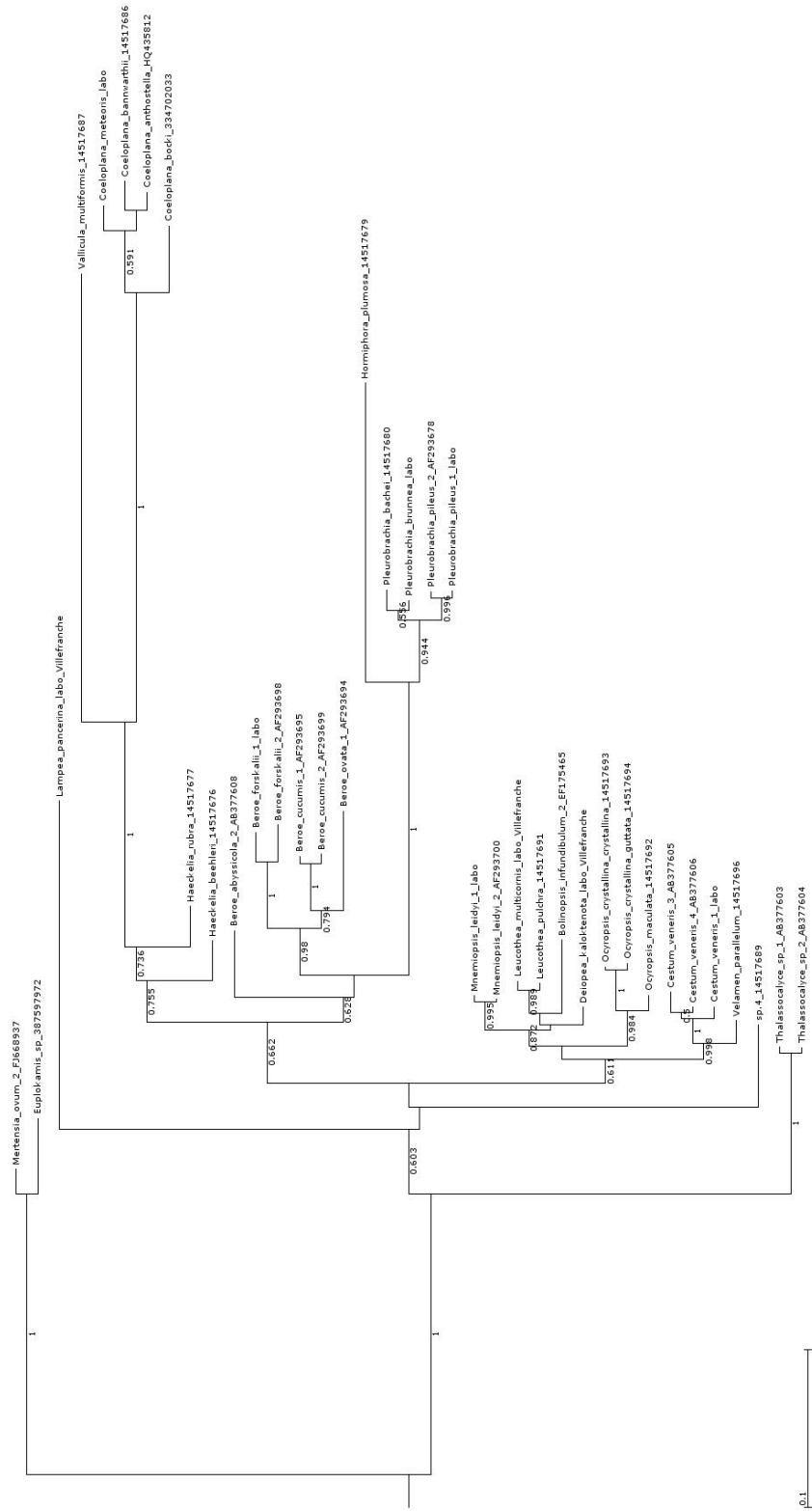


Figure S14 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, GTR+G+I model) from the ITS dataset using no outgroups and without partitioning the ITS1, 5, 8S rRNA and ITS2. The tree is unrooted.

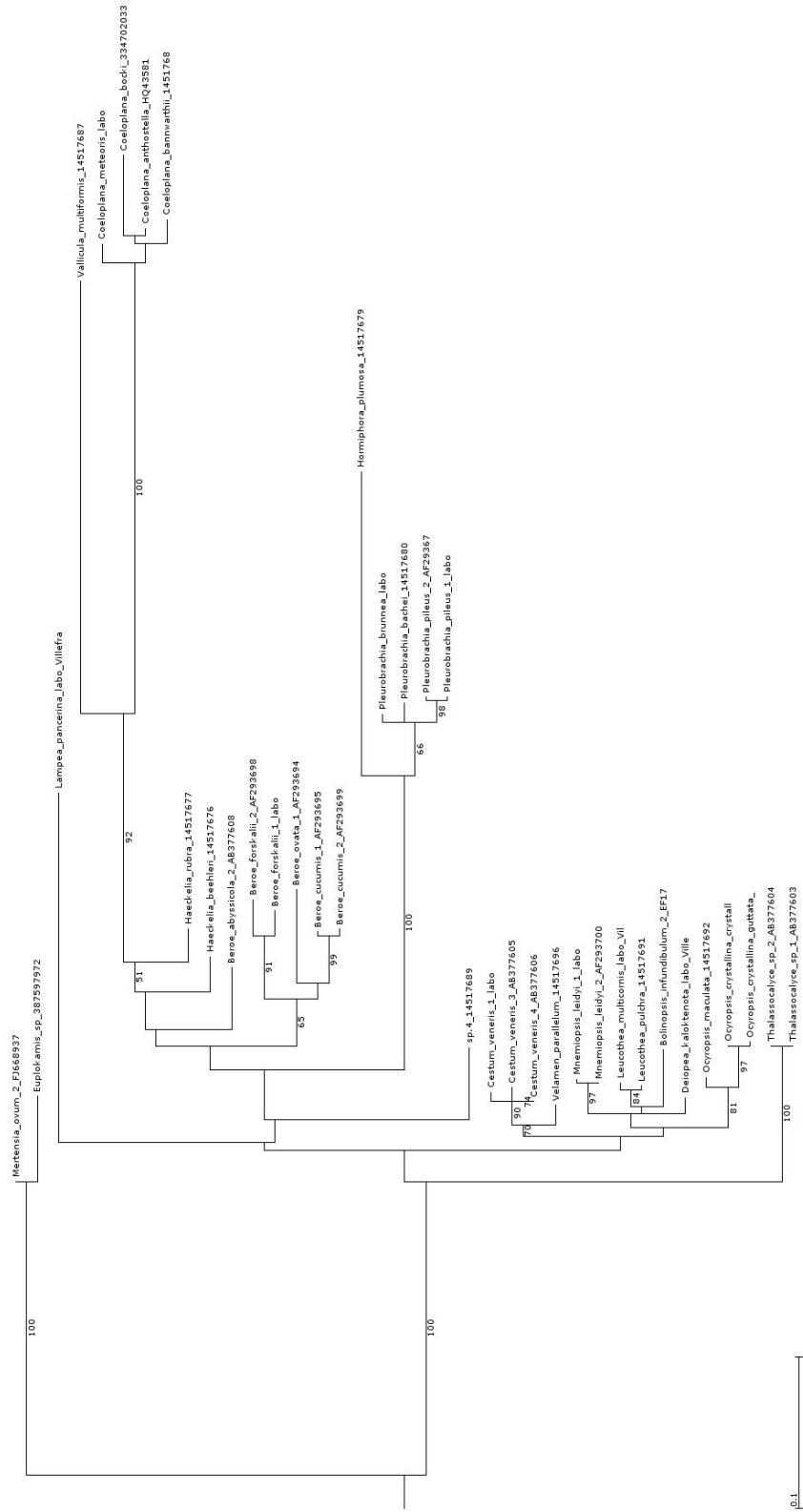


Figure S15 : Ctenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, GTR+G+I model) from the ITS dataset using no outgroups and partitioning the ITS1, 5.8S rRNA and ITS2. The tree is unrooted.

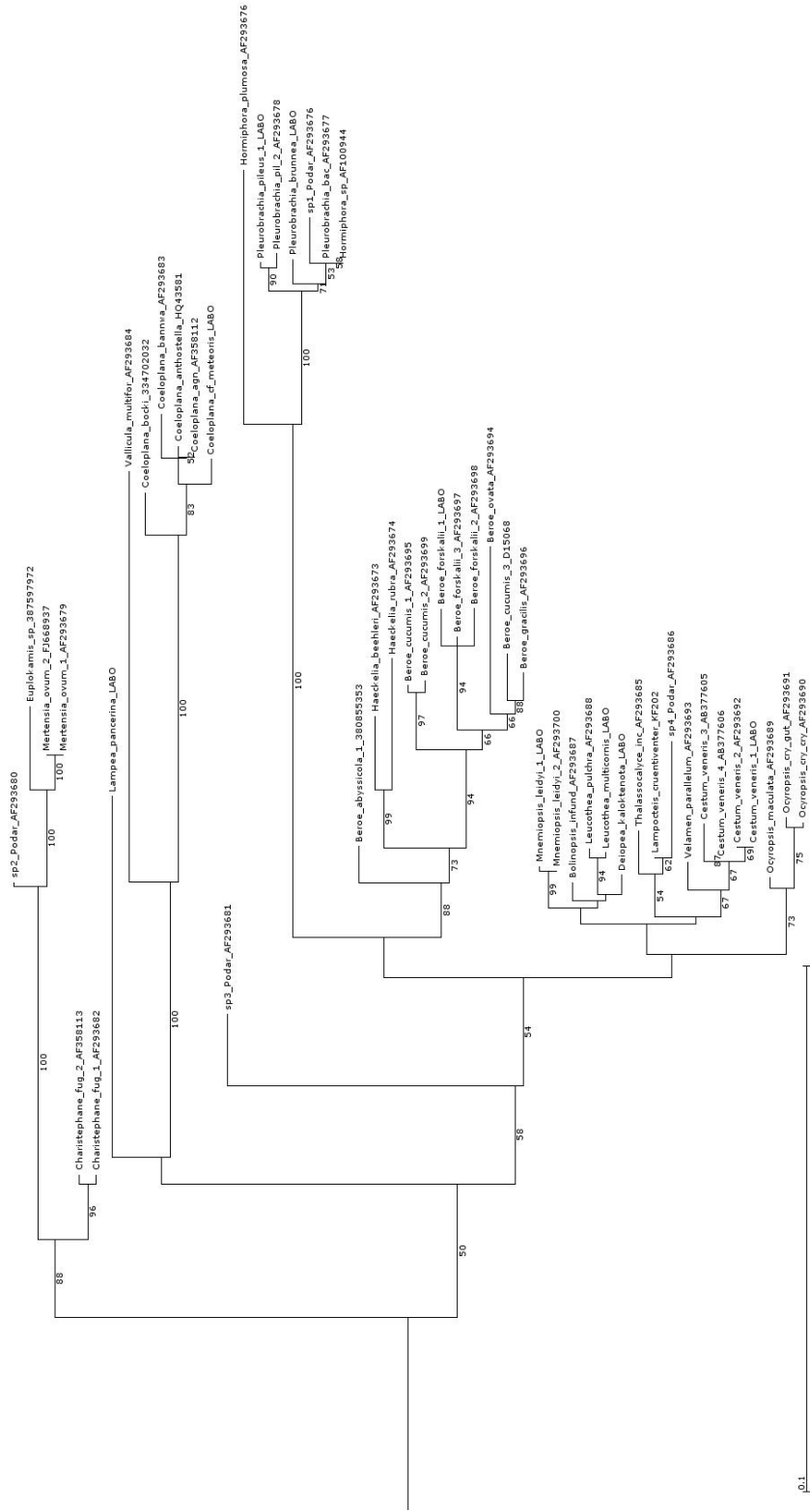


Figure S17 : Ctenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, GTR+G-I model) from the 18S-ITS dataset using all the outgroups without taking into account the 18S rRNA secondary structure nor partitioning the 18S rRNA, ITS1, 5.8S rRNA and ITS2.

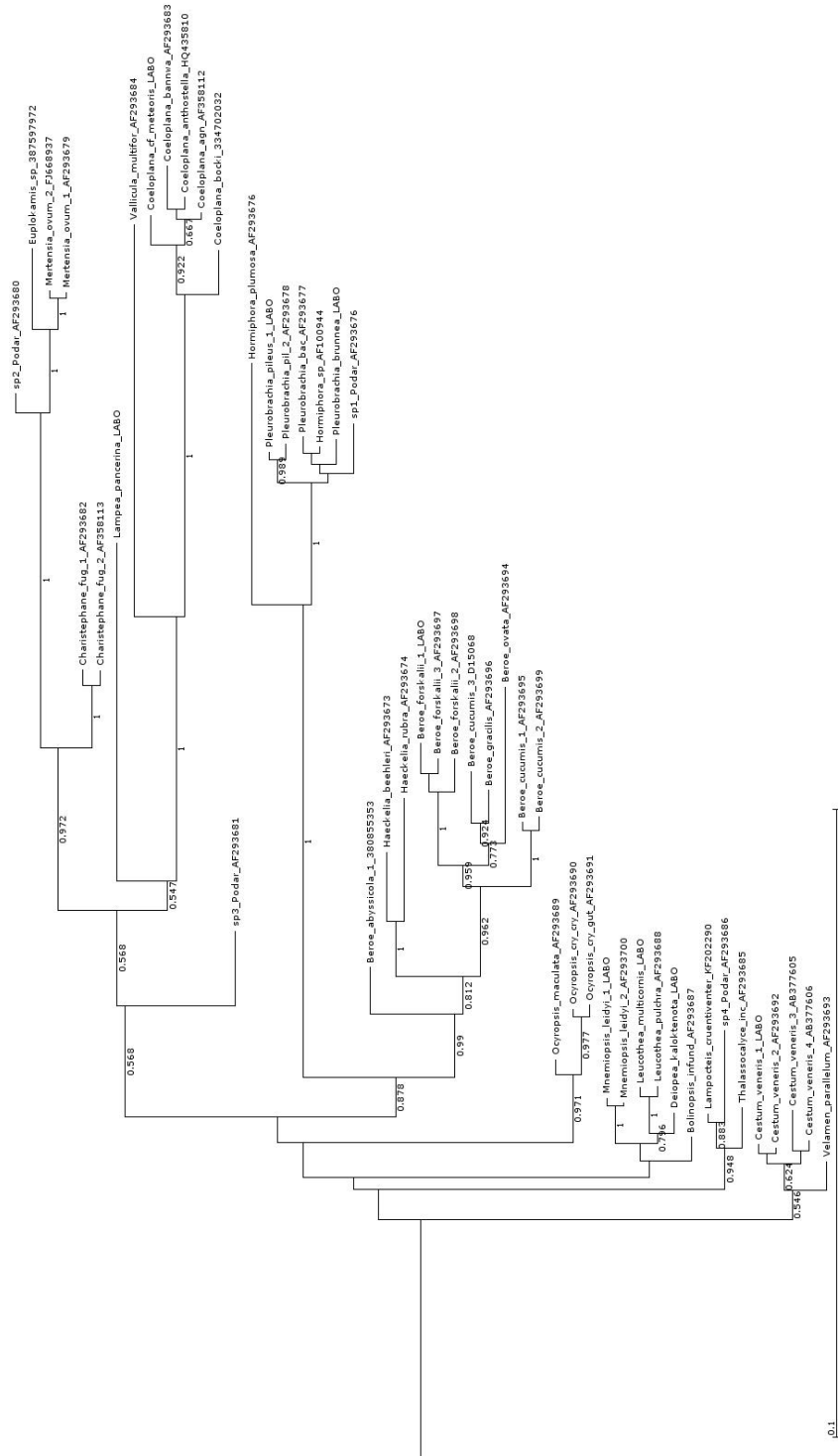


Figure S18 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, GTR+G+I model) from the 18S-ITS dataset using all the outgroups without taking into account the 18S rRNA secondary structure nor partitioning the 18S rRNA, ITS1, 5, 8S rRNA and ITS2.

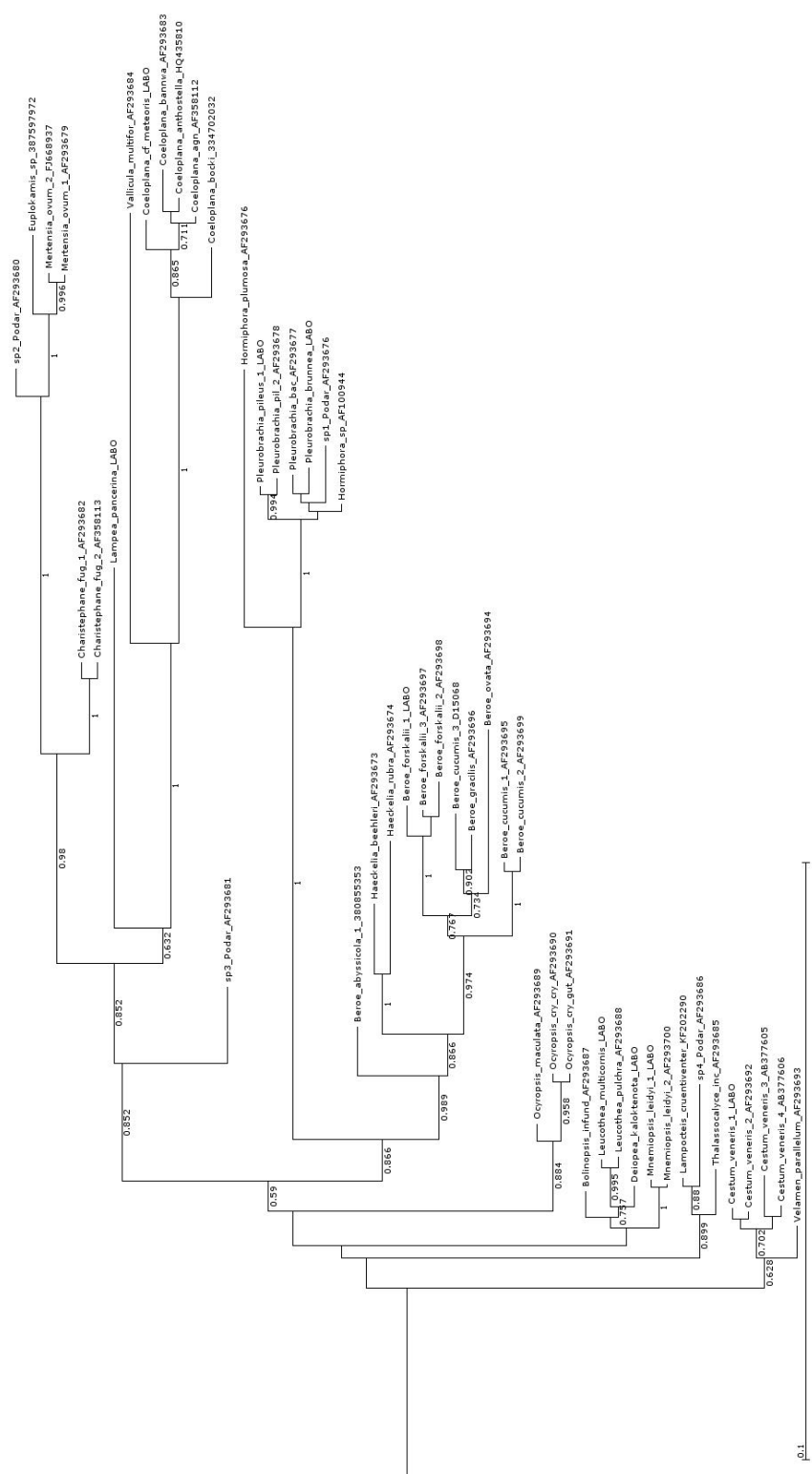


Figure S19 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, 16S doublet model) from the 18S-rRNA, ITS1, 5.8S rRNA and ITS2. account the 18S rRNA secondary structure and partitioning the 18S rRNA, ITS1, 5.8S rRNA and ITS2.

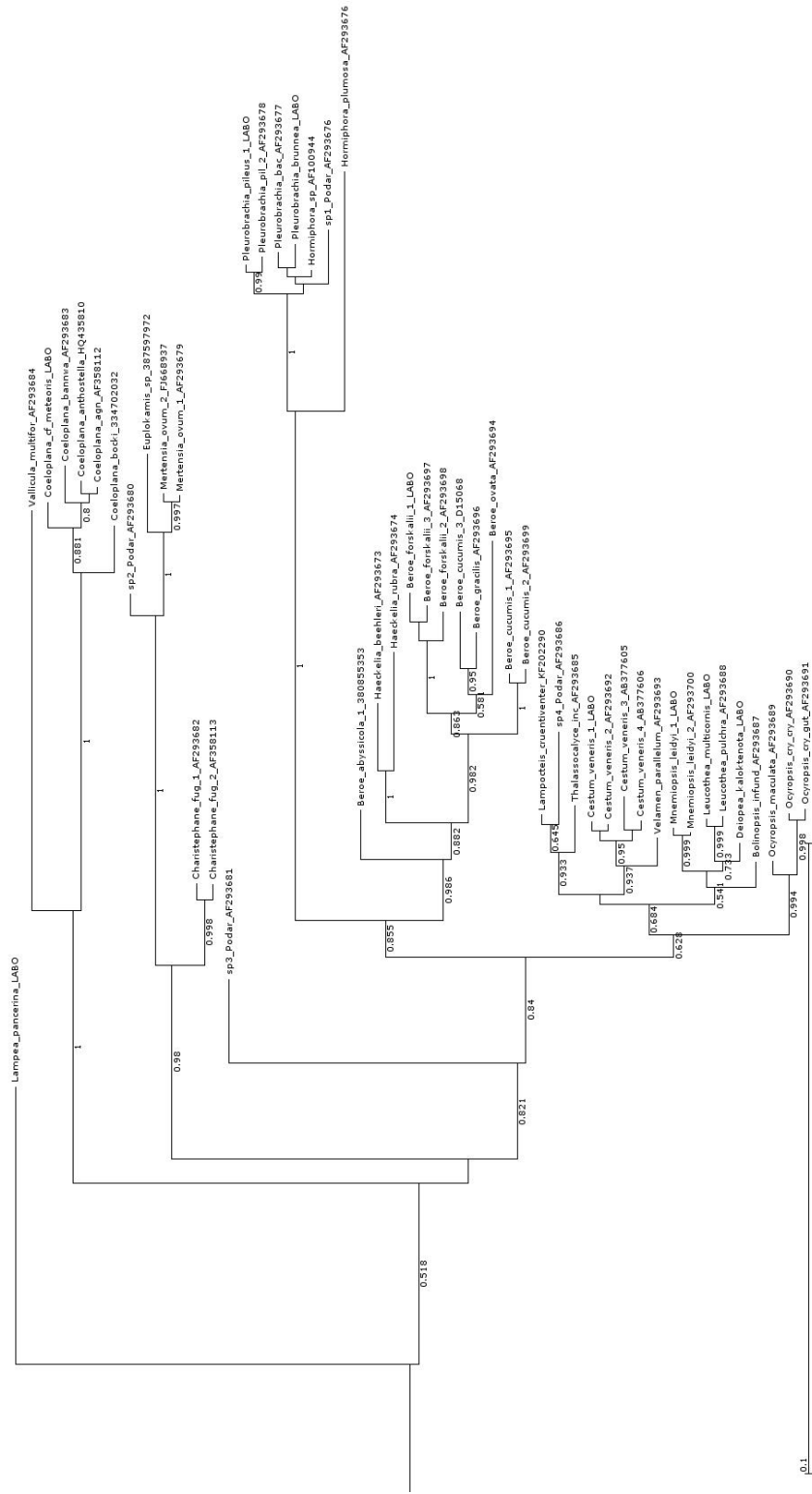


Figure S21 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, GTR+G+I model) from the 18S-ITS dataset using cnidarians as outgroups without taking into account the 18S rRNA secondary structure nor partitioning the 18S rRNA, ITS1, 5.8S rRNA and ITS2.

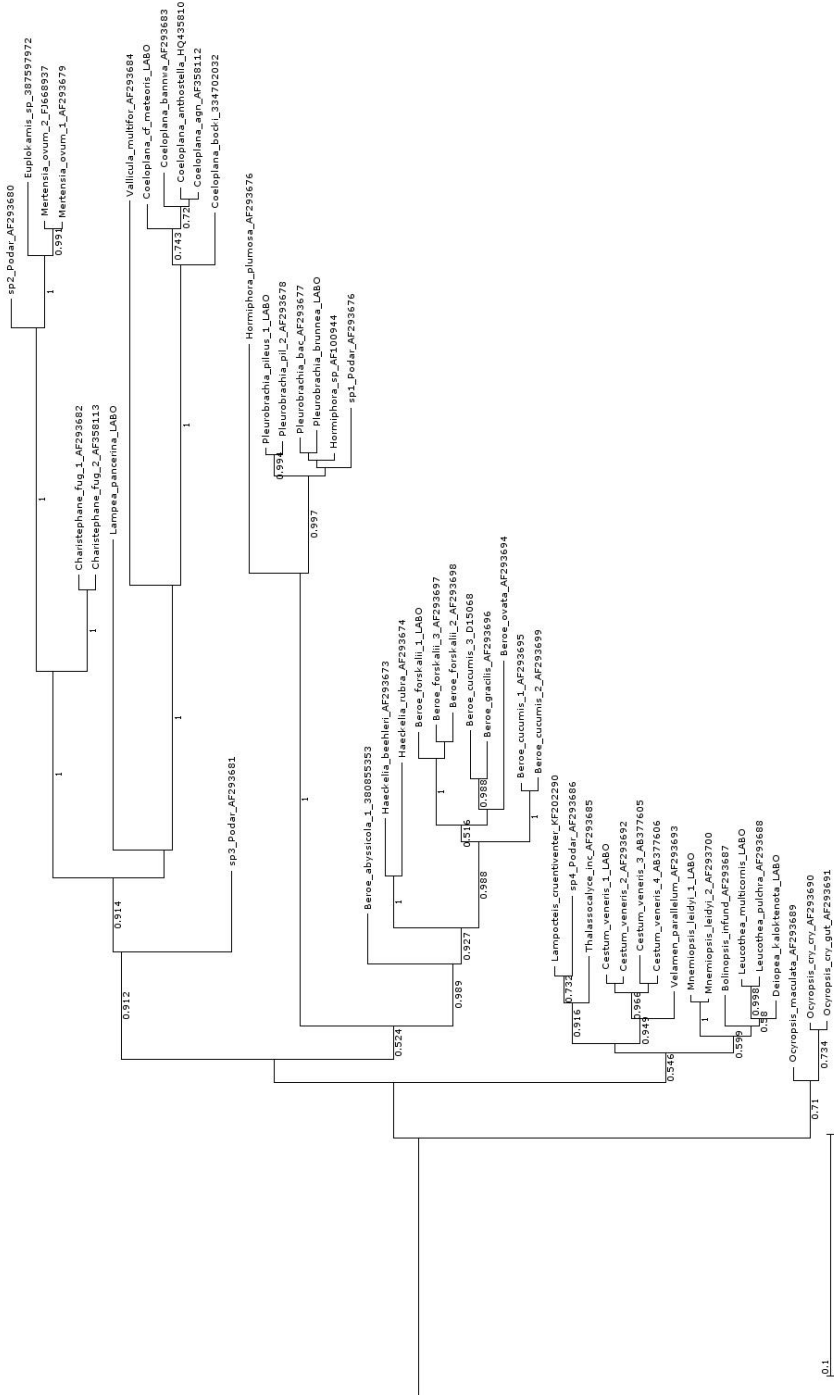


Figure S22 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, 16S doublet model) from the 18S-ITS dataset using cnidarians as outgroups, taking into account the 18S rRNA secondary structure and partitioning the 18S rRNA, ITS1, 5.8S rRNA and ITS2.

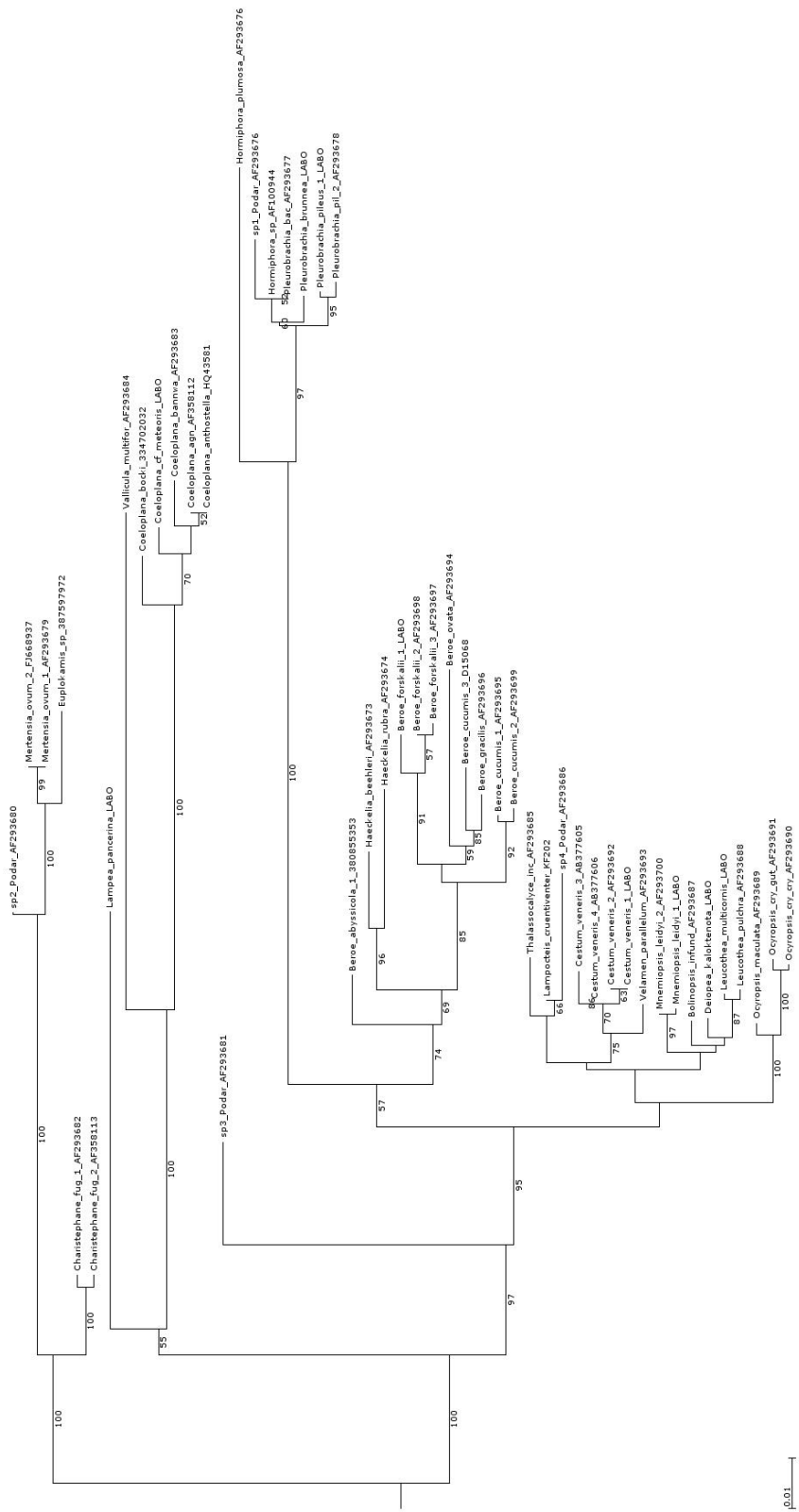


Figure S23 : Ctenophore phylogenetic relationships reconstructed under maximum likelihood (RAxML, GTR+G+I model) from the 18S-ITS dataset using no outgroups and without taking into account the 18S rRNA, ITS1, 5S, ITS2 secondary structure nor partitioning the 18S rRNA, ITS1, 5S, ITS2. The tree is unrooted.

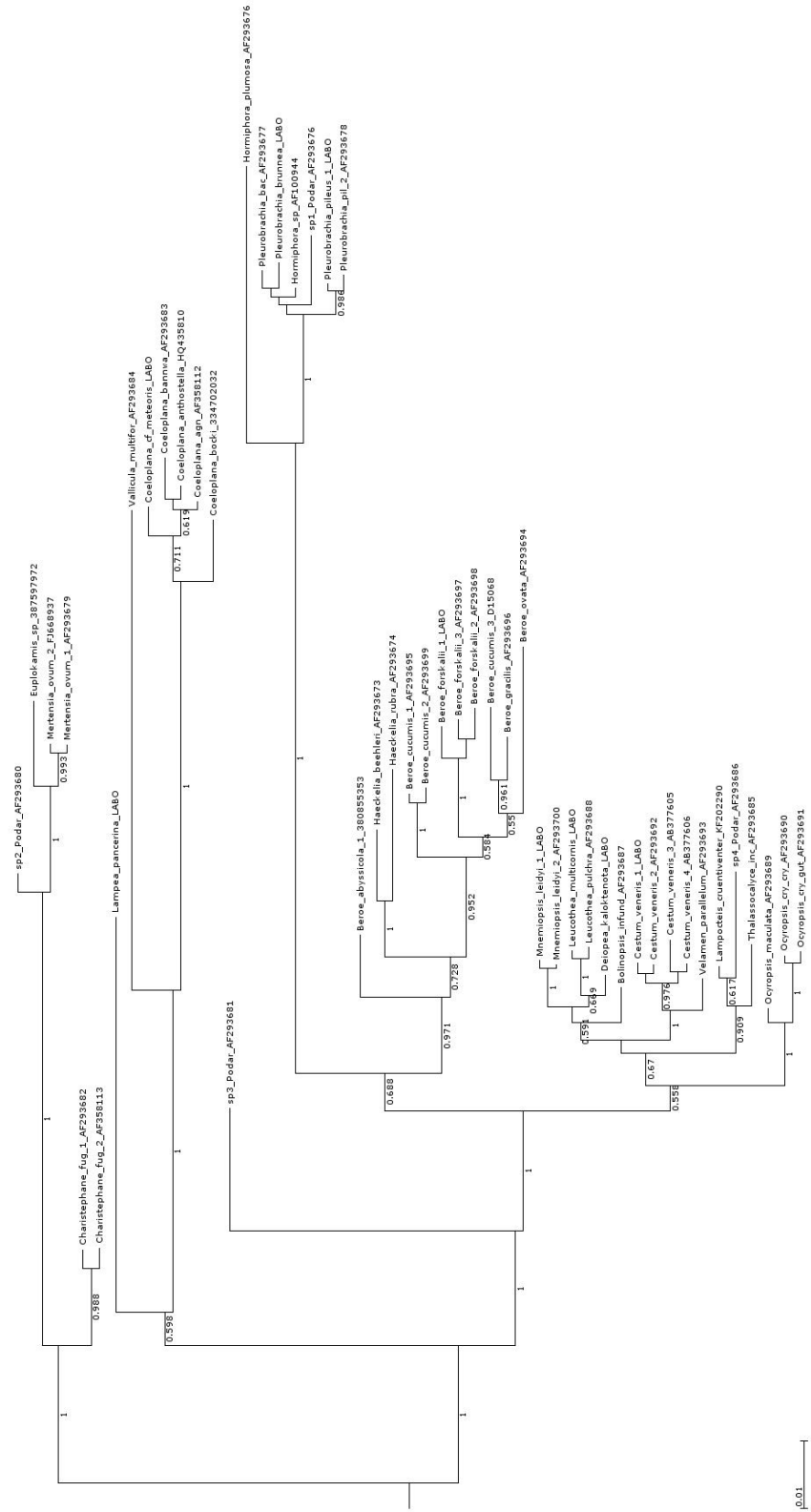


Figure S24 : Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, GTR+G+I model) from the 18S-ITS dataset using no outgroups and without taking into account the 18S rRNA secondary structure nor partitioning the 18S rRNA, ITS1, 5.8S rRNA and ITS2. The tree is unrooted.

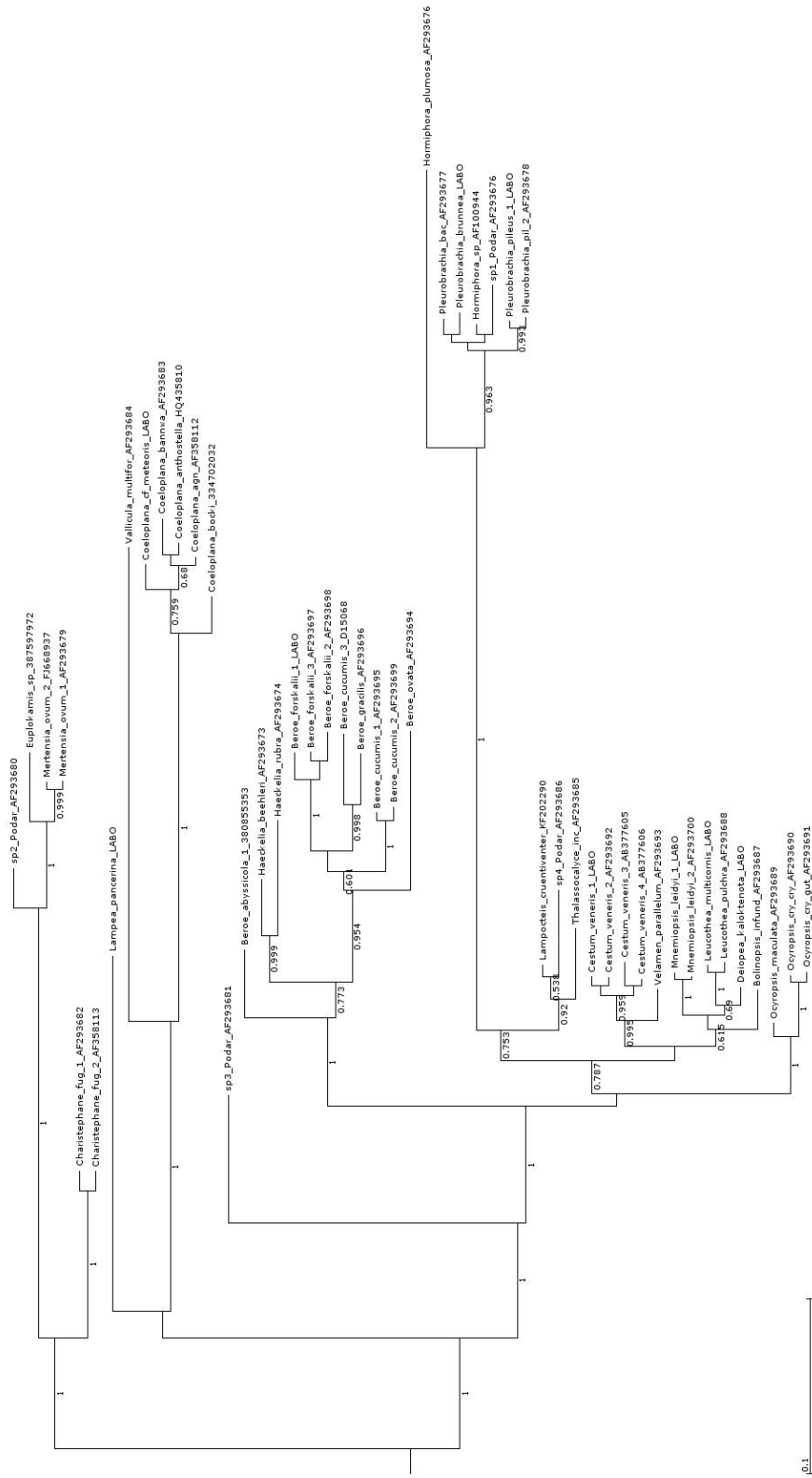


Figure S25: Ctenophore phylogenetic relationships inferred in a Bayesian framework (MrBayes, 16S doublet model) from the 18S-ITS dataset using no outgroups, taking into account the 18S rRNA secondary structure and partitioning the 18S rRNA, ITS1, 5.8S rRNA and ITS2. The tree is unrooted.