

Approche phylogénomique de la dynamique spatiale et temporelle de diversification chez les Lépidoptères Saturniidae

Pierre Arnal

▶ **To cite this version:**

Pierre Arnal. Approche phylogénomique de la dynamique spatiale et temporelle de diversification chez les Lépidoptères Saturniidae. Biodiversité et Ecologie. Museum national d'histoire naturelle - MNHN PARIS, 2020. Français. ⟨NNT : 2020MNHN0012⟩. ⟨tel-03383739⟩

HAL Id: tel-03383739

https://theses.hal.science/tel-03383739v1

Submitted on 18 Oct 2021

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.



HAL Authorization

MUSÉUM NATIONAL D'HISTOIRE NATURELLE

École Doctorale 227
Sciences de la nature et de l'Homme : évolution et écologie

Année 2020 **N°attribué par la bibliothèque**
|_|_|_|_|_|_|_|_|_|_|_|_|

THÈSE

pour obtenir le grade de

DOCTEUR DU MUSÉUM NATIONAL D'HISTOIRE NATURELLE

Spécialité : Écologie Évolutive

présentée et soutenue publiquement par

Pierre ARNAL

le 14 décembre 2020

Approche phylogénomique de la dynamique spatiale et temporelle de diversification chez les Lépidoptères Saturniidae

---

Sous la direction de **Marianne ELIAS** et **Rodolphe ROUGERIE**

Devant le jury composé de

Magalie CASTELIN, Maître de Conférences, MNHN (ISYEB, Paris)	**Examinatrice**
Thibaud DECAËNS, Professeur, Université de Montpellier (CEFE)	**Examinateur**
Marianne ELIAS, Directrice de Recherche, CNRS (ISYEB, Paris)	**Directrice de thèse**
Gaël KERGOAT, Directeur de Recherche, INRAE (CBGP)	**Rapporteur**
Benoît NABHOLZ, Maître de Conférences, Université de Montpellier (ISEM)	**Rapporteur**
Rodolphe ROUGERIE, Maître de Conférences, MNHN (ISYEB, Paris)	**Co-directeur de thèse**

Approche phylogénomique de la dynamique spatiale et temporelle de diversification chez les Lépidoptères Saturniidae

La compréhension des mécanismes évolutifs et écologiques à l'origine des patrons globaux de Biodiversité est une question centrale dans les domaines de l'Écologie et de l'Évolution. Les phylogénies, représentations des liens évolutifs entre les lignées du monde vivant, constituent le socle incontournable permettant d'identifier les processus à l'origine de ces patrons. Cette thèse présente les travaux de recherche que j'ai entrepris sur l'évolution de la diversité de papillons de la famille des Saturniidae Boisduval 1837 (Lepidoptera : Bombycoidea). Cette famille, particulièrement diversifiée biologiquement et morphologiquement, comprend près de 3500 espèces distribuées sur l'ensemble des continents. A l'aide d'approches phylogénomiques, j'ai inféré les relations phylogénétiques entre tous les genres décrits – à partir desquelles j'introduis une nouvelle classification des Saturniidae - et j'ai proposé une phylogénie incluant toutes les espèces du genre Néotropical *Copaxa*. J'ai également conçu un *pipeline* phylogénomique permettant la génération de mégaphylogénies – phylogénies datées de plus de mille feuilles dont la complétion est >50% des espèces – que j'ai appliqué sur un jeu de données combinant des éléments ultra-conservés du génome et des codes-barres ADN pour générer une phylogénie représentant entre 88 et 100% des espèces connues des Saturniidae. Les phylogénies ainsi inférées ont ensuite été utiliser afin d'examiner les dynamiques spatio-temporelles de la diversification des Saturniidae. Dans leur ensemble, les résultats obtenus au cours de ma thèse démontrent l'importance des facteurs biotiques dans la diversification spatiale et temporelle de la famille. J'ai notamment identifié que la capacité de tisser des cocons pleins et denses ainsi qu'un fort degré de polyphagie ont été les clés du succès biogéographique des Saturniidae et que l'hétérogénéité des taux de diversification au sein de la famille s'explique par l'évolution de traits en lien avec la stratégie dite de « *capital breeding* » de ces papillons : augmentation de la taille et de la polyphagie. J'ai également inféré que la niche climatique des *Copaxa*, héritée d'un ancêtre distribué dans la région Holarctique, avait façonné leur patron de diversification au sein de la région Néotropicale : la majorité des espèces volent dans les zones montagneuses, aux climats proches de ceux des zones tempérées, et les deux colonisations indépendantes de la chaine andine ont impliqué des *shifts* positifs des taux de diversification (évènements de *dispersification*). Dans leur ensemble, ces résultats représentent une avancée majeure dans la compréhension de la phylogénie et de l'évolution des Saturniidae, des Lépidoptères et plus généralement constituent un ensemble de supports permettant de mieux comprendre les processus évolutifs qui ont généré l'incroyable diversité des insectes.

Phylogenomic approach of spatial and temporal dynamic of diversification in Saturniidae moths (Lepidoptera)

Understanding the evolutionary and ecological mechanisms governing the global patterns of Biodiversity is a central question in the fields of Ecology and Evolution. Phylogenies, as representations of the evolutionary relationships between lineages of the living world, are a fundamental support to identify the processes at the origin of these patterns. This thesis presents my work on the diversity and evolution of moths in family Saturniidae Boisduval 1837 (Lepidoptera: Bombycoidea). This biologically and morphologically diverse family includes nearly 3,500 species distributed on all continents. Using phylogenomic approaches, I inferred the phylogenetic relationships between all knwown genera – from which I introduce a new classification – and I proposed a phylogeny for all species in the Neotropical genus *Copaxa*. I also designed a phylogenomic pipeline allowing the generation of megaphylogenies – i.e. dated phylogenies with more than a thousand tips and whose completion is >50% of all species – which I applied on a data set combining ultraconserved elements and DNA barcodes to generate a phylogeny including between 88 and 100% of all described saturniid species. The phylogenies thus inferred were used in order to characterize spatial and temporal diversification of Saturniidae. Altogether, my results demonstrate the importance of biotic factors in the spatial and temporal diversification of the family. In particular, I identified that the ability to spin plain and dense cocoons as well as high degrees of polyphagy have been the keys to the biogeographical success of Saturniidae, and that the heterogeneity of the diversification rates within the family was explained by the evolution of traits linked to the capital breeding strategy of these moths: increases in body size and in polyphagy level. I also inferred that the climatic niche of *Copaxa* moths, inherited from an ancestor distributed in the Holarctic region, shaped their diversification within the Neotropical region: the majority of species fly in mountainous areas, the climate of which is similar to those found in temperate areas, and the two independent colonizations of the Andean chain implied positive shifts in diversification rates (dispersification events). Taken together, these results represent a major advance in understanding the evolution of Saturniidae, Lepidoptera and more generally constitute a set of materials allowing a better understanding of the evolutionary processes which have generated the incredible diversity of insects.

Remerciements

Quel chemin parcouru depuis le début de cette thèse. Il est difficile de regarder en arrière et de se souvenir de tous ceux qui vous ont soutenu, encouragé et inspiré. D'avance pardon à ceux que j'oublie, j'espère que vous accepterez d'interpréter ça comme de la maladresse.

J'aimerais d'abord remercier Alice avec qui j'ai la chance de partager ma vie depuis maintenant près d'une dizaine d'années. Merci pour tout ce que tu m'apportes et pour ton soutien au quotidien, pour avoir accepté de m'entendre parler d'évolution, de papillons, de phylogénie.

Merci à mes parents ainsi qu'à mes sœurs pour qui ce travail peut paraître abstrait mais qui ont toujours montré de l'intérêt pour les travaux que j'ai effectué. Merci de m'avoir transmis l'amour de la nature, de la montagne et notamment des Cévennes.

Je profite également de cette page pour témoigner auprès de mes amis à quel point leur amitié m'est essentielle. Merci pour tous ces moments ensemble, à Montpellier, Paris, pour ces semaines à Cocurès. Merci à Amine, Ange, Armel, Marti pour ces parties interminables de Risk GoT. A Marine, François et Florian pour ces week-end au départ de Paris. Merci à tous les participants, et notamment aux co-organisateurs, des deux éditions de la Cocurace que je n'ai pas cité : Laura, Axel, Valentin D., Julien, Johan, Manon, Eva, Tristan, Pauline, Suzanne, Paul, Rémi, Valentin J., Erwan, Angéla, Gabin, Candice, Lorelei, Claire et Florent. Merci pour toutes ces soirées, ces discussions, ces projets et ces débats.

Ces années de thèse m'ont fait changer de maison ou d'appartement de nombreuses fois. Je souhaite chaleureusement remercier ceux qui m'ont accueilli pendant ma première année de thèse. Merci encore à toi Gautier pour ton énergie, ta curiosité et ta bienveillance ; à Jean-Claude et Marie-Angelle pour leur accueil à bras ouverts ; à toi aussi Félix ainsi qu'à toute cette famille que j'adore. Cette première année passée avec vous à s'occuper des poules et du jardin était géniale et voir aujourd'hui la maison continuer à bourdonner, malgré l'incendie, me remplit de bonheur.

Merci à toi Claire pour ton accueil, pour ces discussions et ces parties d'échecs ainsi qu'à vous trois, Jean-Charles, Madeleine et Oscar pour m'avoir accepté dans votre foyer. Sans vous cette période parisienne aurait été bien plus compliquée.

Merci à mes autres colocs successifs, notamment Vicky, Damien, Ludivine et Raphaël. Je reviens de Paris avec d'innombrables anecdotes grâce à vous !

J'aimerais sincèrement ne pas remercier la ville de Paris pour ses voitures, sa pollution, son bruit, ses visages dans le métro, ses légumes calibrés et sa météo.

Je souhaite également remercier ceux qui m'ont aidé, encadré et soutenu lors de mon parcours universitaire. A commencer par Rumsais Blatrix qui me permit d'effectuer une première expérience dans le monde de la recherche. Merci aussi à Emmanuelle Jousselin et Andrea Sanchez-Meseguer pour leur l'encadrement absolument génial et pour m'avoir permis de mettre un premier pas dans le monde de la systématique et de la macroévolution. Merci à Fabien Condamine, l'une des personnalités universitaires les plus inspirantes que j'ai rencontré, pour ces discussions nombreuses et pour le soutien que tu m'as apporté à de nombreuses reprises.

Merci à Orianne, Mélodie et Vincent, les meilleurs co-bureaux que quelqu'un puisse rêver. Merci à l'ensemble des membres du CBGP, où j'ai passé mon année de master 2 ainsi que ma première année

de thèse. Particulièrement aux habitués de la bière party pour les discussions, bières, parties de ping-pong et de baby-foot endiablées, pour ces virés en ville jusqu'au bout de la nuit (merci notamment à Laure pour ces moments de folie, dans la droite lignée des sorties de Master).

Merci à Jérôme, Médine et Joël pour leur bonne humeur et leur accueil au Muséum où l'intégration n'a rien d'aisée, notamment du fait de ces longs couloirs sinueux. Merci à Charline et Camille pour ces verres dans le quartier latin. Merci à Liliana Ballesteros-Mejia pour cette bonne humeur au quotidien et pour l'aide que tu m'as apporté dans mon travail.

Merci à l'ensemble de mes collaborateurs qui ont participé de près ou de loin au travail effectué dans cette thèse. Notamment à tous les passionnés de saturniidés qui ont envoyé des échantillons ; à Jean Haxaire pour son accueil incroyable et pour m'avoir fait vivre ma première chasse de nuit ; à Stefan Naumann également. Je souhaite aussi remercier sincèrement Jean-Yves Rasplus et Astrid Cruaud qui sont à l'origine de l'utilisation des marqueurs UCEs ; merci pour votre aide, notamment en début de thèse. Je remercie également l'ensemble des membres de l'ANR SPHINX et du projet ACTIAS ainsi que l'ensemble des personnes impliquées sur les missions de terrain que j'ai effectué pendant ma thèse.

Enfin, je veux remercier très chaleureusement Marianne Elias et Rodolphe Rougerie, mes directeurs de thèse. D'abord pour m'avoir fait confiance ; après tout nous ne nous connaissions pas (Rodolphe, je ne t'avais vu qu'en visio !!) et vous n'avez pas hésité à me proposer le financement de thèse que Rodolphe venait d'obtenir. Vous m'avez permis de m'éclater à étudier ce modèle merveilleux alors que je n'y connaissais rien avant d'avoir commencé ma thèse. Merci d'avoir été à la fois exigeants, compréhensifs et encourageants, tour à tour, pendant ces années. Ça a été un réel plaisir de travailler et d'apprendre à vos côtés.

Introduction générale ……..... **1**

I] Diversité du monde vivant, émergence des méthodes de classification et avènement de la notion d'évolution 1

II] Les phylogénies, supports incontournables de l'étude du vivant ……..... 5

III] Les Saturniidés ……..... 11

Chapitre 1 L'influence des traits d'histoire de vie sur la diversification spatiale et temporelle des saturniidés (Bombycoidea: Saturniidae) **17**

Introduction ...…….... **19**

Materials and methods …… **21**

Genomic data & Phylogenomic inferences ……. 21

Divergence time estimates ……... 24

Species-level phylogenies inference …… 24

Biogeographical analyses …… 25

Traits measurement, compilation and evolutionary analyses …….. 25

Diversification analyses ……... 27

Results and Discussion …… **28**

Phylogenomics illuminates the evolution and the relationships of extant genera of wild silkmoths 28

A global understanding of wild silkmoths' diversification in space and time ……... 35

Life-history traits played a key role in driving the diversification of wild silkmoths .. 41

Conclusion: Insights into the evolution of capital-breeding insects …….. **48**

Article à soumettre : Life-history innovations drove spatial and temporal diversification in capital-breeding insects 75

Article soumis : A global food plant dataset for wild silkmoths and hawkmoths, and its use in documenting polyphagy of their caterpillars (Lepidoptera: Bombycoidea: Saturniidae, Sphingidae) 99

Chapitre 2 Le conservatisme de niche et des évènements de dispersification façonnèrent la diversification spatiale et temporelle des saturniidés du genre Copaxa **113**

Introduction …….. **116**

Materials and Methods …… **117**

Species diversity in genus Copaxa ……... 117

Molecular data acquisition ……... 119

Phylogenetic inferences ……... 120

Divergence time estimations …… 121

Historical biogeography ……... 122

Altitudinal preferences …… 123

Diversification rates …… 123

Results ……... **124**

Species diversity in genus Copaxa ……... 124

Sequencing of phylogenomic data ……... 125

Phylogenetic inferences ……... 126

Estimates of divergence times ……. 130

Spatial, temporal and altitudinal diversification dynamics of Copaxa …… 130

Discussion …… **134**

Conclusions ……... **137**

Chapitre 3 Inférence de la mégaphylogénie des Saturniidae ……... **155**

Introduction ……..... **157**

Matériels et Méthodes …….. **165**

Définition des Unités Taxonomiques Opérationnelles (OTUs) …… 165

Données nucléotidiques et stratégie de séquençage génomique …….. 166

Définition du backbone et des subtrees …… 168

Inférences de la topologie du backbone et des subtrees ……... 169

Datation et assemblage de l'arbre …… 172

Mesure des taux de diversification ……... 174

Plateforme de calcul ……. 174

Résultats …… **175**

Définition des Unités Taxonomiques Opérationnelles (OTUs) …… 175

Données nucléotidiques ……... 176

Définition du backbone et des subtrees …… 179

Inférences de la topologie du backbone et des subtrees ……... 179

Datation et concaténation de l'arbre …… 186

Mesure des taux de diversification ……... 190

Discussion ……. **192**

Définition des Unités Taxonomiques Opérationnelles …… 192

Topologie du backbone …… 194

Utilisation des UCEs dans l'inférence des nœuds intragénériques ……. 195

De la difficulté de mesurer les taux de diversification ……. 196

Avantages et désavantages du pipeline backbone + subtrees présenté ici …… 197

Conclusion …… **199**

Annexe 1 – Phylogénies des subtrees …… 207

Annexe 2 – Pipeline phylogénomique …… 255

Conclusion générale ……. **285**

I] Avancée des connaissances systématiques sur la famille des Saturniidae & utilisation des marqueurs UCEs 285

II] L'influence des traits d'histoire de vie sur la dynamique spatiale et temporelle des Saturniidae 286

III] Une avancée significative vers la génération d'une phylogénie complète de la famille des Saturniidae 287

Introduction générale

I] Diversité du monde vivant, émergence des méthodes de classification et avènement de la notion d'évolution

L'étude du monde animal

1 552 319. C'est le nombre d'espèces que Zhang (2011) a comptabilisé dans son référencement de la diversité spécifique du monde animal. L'immensité de ce chiffre, obtenu au terme d'un exercice fastidieux, est particulièrement difficile à appréhender. Pourtant, il est très loin de rendre compte de l'ensemble de la diversité du règne animal. De nos jours, près de 20 000 espèces sont décrites chaque année (IISE 2011) par près de 40 000 taxonomistes (Costello et al. 2013) et le nombre d'espèces d'animaux sur Terre a été estimé à près de 7,77 millions (8,7 +/- 1,3 millions d'espèces d'eucaryotes ; Mora et al. 2011). Les générations successives de naturalistes n'auraient donc décrit que près de 18% de la biodiversité animale. Au rythme actuel des descriptions, plus de 300 années de travail seraient nécessaires pour documenter le monde animal qui nous entoure (combien en restera-t-il alors ?). Et encore, ces estimations ne concernent que les espèces vivantes. Selon Newman (1997), les espèces qui vivent aujourd'hui sur Terre ne représenteraient que 1/1000^{e} des espèces ayant vécu (voir aussi Raup 1986), élevant le nombre d'espèces d'animaux existants ou éteints à près de 9 milliards. Ces estimations, qui certes utilisent une notion d'espèce cloisonnée (cela fait-il sens de catégoriser des individus d'une même lignée mais séparés par plusieurs centaines de générations dans des espèces distinctes ?), ont le mérite de rendre compte du fait que la diversité actuelle est le résultat de centaines de millions d'années d'évolution ; là encore, un chiffre impossible à concevoir pour notre esprit.

Sans se rendre compte de cette immensité, l'Homme s'est néanmoins largement intéressé à lever le voile sur les mystères du monde vivant. Mais alors, pourquoi étudier le vivant ? Par besoin de comprendre nos origines, notre physiologie, notre fonctionnement social ? Pour s'en servir de modèle ? Pour s'alimenter, survivre ? Voire, pour s'en protéger ? Certainement pour toutes ces raisons ; mais la réponse que je privilégie est peut-être celle de Jules Renard, écrivain du XIXe siècle, qui disait « La vie, je la comprends de moins en moins, et je l'aime de plus en plus » ; et si nous étudions le monde vivant car il constitue une source d'émerveillement sans fin ? Cette question, vaste et complexe s'il en est, ne sera pas discutée dans cette thèse bien que l'ensemble de mes travaux s'inscrivent dans une tradition naturaliste et scientifique qui fait aujourd'hui l'unanimité et apparait de plus en plus indispensable dans

nos sociétés, profondément enclines à l'accélération et à la détérioration des mondes qui les entourent. Rappelons-nous que l'étude du vivant peut également constituer une source d'humilité : le plus beau fruit des travaux effectués par les générations successives de naturalistes est certainement que notre espèce n'est qu'une simple lignée évolutive, comme il en existe des centaines de milliers.

Classifier le monde vivant

Dès l'antiquité, la diversité des formes biologiques a fasciné et fait l'objet de descriptions et d'analyses qui ont abouti à des regroupements – ou classifications – d'entités biologiques sur la base de caractères communs (par ex. présence d'un squelette osseux interne pour les Vertébrés). Il est probable que les premières classifications soient apparues avec l'émergence même du langage, afin de faire référence à des organismes qui partagent un ou des critère(s) simple(s) (apparence, goût, etc…), et utilisaient un principe naïf d'assimilation (par ex. groupant les oiseaux et chauve-souris ; les cétacés avec les poissons). Les premiers écrits connus proposant une classification du vivant ne remontent cependant qu'à Aristote (385-322 av. J.-C.) qui écrivit une série de neuf livres regroupés dans l'ouvrage zoologique *Histoire des animaux*. Ces textes antiques, décousus et largement superficiels, s'apparentent à une liste d'observations naturalistes (par ex. « La femelle éléphant devient sexuellement réceptive à l'âge de 10 ans pour les plus jeunes, à 15 ans pour les plus vieilles alors que les mâles sont réceptifs sexuellement à partir de 5 ou 6 ans ») qui sont, pour beaucoup, dépassées aujourd'hui. Cette série de livres n'en est pas moins remarquable et mentionne un nombre important de formes d'oiseaux, de mammifères, de poissons et même d'insectes. De façon remarquable, Aristote y définit certains concepts comme celui des vertébrés ou la distinction entre viviparité et l'oviparité qui se révèlent être encore utilisés. Cependant, aucune classification formelle du vivant ne figure dans ses écrits.

Les ébauches de classification du monde vivant furent ensuite nombreuses (*e.g.* celles de Théophraste (371-288 av. J.C.) en botanique et de Barthélémy l'Anglais avec l'ouvrage *Livre des propriétés des choses* publié en 1247) mais c'est le système de nomenclature proposé par Carl von Linné dans son livre *Systema Naturæ* qui révolutionna la manière de classer les entités du monde vivant. Linné proposa de désigner les espèces vivantes à l'aide d'un nom binominal latinisé et de les hiérarchiser au sein de genres, familles, ordres, classes, embranchements et règnes. La force de cette classification réside dans sa simplicité, sa praticité – puisqu'elle peut s'appliquer, sans distinction, à l'ensemble des organismes vivants – et permit aux naturalistes des différents pays de communiquer plus aisément. Homme de son temps, la démarche de Linné avait pour objectif d'identifier et de classifier l'ensemble de l'œuvre de divine. Sa conception des espèces est dite « fixiste » : « Nous comptons autant d'espèces qu'il y a eu au commencement de formes diverses créées ».

Émergence et avènement du concept d'évolution

La description du monde vivant ne consiste cependant pas seulement à compter les espèces. Cela passe aussi par l'identification des différences et des ressemblances entre les entités qui le composent (Figure 1). C'est cette analyse des caractères communs entre les taxons qui permit aux naturalistes de poser les premières pierres de la théorie moderne de l'évolution. Georges-Louis Leclerc de Buffon, dans l'ouvrage *Histoire naturelle, générale et particulière, avec la description du Cabinet du Roy*, publié au XVIIIe siècle, suggérait ainsi que chaque individu était issu d'un modèle originel (divin) propre à son espèce et que les variations que chacun pouvait constater étaient le résultat « d'améliorations ou de perfections » accumulées au fil des générations. Un des exemples qu'il choisit est celui de l'âne, dont Buffon estima qu'il était probablement un cheval « dégénéré », résultat de changements accumulés au cours des générations. Buffon ira d'ailleurs plus loin dans ses observations en soulignant les ressemblances existantes entre l'Homme et les animaux. Il compara par exemple l'Homme au cheval (encore une fois) : « Prenez le squelette de l'Homme, inclinez les os du bassin, accourcissez les os des cuisses, des jambes et des bras, allongez ceux des pieds et des mains, soudez ensemble les phalanges, allongez les mâchoires et raccourcissant l'os frontal, et enfin allongez aussi l'épine du dos, ce squelette cessera de représenter la dépouille d'un Homme, et sera le squelette d'un cheval ». Plus tard, au début du XIXe siècle, Jean-Baptiste de Lamarck, le premier scientifique à considérer la biologie comme science à part entière, conceptualisa plus formellement ces observations et théorisa que l'évolution des organismes allait de pair avec leur complexification. Allant à l'encontre du dogme ecclésiastique, Lamarck alla jusqu'à affirmer que des formes de vie primaires apparaissaient continuellement, que la vie n'était pas une œuvre divine mais le résultat de milliers de génération de complexification. Cette complexification, selon Lamarck, permettrait aux lignées de s'adapter à leur environnement, les générations successives transmettant les adaptations qu'elles ont pu accumuler à leur descendance (un processus nommé plus tard 'hérédité des caractères acquis', par August Weismann).

La théorie de Lamarck ne permet cependant pas d'expliquer l'existence de certaines formes rudimentaires de vie dont le fort apparentement avec d'autres entités biologiques plus complexes ne fait pas de doute. Prenons l'exemple des poux (Insecta : Phthiraptera). Ces derniers sont aptères (absence d'ailes), un caractère généralement attribué à des lignées peu complexes d'insectes, pourtant ils partagent la même articulation mandibulaire dicondylienne que les Hyménoptères ou les Coléoptères. Selon Lamarck, les formes de vie les plus simples seraient issues de générations spontanées (cette perception est comparable à la théorie de la « soupe primitive » introduite au XXe siècle) qui ont accumulé encore peu de complexité. Dans notre cas, cela impliquerait que les poux aient remarquablement convergé, à partir d'une génération spontanée, vers un plan d'organisation similaire aux Dicondylia sans pour autant avoir acquis la capacité de voler. Ce genre d'exemples met à mal la théorie lamarckienne ; en réalité, la notion de complexification qu'il introduisit ne repose que sur une

vision subjective des caractères qui place l'Homme au sommet absolu de la pyramide de l'évolution et de la complexité.

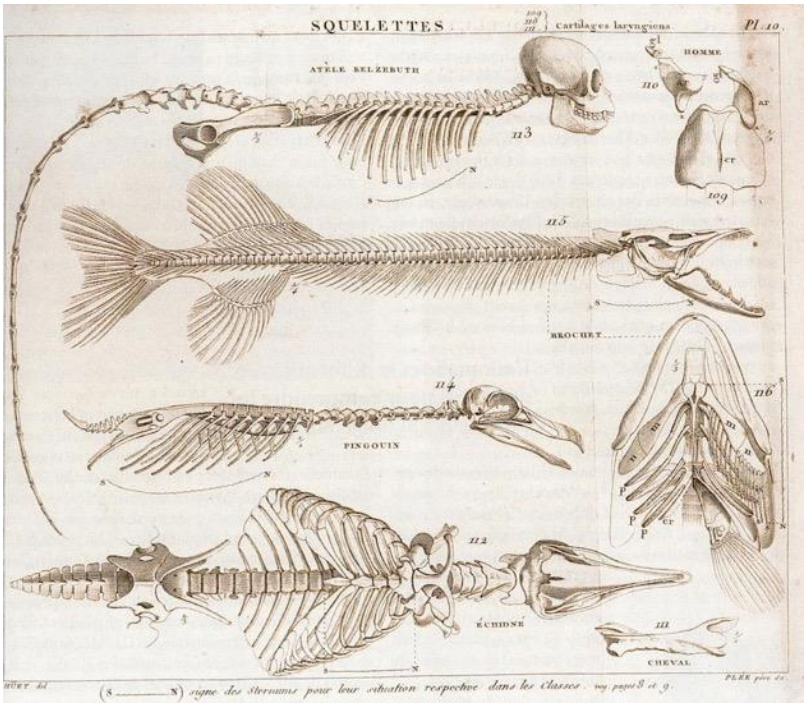


Figure 1 – Planche d'anatomie comparée issue de l'ouvrage Philosophie anatomique des organes respiratoires sous rapport de la détermination et de l'identité de leur pièces osseuses – Planche 10. Du haut vers le bas, un singe atèle, un brochet, un pingouin et un échidné. Étienne Geoffroy Saint-Hilaire, 1818.

Ce n'est que près de 50 ans après le livre *Philosophie zoologique* de Lamarck qu'une autre conception de l'évolution fut introduite par Charles Darwin et Alfred Wallace (1858), puis largement détaillé dans l'ouvrage de Darwin *On the Origin of Species*, en 1859. Selon Darwin, le moteur de l'évolution (ou plutôt de la « descendance avec modification », terme qu'il remplaça plus tard par « évolution ») est la sélection naturelle. La théorie de Darwin repose sur le fait que tous les individus d'une même espèce diffèrent légèrement entre eux (la variation phénotypique) et que seule une partie de ces individus, ceux qui sont les mieux adaptés à leur environnement, réussissent à survivre et se reproduire (ils ont une meilleure valeur sélective), transmettant à leur descendance les caractères responsables de leur succès (héritabilité des traits). Selon Darwin, les individus d'une même espèce sont en compétition pour les ressources et pour se reproduire dans un environnement dont les ressources sont limitées (autre condition *sine qua non* pour que les lignées évoluent). Sa théorie permet, contrairement aux idées de Lamarck, de comprendre l'évolution vers des formes de vie plus simple ainsi que l'apparition de traits extravagant : bien que ceux-ci n'apportent pas une meilleure probabilité de survie aux individus qui l'arborent (par ex. les queues du paon), ces traits peuvent conférer un avantage dans la recherche de partenaire et donc induire un meilleur succès reproducteur (ce que Darwin appela la sélection sexuelle). La théorie de « la descendance avec modification » eut un impact considérable dans les domaines de la paléontologie (par

ex. travaux de Charles Lyell), de l'agronomie, de l'écologie, de la sociologie et dans la société dans son ensemble.

II] Les phylogénies, supports incontournables de l'étude du vivant.

Premières phylogénies et données morphologiques

Dès l'émergence des notions d'évolution, les scientifiques utilisèrent la forme d'un arbre pour représenter les liens de parenté entre les entités du vivant. Les premiers arbres étaient hiérarchiques et sous-tendaient que les formes de vie les plus complexes étaient issues d'être primitifs. C'est par exemple le cas de la représentation en « arbre-tableau » de Lamarck (dans l'ouvrage *Philosophie Zoologique*, 1809). Darwin y préféra une représentation d'arbre dans lequel les lignées existantes sont au même niveau (Figure 2), récusant toute hiérarchie entre les espèces, alors que les lignées éteintes sont à des niveaux inférieurs : « Les rameaux verts et bourgeonnants peuvent représenter les espèces existantes […]. (Ces) nombreuses branches […] tombées, de taille variable, peuvent représenter ces ordres, ces familles et ces genres tout entiers, qui n'ont plus de représentants vivants ».

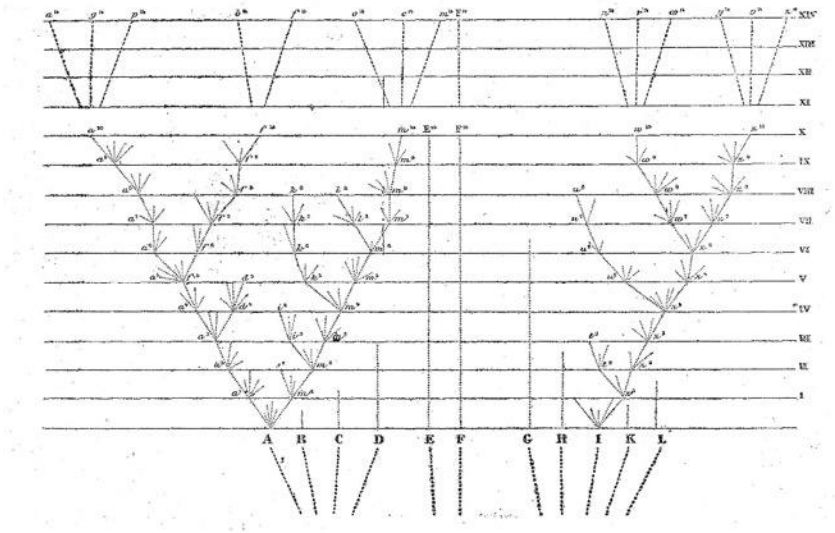


Figure 2 – Arbre phylogénétique illustré dans On the Origin of Species, Darwin (1859). Darwin illustre ici les histoires évolutives de 11 lignées évolutives qui ont eu différents succès évolutifs. Il est remarquable que, déjà dans cette illustration, la majorité des lignées soient représentées comme disparues (lignées B, C, D par exemple, mais aussi d, s, i) ; les seules lignées existantes sont celles dont les adaptations ont permis la survie. Darwin pose ici les bases de la classification moderne «Hence the six new species descended from (I), and the eight descended from (A), will have to be ranked as very distinct genera, or even as distinct sub-families ».

La théorie de Darwin et la représentation des apparentements entre les entités du vivant qu'il introduisit entrainèrent un changement fondamental, mais tardif, dans les méthodes de classification des espèces. Ce n'est effectivement qu'en 1950[1] que furent établies les bases de la systématique moderne, la

[1] Si Willi Hennig définit les concepts de la cladistique dès 1950 dans l'ouvrage *Grundzüge einer Theorie der phylogenetischen Systematik,* celui-ci ne fut traduit qu'en 1966. A sa publication, il devint l'ouvrage de référence des systématiciens.

cladistique, par Willi Hennig dans son ouvrage *Grundzüge einer Theorie der phylogenetischen Systematik* (« Fondements d'une théorie de la systématique phylogénétique »). Cette méthode de classification ne considère que les clades constitués d'un ancêtre et de l'ensemble de ses descendants – *i.e.* les groupes monophylétiques – invalidant de fait des classes historiques telles que celle des poissons ou des reptiles. Les classifications cladistiques considèrent donc que les entités biologiques sont des éléments interconnectés et indissociables et sont le reflet des millions d'années d'évolution qui façonnèrent les organismes ; leur établissement repose dorénavant sur l'identification de caractéristiques homologues (*i.e.* issue d'un même ancêtre, en opposition à la convergence) : les synapomorphies. Les principes cladistiques firent l'unanimité au cours du XXe et ancrèrent l'utilisation des arbres phylogénétiques pour représenter le concept central de l'évolution, la « descendance avec modifications ».

Pendant la majorité des XIXe et XXe siècles l'établissement de phylogénies reposaient sur des critères morphologiques, comportementaux et géographiques. Bien que l'identification de caractères morphologiques indépendants soit particulièrement complexe, les arbres phylogénétiques s'étoffèrent et se précisèrent grâce aux travaux des systématiciens qui, au fil des années, développèrent des techniques de mesures (par ex. morphométrie géométrique ; Adams *et al.* 2004) et de reconstruction phylogénétique de plus en plus élaborées (*e.g.* phénogrammes, méthodes d'agrégation). De nos jours, si les données morphologiques sont encore utilisées afin de construire des arbres phylogénétiques, notamment dans l'inférence de chronogrammes afin d'inclure des représentants fossiles (*i.e. total evidence dating*; voir par exemple Ronquist *et al.* 2012), l'avènement et le développement rapide des techniques moléculaires et des méthodes analytiques associées ont révolutionné notre façon d'inférer des phylogénies.

Utilisation des données moléculaires dans l'inférence des phylogénies

L'utilisation de données moléculaires dans l'inférence de relations phylogénétiques (par ex. Sibley & Ahlquist 1984, 1990) constitue un point de bascule majeur dans le domaine de l'Écologie et de l'Évolution. Similairement aux approches utilisant des données morphologiques, il est effectivement possible d'estimer les relations entre les entités du vivant en fonction de leur similarité moléculaire. Dans ce cas, une séquence ADN est assimilée à une matrice morphologique : chaque paire de base constituent un caractère indépendant et les 4 bases nucléotidiques (adénosine (A), thymine (T), guanine (G) et cytosine (C)) sont les 4 états de caractères possibles. La force principale de cette approche est l'universalité des macromolécules biologiques que sont l'ADN, l'ARN ou les protéines. L'utilisation de ces séquences a permis une multiplication du nombre de caractères étudiés en phylogénie (d'une centaine à plus d'un millier ; Delsuc *et al.* 2005) dont l'hétérogénéité des taux d'évolution permet l'inférence de relations phylogénétiques entre des taxons particulièrement divers, éloignés comme très apparentés (par exemple appartenant à une même espèce).

Depuis l'utilisation de ces séquences moléculaires, les scientifiques n'ont cessé de progresser dans la compréhension de l'arbre du vivant. Les phylogénies inférées à partir de telles données ont par exemple permis de redéfinir la position des Gnétophytes (Chaw *et al.* 2000) au sein des Gymnospermes, pourtant longtemps considérées comme groupe frère des Angiospermes sur la base de caractères communs comme la double fécondation (Friedman *et al.* 1996). Après examen, il s'est avéré ensuite que la double fécondation présente au sein des Gnétophytes était sensiblement différente de celle des Angiospermes : chez ces derniers, un gamète engendre le zygote avec l'oosphère tandis qu'un autre gamète se développe, avec les noyaux polaires, en albumen (tissu de réserve) ; chez les Gnétales, le zygote surnuméraire avorte. L'étude des données moléculaires permit également de mesurer le très fort apparentement de l'espèce humaine avec les grands singes (<2% ; Chen & Li 2001 de divergence moléculaire dans les marqueurs considérés entre *Pan* et *Homo*), autrefois classés au sein de la famille des Pongidae. Aujourd'hui, l'orang-outan, le gorille, le chimpanzé et le genre *Homo* appartiennent à la même famille des Hominidae

L'utilisation de marqueurs moléculaires pour construire des matrices de caractères a été, dans un premier temps, largement généralisée grâce au développement des techniques de séquençage Sanger (notamment l'introduction de traceurs fluorescents en lieu et place de marqueurs radioactifs) et à l'adaptation des techniques de Réaction en Chaîne par Polymérase (PCR en anglais) pour le séquençage. Cependant, cette technique nécessite des temps de manipulation importants et a un coût élevé, ce qui explique le nombre limité d'échantillons étudiés au début des années 2000 ainsi que la faible quantité et variété des marqueurs utilisés. Malgré les nombreux succès qu'a eu l'utilisation de ce type de données, ces dernières ne sont pas exemptes de biais et le signal phylogénétique qu'elles contiennent s'est parfois avéré être insuffisant (puisqu'il n'y a que 4 états de caractère, la probabilité de retrouver des états de caractères homoplasiques[2] est élevée) pour résoudre certains nœuds. C'est le cas de la monophylie des rongeurs, qui fut établie avec de forts supports par Murphy *et al.* (2001) à partir d'une matrice de près de 10 000 paires de bases pour 64 espèces alors qu'elle avait été rejetée 5 années auparavant par D'Erchia *et al.* (1996) avec une matrice mitochondriale comprenant 16 espèces de mammifères. Un tel exemple démontre l'importance de la taille et de la complétion du jeu de données dans les inférences phylogénétiques.

[2] L'homoplasie est la similitude d'un état de caractère (nucléotide en biologie moléculaire) chez différents taxons qui, contrairement à l'homologie, ne provient pas d'un ancêtre commun. Un fort taux d'homoplasie altère le signal phylogénétique et peut impliquer l'inférence de nœuds incorrects.

Techniques de séquençage de nouvelle génération et leur application en phylogénomique

La généralisation de la technique de séquençage conçue par Sanger (1977) et le développement de protocoles optimisés ont permis une réduction considérable des coûts de séquençage dans les années 2000 (Church 2006) et la constitution de larges jeux de données génétiques (par ex. Bininda-Emonds *et al.* 2007). Mais c'est l'émergence des nouvelles techniques de séquençages, dites « haut-débit », qui marqua le basculement dans « l'ère phylogénomique » (Delsuc *et al.* 2005 ; Von Bubnoff 2008). Ces techniques, commercialisées à partir de 2005, reposent toutes sur une succession de cycles de lavage et d'identification des bases nucléotidiques à l'aide de signaux lumineux (voir Metzker 2010 pour plus de détails), contrastant avec la méthode Sanger qui s'appuie sur une électrophorèse. Ces avancées technologiques eurent un écho considérable au sein de la communauté des chercheurs en Écologie et Évolution qui se les approprièrent et développèrent une multitude de méthodes pour isoler et amplifier des fragments génomiques informatifs dans l'inférence de phylogénies.

Le choix des fragments génomiques dépend alors de la question (par ex. pour étudier les informations phylogénétiques discordances des différents *loci*), du cadre taxonomique (par ex. intraspécifique) et de la faisabilité. Par exemple, le séquençage de l'ADN associé aux sites de restrictions (RADseq) permet la génération de milliers de *loci* indépendants recouvrant l'entièreté du génome (Lewis *et al.* 2007 ; Miller *et al.* 2007) ; une technique le plus souvent utilisée pour inférer des relations intra-spécifiques ou entre espèces proches (Emerson *et al.* 2010 ; Dupuis *et al.* 2018) car le nombre de *loci* partagés décroit avec la distance entre les taxons (McCormack, *et al.* 2012). Le séquençage de RAD est particulièrement intéressant car il ne nécessite pas de disposer de génomes de références de taxons apparentés grâce à l'universalité des sites de restriction, ce qui contraste avec les approches utilisant les techniques d'hybridation ciblée (« Hybridization-Based Target Enrichment » ; Figure 3). Celles-ci reposent effectivement sur l'alignement de plusieurs génomes de référence dans l'identification de séquences adaptées au cadre taxonomique d'intérêt. Malgré ce prérequis, leur succès grandissant s'explique par l'augmentation du nombre de génomes publiés et, surtout, par la possibilité de générer des marqueurs très divers : gènes, *loci* très conservés du génome (*Ultra Conserved Elements* - UCEs), séquences riches en SNPs, etc… Contrairement aux approches dites de « *shotgun* », elles permettent de concentrer l'effort de séquençage sur les régions génomiques d'intérêt et ainsi d'augmenter le nombre d'échantillons traités simultanément (voir par ex. Cruaud *et al.* 2019) tout en allégeant le traitement bioinformatique grâce à l'identification simple des régions orthologues par comparaison aux séquences de référence. Par ailleurs, il a été montré que ces techniques sont très efficaces dans le traitement d'échantillons anciens (par ex. Briggs *et al.* 2009), conservés depuis plusieurs décennies (Yeates *et al.* 2016) et nécessitent une très faible quantité initiale d'ADN (Cruaud *et al.* 2019). Leur utilisation, qui croît d'années en années, permit par exemple d'inférence des relations phylogénétiques profondes au

sein des papillons (Papilionoidea ; Espeland *et al.* 2018) ou entre les grands clades d'Oiseaux (Gilbert *et al.* 2018).

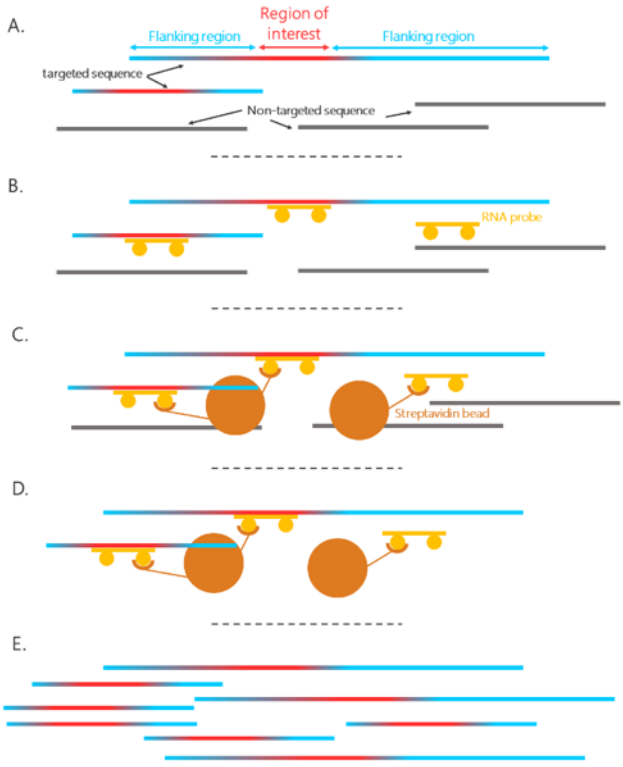


Figure 3 – Schéma présentant les étapes des méthodes « Hybridization-Based Target Enrichment ». (A) ADN double brin préparé à partir d'ADN génomique. (B) Introduction et hybridation des probes *ARN aux séquences d'intérêt. (C) Introduction de billes magnétiques de streptavidine. (D) Isolement des séquences d'intérêt à l'aide d'un aimant et de lavages successifs. (E) Désolidarisation des billes de streptavidine et amplificatin des séquences d'intérêt par PCR.*

La génération de jeux de données comprenant plusieurs centaines de marqueurs permet aussi l'étude des incompatibilités entre les topologies inférées à partir des différents *loci*, déjà identifiées avant l'ère phylogénomique (Delsuc *et al.* 2005). L'analyse de ces incompatibilités, l'identification des processus qui les génèrent et de leur impact sur l'inférence des phylogénies constituent aujourd'hui un champ de recherche particulièrement riche qui permet de redéfinir les processus de spéciation non pas comme un évènement soudain, binaire, mais comme un processus long, impliquant une multitude de flux géniques.

L'étude de l'évolution du vivant à partir des phylogénies

En plus de leur apport à la systématique, les phylogénies constituent un support matériel fondamental dans la compréhension de l'évolution et de la distribution des entités du vivant. Lorsqu'elles sont datées (chronogrammes), les branches des phylogénies ne sont alors plus proportionnelles aux taux d'évolution des séquences étudiées mais représentent les temps de divergence entre les entités considérées. Les questions pouvant être étudiées à partir de tels supports sont alors extrêmement variées, couvrant à la fois l'échelle intra-spécifique – on parle alors de microévolution – et inter-spécifique – on utilise alors le terme de macroévolution. De façon similaire à ce qu'il s'est passé pour les méthodes de reconstruction

phylogénétiques, l'augmentation du nombre de phylogénies s'accompagna du développement d'outils d'analyses macroévolutifs permettant de mesurer les taux de spéciation et d'extinction des lignées à partir de chronogrammes (Nee *et al.* 1994 ; Alfaro *et al.* 2009 ; Morlon *et al.* 2011 ; Stadler 2011 ; Rabosky *et al.* 2013 ; Höhna *et al.* 2019 ; Maliet *et al.* 2019). Avec ces méthodes, la communauté scientifique dispose ainsi d'outils puissants pour tester des hypothèses historiquement considérées dans la littérature (par ex. Valentine 1968) et pour lier les patrons de diversité existants à des processus évolutifs. Considérons par exemple le gradient latitudinal de la biodiversité (Willig *et al.* 2003 ; Hillebrand 2004 ; Barthlott *et al.* 2007 ; Mannion *et al.* 2014), *i.e.* l'augmentation du nombre d'espèces des pôles vers l'équateur (Latitudinal Diversity Gradient, LDG, en anglais). L'utilisation des méthodes macroévolutives permit de comprendre que les lignées tropicales de plusieurs groupes taxonomiques se sont diversifiées plus rapidement que les lignées tempérées apparentées (par ex. Cardillo *et al.* 2005 ; Jetz *et al.* 2012), expliquant, au moins en partie, le LDG (voir cependant Rabosky *et al.* (2018) qui inférèrent un patron inverse chez les Actinoptérygiens). En revanche, de tels modèles se limitent à la mesure des taux de spéciation et d'extinction des lignées du vivant, sans chercher à formellement tester l'effet de facteurs biotiques ou abiotiques sur leur diversification ; la diversification et les traits distinctifs des organismes ne peuvent ainsi qu'être corrélés à postériori en utilisant de telles méthodes. De fait, une étape importante dans le développement de ces méthodes macroévolutives fut l'intégration de l'effet de facteurs biotiques (Maddison *et al.* 2007 ; Fitzjohn *et al.* 2009 ; Fitzjohn 2010 ; Etienne & Rosindell 2012) et abiotiques (Condamine *et al.* 2013) dans les calculs des taux de diversification. Ainsi, les processus évolutifs pouvaient être testés de façon formelle et non plus sur la base de corrélations à posteriori. Par exemple, à partir de tels modèles, Lagomarsino *et al.* (2016) ont pu montrer que le grand succès évolutif des campanules dans les Andes s'expliquait à la fois par des facteurs biotiques – *i.e.* l'évolution vers des fruits de type baies et vers une association avec les oiseaux et les chauves-souris pour la pollinisation – et abiotiques – *i.e.* le refroidissement global au cours du Cénozoïque.

A partir des données de distribution, des liens phylogénétiques et des temps de divergence, il est également possible d'inférer l'histoire biogéographique des lignées du vivant (Ree 2005 ; Ree & Smith 2008 ; Matzke 2013 ; Beeravolu & Condamine 2016). Les méthodes d'inférence de biogéographie historique permettent de comprendre l'origine des espèces ou de clades ainsi que de mieux identifier quels sont les facteurs ayant façonné la distribution actuelle de la biodiversité. Par exemple, Meseguer *et al.* (2014) inférèrent l'histoire biogéographique complexe des millepertuis (Clusiaceae : *Hypericum*), en incorporant la distribution de fossiles connus, et estimèrent une origine holarctique de ce groupe mondialement distribué. Mais contrairement aux méthodes d'inférence des taux de diversification, les méthodes de reconstruction des zones de distribution ancestrales ne tiennent, dans leur majorité, pas encore explicitement en compte les facteurs biotiques dans leurs calculs, limitant de fait le pouvoir explicatif des analyses effectuées (voir cependant Klaus & Matzke 2020 ; Sukumaran *et al.* 2016 pour le développement récent de telles méthodes, encore marginalement applicables).

III] Les Saturniidés



afin d'effrayer un prédateur. © Armin Dettz

Les Saturniidae Boisduval 1837 (superfamille des Bombycoidea) constituent une famille de Lépidoptères particulièrement diversifiée taxonomiquement, morphologiquement, spatialement et biologiquement. Historiquement, les saturniidés s'imposèrent comme un groupe emblématique d'insectes, auquel de nombreux naturalistes, amateurs, universitaires se sont intéressés pendant des siècles. Cet intérêt s'explique en grande partie par leur taille remarquable : certains représentants de la famille des Saturniidae font partie des plus grandes espèces de Lépidoptères existants (au sein des genres *Attacus*, *Coscinocera*, *Arsenura* notamment). Les Saturniidae se distinguent aussi des autres Bombycoidea par la diversité exceptionnelle de leurs formes, des motifs alaires et des couleurs que certains arborent : de nombreuses espèces ont par exemple développé de longues queues sur leurs ailes postérieures (jusqu'à plus de 10cm pour certaines espèces), d'autres déploient des couleurs vives (par ex. *Eochroa trimenii*), des ocelles comme *Antherina suraka* (cf. photo de Armin Dettz © en haut à droite de la page) ou des ailes aux contours admirables comme *Almeidaia aidae* ou *Loxolomia johnsoni*. De tels exemples ne doivent cependant pas faire oublier la diversité exceptionnelle de formes cryptiques au sein de la famille, moins admirable mais tout aussi pertinentes d'un point de vue évolutif. Les saturniidés ont également développé des formes larvaires tout aussi exceptionnelles de par leurs couleurs et leurs stratégies d'évitement des prédateurs : certaines espèces se sont spécialisées, là encore, dans les formes cryptiques, maximisant le camouflage dans la végétation ou la litière alors que d'autres ont évoluées vers des stratégie grégaire (par ex. les chenilles du genre *Hylesia*) ou ont développé des scolis venimeux (par ex. le contact avec certaines chenilles d'espèces du genre *Lonomia* peut être mortel pour l'Homme). Enfin, plusieurs espèces de Saturniidae, sur différents continents, sont connues pour les soies dites « sauvages » produites à partir de leurs cocons et utilisées pour produire des textiles depuis plusieurs siècles (Peigler & Maldonado 2005). L'ensemble de ces caractéristiques, et le fait que ces papillons, dans leur majorité, sont actifs la nuit et sont relativement faciles à échantillonner à l'aide d'un piège lumineux, explique la popularité de la famille des Saturniidae.

Cette popularité a résulté en de nombreuses descriptions d'espèces et des classifications établies par les naturalistes et scientifiques du monde entier (par ex. Packard 1895 ; Bouvier 1927 ; Michener 1952 ; Rougeot 1955 ; Lemaire 1978, 1980, 1988, 2002 ; Bénéluz 1986 ; Naumann & Peigler 2001). De larges collections, privées et publiques, ont ainsi été établies et constituent aujourd'hui une source inestimable

d'informations ; la collection du Muséum national d'Histoire naturelle de Paris, qui renferme plus de 60 000 spécimens de Saturniidae, en est un exemple. La diversité des Saturniidae est, par conséquent, remarquablement bien documenté en comparaison à d'autres groupes d'insectes. Dans une démarche de synthèse des descriptions effectuées au sein des Bombycoidea, Kitching, Rougerie *et al.* (2018) listèrent 3454 espèces de Saturniidae, classées dans 180 genres, 11 tribus et 8 sous-familles. Nous estimons qu'aujourd'hui une très grande majorité de la diversité existante des Saturniidae a été décrite, notamment grâce à l'utilisation des code-barres ADN qui permirent de redéfinir certaines lignées cryptiques (Hebert *et al.* 2003, 2004 ; Ratnasingham & Hebert 2007) ; ainsi le déficit linnéen (manque de connaissance taxonomique ; Brito 2010) qui caractérise la majorité des groupes d'insectes est proche d'être résolu pour les Saturniidae. Un autre déficit, concernant les données de distribution (déficit wallacéen ; Diniz-Filho *et al.* 2010) est quant à lui toujours prononcé, bien que des campagnes de géoréférencement à partir des échantillons de musées soient en cours, notamment dans le cadre du projet ACTIAS (financement FRB/CESAB ; PI : Rodolphe Rougerie). Les données assemblées dans le cadre de ces campagnes sont compilées sur la plateforme BOLD (www.boldsystems.org ; Ratnasingham & Hebert 2007) et tendent à combler ce déficit wallacéen pour les Saturniidae. Enfin, la position phylogénétique de la famille des Saturniidae apparaît aujourd'hui résolue. Au sein des Lépidoptères, la superfamille des Bombycoidea, à laquelle appartiennent les Saturniidae, se positionne comme groupe frère des Lasiocampoidea qui ne comprennent qu'une seule famille, les Lasiocampidae (Figure 4). Avec les superfamilles Geometroidea et Noctuoidea, ils forment ensemble le clade dit des « macromoths », qui divergea il y a près de 92 millions d'années du reste des Lépidoptères. Au sein des Bombycoidea, les Saturniidae sont le groupe frère de la famille des Sphingidae (Figure 5) et forment ensemble un clade particulièrement riche à la base duquel Hamilton *et al.* (2019) inférèrent un shift positif des taux de diversification. Mais si nos connaissances sur la position phylogénétique de la famille des Saturniidae ont particulièrement progressé ces dernières années, celles concernant la phylogénie de la famille même demeurent toujours superficielles, se limitant à une l'analyse des positions d'un nombre limité de genres (Regier *et al.* 2002, 2008 ; Barber et al. 2015 ; Rubin *et al.* 2018 ; Hamilton et al. 2019), et révèlent un grand déficit darwinien (*i.e.* le manque de données phylogénétiques ; Diniz-Filho *et al.* 2013).

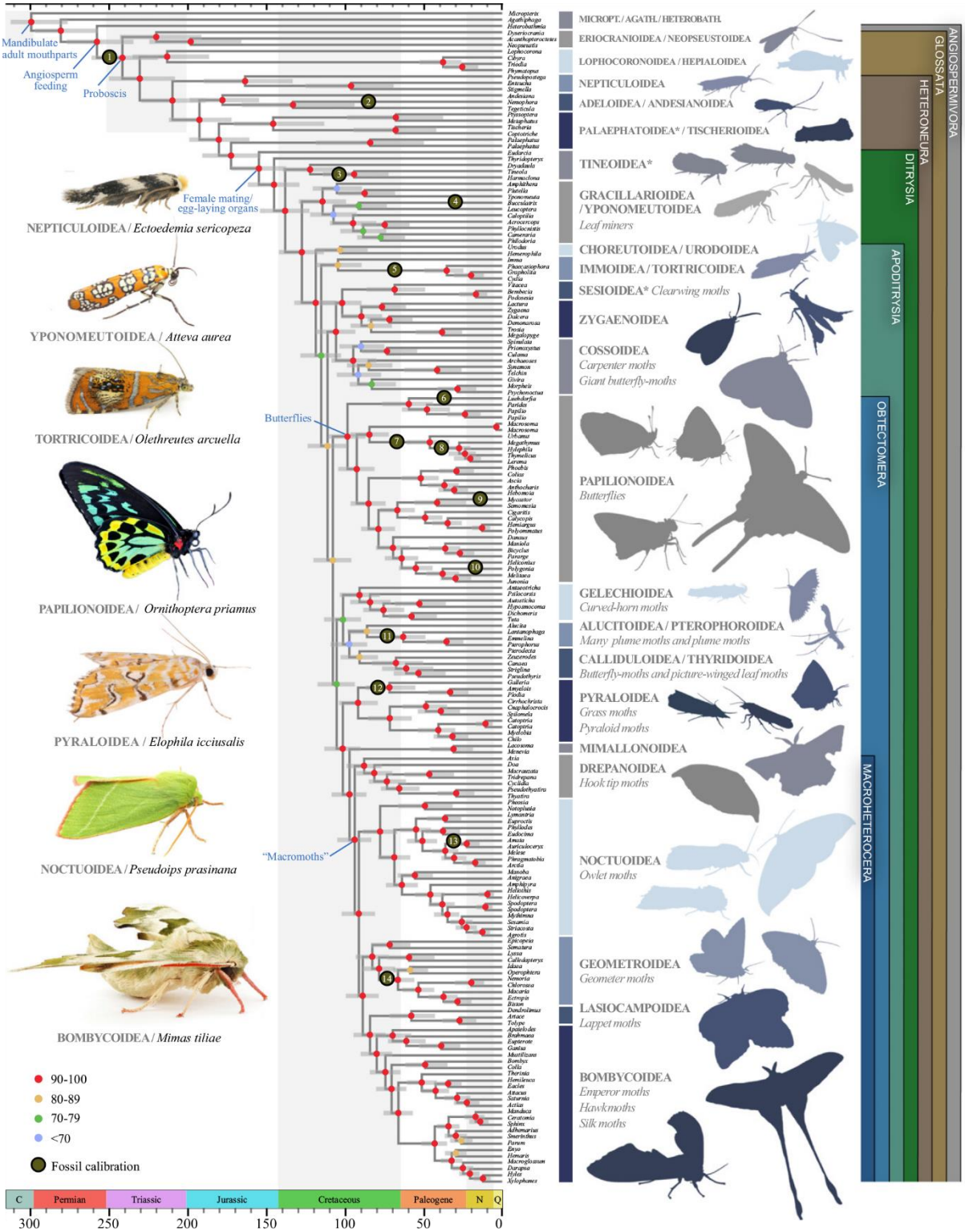


Figure 4 – Phylogénie datée des Lépidoptères telle qu'inférée par Kawahara et al. (2019). 6 espèces de Saturniidae furent inclues dans l'échantillonnage.

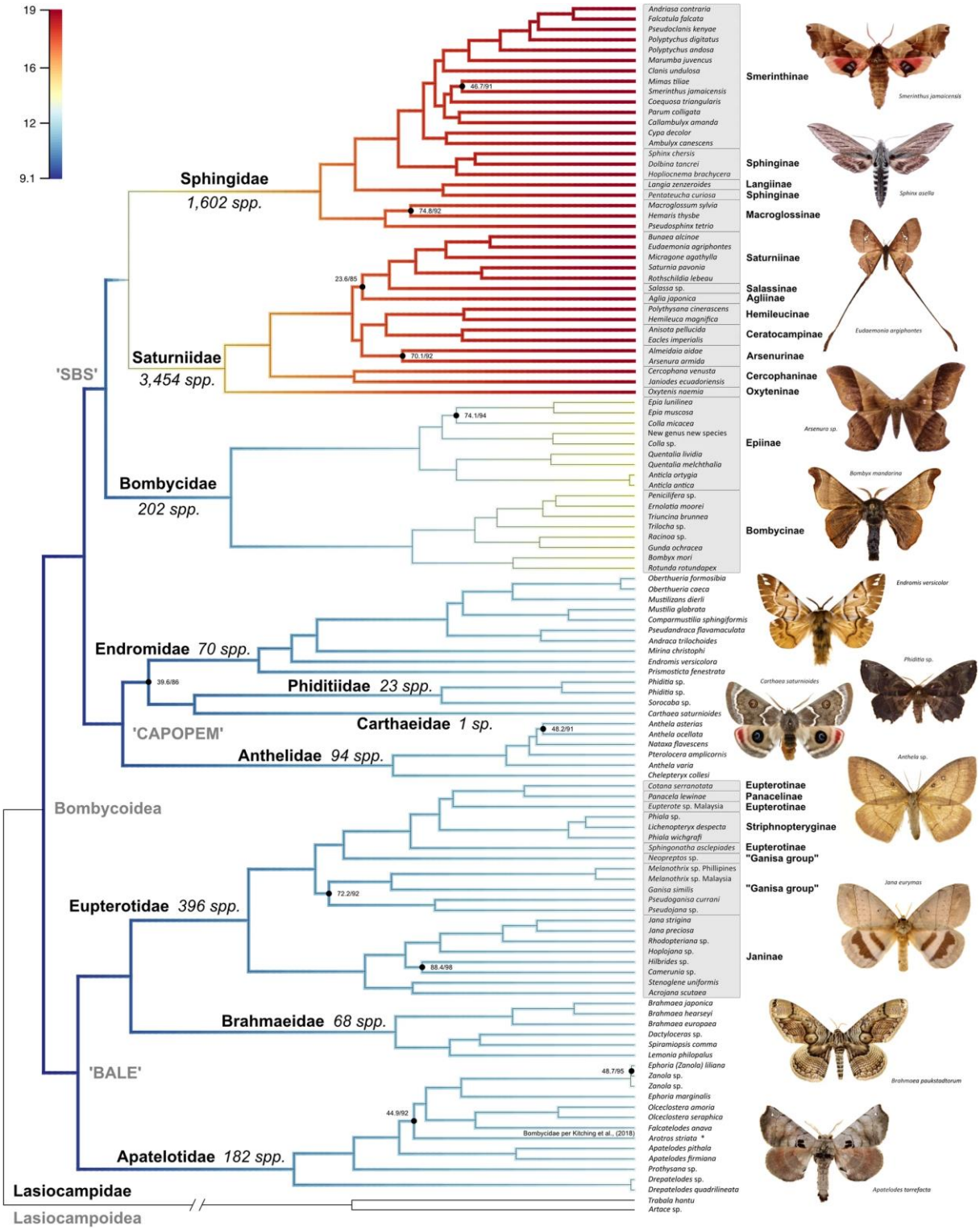


Figure 5 – Phylogénie des Bombycoidea telle qu'inférée par Hamilton et al. 2019. 16 espèces de Saturniidae furent inclues dans l'échantillonnage.

Objectifs de la thèse et présentation des chapitres

La remarquable diversité de la famille des Saturniidae, sa distribution mondiale et l'hétérogénéité des traits arborés par les différentes espèces soulèvent de multiples questions auxquelles j'ai essayé d'apporter des réponses lors de ma thèse :

- Chapitre 1 : Au sein des Lépidoptères, le fort degré de polyphagie des chenilles et leur faculté à tisser des cocons très denses, la faible capacité de dispersion et la très grande taille corporelle sont autant de traits présents chez les Saturniidae qui contrastent fortement avec le reste de l'ordre. Au cours de ce Chapitre, mon objectif fut de documenter l'histoire évolutive des Saturniidae et d'explorer l'effet de ces traits d'histoire de vie sur la diversification spatiale (histoire biogéographique) et temporelle (taux de spéciation et d'extinction).
- Chapitre 2 : Les *Copaxa* constituent un genre de Saturniini (Saturniinae : Saturniidae) particulièrement diversifié au sein de la région Néotropicale. Dans le Chapitre 1, j'ai montré que leur origine commune avec les représentants du genre *Saturnia* est Holarctique ; les *Copaxa* constituent donc un groupe pertinent pour étudier les patrons temporels de colonisation des différentes régions des Néotropiques dans le contexte d'une diversification dite « Into the tropics ». Puisque les espèces actuelles sont majoritairement distribuées dans les montagnes d'Amérique centrale et des Andes, j'ai notamment exploré le rôle de l'adaptation à l'altitude dans les dynamiques de diversification du groupe.
- Chapitre 3 : Les deux premiers Chapitres démontrent l'intérêt de travailler à des échelles différentes afin de pouvoir étudier des variations phénotypiques et biologiques d'amplitudes distinctes dans un contexte évolutif. Le développement d'une phylogénie comprenant l'ensemble des espèces de Saturniidae permettrait de pouvoir travailler à partir des deux échelles simultanément et de mieux appréhender l'origine de ces variations et leurs conséquences à différentes échelles. Faisant face à des défis d'ordre méthodologiques et opérationnels pour traiter des jeux de données génomiques et génétiques, mon objectif était ici de mettre en place une approche efficace et robuste pour générer ce type de « mégaphylogénie », encore inédite chez les insectes.

Chapitre 1

L'influence des traits d'histoire de vie sur la diversification spatiale et temporelle des saturniidés (Bombycoidea: Saturniidae)

A global analysis reveals that life-history innovations drove spatial and temporal diversification in capital-breeding wild silkmoths (Bombycoidea: Saturniidae)

Préambule

Dans ce premier Chapitre, j'ai d'abord généré un jeu de données génomiques (*Ultra Conserved Elements* - UCEs) comprenant l'ensemble des genres de Saturniidae décrits, inféré une série de phylogénies et proposé une redéfinition la classification de la famille des Saturniidae. A partir de ces résultats, j'ai ensuite estimé les temps de divergence, mesuré les taux de spéciation et d'extinction, inféré l'histoire biogéographique et exploré l'influence des traits biotiques (*i.e.* la taille corporelle, la capacité de dispersion, le degré de polyphagie et le mode de nymphose) et des évènements abiotiques majeurs sur les dynamiques spatiales et temporelles de diversification au sein de la famille des Saturniidae. Afin de limiter les biais d'inférence liés à l'utilisation d'une phylogénie de niveau supérieur (*higher-level phylogeny*), j'ai généré une phylogénie à l'espèce en utilisant une méthode de greffage (*grafting*) et développé une approche utilisant la randomisation pour prendre en compte l'incertitude sur l'estimation des aires géographiques ancestrales. Dans leur ensemble, les résultats présentés ici représentent une avancée majeure dans la compréhension de l'évolution des Saturniidae et, plus généralement, des Lépidoptères dont la reproduction est basée sur une stratégie de type « capital breeding », *i.e.* dont la reproduction est rapide et la durée de vie courte.

Les travaux effectués dans ce Chapitre vont faire l'objet d'une soumission dans un journal généraliste dont le format, particulièrement concis, ne permet pas de discuter de l'ensemble des résultats (notamment les résultats phylogénétiques et biogéographiques). Ce Chapitre présente donc, dans un premier temps, une restitution complète des travaux effectués sur ce sujet. Le manuscrit, tel que nous souhaitons le soumettre dans le journal Proceedings of the National Academy of Sciences (PNAS), après relecture de l'ensemble des coauteurs, est intégré en fin de Chapitre.

Abstract

Today's uneven geographical and phylogenetic distribution of biodiversity reflects complex underlying temporal and spatial dynamics. The development of large and robust phylogenies, the assembly of distribution and trait databases, and new analytical methods are shedding light on the respective roles of abiotic and biotic factors on patterns of diversification and emphasize the importance of adaptive traits. Yet, few studies have investigated the role of these factors on diversification dynamics at a global scale, and nearly none in insects, the most diverse group of organisms on Earth. Using a combination of phylogenomics, diversification and historical biogeography analyses, we propose a comprehensive account of the spatial and temporal diversification dynamics in wild silkmoths (Saturniidae). These moths are typical capital-breeding insects in which non-feeding adults have short lifespan entirely devoted to reproduction. Considering several key life-history traits analyzed using trait-dependent models of diversification, we found that wild silkmoths overall evolved toward larger body-size and increased level of polyphagy of their caterpillars, in contrast to most other insects. Polyphagy, pupation mode and dispersal capacities positively impacted diversification rates, but only the first two traits, not dispersal capacities, likely favored the mobility of lineages between distant biogeographical regions. Our work demonstrates the importance of life-history innovations in driving the dynamics of diversification in a family of moths. Importantly, the key traits identified here are all driven by the mode of reproduction, as capital-breeders, of these organisms.

Introduction

How ecological and evolutionary processes shape extant biodiversity patterns remains one of the most elusive question in biology (Benton 2016). Empirical results from evolutionary, historical biogeography and macroecology studies have revealed that Earth's biodiversity is unevenly distributed among taxa and geographical regions of the world (Myers et al. 2000; Barthlott et al. 2007; Jetz et al. 2012). This is understood as the result of millions of years of evolution of Earth's organisms, governed by natural selection and further shaped by environmental conditions, and stochastic processes. The former two, divided into biotic and abiotic factors (Barnosky 2001; Benton 2009), both act on the spatial distribution of organisms (e.g. abiotic: land masses movements, oceanic currents, winds; biotic: individual dispersal capacity, phoresy) and on their reproductive success (e.g. abiotic: climatic optimum, atmosphere composition; biotic: competition, predation). Thanks to the development of large molecular datasets (the phylogenomic era; Delsuc et al. 2005), of advanced phylogenetic methods (Minh et al. 2020) and of recent macroevolutionary and biogeography analytical tools (Morlon 2014; Beeravolu & Condamine 2016; Maliet et al. 2019), our ability to simultaneously address the role of both biotic and abiotic forces in empirical studies has considerably improved over the years (*e.g.* Ezard et al. 2011; Condamine et al. 2012; 2018). The relative roles of those forces were also addressed through simulations (Aguilée et al. 2018) and meta-analyses (Morlon et al. 2010), or theoretically discussed in a temporal framework (Moen & Morlon 2014). However, such questions are marginally addressed analytically in a spatial context (but see Klaus & Matzke 2020; Matos-Maravi et al. 2018; Sukumaran et al. 2016; Sukumaran & Knowles 2018), despite conspicuous relationships between historical biogeography patterns and biotic (e.g. dispersal capacities, food plants for phytophagous insects) or abiotic (e.g. positions of landmasses positions, climatic barriers) factors.

Large-scale studies of diversification investigating those questions were generally carried out in some of the most documented groups of organisms, such as birds (Claramunt & Cracraft 2015; Cooney et al. 2017; Jetz et al. 2012), amphibians (Bonetti & Wiens 2014) or plants (Antonelli et al. 2015; Fiz-Palacios et al. 2011; Silvestro et al. 2015). However, few of them have empirically addressed the role of adaptive traits in diversification dynamics for geographically widespread taxa (but see Burin et al. 2016; Cooney et al. 2017; Modica et al. 2020; Price et al. 2012). In insects, which form the bulk of Earth's biodiversity, investigations of diversification dynamics are strongly impeded by major shortfalls (Diniz-Filho et al. 2010; Cardoso et al. 2011; Hortal et al. 2015) in the knowledge of their diversity (Linnean shortfall), their distribution (Wallacean shortfall), their evolution (Darwinian shortfall), their traits (Raunkiærian shortfall) or biotic interactions (Eltonian shortfall). Studies of insect diversification are rarely global (but see Economo et al. 2018; 2019), or often focused on higher taxonomic ranks (Condamine et al. 2016; Misof et al 2014; Rainford et al. 2016) with limited insights into the drivers of diversification. Those drivers were more thoroughly investigated at regional scale, considering the role of ecological

innovations in a few insect orders and showing the significant influence of food plant shifts in butterflies (Condamine et al. 2018), or habitat specialization in ants (Matos-Maravi et al. 2018) and mayflies (Cozzarolo et al. 2019).

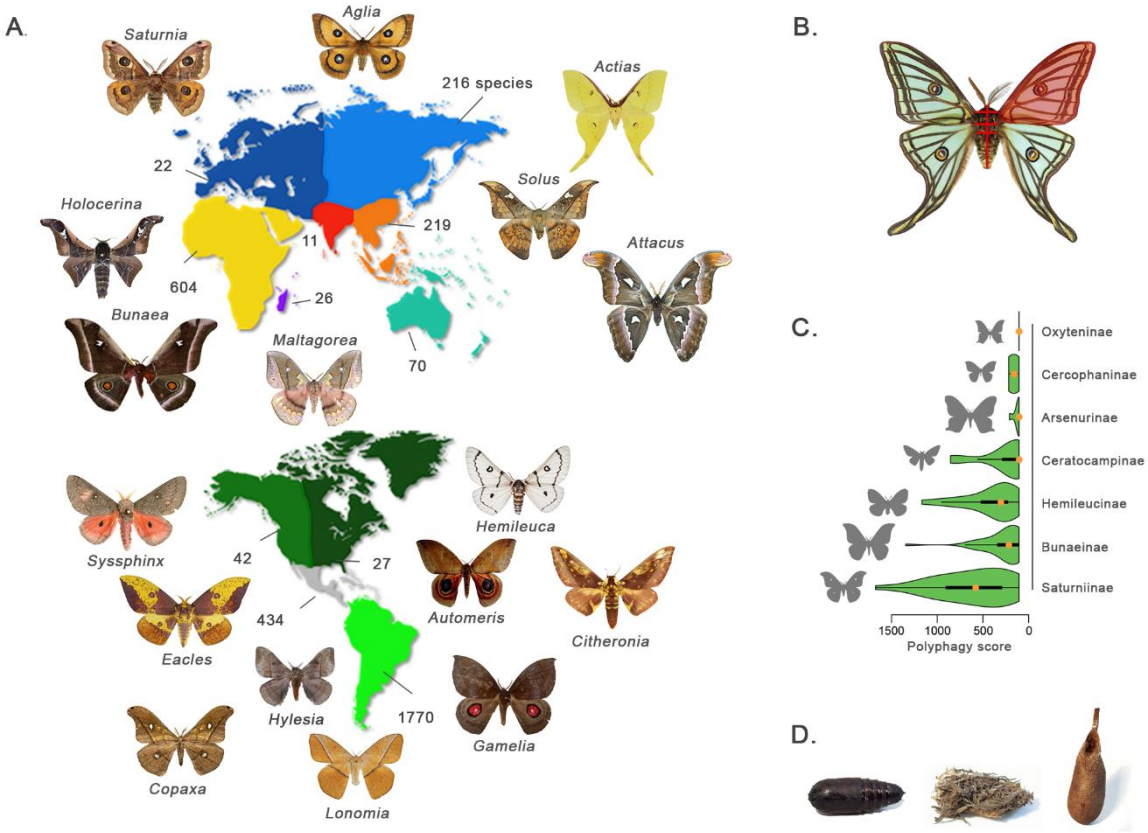


Figure 1 - Saturniidae life-history traits and distribution. (A) Saturniid diversity. Number of saturniid species per biogeographical area are depicted along with representative photos of several genera. (B) Morphological measurements considered in this study: thorax width (three measurements), body length and forewing surface. (C) Polyphagy score in the main Saturniidae subfamilies as estimated with the Phylogenetic Diversity (PD) metric. (D) The three modes of pupation in Saturniidae, i.e., from left to right, pupation underground, loose cocoon and plain cocoon.

In this Chapitre, I produced the first global scale documentation of spatial and temporal diversification dynamics in a group of worldwide distributed insects, wild silkmoths (family Saturniidae), comprising about 3500 species. These spectacular and popular moths have attracted attention of naturalists and biologists for several centuries and as a result their diversity, distribution and biology are unequally well documented among insects. Their diversity is unevenly distributed, both phylogenetically (more than 4/5th of all species belong to 2 of the 8 recognized subfamilies) and geographically (2/3rd of species inhabit the Neotropical region; Figure 1). Saturniids represent typical capital-breeding insects, a reproductive strategy lying at the extreme end of a continuum toward income-breeding strategies where adult-derived resources are used for reproduction (Stephens et al. 2009; Davis et al. 2016). Life as a capital-breeder in Lepidoptera has been associated to a set of life-history characteristics (Davis et al. 2016; Jervis et al. 2006; Prinzing 2003; Tammaru & Haukioja 1996; Figure 1) that are broadly recognized (even paroxysmal in some species) within the Saturniidae family (de Camargo et al. 2017;

Janzen 1984; Michener 1952; Oberprieler & Nässig 1994): large size (with marked sexual dimorphism, females having large abdomens carrying reserves from larval stages in the form of pre-matured eggs), short lifespan (<7d on average, because only relying on reserves acquired during larval stages), low mobility (short period of activity at night), and larval polyphagy (i.e. plasticity in food plant use by caterpillars for completing their development, which reduces the time needed by females to locate a suitable food plant and ensures high availability of food). Here, we used a phylogenomic approach to produce a spatially and temporally resolved evolutionary history of the entire family, and we investigated what role geological and climatic changes, on one hand, and life-history traits, on the other hand, may have played in driving the spatial and temporal diversification dynamics of wild silkmoths.

Materials and methods

Phylogenomic inferences

In order to resolve the Saturniidae genus-level phylogeny, we sequenced hundreds of Ultra Conserved Elements (UCEs) and inferred phylogenies with both maximum-likelihood (ML) and multispecies coalescent (MSC) approaches. UCEs, initially proposed as phylogenomic markers by Faircloth et al. (2012), have been successfully used in the inference of phylogenies in various organisms (Blaimer et al. 2016; Andersen et al. 2019; Friedman et al. 2019). Because one set of ultraconserved loci, identified from highly divergent reference genomes, can be applied to phylogenetically distant taxa, the capture of UCEs is very relevant when inferring deep phylogenetic relationships (Faircloth et al. 2012, 2020). Also, thanks to their increased variability toward their flanking regions (Faircloth et al. 2012; Tagliacollo & Lanfear 2018), UCE markers can be used at recent evolutionary timescales to resolve intra-generic and intra-specific nodes (Prebus et al. 2017; French et al. 2019).

Taxon sampling

Our sampling includes 240 species in 180 genera (Table S2). It comprises all genera of Saturniidae (Kitching et al. 2018) as well as all subgenera within *Psilopygida*, *Meroleuca*, *Gonimbrasia* and *Antheraea*; all subgenera but one in genus *Saturnia* (subgenus *Perisomena*); and two out of six subgenera in *Hylesia*. Fifteen outgroups were used, representing four other families of Bombycoidea: *Quentalia mailynae* in Bombycidae; *Acanthobrahmaea europaea* in Brahmaeidae; *Endromis versicolora* in Endromidae; *Lemonia taraxaci* in Lemoniidae and 11 Sphingidae species of the Smerinthini, a tribe for which a fossil has been described (Kitching & Sadler 2011).

DNA extraction and library preparation.

Tissues were obtained from specimens preserved either in 80%–100% ethanol and stored at -35°C, stored dry in envelopes, or pinned. Voucher information can be found in Table S1. DNA was extracted from legs or thoracic muscles using the Qiagen DNeasy Blood and Tissue kit using an overnight lysis. Library preparation followed Cruaud et al. (2019). Briefly, input DNA was sheared to a size of *ca.* 400bp using the Bioruptor® Pico (Diagenode). End repair, 3'-end adenylation, adapter ligation and PCR enrichment were then performed with the NEBNext Ultra II DNA Library prep kit for Illumina (NEB). We used adapters that contained amplification and Illumina sequencing primer sites, as well as nucleotide barcodes of 5 or 6bp long for samples demultiplexing. After quantifying DNA with a Qubit® 2.0 Fluorometer (Invitrogen), we pooled samples at equimolar ratio. The number of individual libraries per pool ranged from 6 to 16. Each pool was enriched using the Lepidoptera probe set designed by Faircloth (2017) (14,363 probes targeting 1,381 conserved loci) using a MYbaits kit (Arbor Biosciences) and following manufacturer's protocol. The hybridization reaction was run for 24h at 65°C. Post enrichment amplification was performed on beads with the KAPA Hifi HotStart ReadyMix. The enriched libraries were all quantified with a Qubit® 2.0 Fluorometer, a 2100 Agilent Bioanalyzer and qPCR using the Library Quantification Kit - Illumina/Universal from KAPA (KK4824). They were then pooled at equimolar ratio. Paired-end sequencing (2*300bp) was performed on an Illumina Miseq at UMR AGAP laboratory (CIRAD, Montpellier, France).

Raw data cleaning

Raw data cleaning followed Cruaud et al. (2019). In a nutshell, quality controls were performed with FastQC v.0.11.2 (Andrews 2010), quality filtering and adapter trimming with Trimmomatic-0.36 (Bolger et al. 2014). Overlapping reads were merged using FLASH-1.2.11 (Magoč & Salzberg 2011). Demultiplexing was performed using a bash custom script (no mismatch in barcode sequences allowed). Assembly of resulting reads was performed using CAP3 (Huang & Madan 1999). The probes designed by Faircloth (2017) were assembled into a set of non-overlapping reference UCEs (n=1381) using Geneious 8.1.8 (Kearse et al. 2012). Contigs were aligned to the reference UCEs using LASTZ 1.02.00 (Harris 2007); those that aligned with more than one UCE were filtered out using Geneious 8.1.8 as well as UCEs loci matched by different contigs.

Phylogenetic analyses and Exploration of systematic bias

We only considered the loci obtained for at least 50% of the samples. Alignments were performed with MAFFT v7.245 (Katoh & Standley 2013) (*-linsi* option) for each locus independently. Ambiguously aligned blocks were removed using Gblock_0.91b with relaxed constrains (-t=d -b2=b1 -b3=10 -b4=2 -b5=h) (Talavera & Castresana 2007). Gene trees were then inferred with RAxML 8.2v (Stamatakis 2014) and node supports were assessed with 100 bootstrap replicates. We used TreeShrink v1.3.1 (Mai

& Mirarab 2018) to identify and remove erroneous sequences causing unexpectedly long branches in the inferred gene trees: two successive analyses were run with k=1 (one sample removed at each analysis maximum). Once outlier sequences were removed, we re-performed the alignment, Gblock step and the gene trees analyses again with RAxML. The resulting clean dataset (1171 loci in total) was analyzed with coalescent-based and supermatrix approaches. ASTRAL-III v5.5.6 (Zhang et al. 2018) was used to infer a species tree from individual gene trees. To improve accuracy (Zhang et al. 2018) nodes with UFBoot<10 were collapsed in individual gene trees with the Perl script AfterPhylo.pl v0.9.1 (https://github.com/qiyunzhu/AfterPhylo). Node supports were evaluated with local posterior probabilities (local PP). Clades were considered supported when their localPP were ≧ 0.95 (Sayyari & Mirarab 2016). The alignment resulting from concatenation of all datasets for individual loci was analyzed with IQ-TREE v1.6.3 (Nguyen et al. 2014) using different partitioning strategies. First, an unpartitioned dataset was analyzed with the GTR+F model and the most likely number of FreeRate categories of rate heterogeneity estimated by ModelFinder (Kalyaanamoorthy et al. 2017) using BIC scores (Yang 1995; Soubrier et al. 2012). FreeRate generalizes the +G model by relaxing the assumption of Gamma-distributed rates and is recommended for the analysis of large matrices (Kalyaanamoorthy et al. 2017). Secondly, the concatenated dataset was also partitioned using the method of Tagliacollo & Lanfear (2018) with the Sliding-Window Site Characteristics algorithm using sites entropy (SWSC-EN). It consisted in splitting each ultraconserved locus into three parts: one core and two flanking regions. The number of partitions was then reduced using PartitionFinder 2.1.1 (Lanfear et al. 2017) considering branch lengths as linked, the corrected Aikaike Information Criterion (AICc) for model selection and the *rclusterf* algorithm for partitioning scheme comparison. Once the partition scheme established, we identified the best model for every partition with ModelFinder. For all IQ-TREE analyses, node supports were evaluated with UltraFast Bootstraps (UFBoot, 1,000 replicates) and SH-aLRT tests (1,000 replicates; see Guindon et al. 2010). Clades were considered supported when SH-aLRT was ≧ 80 and UFboot was ≧ 95%.

Variable evolutionary rates among taxa and base composition heterogeneity among genes are important sources of systematic bias (Brinkmann et al. 2005; Philippe et al. 2017; Romiguier & Roux 2017). Notably, it has been shown that GC content of UCEs and branch length heterogeneity can bias inferences towards highly supported but incorrect topologies (Bossert et al. 2017; Cruaud et al. 2020). To test for possible bias in our inferences, trees were also inferred from data subsets in which a given percentage of UCEs with the highest GC content or with the highest Long Branch Heterogeneity (LBH) scores were discarded. GC content was calculated with AMAS (Borowiec 2016) and LBH scores were calculated with TreSpEx (Struck 2014). Different percentages were considered: 2, 5, 10, 20, 30, 40 and 50%, hereafter referred to as GC_X and LB_X data subsets, where X represents the percentage used. ASTRAL and IQ-TREE analyses with the two partitioning strategies were performed on these data subsets.

Divergence time estimates

Overall, we performed the following analyses using only one sample per genus, except in genera *Antheraea* and *Saturnia* for which all subgenera were included, and in genus *Actias* for which we included eight species because of the very broad geographical range of this genus, spanning over several biogeographical regions. The non-Sphingidae outgroup were removed. In total 200 terminal taxa were included in the datation analyses. The massive amount of data analyzed makes the calibration step particularly time consuming. In order to reduce the computation time, we thus considered different data subsets of the LB_0.05 alignment, in which 5% of the loci with the highest LB score were discarded. In a first sub-dataset (named A) we selected the 50 loci that produced the more supported gene trees (referred as BP data subset). In a second data subset (B) we selected the 50 loci that produced the trees with the lowest Robinson-Fould distance compared to the tree inferred with the ASTRAL tree (RF data subset). A third data subset (C) was also based on a Robinson-Fould metric selection, but considering only the loci with at least 500bp (RF-size data subset). Finally, the data subsets D and E were composed of 50 randomly selected loci. The five datasets were built ensuring that the minimum number of loci per species is 5. If it was not the case, then the locus ranked 50th was discarded and the locus ranked 51th included in the sub-dataset until the fulfilment of that condition. Because ASTRAL accounts for gene tree discordance (Degnan & Rosenberg 2009), it is supposed to better estimate topology than concatenation approaches (Mirarab et al. 2014; Mirarab & Warnow 2015) when Incomplete Lineage Sampling (ILS) effect is important. We thus used the topology inferred with ASTRAL on the LB_0.05 dataset as constraint during the divergence times estimation. To date the phylogeny, we used MCMCTREE in PAML v4.8 (Yang 1997), using 2 fossils and 6 secondary calibrations with uniform prior distributions and soft bounds (Table S2). Because calculation of the likelihood function during the MCMC is computationally intense for large alignments, we first approximated the branch lengths with a maximum likelihood method introduced by Thorne et al. (1998) and implemented in PAML. The method produces a gradient and a Hessian matrix that contain information about the curvature of the likelihood surface. Using these approximations, we then estimated the divergence times using MCMCtree with 5 millions of generations (1 million discarded as burn-in). We ran these analyses twice for each data subset. The convergence was checked ensuring that estimates were similar from the two analyses and that Effective Sampling Size (ESS) of each parameter was over 200. We finally combined the runs with logCombiner (Drummond & Rambaut 2007) and summarized the parameters with MCMCtree and the *print = -1* option.

Species-level phylogenies inference

Because most of the tools used for diversification analyses are designed for species-level phylogenies, we generated a species level phylogeny using PASTIS (Thomas et al. 2013) and MrBayes (Ronquist et al. 2012). We did not use any input alignment but rather fixed every node of the phylogeny to match the

topology and the median ages as estimated with ASTRAL and MCMCtree. We also constrained every genus or subgenus to be monophyletic. We used one chain of 10 million generations and discarded 25% of generations as burnin before using the *sumt* command that summarize the trees sampled during the MCMC analysis.

Biogeographical analyses

Species distributions were categorized into 11 major biogeographical areas. We considered South America (SA) and Central America (CA), because such distinction allows to infer the timing of dispersions across Panama strait and isthmus. We separated Nearctic and Palearctic into Eastern (EN and EP) and Western (WN and WP) parts so that we can test whether dispersal events between Holarctic regions happened through the Bering strait or the Thulean route (Nilsen 1978). Madagascar (MD) was distinguished from Africa (AF) because of its particular fauna and India (IN) was also considered distinct because it collided with the Eurasian continent around 50Ma (Patriat & Achache 1984). Finally, we considered the Malayan region (WA) separated from the Australian (AU) region at the Weber line. Ancestral area reconstruction analyses were performed using DECX (Beeravolu & Condamine 2016) on the species-level phylogenies generated with MrBayes. To account for uncertainty associated with ancestral range estimation at the inter-generic nodes, we ran 1000 different analyses on different trees sampled in the posterior distribution of the MrBayes analysis and averaged the results obtained at each node of interest. Because the intra-generic phylogenetic relationships were unknown, we randomized the positions of the species within each genus. We divided Cenozoic into 6 time periods: Paleocene (66-56Ma), Eocene (56-34Ma), Oligocene (34-23Ma), early Miocene and Langhian (23-14Ma), Serravallian and late Miocene (14-5Ma) and Pliocene and Quaternary (5Ma to present). We considered adjacency as well as dispersal rate matrices (DRM, see Supplementary Material section).

Traits measurement, compilation and evolutionary analyses

Morphological measurements

In order to understand the evolution of body size throughout the evolution of Saturniidae, we measured morphological traits from images of 2577 male specimens representing all existing genera (or subgenera included in the dated phylogeny) but two, *Mielkesia* and *Jaiba*. We focused on the following measurements: *body length* considered from the top of the head to the end of the abdomen; *thorax width*, as measured (i) between the junction points of thorax and forewings, (ii) in the middle of the thorax, (iii) between the junction points of thorax and hindwings; and *forewing surface*. These measurements were carried out using an image annotation tool in the Barcode of Life Datasystems (www.boldsystems.org; Ratnasingham & Hebert 2007). We then approximated *body size* as being the product of *thorax width* and *body length* and *wingload* as being the ratio between *body size* and the *forewing surface*. Forewings being the most important pair of wings involved in flight performance (Le

Roy et al. 2019), we consider *wingload* as an indicator of flight capacities and of species dispersal capacities (Dudley & Srygley 1994; Le Roy et al. 2019). Ancestral values of *body size* and *wingload* were estimated using the *ace* function in the R package *ape* (Paradis & Schliep 2019). Phylogenetic signal was measured with Blomberg's K metric (Blomberg et al. 2003) using the *phylosig* function of the R (R code Team 2019) package *phytools* (Revell 2012).

Polyphagy estimates

To investigate the role of caterpillar polyphagy during the evolutionary history of saturniid moths, we used a newly compiled hostplant database for Saturniidae (Ballesteros-Mejia et al. submit., see attached articles) that includes 6923 hostplant records for 605 species of Saturniidae, mostly derived from literature, observations and rearing. We uniformized the higher-level taxonomy of plants following Magallon et al. (2015) and only considered plants eaten by saturniids *in natura* (i.e. captive records were excluded). Even though captive breeding could be informative to understand diet breadth, these records are very heterogeneous and biased toward spectacular and usually rather common species. We considered plant identification at family level, a rank at which we can be confident that we have a fairly representative hostplant dataset. As a metric of the degree of polyphagy level of each genus, we used Magallón et al. (2015) dated angiosperm phylogeny and the Phylogenetic Diversity (PD; Faith 1992) metric, *i.e.* the total length of the phylogenetic tree branches connecting the different families of plants a given species eats. PD scores were calculated for every saturniid species individually and averaged within each genus. In the Magallón et al. (2015) phylogeny, there is a considerable phylogenetic distance between Gymnosperms and Angiosperms. Because it would have biased PD score for a very limited number of genera, we discarded any Gymnosperm records of the host-plant database. We estimated ancestral diet breadth scores using the *ace* function in the *ape* R package. Phylogenetic signal was measured similarly to *body size* and *wingload*.

Pupation mode

The pupa of Lepidoptera represents a rather vulnerable stage, because of its immobility. As protection against predators and parasitoids, the caterpillars of saturniid moths have adopted three main strategies: some spin silk threads to encase themselves in a loose cocoon through which the pupa can be seen; others spin a plain cocoon, sometimes hard-shelled or even multi-layered; finally, some simply find shelter and pupate underground (or occasionally in the litter) without spinning any silk. We reconstructed the evolution of pupation strategy in saturniid moths using maximum likelihood with the *ace* function of the *ape* R package. We considered a model in which the probability of evolution from a plain cocoon to an underground pupation was null, and reversely; *i.e.* such transition had to evolved through an intermediate loose cocoon condition.

Diversification analyses

In order to document diversification dynamics throughout the evolution of Saturniidae, we used two approaches: (i) the Bayesian Analysis of Macroevolutionary Mixture (BAMM v2.5, Rabosky et al. 2013) to identify diversification rate shifts without any a priori and to test how biotic or abiotic factors may have affected saturniid diversification; (ii) a trait-dependent diversification approach (Maddison et al. 2007; FitzJohn et al. 2009) that combines Maximum Likelihood (ML) and Bayesian through Markov Chain Monte Carlo (MCMC) methods and is implemented in the R-package *diversitree* v.0.9-8 (FitzJohn 2012).

Before launching the analyses in BAMM, we set proper priors for the extinction, speciation and rate shift parameters with the *setBAMMpriors* function implemented in the BAMMtools 2.1.6v R package (Rabosky et al. 2014). We ran multiple analyses using different values of the *expectedNumberOfShifts* parameter, ranging from 0.01 to 50. We ran each BAMM analysis for 10M of generations on 4 chains. After discarding 2.5M of generations as burnin, we checked for the likelihood convergence assessing that the Effective Sample Size (ESS), calculated with the *effectiveSize* function of the *coda* R package (Plummer et al. 2006) on the Likelihood of the analysis, was above 200. When it was not, we launched the analysis again, doubling the number of generations (20M). Because of the large size of our phylogenetic tree, the number of equiprobable shift configurations is considerable; we thus considered maximum shift credibility configuration, as estimated with the *maximumShiftCredibility* function (BAMMtools R package).

Trait-dependent diversification models as designed by Maddison et al. (2007) simultaneously model trait evolution and its impact on diversification. We used these models, as implemented in the *diversitree* R package, to analyze the influence of traits on the evolution of Saturniidae. We ran the models independently on four different traits dataset: body size, dispersal capacities, polyphagy level, and pupation strategy. For each trait dataset, we considered a pruned Saturniidae species-level phylogenetic tree in which the genera for which we were not able to compile data were removed. We discretized the first three datasets (pupation modes being already discretized) into three categories and applied the Multi-State Speciation and Extinction model (MuSSE, FitzJohn et al. 2009) independently on each trait dataset. In every MuSSE model, we did not allow an evolution from state 1 to state 3 and reciprocally ($q_{13} \sim 0$, $q_{31} \sim 0$). We first fitted a 'null model' in which the speciation, extinction and transition rates were respectively equal ($\lambda_1 \sim \lambda_2$, $\lambda_1 \sim \lambda_3$; $\mu_1 \sim \mu_2$, $\mu_1 \sim \mu_3$; $q_{21} \sim q_{12}$, $q_{23} \sim q_{12}$, $q_{32} \sim q_{12}$). We also considered a model where the speciation and extinction rates were fixed but the transition rate were free to differ ('qfree model'). Finally, we inferred a model, hereafter referred as 'full model', in which every parameter, but q_{13} and q_{31} that were fixed to zero, were free to vary. The null and full models were then compared using AIC and ΔAIC: the model supported with the lowest AIC was considered the best (only if ΔAIC >2 against the second-best model; otherwise the model with the lowest number of parameters

was instead considered the best). In order to compare the different variables and to understand which of them influence Saturniidae diversification the most, we divided the ΔAIC per the AIC of the null model. Finally, to test if the level of polyphagy may explain the heterogeneity of diversification rate within each subfamily, we applied polyphagy-dependent models on the most diversified subfamilies: Bunaeinae, Ceratocampinae, Hemileucinae and Saturniinae. For these models, we discretized the polyphagy score into three categories determined by quantiles 0.33 and 0.66. Models were fitted using ML methods as implemented in the *diversitree* R package. We then used 10,000 MCMC generations (10% as burnin) to examine the confidence interval of the parameter estimates using the parameters obtained by ML as priors (FitzJohn 2012).

Results and Discussion

Phylogenomics illuminates the evolution and the relationships of extant genera of wild silkmoths

To reconstruct a comprehensive and robust phylogenetic hypothesis for family Saturniidae, we generated a genomic dataset consisting of 1381 Ultra Conserved Elements (UCEs) for 255 samples, representing 15 outgroup species (in families Bombycidae, Brahmaeidae, Endromidae, Sphingidae) and 234 Saturniidae species (Table S1). All 180 genera and 18 of the 25 subgenera currently recognized are represented, in all 8 subfamilies and 11 tribes of the family (Kitching et al. 2018) (Table S1). On average, 909 loci were sequenced per sample [138 to 1,193] and loci were 557 bp long. We considered the loci for with more than 50% of the samples were available. The concatenated matrix used in the following analyses was 788,018 bp long, totalizing more than 124M nucleotides and less than 6% gaps.

Exploration of systematic bias & deep phylogenetic relationships

We recovered a negatively skewed distribution of the Long Branch Heterogeneity scores (Figure S1), solve when removing 5% of the UCEs with the highest LBH scores. If we do not consider differences caused by the inference methods used (concatenated versus coalescent), the trees inferred with the GC_X datasets were overall very congruent with only few superficial intergeneric changes (Figures S2, S3 - Online resources). The results obtained with the different LB_X datasets were more distinct, both in terms of tree topology and branch lengths (Figures S2, S3). Removing a large proportion (≥20%) of the loci with the highest LBH scores led to relative shorter internal branches and longer terminal ones, and the topologies inferred with such datasets also included some dubious clades. For example, *Cricula* was recovered as sister clade of Attacini and Saturniini in the IQTREE LB_0.4 and LB_0.5 analyses (Figure S3), or *Hirpida* was sister to Arsenurinae in the IQTREE and ASTRAL LB_0.3, LB_0.4, LB_0.5 and IQTREE SWSC-EN LB_0.5 analyses. If discarding loci with the highest LBH scores could reduce or remove artefactual phylogenetic signal (Struck 2014; Cruaud et al. 2020), we consider that this

approach should be limited to the correction of the tail-end of the LBH distribution. In our case, discarding more than 20% of the loci with the highest LBH scores was excessive. Taken together, the analyses we performed with the different datasets were very consistent (Figures 2, S2) and confirmed the relevance of the use of Ultra Conserved Elements when inferring deep phylogenetic relationships.

The relationships among subfamilies are all strongly supported (Figures 2, S2, S3), except for the position of the Agliinae, a mono-generic Palearctic subfamily whose position varies with the methods used to build the trees, but that remains robust to our tests of possible biases caused by LBH scores and GC contents of the UCE loci analyzed (Figures 1 and S2). Interestingly, this lack of resolution regarding Agliinae is a persistent problem in previous analyses based on completely independent datasets (Anchor Hybrid Enrichment of exons in Barber et al. (2015) and Hamilton et al. (2019); 5 nuclear genes in Regier et al. (2008)), suggesting that we are facing a "hard" polytomy (Hoelzer & Melnick 1994) resulting from nearly simultaneous speciation of the ancestors to Agliinae, Hirpidinae and Arsenurinae subfamilies. Whereas extending taxon sampling is unlikely to improve much the resolution here, because all genera are represented and are rather species-poor, it might be that the combination of existing genomic datasets helps establish the relationships between these taxa. Here, we give preference to the ASTRAL topology placing Agliinae sister to Salassinae + Saturniinae, because of its congruence with that of Hamilton et al. (2019) and because it implies a more parsimonious hypothesis regarding the colonization of the Old World by wild silkmoths, only once by the ancestor to these three subfamilies versus two times if Agliinae is sister to all new world subfamilies. In the following analyses, we therefore considered the phylogeny inferred with ASTRAL from the LB_0.05 genomic dataset. The Figure 3 depicts this topology with branch lengths estimated with IQ-TREE (as ASTRAL does not infer terminal branch lengths).



Figure 2 – Phylogenetic supports obtained with the different analyses (in column) for the main Saturniidae clades (in row). IQ-TREE or ASTRAL refer to the phylogenetic software used and SWSC-EN indicate when, for the IQ-TREE analyses, the Tagliacollo & Lanfear (2018) partitioning method was used. These results were inferred from the original genomic dataset and from the datasets in which 5% of the loci were discarded because of their LBH score (LB_0.05) or their GC composition (GC_0.05). A blue cell indicates that the monophyly of the clade was strongly supported; cells are colored yellow or orange when clades are weakly supported or absent, respectively. For IQ-TREE analyses, branch supports are indicated in cells as: SH-aLRT/UFBoot; for the ASTRAL analyses, the indicated supports are localPP.

Phylogenetic relationships and revised classification of the Saturniidae

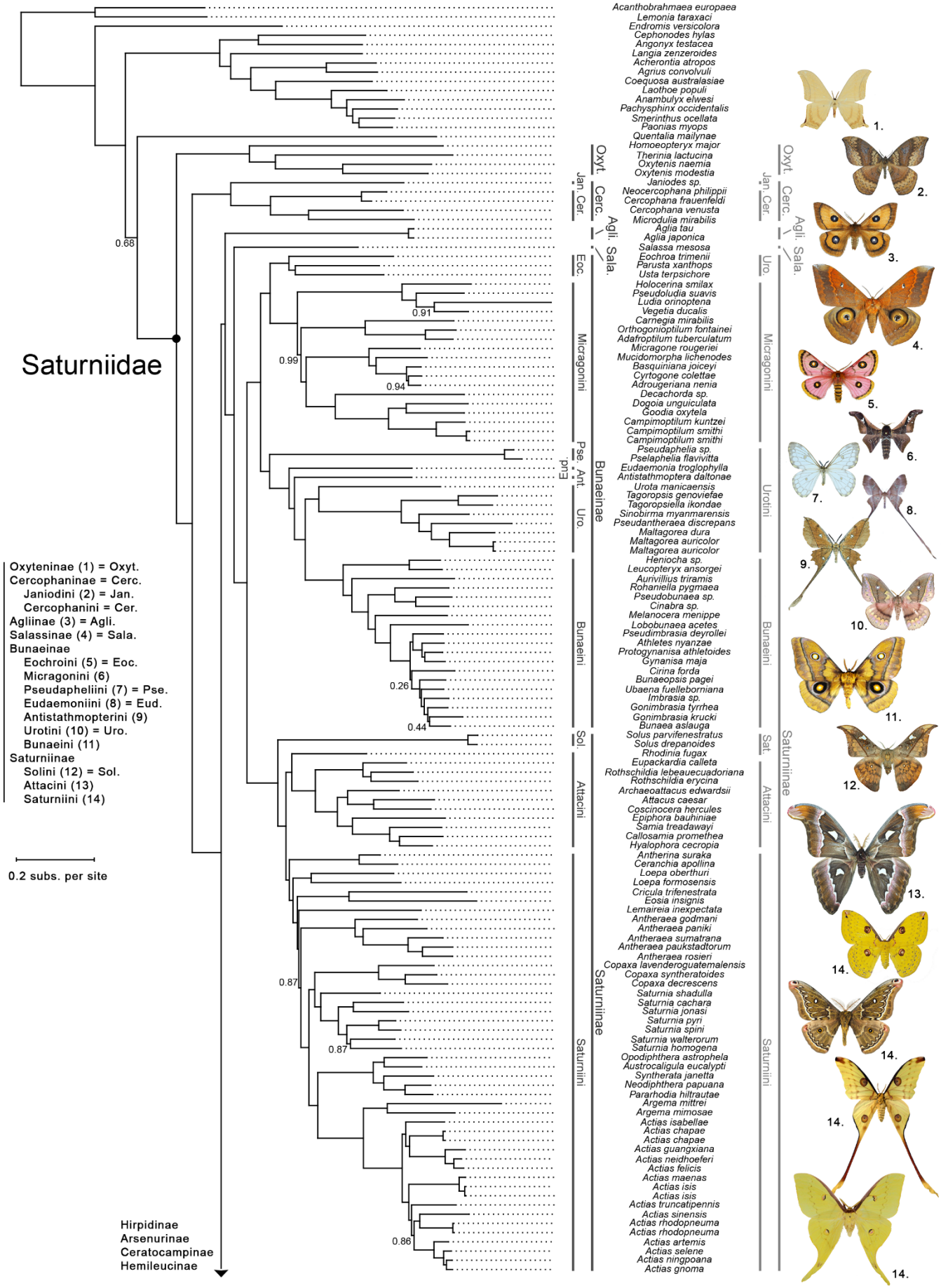
Our study provides a very comprehensive and well supported genus-level phylogeny of family Saturniidae. Its monophyly was recovered with strong support in all analyses, and 7 of the 8 recognized subfamilies were strongly supported as monophyletic in their current definition (Figures 2, 3, S2; Kitching et al. 2018). Traditionally recognized tribes were also strongly supported as monophyletic in all analyses, except for the Hemileucini, Urotini and Saturniini (Figure S2). Overall, our results are well resolved and because all currently recognized genera are represented, we consider that they offer a solid

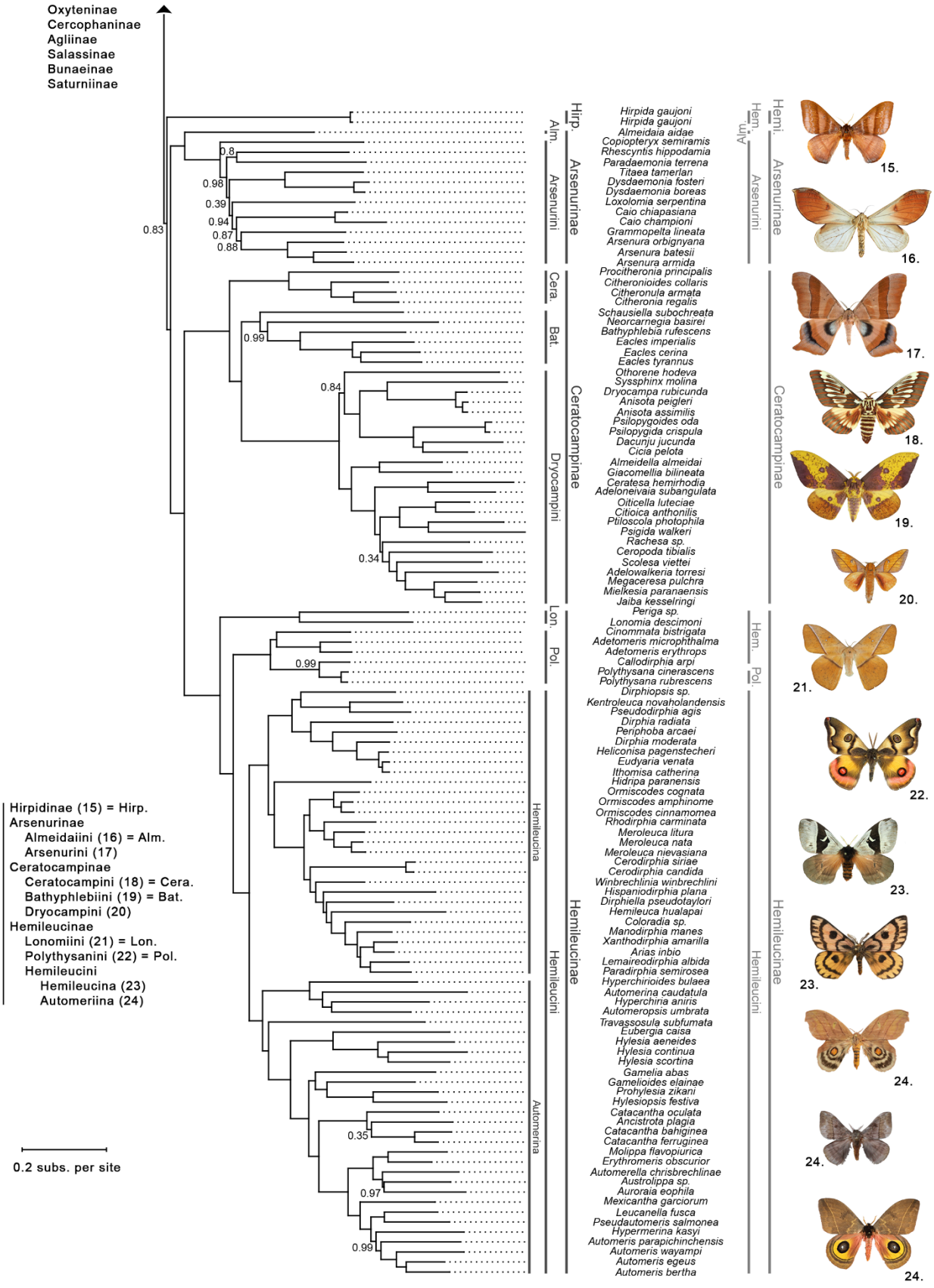
ground to propose a revised classification for the family. This new classification is preliminarily presented and illustrated here (Figure 3; Table 3); it will the object of a formal independent publication where the results from our phylogenomic analyses will be combined to a comparative morphological analysis for a better definition of the groups and to bring support for the changes proposed.

Within the New-World lineages, our results unambiguously revealed the paraphyly of subfamily Hemileucinae as previously defined, because of the position of genus *Hirpida*. This genus was considered until now as being part of the Hemileucinae (Michener 1952; Lemaire 2002), albeit recognized as being of unclear affinities within this subfamily. Our results will require the use of a new subfamily Hirpidinae, including all currently known species in genus *Hirpida*[3]. Within the Hemileucinae, we propose to add one new tribe, namely Lonomiini, to the two currently recognized tribes: Polythysanini and Hemileucini. Tribe Lonomiini (already proposed by Bouvier (1930) as "Lonomiicae") comprises the genera *Lonomia* and *Periga*, currently considered as belonging to Hemileucini. Furthermore, it is necessary to redefine the Polythysanini (previously limited to the sole genus *Polythysana*) as including the genera *Adetomeris*, *Callodirphia*, *Cinommata* and *Polythysana*. All other Hemileucinae genera belong to tribe Hemileucini (except for genus *Catharisa*, to be transferred to the Ceratocampinae – see below). We also propose to introduce two subtribes within the Hemileucini: Hemileucina and Automeriina (see Figure 3) to account for the two major hemileucine lineages revealed by our analyses. Within the Ceratocampinae subfamily, we propose to distinguish three tribes forming three clearly distinct lineages (Figure 3): Ceratocampini, Bathyphlebiini and Dryocampini.

Figure 3 (two following pages) - Genus-level phylogeny of the Saturniidae family inferred from the LB_0.05 dataset with IQTREE, constraining the topology to match the one obtained with ASTRAL. LocalPP support values estimated with ASTRAL (unless equal to 1) are given below branches. Scale bar (left side) represents the mean number of substitutions per site. Species names are indicated on the right side of the tree, bordered by the newly proposed (left) and current (right) Saturniidae classifications. Numbered photos refer to the different saturniid clades listed on the left side of the phylogeny.

[3] Recently, a new genus, related to *Hirpida*, has been described by Brechlin (2019): *Hirpsinjaevia*. We inferred in the Chapitre 3 that this genus is part of the Hirpidinae subfamily. This new taxon is however not included in our analyses that were performed before its description.


Saturniidae
Oxyteninae (1) = Oxyt.
Cercophaninae = Cerc.
Janiodini (2) = Jan.
Cercophanini = Cer.
Agliinae (3) = Agli.
Salassinae (4) = Sala.
Bunaeinae
Eochroini (5) = Eoc.
Micragonini (6)
Pseudapheliini (7) = Pse.
Eudaemoniini (8) = Eud.
Antistathmopterini (9)
Urotini (10) = Uro.
Bunaeini (11)
Saturniinae
Solini (12) = Sol.
Attacini (13)
Saturniini (14)
0.2 subs. per site
Hirpidinae
Arsenurinae
Ceratocampinae
Hemileucinae
0.68
0.91
0.99
0.94
0.26
0.44
0.87
0.87
0.86
Acanthobrahmaea europaea
Lemonia taraxaci
Endromis versicolora
Cephonodes hylas
Angonyx testacea
Langia zenzeroides
Acherontia atropos
Agrius convolvuli
Coequosa australasiae
Laothoe populi
Anambulyx elwesi
Pachysphinx occidentalis
Smerinthus ocellata
Paonias myops
Quentalia mailynae
Homoeopteryx major
Therinia lactucina
Oxytenis naemia
Oxytenis modestia
Janiodes sp.
Neocercophana philippii
Cercophana frauenfeldi
Cercophana venusta
Microdulia mirabilis
Aglia tau
Aglia japonica
Salassa mesosa
Eochroa trimenii
Parusta xanthops
Usta terpsichore
Holocerina smilax
Pseudoludia suavis
Ludia orinoptena
Vegetia ducalis
Carnegia mirabilis
Orthogoniоptilum fontainei
Adaftoptilum tuberculatum
Micragone rougeriei
Mucidomorpha lichenodes
Basquiniana joiceyi
Cyrtogone colettae
Adrougeriana nenia
Decachorda sp.
Dogoia unguiculata
Goodia oxytela
Campimoptilum kuntzei
Campimoptilum smithi
Campimoptilum smithi
Pseudaphelia sp.
Pselaphelia flavivitta
Eudaemonia troglophylla
Antistathmoptera daltonae
Urota manicaensis
Tagoropsis genoviefae
Tagoropsiella ikondae
Sinobirma myanmarensis
Pseudantheraea discrepans
Maltagorea dura
Maltagorea auricolor
Maltagorea auricolor
Heniocha sp.
Leucopteryx ansorgei
Aurivillius triramis
Rohaniella pygmaea
Pseudobunaea sp.
Cinabra sp.
Melanocera menippe
Lobobunaea acetes
Pseudimbrasia deyrollei
Athletes nyanzae
Protogynanisa athletoides
Gynanisa maja
Cirina forda
Bunaeopsis pagei
Ubaena fuelleborniana
Imbrasia sp.
Gonimbrasia tyrrhea
Gonimbrasia krucki
Bunaea alsauga
Solus parvifenestratus
Solus drepanoides
Rhodinia fugax
Eupackardia calleta
Rothschildia lebeauecuadoriana
Rothschildia erycina
Archaeoattacus edwardsii
Attacus caesar
Coscinocera hercules
Epiphora bauhiniae
Samia treadawayi
Callosamia promethea
Hyalophora cecropia
Antherina suraka
Ceranchia apollina
Loepa oberthuri
Loepa formosensis
Cricula trifenestrata
Eosia insignis
Lemaireia inexpectata
Antheraea godmani
Antheraea paniki
Antheraea sumatrana
Antheraea paukstadtorum
Antheraea rosieri
Copaxa lavenderoguatemalensis
Copaxa syntheratoides
Copaxa decrescens
Saturnia shadulla
Saturnia cachara
Saturnia jonasi
Saturnia pyri
Saturnia spini
Saturnia walterorum
Saturnia homogena
Opodiphthera astrophela
Austrocaligula eucalypti
Syntherata janetta
Neodiphthera papuana
Pararhodia hiltrautae
Argema mittrei
Argema mimosae
Actias isabellae
Actias chapae
Actias chapae
Actias guangxiana
Actias neidhoeferi
Actias felicis
Actias maenas
Actias isis
Actias isis
Actias truncatipennis
Actias sinensis
Actias rhodopneuma
Actias rhodopneuma
Actias artemis
Actias selene
Actias ningpoana
Actias gnoma
Oxyt.
Cerc.
Jan.
Cer.
Agli.
Sala.
Eoc.
Micragonini
Pse.
Ant.
Eud.
Uro.
Bunaeini
Sol.
Attacini
Saturniini
Bunaeinae
Saturniinae
Urotini
Micragonini
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
14.
14.
14.


Oxyteninae
Cercophaninae
Agliinae
Salassinae
Bunaeinae
Saturniinae
0.83
0.8
0.98
0.39
0.94
0.87
0.88
0.99
0.84
0.34
0.99
0.35
0.97
0.99
Hirp.
Alm.
Arsenurini
Arsenurinae
Cera.
Bat.
Dryocampini
Ceratocampinae
Lon.
Pol.
Hemileucina
Hemileucini
Automerina
Hemileucinae
Hirpida gaujoni
Hirpida gaujoni
Almeidaia aidae
Copiopteryx semiramis
Rhescyntis hippodamia
Paradaemonia terrena
Titaea tamerlan
Dysdaemonia fosteri
Dysdaemonia boreas
Loxolomia serpentina
Caio chiapasiana
Caio championi
Grammopelta lineata
Arsenura orbignyana
Arsenura batesii
Arsenura armida
Procitheronia principalis
Citheronioides collaris
Citheronula armata
Citheronia regalis
Schausiella subochreata
Neorcarnegia basirei
Bathyphlebia rufescens
Eacles imperialis
Eacles cerina
Eacles tyrannus
Othorene hodeva
Syssphinx molina
Dryocampa rubicunda
Anisota peigleri
Anisota assimilis
Psilopygoides oda
Psilopygida crispula
Dacunju jucunda
Cicia pelota
Almeidella almeidai
Giacomellia bilineata
Ceratesa hemirhodia
Adeloneivaia subangulata
Oiticella luteciae
Citioica anthonilis
Ptiloscola photophila
Psigida walkeri
Rachesa sp.
Ceropoda tibialis
Scolesa viettei
Adelowalkeria torresi
Megaceresa pulchra
Mielkesia paranaensis
Jaiba kesselringi
Periga sp.
Lonomia descimoni
Cinommata bistrigata
Adetomeris microphthalma
Adetomeris erythrops
Callodirphia arpi
Polythysana cinerascens
Polythysana rubrescens
Dirphiopsis sp.
Kentroleuca novaholandensis
Pseudodirphia agis
Dirphia radiata
Periphoba arcaei
Dirphia moderata
Heliconisa pagenstecheri
Eudyaria venata
Ithomisa catherina
Hidripa paranensis
Ormiscodes cognata
Ormiscodes amphinome
Ormiscodes cinnamomea
Rhodirphia carminata
Meroleuca litura
Meroleuca nata
Meroleuca nievasiana
Cerodirphia siriae
Cerodirphia candida
Winbrechlinia winbrechlini
Hispaniodirphia plana
Dirphiella pseudotaylori
Hemileuca hualapai
Coloradia sp.
Manodirphia manes
Xanthodirphia amarilla
Arias inbio
Lemaireodirphia albida
Paradirphia semirosea
Hyperchirioides bulaea
Automerina caudatula
Hyperchiria aniris
Automeropsis umbrata
Travassosula subfumata
Eubergia caisa
Hylesia aeneides
Hylesia continua
Hylesia scortina
Gamelia abas
Gamelioides elainae
Prohylesia zikani
Hylesiopsis festiva
Catacantha oculata
Ancistrota plagia
Catacantha bahiginea
Catacantha ferruginea
Molippa flavopiurica
Erythromeris obscurior
Automerella chrisbrechlinae
Austrolippa sp.
Auroraia eophila
Mexicantha garciorum
Leucanella fusca
Pseudautomeris salmonea
Hypermerina kasyi
Automeris parapichinchensis
Automeris wayampi
Automeris egeus
Automeris bertha
Hemi.
Alm.
Arsenurini
Arsenurinae
Ceratocampinae
Hem.
Pol.
Hemileucini
Hemileucinae
15.
16.
17.
18.
19.
20.
21.
22.
23.
23.
24.
24.
24.
Hirpidinae (15) = Hirp.
Arsenurinae
Almeidaiini (16) = Alm.
Arsenurini (17)
Ceratocampinae
Ceratocampini (18) = Cera.
Bathyphlebiini (19) = Bat.
Dryocampini (20)
Hemileucinae
Lonomiini (21) = Lon.
Polythysanini (22) = Pol.
Hemileucini
Hemileucina (23)
Automeriina (24)
0.2 subs. per site

Within the Old-World lineages, as already found in previous molecular analyses (Barber et al. 2015; Rubin et al. 2018), the Urotini tribe is polyphyletic and has to be redefined. We propose here to define five monophyletic tribes to account for the newly revealed phylogenetic relationships of all genera previously placed in tribe Urotini: Eochroini (*Eochroa* (previously a member of tribe Bunaeini (Kitching et al. 2018), *Parusta* and *Usta* genera), Pseudapheliini (*Pseudaphelia* and *Pselaphelia*), Eudaemoniini (*Eudaemonia*), Antistathmopterini (*Antistathmoptera*) and Urotini (*Maltagorea*, *Pseudantheraea*, *Sinobirma*, *Tagoropsiella*, *Tagoropsis* and *Urota*) (Figure 3). The position of genus *Rhodinia* with respect to tribes Attacini and Saturniini has been the object of debate (Regier et al. 2008). In all our analyses, it is recovered sister to Attacini (Figures 3, S3), a position proposed by some authors on the basis of morphological (Bouvier 1936; Oberprieler & Nässig 1994) or molecular analyses (Friedlander et al. 1998; Regier et al. 2008; Rubin et al. 2018). This relationship is supported by several apomorphies that were previously considered as homoplastic characters (Peigler 1989): presence of an open discoidal cell (no disco-cellular nervule) and of a large triangular hyaline area at the end of this cell as well as by a fiber structure of the silk of *Rhodinia* which appears more closely related to those of Attacini (Ramos & Peigler 1999). Consequently, we consider here that the position of genus *Rhodinia* is well supported inside the Attacini tribe. Beside the Saturniini and Attacini tribes, we propose the introduction of a new tribe within the subfamily Saturniinae: tribe Solini, comprising the sole genus *Solus* (Figure 3); it is sister to Attacini and Saturniini. Our results revealed that genus *Solus* indeed represents a rather isolated lineage, sister to two large clades comprising all genera of the Saturniini and Attacini tribes, respectively. Because of the major biogeographical and biological contrasts between the Solini, Attacini and Saturniini on one side and the previously recognized African tribes of Saturniinae (Bunaeini, Micragonini and Urotini) on the other side, we propose, following Nässig et al. (2015) to regroup the latter in a subfamily Bunaeinae that now comprises seven tribes: Antistathmopterini, Bunaeini, Eochroini, Eudaemoniini, Micragonini, Pseudapheliini, and Urotini and tribes. Because of this change, the subfamily Saturniinae now comprises three tribes: Attacini, Saturniini, and Solini.

Although taxon sampling did not allow us to test the monophyly of most genera, our results highlighted some cases of paraphyly that would require re-assignment of several genera to respect the criterion of monophyly in the classification: *Graellsia* is to be synonymized with *Actias*; *Catharisa*, known from a single species *C. cerina*, was previously considered to belong to subfamily Hemileucinae, but proved here to be a very unique representative of genus *Eacles* within the Ceratocampinae subfamily; the genus *Eubergioides* will also need to be reconsidered in the light of phylogenetic relationships among representatives of genus *Automeris*. Within the Cercophaninae, the genus *Cercophana*, comprising only two species, is paraphyletic and will need to be redefined. This is also the case of the genus *Psilopygida*, with its two subgenera branching in two very distinct positions of our tree. This issue of paraphyly of genus *Psilopygida* could be resolved by raising subgenus *Psigida* to genus level. Our results also suggest that *Catacantha* could be paraphyletic because of the position of genus *Ancistrota*. However, because

of poor supports, further analyses with denser sampling will be needed to address the relevance of distinguishing the two genera *Catacantha* and *Ancistrota*. Finally, the large genus *Dirphia* was unambiguously found to be paraphyletic with respect to the position of representatives of four other genera genera *Ithomisa*, *Heliconisa*, *Eudyaria* and *Periphoba*. Denser sampling is needed to better circumscribe the newly revealed lineages and to figure the best way to redefine genera among them. In the following analyses of this chapter, we will consider *Ancistrota*/*Catacantha* and *Ithomisa*/*Heliconisa*/*Eudyaria*/*Periphoba*/*Dirphia* as unique lineages.

Overall, our results lead us to recognize 10 subfamilies, 18 tribes, and 2 subtribes (see Figures 3, S2, Table S3).

Estimation of divergence times with phylogenomic data

All datation analyses ran with MCMCtree converged: the Effective Sample Size (ESS) of all parameters were superior to 200 and the two independent runs launched with the different sub-datasets led to identical estimates of divergence times (Figure S4). The four sampling strategies used to reduce the full phylogenomic dataset to subsets of 50 loci resulted in consistent results. Overall, these results suggest that divergence times can be reliably estimated through reduced data subsets, thus avoiding endless computation times implied by the use of complete genomic matrices. Because the BP sub-dataset (bootstrap-based selected loci) provided intermediate estimates, we considered the divergences times estimated with this sub-dataset in subsequent analyses and in our discussion.

A global understanding of wild silkmoths' diversification in space and time

Using the dated phylogeny derived from the ASTRAL analysis of the LB_0.05 matrix, we performed historical biogeography as well as diversification analyses to document the spatial and temporal dynamics of the evolution of saturniid moths. Our results are summarized in Figure 4 (left panel), representing a synthetic view of the evolution of wild silkmoths, in space and time (see the complete tree in Figure S5).

A journey through the world

Ancestral range estimations unambiguously support a neotropical origin of wild silkmoths, shortly after the K-T extinction event, during the Paleocene (60.3Ma, 95% credibility interval: [66.3-54.6Ma]). Whereas several lineages subsequently diverged and diversified in the Neotropics to represent the extant hotspot of saturniid diversity, a unique and ancient dispersal event from the New World to the Palearctic region occurred: the common ancestor to the Agliinae, Salassinae, Bunaeinae and Saturniinae subfamilies colonized the Eastern Palearctic region through Beringia in the early Eocene period (around 47Ma). This passage, all the way from the Neotropics to the Palearctic region could have been enabled

by the island arc (or "landspan"; Iturralde-Vinent & MacPhee 1999) formed by the Caribbean plate between the late Cretaceous and the middle Eocene and connecting Yucatan and Colombia (Pindell et al. 1988; Morley 2003). This connection was hypothesized to have permitted dispersion of various snakes, lizards or dinosaurs (Bonaparte 1984) and could have also enabled Saturniidae to expand their range northward in the Nearctic region where warmer climates of the Eocene period were more hospitable to megathermal lineages (Zachos et al. 2008). From there, dispersal from North-America to the Palearctic region likely occurred through the Bering land bridge that permitted biotic interchange between North-America and Eurasia for most of the time since the Cretaceous (Tiffney 1985; Condamine et al. 2013; Jiang et al. 2019). The presumed existence of a widespread boreotropical northern hemisphere forest in the Eocene (Wolfe 1975; Tiffney 1985; Baskin & Baskin 2016) and of a lush deciduous forest extending beyond the Arctic Circle (Jahren 2007) indeed supports the hypothesis that environments of the Nearctic and Palearctic regions were suitable for saturniid moths, with most of the present-day hostplants (e.g. oak, beech, maple, plum, sumac, etc.) of the caterpillars of early branching lineages (i.e. Agliinae, Salassinae, early branching genera of Saturniinae like *Solus* and *Rhodinia*) present. Interestingly, all these lineages but Agliinae are now mostly restricted to the Sino-Himalayan mountains where the Eocene–Oligocene climatic deterioration (i.e. the "Grande Coupure", Zachos et al. 2001) likely "pushed" tree taxa from the disrupted belt of boreotropical forests (LePage et al. 2005).

Eastern Palearctic then appears to have acted as a "biogeographical hub" from where several lineages colonized all regions of the Old World and re-colonized the New World on several occasions. The Agliinae and Salassinae subfamilies are early-diverged lineages of Old World saturniid moths currently only represented by a small number of species in the Palearctic region for the former, and in the Eastern Palearctic and Oriental regions for the later. A third lineage (Bunaeinae + Saturniinae), experienced a more successful spatial diversification, diverging soon after the colonization of the Old World (*ca.* 44Ma) and expanding its range to the Western Palearctic and African regions. From this point in time, two lineages diverged and diversified to represent the bulk of saturniid diversity in the Old World. One of these – the Bunaeinae – thrived in the Afrotropics (Africa + Madagascar), a region from where it never successfully expanded or dispersed, with the noticeable exception of one genus (*Sinobirma* Bryk 1944) in the bunaeine tribe Urotini (Rougerie et al. 2012). The second lineage – the Saturniinae – is now mainly represented by two large tribes – the Attacini and Saturniini – that remarkably colonized all main land masses, including the New World. The former dispersed into the Nearctic region through Bering *ca.* 30.0Ma, then subsequently into the Neotropics; it also dispersed into Africa during the Neogene and into the Australian region *ca.* 15Ma, crossing the Weber line. The Saturniini lineages also experienced outstanding biogeographical dynamics, colonizing the New World through Bering on four independent occasions between the late Oligocene and late Miocene (see Chapitre 2 for a detailed analysis of one of

these events), dispersing into the Australian region two times and also expanding again into the African continent twice, and from there to Madagascar on two occasions as well.

Biotic interchanges across the Bering land bridge appear strongly asymmetric, in line with observations in other taxa (Jiang et al. 2019). As detailed above, saturniid moths recolonized the New World from the Eastern Palearctic region through Bering on five independent occasions, which is in contrast with the only dispersal in the reverse direction, i.e. the initial colonization of the Old World by saturniid moths during the early Eocene. Recent analyses focusing on the *Saturnia* genus (Rubinoff & Doorenweerd 2019) suggested a possible recent dispersal through the North Atlantic Land Bridge from Eastern North-America to Western Palearctic of the ancestor to subgenera *Saturnia* and *Eudia*, an hypothesis based on incompletely resolved phylogenetic results and that is contradicted by both our phylogenomic and historical biogeography analyses. The two most ancient colonization events of the New World happened *ca.* 32.8-26.5 Ma [35.4-23.5 Ma] in the ancestors of genera *Copaxa* and *Rothschildia*, both mostly distributed nowadays in Central and South-America where they probably dispersed and diversified after the formation of the Central American volcanic arc and the closure of the Central American Seaway during the middle Miocene (Coates & Stallard 2013, Montes et al. 2012; see Chapitre 2), a connection that also allowed dispersal of neotropical lineages of saturniid moths in the reverse direction (see below). Interestingly many of the *Copaxa* and *Rothschildia* species diversified on Andean slopes where the climate and environment are more similar to the conditions they encountered during their journey through the Nearctic region, suggesting some degree of niche conservatism (see Chapitre 2). The timing of the colonization of the New World by genus *Antheraea* remains uncertain (18.0-0.0 Ma [21.3-0.0 Ma]) because of the genus-level sampling strategy, but its lower diversity (only 4 species, vs. 128 and 79 in *Copaxa* and *Rothschildia*, respectively) and its very limited range in South-America could be the results of a more recent dispersal into environments already inhabited by large Saturniinae representatives. The last two dispersal events through Bering (*Hyalophora*/*Callosamia* and in genus *Actias*) were more recent, *ca.* 11.6-8.4 Ma [13.9-6.5 Ma], and did not lead to the colonization of the South American continent.

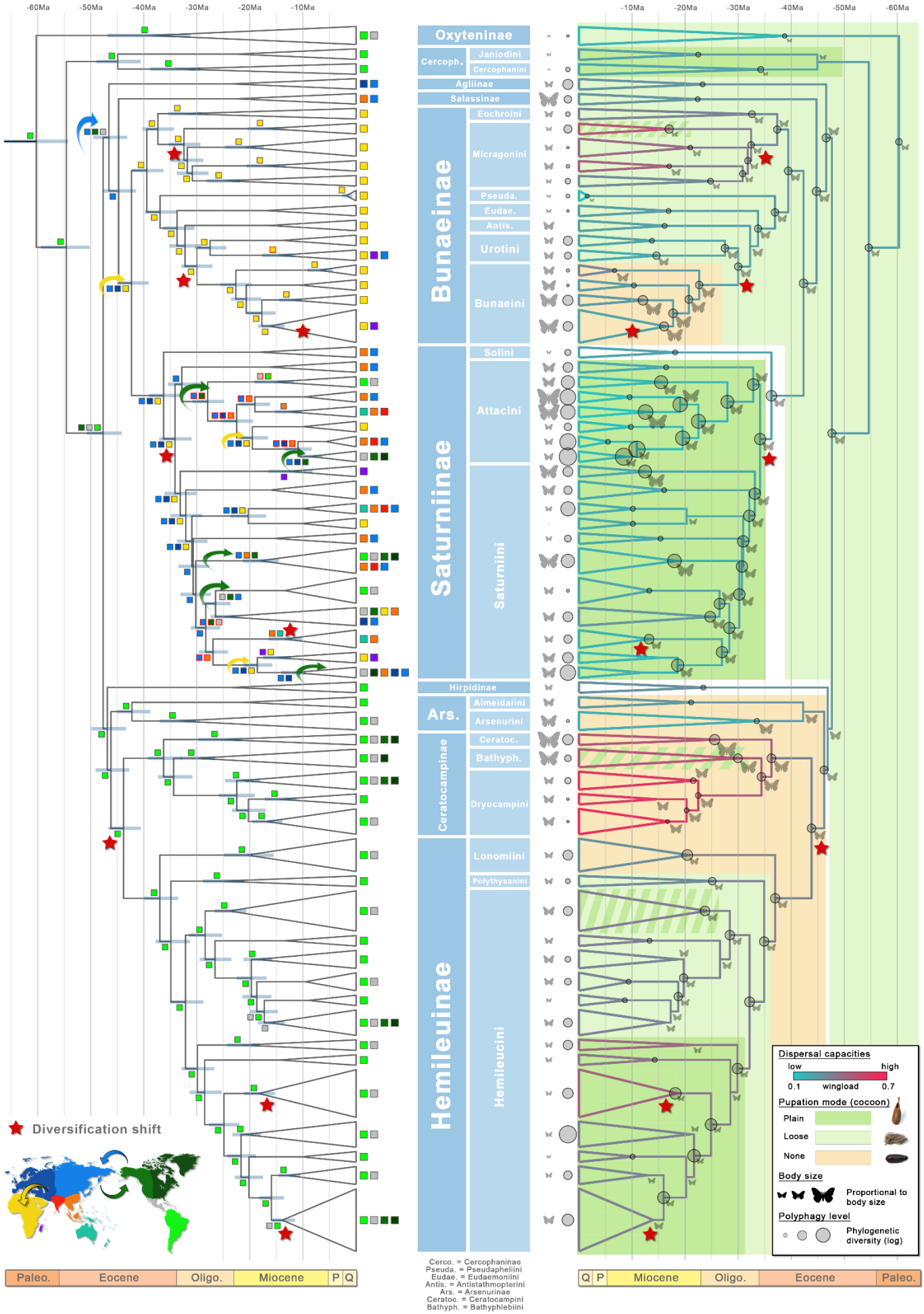
The first colonization of Africa by the ancestor to Bunaeinae + Saturniinae happened between 44.8 and 39.4 Ma [47.7-35.7 Ma], during the middle Eocene climatic optimum (MECO, ~40 Ma), a period of global warming that interrupted the Cenozoic cooling trend (Zachos et al 2001) during which several groups of Asian vertebrates dispersed across the Tethys Sea to colonize Africa (Chaimanee et al. 2012; Huchon et al. 2007). Saturniidae probably also dispersed to Africa through the Alboran and Apulian platforms that were largely emerged during the late Lutetian (44-41 Ma) (Vandenberghe et al. 2012) and the broad-leaved mixed deciduous and evergreen forests occurring from central China to the Tethyan islands (Axelrod 1975; Dutta et al. 2011) might have played a significant role in the successful dispersals of Saturniidae. Three other, more recent, colonization events of Africa were recovered. Whereas *Saturnia* colonized North Africa recently (<5.8 Ma), most likely from Western Europe, the

colonization of Africa both by *Epiphora* (22.5-19.5 [25.4-16.6 Ma]) and by the ancestor to *Argema* and *Actias* (27.0-18.6 [29.4-15.7 Ma]) appear to be concomitant with similar dispersal events identified in other organisms (e.g. pliopithecine primates, Harrison & Yumin 1999).

In the Oriental region, saturniid moths crossed the Weber line several times, colonizing the Australian region. All but one of these events are recent and cannot be precisely dated here as they occurred within extant genera. The older colonization within the Saturniini tribe occurred *ca.* 27.0-13.2 Ma [29.4-10.0 Ma]. Although that timing remains imprecise, it agrees with the northward movement of the Australian plate and this dispersal event likely happened after the initiation of the collision of the Sula Spur promontory with the Sundaland south margin, *ca.* 23 Ma (Hall 2011; de Bruyn et al. 2014).

Concomitantly to the complex history of Old World saturniids described above, several lineages of these moths diverged in South-America in the early times of the diversification of the family (47-36Ma). The bulk of the diversity of subfamilies Hemileucinae, Ceratocampinae and Arsenurinae remains distributed in this region, but several independent dispersal events into Central America and the Nearctic region occurred. Our results reveal that most of these events occurred at terminal branches (within genera) and thus, are relatively recent (probably during Pliocene or Quaternary, after the Panama isthmus formation, Coates & Stallard 2013), though two nearly simultaneous dispersal events from South to Central America occurred during the lower Miocene in both Hemileucinae and Ceratocampinae subfamilies (Figures 4 and S5, left panels) possibly enabled by the formation of the Central American volcanic arc and the closure of the Central American Seaway (CAS; Sepulchre et al. 2014; Montes 2015), thus crossing the route of Saturniinae genera *Rothschildia* and *Copaxa*, making their way from Central to South-America at the same period.

Figure 4 (following page) – Historical biogeography and traits evolution of Saturniidae. Tips were artificially added and nodes collapsed to reflect clade species richness (no proportionality). Red stars depict the diversification rate shifts as estimated with BAMM. Left panel: biogeographic history of Saturniidae as estimated with DECX on 1000 PASTIS, species-level trees. Only the best ranges are represented. At nodes, red border was used when the confidence about the estimation was <50%. Arrows highlight the main dispersal events as represented in the World map at the left bottom. Blue bars indicate node age 95% confidence intervals. Right panel: saturniid life traits evolution. Branches were colored according to wingload values. Background colors represent the pupation modes. Shaded silhouette size are proportional to body size (log-normalized). Polyphagy level is depicted at tips and nodes with circles which size is proportional to the estimated Polyphagy Diversity score (log-normalized). The full phylogeny is represented in the Figure S5.


-60Ma
-50Ma
-40Ma
-30Ma
-20Ma
-10Ma
Oxyteninae
Cercoph.
Janiodini
Cercophanini
Agliinae
Salassinae
Bunaeinae
Eochroini
Micragonini
Pseuda.
Eudae.
Antis.
Urotini
Bunaeini
Saturniinae
Solini
Attacini
Saturniini
Hirpidinae
Ars.
Almeidaiini
Arsenurini
Ceratocampinae
Ceratoc.
Bathyph.
Dryocampini
Hemileuinae
Lonomiini
Polythysanini
Hemileucini
Diversification shift
Dispersal capacities
low
high
0.1
wingload
0.7
Pupation mode (cocoon)
Plain
Loose
None
Body size
Proportional to body size
Polyphagy level
Phylogenetic diversity (log)
Paleo.
Eocene
Oligo.
Miocene
P
Q
Cerco. = Cercophaninae
Pseuda. = Pseudapheliini
Eudae. = Eudaemoniini
Antis. = Antistathmopterini
Ars. = Arsenurinae
Ceratoc. = Ceratocampini
Bathyph. = Bathyphlebiini

Shifts in diversification rates are linked to major environmental changes

Our investigation of diversification rates using *BAMM* showed that the prior applied on the number of shifts influenced the number of recovered shifts when using values lower than 5 (Figure S9). However, when considering higher prior values, the number of recovered shifts was no more influenced significantly and the detected shifts were consistent (Figure S10, Online resources). We here considered the results obtained with *expectedNumberOfShifts*=5 because this value is the closest to the one suggested by the *setBAMMpriors* function of the BAMMtools R package (*expectedNumberOfShifts*=1) that led to steady estimates. We recovered 19 shifts that all implied an increase in diversification rates. Some of those shifts should however be considered cautiously. Since we assumed the monophyly of sampled genera when building the species-level phylogeny, our estimates of crown ages and divergence times may have been biased if this condition was not respected, which would then lead to the identification of erroneous shifts in diversification rate. To avoid these biases, we therefore discarded intra-generic shifts (but one, see further) as well as shifts identified at nodes grouping only two genera. However, if the monophyly of a genus was well established, the estimate of its crown age would be biased toward its stem age and the recovered shift would be underestimated. This is why we considered as valid the shift identified within the genus *Hylesia*, whose monophyly is well established.

That being said, our analyses revealed several shifts throughout the evolutionary history of the family. The earliest shift occurred *ca.* 44 Ma in the branch leading to extant Ceratocampinae and Hemileucinae subfamilies, in South-America. Within the Hemileucinae, two other shifts were identified, both *ca.* 16 Ma. The first happened along the branch leading to genera *Leucanella*, *Pseudautomeris*, *Hypermerina* and *Automeris*, and the second happened within the *Hylesia* genus. These shifts may be linked to the Middle Miocene Climatic Optimum and the rise of the Northern Andes that initiated about 23 Ma and accelerated about 15 Ma (Garzione et al 2008 2014; Hoorn et al. 2010) fundamentally changing the neotropical ecosystems. All other diversification rate shifts we identified occurred in Old World lineages. The most ancient happened *ca.* 35 Ma along the branch of the tree leading to the Attacini and Saturniini tribes. Within Saturniini, we identified another shift *ca.* 12 Ma when the lineage leading to genera *Neodiphthera* and *Syntherata* emerged in the Oriental and Australian regions. Furthermore, three shifts were detected within the Bunaeinae subfamily, on the African continent. The first two happened *ca.* 35 Ma and 31 Ma on the branches leading to the Micragonini tribe and to tribes Bunaeini and Urotini, respectively. They were concomitant with the global climate transition towards a cooler climate that happened at the Eocene–Oligocene boundary (~33.5–26 Mya; Zachos et al. 2001) that induced a dramatic faunal and floral turnover all over the globe (Ivany et al. 2000; Coxall & Pearson 2007; Seiffert 2007). A third, more recent shift in diversification rates also occurred within the Bunaeini tribe on the African continent, *ca.* 10 Ma [14.8-7.1 Ma]), and may be linked to the aridification of North Africa (Zhang et al. 2014), the emergence of C4 plants and the fragmentation of forest habitats in the Eastern side of the continent (Cerling et al. 1997; Ségalen et al. 2007).

Life-history traits played a key role in driving the diversification of wild silkmoths

The results presented and discussed above depict the long and complex dynamics of diversification of wild silkmoths through time and across biogeographical regions of the world. This work represents one of the very few analyses of diversification at global scale in a diverse insect group (but see Economo et al. 2018, 2019; Chazot et al. 2020) and it is remarkable that these spatial and temporal patterns combine both marked structuration and striking heterogeneity, suggesting that beyond contingencies caused by abiotic events having affected the planet in its past 60 Myr, biotic features or traits specific of certain lineages likely played a major role in driving the diversification dynamics of these insects. We therefore investigated the role of biotic traits in the diversification of wild silkmoths. Our analyses focused on four main life-history traits which together make the family rather distinctive within lepidopterans: (i) body size; (ii) dispersal capacity; (iii) degree of polyphagy; (iv) pupation mode. Overall, all four life-history traits considered explained significantly better the diversification dynamics than null models in which the diversification rates are constant across the phylogeny (Table 1).

Table 1 – Results of the MuSSE analysis of trait-dependent diversification models fitted to the phylogeny of Saturniidae. For each life-history trait, we compared three models: (i) a null model in which the diversification rates were homogenous across the phylogeny, (ii) a qfree model in which the transition rates between the different categories were not fixed as equal (but $q_{13} \sim 0$, $q_{31} \sim 0$), (iii) a full model in which we estimated distinct diversification rates for the three categories as well as distinct transition rates. For all traits, the full model was the best model identified by the AIC scores. ΔAIC indicates the AIC difference to the best model. Df is the number of degrees of freedom for each model. AIC impr. charcterizes the relative improvement of AIC score of the full model when compared to the null model.

		Df	lnLik	AIC	ΔAIC	AIC impr. (%)
Body Size 3448 tips	null model	3	-10132	20270	32	
	qfree model	6	-10123	20257	19	
	full model	10	-10109	20238	0	0,16
Dispersal capacities 3448 tips	null model	3	-10006.5	20019	74	
	qfree model	6	-10004.3	20021	72	
	full model	10	-9962.6	19945	0	0,37
Polyphagy level 3056 tips	null model	3	-8836.2	17678	194	
	qfree model	6	-8825.6	17663	179	
	full model	10	-8731.9	17484	0	1,10
Pupation mode 2951 tips	null model	3	-8402.0	16810	75	
	qfree model	6	-8392.9	16798	63	
	full model	10	-8357.6	16735	0	0,45

The body size conundrum of capital-breeding insects

The analyses of measurements from images of 1576 species in 172 genera (2577 specimens, and 12,850 measurements in total; see Table S4 - Online resources) reveal that the common ancestor to all wild silkmoths was small, falling into the bottom tier of the smallest extant saturniids. Body size carries phylogenetic signal (Bloomberg's K=0.84; p=0.001), but the evolution toward smaller or larger body sizes occurred several times throughout the evolution of the family. Salassinae, Bunaeini, Attacini, Bathyphlebiini and Citheroniini are striking examples of independent evolution of gigantism, whereas

Pseudapheliini, Micragonini, Oxyteninae, and several genera in Dryocampini, Hemileucini and Saturniini are small moths that can be as much as 20 times smaller than the largest members of the family. Our results reveal an overall high heterogeneity in the evolution of this trait, with similar transition rate toward larger or smaller body size (Figure 5E, Table S5). Interestingly, this contrasts with the negatively skewed distribution of saturniid body size (log-transformed; Figure S6), showing that larger saturniids are more numerous than smaller ones. Furthermore, large body-sized moths (Figure 5A, Table S5) had higher diversification rates than small ones, whereas lineages of medium-sized moths have intermediate values. These results suggest that, for saturniids, evolution toward smaller body size can be an evolutionary dead-end, an observation contrasting with previously proposed models in other groups of animals (Hutchinson & MacArthur 1959; Maurer 1998), including insects (Rainford et al. 2016; but see Misof 2002 for a similar pattern in Anisoptera), and that may reflect the importance of being large when you are a capital-breeder. The capital-breeding reproductive strategy of wild silkmoths indeed implies storage by the larvae of the important resources needed by the adults for their reproduction and for the formation of a large amount of formed eggs (Tammaru & Haukioja 1996; Davis et al. 2016).

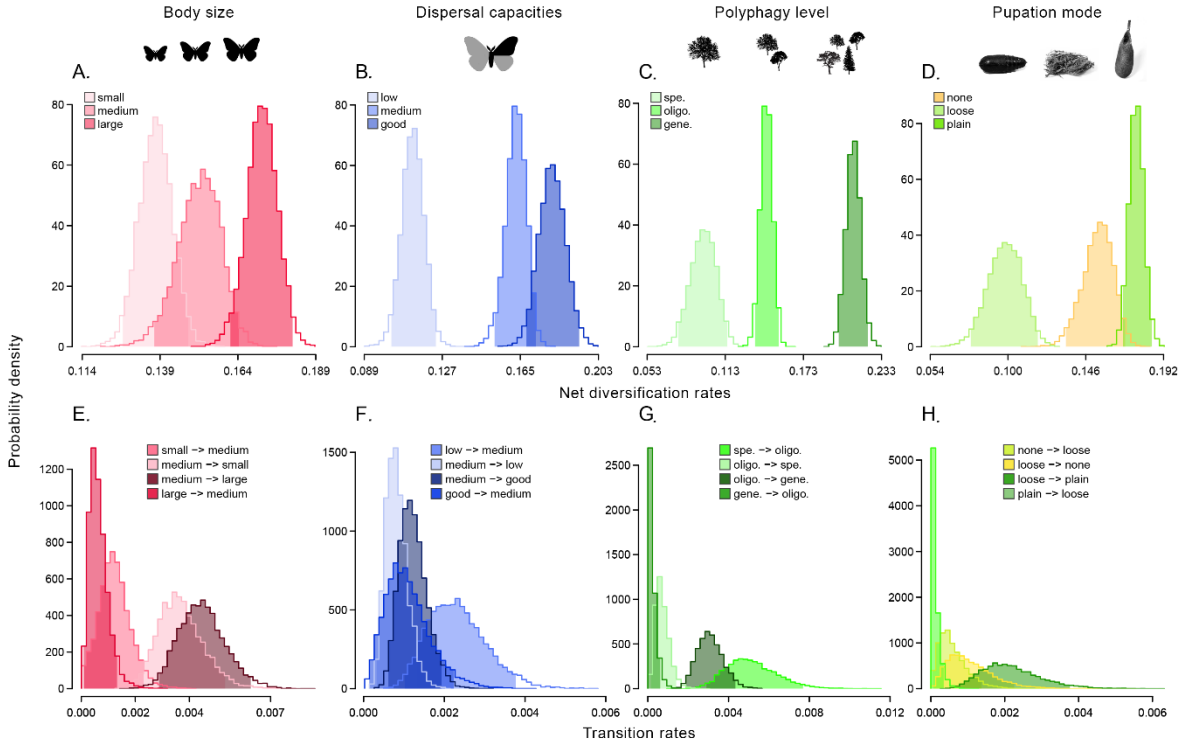


Figure 5 - Posterior probability distributions of parameters obtained from the MuSSE analysis of trait dependent diversification models. Upper row (A, B, C, D): net diversification rate (λ-μ); lower row (E, F, G, H): transition rates. Shaded areas correspond to the 95% credibility intervals.

Dispersal capacity promotes diversification, but not biogeographical interchanges

From thousands of measurements of body and wing metrics (Table S4), we calculated wingload ratio to approximate the relative dispersal capacity within Saturniidae. Overall, most saturniid lineages are poor

dispersers with a rather low wingload ratio (Figure 4); our reconstruction of ancestral states for this trait suggests that this was the condition in the ancestor to all Saturniidae, which was also of small size. A significant increase in dispersal capacity happened twice independently during the evolution of the family: in the Neotropical subfamily Ceratocampinae (especially in the Dryocampini tribe) and, to a lesser extent, within the African tribe Micragonini (Figures 4, S5). In both cases these moths have a rather broad thorax and short "triangular" wings allowing high wing beat frequency, and their high wingload ratio reflects field observations that these moths are strong and fast fliers. Overall, the phylogenetic signal of wingload appears to be strong, following rather tightly the evolution of the family (Bloomberg's K=1.04; p=0.001). The diversification rates in lineages with high or medium dispersal capacity (i.e. with high wingload values) are higher than in lineages with low dispersal capacities (Figure 5B, Table S5); this agrees with observations in other groups (e.g. Willis et al. 2014) and with predictions that higher dispersal ability increases the probability of isolation following colonization of novel habitats (Faurby et al. 2019). Interestingly, when considering diversity of saturniid moths in the Andes, which represent the largest hotspot of diversity for the family, lots of species are local endemics with very narrow ranges; this pattern is generally understood as the result of the very limited dispersal ability of wild silkmoths, which could then also be seen as responsible for increased diversification rates, under a purely neutral model, because lower gene flow promotes population fragmentation, genetic divergence and local adaptation, leading to speciation and prevalence of endemism and small range species (Faurby & Antonelli 2018). One way to reconcile this apparent contradiction may in fact be considering all saturniid moths as poor dispersers, a trait again linked to their reproductive strategy as capital-breeders, with short-lived non-feeding adults; this promotes disruptions to gene flow, which in turn are likely to happen more frequently in those saturniids that have higher dispersal abilities and can occasionally cross barriers or expand into novel ecological habitats, especially, for instance, in mountain environments where these moths are especially diverse. One final consideration with respect to dispersal capacity is that it does not appear to have played an important role in promoting long distance dispersal of wild silkmoths (Figure 4); as mentioned before, the group with the highest dispersal capacity, *i.e.* the members of the Ceratocampinae subfamily and the Micragonini tribe, never dispersed outside the New World and the African continent, respectively (Figure 4).

Polyphagy – the engine of wild silkmoths' diversification

In contrast with nearly all other diverse groups of insects, especially those whose diversity peaks in the inter-tropical region, our current knowledge of the biology of wild silkmoths is rather well documented thanks to the long-lasting popularity of these often-spectacular moths and caterpillars. Food plant information were compiled for 586 species of Saturniidae representing 121 genera in all subfamilies and tribes but Hirpidinae and Antistathmopterini; overall, we retrieved 2568 distinct records of associations between saturniid genera and the plant families consumed by their caterpillars (Table S6, Online resources). Polyphagy appears to be very variable within the family (Figures 1, 4 and S5); it ranges from

genera with specialized species known to feed on a single plant family (e.g. genus *Tagoropsis* with all five documented species out of six feeding on Sapindaceae) to some mostly comprising species known as generalists, feeding on several dozens of plant families (e.g. *Actias luna* caterpillars feeding on >16 different families of plants). Alike body size and wingload, there is a phylogenetic signal in the evolution of polyphagy in saturniid moths (Bloomberg's K= 0.547; p=8.0e-04). We inferred that the ancestor of all Saturniidae was oligophagous, its degree of polyphagy being close to the median of the polyphagy scores measured for extant genera. The degree of polyphagy increased independently several times throughout the evolution of the family; the ancestor to tribes Attacini and Saturniini was polyphagous and all extant genera in the former tribe are either highly polyphagous (e.g. *Hyalophora*) or oligophagous (e.g. *Epiphora*), whereas several extant genera of the Saturniini specialized on a limited number of plant families (e.g. *Ceranchia*, *Copaxa*). The Bunaeinae subfamily mostly comprises genera with caterpillars feeding on few plant families, though several genera in the Bunaeini tribe became quite generalists, with genus *Bunaea* being the second most polyphagous Saturniidae genus. Among the three main New World lineages of Saturniidae, subfamilies Arsenurinae and Ceratocampinae (with few exceptions therein) have specialized caterpillars, whereas in the Hemileucinae most genera are either oligophagous or polyphagous (Figures 1, 4, S5). The most diverse saturniid clades are polyphagous (e.g. *Hylesia*, *Automeris*, Attacini and Saturniini), while in contrast several specialized clades are poorly diversified (e.g. Dryocampini, Oxyteninae or Arsenurinae). Importantly, our results revealed that lineages with highly polyphagous caterpillars diversified 66% and 46% faster than specialized and oligophagous lineages, respectively, and that the level of polyphagy is the trait that best explains the diversification dynamics of the family (Figure 5C, Table 1). Furthermore, the transition rates toward specialization are significantly lower than the rates toward generalism, suggesting that the dominant evolutionary trend was toward an increase of the polyphagy level (Figure 5G, Table S5). Plasticity in the use of food plants also explains the heterogeneity of diversification rates within the four most diversified subfamilies of wild silkmoths (Table 2 and Figure S7 - Online resources): the polyphagy-dependent models indeed show that this trait positively influenced the diversification rate in the Hemileucinae, Bunaeinae and Saturniinae, three subfamilies characterized by high levels of polyphagy, but not in the Ceratocampinae, whose caterpillars generally feed on a limited set of food plants.

Table 2 – Results of the MuSSE analyses of polyphagy-dependent diversification models fitted to the phylogeny of the four most diverse subfamilies of wild silkmoths. For each subfamily, we compared three models: (i) a null model in which the diversification rates were homogenous across the phylogeny, (ii) a qfree model in which the transition rates between the different polyphagy levels were not fixed as equal (but $q_{13} \sim 0$, $q_{31} \sim 0$), (iii) a full model in which we estimated distinct diversification rates for the three polyphagy categories as well as distinct transition rates. In subfamilies Bunaeinae, Hemileucinae, and Saturniinae, the full model was ranked first according to AIC scores. In Ceratocampinae (Cerato.), polyphagy level did not better explain species diversity. ΔAIC indicates the AIC difference to the best model. Df is the number of degrees of freedom for each model. The evolution of traits and diversification rates throughout the saturniid evolution can be seen in Figure S7 (Online resources).

Polyphagy-dependent diversification models		Df	lnLik	AIC	ΔAIC
Bunaeinae	null model	3	-1203.7	2413	17
	qfree model	6	-1199.5	2411	15
	full model	10	-1183.1	2396	0
Cerato.	null model	3	-709.94	1426	1
	qfree model	6	-706.64	1245	0
	full model	10	-704.39	1429	3
Hemileuc.	null model	3	-3605.1	7216	26
	qfree model	6	-3596.7	7205	15
	full model	10	-3585.2	7190	0
Saturniinae	null model	3	-1897.7	3801	4
	qfree model	6	-1896.6	3805	8
	full model	10	-1888.5	3797	0

Higher diversification rates in polyphagous lineages is one of the corollaries of the *oscillation hypothesis* proposed by Janz & Nylin (2008; see also Jousselin & Elias 2019) to explain how host preferences can drive speciation and diversification in herbivore insects. According to this hypothesis, generalist feeding diet is a transient condition stage where generalization opens novel ecological opportunities and fuels diversification through the specialization process. This however implies that groups whose diversification is driven by food plant use tend to become specialists, which is not what we observe here in wild silkmoths that instead display higher rates toward generalism (Figure 5G, Table S5). Wang et al. (2017) recently claimed that they found support for the *oscillation hypothesis* in a group of Lepidoptera with generalist caterpillars (tussock moths, Lymantriinae), observing that the most polyphagous lineages were also the most speciose, but there is still ample debate on the predictions linked to this hypothesis and to alternative ones (Hardy & Otto 2014; Jousselin & Elias 2019). Specialization is hypothesized to confer higher physiological efficiency, enemy-free space (e.g. physical or chemical crypsis, toxicity gained from the plant), as well as optimal foraging conditions (Singer 2008). However, while most herbivorous insects tend to evolve toward specialization over evolutionary time, thus gaining greater fitness, wild silkmoths follow the opposite path toward generalism. Such rare pattern was documented in fruit flies (Clarke 2017) and shown to be a "probable" evolutionary outcome driven by a number of ecological trade-offs linked to the life history and biotic interactions, both with their host and with their parasitoids. Polyphagy was proposed as a corollary of capital-breeding reproductive strategy, because short female adult life-span and its large body size reduces its motility; females are "pressed for time" and then benefit from being less selective for the plant they oviposit on (Prinzing 2003; Jervis et al. 2006; Davis et al. 2013). In a similar way as capital-breeding strategy was

shown to drive saturniid moths toward becoming larger, it is hypothesized here as the main force driving their evolution toward increased polyphagy of their caterpillars. In line with the first stage of the *oscillation hypothesis* discussed above, high level of polyphagy in saturniid caterpillars can create novel ecological opportunities and favor geographical range expansion and successful establishment in newly colonized areas. As a result, the probability of population fragmentation increases, especially considering the low dispersal abilities of adult saturniids, and this can lead to isolation and speciation. Polyphagy then acts as a "diversification engine" that is not necessarily fueled by specialization; this subsequent predicted stage of the *oscillation hypothesis* is indeed antagonized by the ecological trade-offs and selective pressures induced by the capital-breeding reproductive strategy of these insects. In other words, and in contrast with some previous studies on butterflies or moths (Janz & Nylin 2008; Wang et al. 2017), our results suggest that specialization or shift to new food plants are not major drivers of diversification in wild silkmoths; the level of polyphagy however plays a key role in promoting speciation and their diversification. Also, though not all highly polyphagous lineages of Saturniidae successfully expanded their ranges beyond their biogeographical region, our results show that the two tribes of Saturniinae – Attacini and Saturniini – displaying the highest frequency of long-distance dispersal events are also those comprising the most polyphagous genera in the family (Figures 4 and S5). The ability of caterpillars to use diverse food plants may have greatly facilitated the establishment and adaptation of populations in new environments with different plant communities.

Protection of pupae promoted diversification

The large-sized and motionless pupae of saturniid moths represent a defenseless source of proteins for predators or parasitoids. It is also the most common diapausing stage within the family when individuals need to pass cold or dry seasons. Cocoons of wild silkmoths form remarkable cases surrounding the pupa and conferring both physical and chemical protection (Gross 1993; Nirmala et al. 2001); some are very strong and several species are actually well known from the silk produced by their caterpillars used as an alternative to that produced by domesticated silkworms (Sutherland et al. 2010; Peigler & Oberprieler 2017). Not all saturniids however do spin strong cocoons, and some even do not spin at all and pupate underground. Expecting that this life-history trait may have played an important role in the evolution of wild silkmoths, we compiled information about the pupation mode of 139 Saturniidae genera or subgenera representing all subfamilies but Hirpidinae, and all tribes but Antistathmopterini and Solini (Table S7 - Online resources). The caterpillars of about one third of these documented taxa actually pupate underground, without spinning a cocoon, while the other two thirds pupate on the ground or in the vegetation, inside a cocoon spun with silk and that shows different degrees of elaborateness, from a very loose net of silk threads to plain and strong cocoons encasing the pupa. Phylogenetic signal in pupation mode is very strong and obvious (Figures 4 and S5). The ancestor to all extant wild silkmoths must have been pupating in a loose cocoon and the evolution of a plain cocoon, offering a better protection to the pupal stage, occurred several times independently in the ancestors to Cercophaninae,

some Micragonini, Saturniinae, some Hemileucinae, and in one isolated genus (*Neorcarnegia*) of Ceratocampinae. Our results show that the shift to pupation underground, where the caterpillar finds shelter, occurred independently twice during the evolution of the family: first in the ancestor to extant New World lineages (Arsenurinae, Ceratocampinae and Hemileucinae), *ca.* 47Ma, and then a second time *ca.* 25Ma, in the ancestor to the Bunaeini tribe, exclusively Afrotropical. The causes triggering this important behavioral shift remain intriguing and warrant further study; we note however that in both cases here the lineages branching early after the shift to pupation underground have all rather specialized caterpillars feeding on a very limited set of food plants (Figures 4 and S5). Habitat characteristics such as low plant diversity and frequent exposure to harsh environmental conditions (drought, fires), may have driven this shift. In Africa, the caterpillars of early branching lineages of the Bunaeini tribe mostly feed on *Acacia* trees, suggesting that pupation underground might have been an adaptation against desiccation in hot and dry savanna environments where also seasonal loss of leaves and pressure from herbivorous megafauna selected against pupation in the vegetation above ground. On the other hand, it is noteworthy that the capacity to spin a cocoon, either loose or plain, was gained again twice independently within the New World lineages that had shifted, earlier in their evolution, to pupating underground. This happened in genus *Neorcarnegia* of the Ceratocampinae subfamily, and in all Hemileucinae lineages but the Lonomiini; in this tribe, which is sister to all other Hemileucinae, caterpillars actually pupate on the ground in the litter, without spinning any cocoon.

The results of our analyses of trait-dependent diversification models show that wild silkmoths diversified significantly slower in lineages with caterpillars spinning loose cocoons than in those where they pupate underground or spin a plain cocoon (Figure 5D, Table S5). The latter two conditions are efficient strategies to escape predators and parasites (Gross 1993) and survive unfavorable (drought, cold) or even hostile (fire, flooding) environmental conditions (Danks 2004). Better protection of the pupa may thus have increased diversification rate both through decreasing extinction risks – because of higher probabilities for pupae to develop into adults, leading to larger population size – as well as increased speciation rates, because protection of the pupa from environmental hazards promotes range expansion and ecological opportunities, possibly leading to fragmentation and eventually reproductive isolation. Interestingly, the positive diversification shift we identified along the branch leading to tribes Saturniini and Attacini is congruent with a transition from pupation in a loose cocoon to a condition where caterpillars spin dense plain cocoons conferring higher level of protection. Similar transitions however happened independently in several other clades (Figures 4 and S5) without causing an apparent shift in diversification rate, emphasizing that pupation mode, if a key driver of the diversification in some lineages, acts along with other factors and traits to impact evolutionary dynamics. It is also worth noting that further understanding of the role of this trait will likely require a more thorough account of its conditions; in particular, “plain cocoons” as understood in our study refers to strong silk cases (as opposed to loose cocoons or silk nets), but there are conspicuous differences in the strength of these

structures between for instance papery, thin cocoons of many Hemileucinae, and the very strong cocoons of Saturniini and Attacini. Considering that these last two tribes are the most represented in the temperate regions of the globe, strong plain cocoons might have been key to surviving cold seasons and to successfully colonize distant regions as is the case in genera *Rothschildia* and *Copaxa*, all the way from the Eastern Palearctic to Neotropical mountains (both Central American and Andean cordilleras), through the Nearctic region.

Conclusion: Insights into the evolution of capital-breeding insects

Our study offers one of the first comprehensive overview of the temporal and spatial dynamics of diversification in a globally distributed diverse group of phytophagous insects. This reveals a remarkably clear picture of the main events that have shaped the distribution of extant diversity in these moths. Throughout their evolutionary history, wild silkmoths seized multiple opportunities linked to changes in environmental conditions (climate, topology) to expand geographically, from a Neotropical origin during the Paleocene to a global distribution today and proving capable of dispersing across landmasses repeatedly. We show however that not all lineages experienced similar "success" in terms of diversification and range expansion. In particular, our results emphasize the importance of biological traits in shaping the evolution of these moths, both in terms of diversification and geographical mobility. Neither dispersal capacity, nor body size appear to be good predictors of long-distance dispersal; instead the features that seem to be the most critical in successfully colonizing new biogeographical regions are both the ability to exploit a diversity of food plants (polyphagy) and to spin plain cocoons.

Furthermore, the reproductive strategy of wild silkmoths, as capital-breeding insects with non-feeding short-lived adults, seems to represent the major constraint that drove the evolution of life-history traits in directions opposite to those documented to date in other groups of insects. Saturniid moths generally evolved toward larger size, and toward higher polyphagy level. These two characteristics, along with a better protection of the pupa – i.e. the most defenseless stage of development – have promoted diversification in the family. In other words, lineages that best adapted to coping with large body size through increased polyphagy and increased pupal protection diversified more and were more able to colonize biogeographical regions. Tracing back the evolution of wild silkmoths to their split from their sister-group, family Sphingidae (hawkmoths; see Hamilton et al. 2019), it is interesting to observe that these two families followed distinct evolutionary paths, already brought into light by Janzen (1984). In this essay on "two ways to be a tropical big moth" in Costa Rican Santa Rosa national park, he hypothesized that the shift to a capital-breeding reproductive strategy, as opposed to the income-breeding strategy of sphingid moths, was an adaptive response in environments where adult moths are exposed to high predation pressure. As the time of divergence of the two families (*ca.* 60Ma; see also

Kawahara et al. 2019) matches the origin and diversification of bats (Shi & Rabosky 2015; Lei & Dong 2016), one emerging view from the present work is that while most moths and nocturnal insects engaged into an “arm race” with these new major nocturnal insect predators through an arsenal of morphological and/or behavioral adaptations (Kristensen 2012; Kawahara & Barber 2015; Rubin et al. 2018), wild silkmoths instead reduced their exposure to predation through short adult activity, devoted to reproduction, and capitalization on resources accumulated during larval stages.

In that context, the diversification of these capital-breeding moths over macro-evolutionary scale is then better understood primarily as the diversification of their slow-growing large caterpillars and their large pupae that are the most exposed stages of development in terms of both abiotic and biotic interactions. This view then brings new light into further observations and hypotheses, expanding Janzen’s (1984) local appraisal into a broader macroevolutionary scale. One of these is that saturniid caterpillars are overall much more diverse in their forms and defensive mechanisms (e.g. urticating hairs, poisonous spines) than their counterparts in the Sphingidae family whose aspect is overall very homogeneous (Janzen 1984). This can be seen as a consequence of having to retain high level of polyphagy through evolutionary times, thus not evolving toward enemy-free and physiological efficiency advantages of specialization. On the other hand, short-lived adult saturniids have very limited exposure time to predation by bats; females are highly sedentary and males have a very limited period of activity at night (Lamarre et al. 2018), probably not exceeding an hour. This may explain why these moths lack the elaborated morphological adaptations (e.g. tympanal organs) found in many other families of moths (Kristensen 2012) to counter predation by bats, though a few lineages (anecdotical in terms of species diversity within the family) developed rather unique spectacular adaptations against echolocation by bats in the form of long hindwing tails whose tips act as defectors to attacks by bats (Barber et al. 2015). Instead, saturniids are remarkable, when compared to Sphingidae or many other moth families, by the diversity of shapes and patterns of their wings, which may be explained by a relaxed selective pressure on flight capacities (Janzen 1984; Hamilton et al. 2020) and high diurnal predation rate on these large preys, promoting the selection of diverse shapes and patterns to deter or avoid predators through crypsis, aposematic or intimidating patterns (e.g. eyespots; Blest 1957).

References

Aguilée, R., Gascuel, F., Lambert, A., & Ferriere, R. (2018). Clade diversification dynamics and the biotic and abiotic controls of speciation and extinction rates. Nature communications, 9(1), 1-13.

Andersen, M. J., McCullough, J. M., Friedman, N. R., Peterson, A. T., Moyle, R. G., Joseph, L., & Nyári, Á. S. (2019). Ultraconserved elements resolve genus-level relationships in a major Australasian bird radiation (Aves: Meliphagidae). Emu-Austral Ornithology, 119(3), 218-232.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Antonelli, A., Zizka, A., Silvestro, D., Scharn, R., Cascales-Miñana, B., & Bacon, C. D. (2015). An engine for global plant diversity: highest evolutionary turnover and emigration in the American tropics. Frontiers in Genetics, 6, 130.

Axelrod, D. I. (1975). Evolution and biogeography of Madrean-Tethyan sclerophyll vegetation. Annals of the Missouri Botanical Garden, 62, 280-334.

Ballesteros-Mejia, L., Arnal, P., Hallwachs, W., Haxaire, J., Janzen, D. H., Kitching, I. J. & Rougerie R. (submitted). A global food plant dataset for wild silkmoths and hawkmoths, and its use in documenting polyphagy of their caterpillars (Lepidoptera: Bombycoidea: Saturniidae, Sphingidae). Biodiversity Data Journal.

Barber, J. R., Leavell, B. C., Keener, A. L., Breinholt, J. W., Chadwell, B. A., McClure, C. J., ... & Kawahara, A. Y. (2015). Moth tails divert bat attack: evolution of acoustic deflection. Proceedings of the National Academy of Sciences, 112(9), 2812-2816.

Barnosky, A. D. (2001). Distinguishing the effects of the red queen and court jester on Miocene mammal evolution in the northern Rocky Mountains. Journal of Vertebrate Paleontology, 21(1), 172-185.

Barthlott, W., Hostert, A., Kier, G., Küper, W., Kreft, H., Mutke, J., ... & Sommer, J. H. (2007). Geographic patterns of vascular plant diversity at continental to global scales (Geographische Muster der Gefäßpflanzenvielfalt im kontinentalen und globalen Maßstab). Erdkunde, 305-315.

Baskin, J. M., & Baskin, C. C. (2016). Origins and Relationships of the Mixed Mesophytic Forest of Oregon–Idaho, China, and Kentucky: Review and Synthesis1. Annals of the Missouri Botanical Garden, 101(3), 525-552.

Beeravolu, C. R., & Condamine, F. L. (2016). An extended maximum likelihood inference of geographic range evolution by dispersal, local extinction and cladogenesis. BioRxiv.

Benton, M. J. (2009). The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. Science, 323(5915), 728-732.

Benton, M. J. (2016). Origins of biodiversity. PLoS Biology, 14(11), e2000724.

Blaimer, B. B., Lloyd, M. W., Guillory, W. X., & Brady, S. G. (2016). Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PloS one, 11(8), e0161531.

Blest, A. D. (1957). The evolution of protective displays in the Saturnioidea and Sphingidae (Lepidoptera). Behaviour, 11(4), 257-309.

Blomberg, S. P., Garland Jr, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution, 57(4), 717-745.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.

Bonaparte, J. F. (1984). Late Cretaceous faunal interchange of terrestrial vertebrates between the Americas. Reif W. E. & Westphal F. (eds.), Third Symposium on Mesozoic Terrestrial Ecosystems, Short Papers. Tiibingen: Attempto Verlag, 19-24.

Bonetti, M. F. & Wiens, J. J. (2014). Evolution of climatic niche specialization: a phylogenetic analysis in amphibians. Proceedings of the Royal Society B: Biological Sciences, 281(1795), 20133229.

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation & computing of summary statistics. PeerJ, 4, e1660.

Bossert, S., Murray E. A., Blaimer B. B. & Danforth B. N. (2017). The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. Molecular Phylogenetics and Evolution, 111, 149-157.

Brechlin, R. (2019). *Hirpsinjaevia* gen. nov. *viksinjaevi* sp. nov., eine neue Saturniide aus Peru (Lepidoptera). Entomo-Satsphingia, 12(1), 65-68.
Brinkmann, H., van der Giezen, M., Zhou, Y., de Raucourt G. P., & Philippe H. (2005). An Empirical Assessment of Long-Branch Attraction Artefacts in Deep Eukaryotic Phylogenomics. Systematic Biology, 54(5), 743-757.
Burin, G., Kissling, W. D., Guimarães, P. R., Şekercioğlu, Ç. H., & Quental, T. B. (2016). Omnivory in birds is a macroevolutionary sink. Nature Communications, 7(1), 1-10.
Cardoso, P., Erwin, T. L., Borges, P. A., & New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. Biological Conservation, 144(11), 2647-2655.
Cerling, T. E., Harris, J. M., MacFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V., & Ehleringer, J. R. (1997). Global vegetation change through the Miocene/Pliocene boundary. Nature, 389(6647), 153-158.
Chaimanee, Y., Chavasseau, O., Beard, K. C., Kyaw, A. A., Soe, A. N., Sein, C., ... & Rugbumrung, M. (2012). Late Middle Eocene primate from Myanmar and the initial anthropoid colonization of Africa. Proceedings of the National Academy of Sciences, 109(26), 10293-10297.
Chazot, N., Condamine, F., Dudas, G., Peña, C., Matos-Maraví, P., Freitas, A. V., ... & Lohman, D. J. (2020). The latitudinal diversity gradient in brush-footed butterflies (Nymphalidae): conserved ancestral tropical niche but different continental histories. BioRxiv.
Claramunt, S., & Cracraft, J. (2015). A new time tree reveals Earth history's imprint on the evolution of modern birds. Science advances, 1(11), e1501005.
Clarke, A. R. (2017). Why so many polyphagous fruit flies (Diptera: Tephritidae)? A further contribution to the 'generalism' debate. Biological Journal of the Linnean Society, 120, 245-257.
Coates, A. G. & Stallard R. F. (2013). How old is the Isthmus of Panama? Bulletin of Marine Science, 89(4), 801-813.
Condamine, F. L., Silva-Brandão, K. L., Kergoat, G. J., & Sperling, F. A. (2012). Biogeographic and diversification patterns of Neotropical Troidini butterflies (Papilionidae) support a museum model of diversity dynamics for Amazonia. BMC Evolutionary Biology, 12(1), 82.
Condamine, F. L., Sperling, F. A., & Kergoat, G. J. (2013). Global biogeographical pattern of swallowtail diversification demonstrates alternative colonization routes in the Northern and Southern hemispheres. Journal of Biogeography, 40(1), 9-23.
Condamine, F. L., Clapham, M. E., & Kergoat, G. J. (2016). Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? Scientific Reports, 6, 19208.
Condamine, F. L., Rolland, J., Höhna, S., Sperling, F. A., & Sanmartín, I. (2018). Testing the role of the Red Queen and Court Jester as drivers of the macroevolution of Apollo butterflies. Systematic biology, 67(6), 940-964.
Cooney, C. R., Bright, J. A., Capp, E. J., Chira, A. M., Hughes, E. C., Moody, C. J., ... & Thomas, G. H. (2017). Mega-evolutionary dynamics of the adaptive radiation of birds. Nature, 542(7641), 344-347.
Coxall, H. K., Pearson P. N., Williams, M., Haywood, A. M., Gregory, F. J., Schmidt, D. N. (2008). The Eocene-Oligocene transition, Deep time perspectives on climate change: marrying the signal from computer models and biological processes. London (UK), Geological Society London, 351-387
Cozzarolo, C. S., Balke, M., Buerki, S., Arrigo, N., Pitteloud, C., Gueuning, M., ... & Alvarez, N. (2019). Biogeography and ecological diversification of a Mayfly Clade in New Guinea. Frontiers in Ecology and Evolution, 7, 233.
Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., ... & Rasplus, J. Y. (2019). Optimized DNA extraction and library preparation for minute arthropods: application to target enrichment in chalcid wasps used for biocontrol. Molecular Ecology Resources, 19(3), 702-710.
Cruaud, A., Delvare, G., Nidelet, S., Sauné, L., Ratnasingham, S., Chartois, M., ... & van Noort, S. (2020). Ultra-Conserved Elements and morphology reciprocally illuminate conflicting phylogenetic hypotheses in Chalcididae (Hymenoptera, Chalcidoidea). Cladistics.
Danks, H. V. (2004). The roles of insect cocoons in cold conditions. European Journal of Entomology, 101(3), 433-438.

Davis, R. B., Õunap, E., Javoiš, J., Gerhold, P., & Tammaru, T. (2013). Degree of specialization is related to body size in herbivorous insects: a phylogenetic confirmation. Evolution: International Journal of Organic Evolution, 67(2), 583-589.

Davis, R. B., Javoiš, J., Kaasik, A., Õunap, E., & Tammaru, T. (2016). An ordination of life histories using morphological proxies: capital vs. income breeding in insects. Ecology, 97(8), 2112-2124.

De Bruyn, M., Stelbrink, B., Morley, R. J., Hall, R., Carvalho, G. R., Cannon, C. H., ... & Maiorano, L. (2014). Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. Systematic Biology, 63(6), 879-901.

de Camargo, N. F., de Camargo, W. R., do CV Corrêa, D., de Camargo, A. J., & Vieira, E. M. (2016). Adult feeding moths (Sphingidae) differ from non-adult feeding ones (Saturniidae) in activity-timing overlap and temporal niche width. Oecologia, 180(2), 313-324.

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution, 24(6), 332-340.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics, 6(5), 361-375.

Diniz-Filho, J. A. F., De Marco P. & Hawkins B. A. (2010). Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. Insect Conservation and Diversity, 3, 172-179.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7(1), 1-8.

Dudley, R. & Srygley R. (1994). Flight physiology of neotropical butterflies: allometry of airspeeds during natural free flight. Journal of Experimental Biology, 191(1), 125-139.

Dutta, S., Tripathi, S. M., Mallick, M., Mathews, R. P., Greenwood, P. F., Rao, M. R., & Summons, R. E. (2011). Eocene out-of-India dispersal of Asian dipterocarps. Review of Palaeobotany and Palynology, 166(1-2), 63-68.

Economo, E. P., Narula, N., Friedman, N. R., Weiser, M. D., & Guénard, B. (2018). Macroecology and macroevolution of the latitudinal diversity gradient in ants. Nature communications, 9(1), 1-8.

Economo, E. P., Huang, J. P., Fischer, G., Sarnat, E. M., Narula, N., Janda, M., ... & Knowles, L. L. (2019). Evolution of the latitudinal diversity gradient in the hyperdiverse ant genus *Pheidole*. Global Ecology and Biogeography, 28(4), 456-470.

Ezard, T. H., Aze, T., Pearson, P. N., & Purvis, A. (2011). Interplay between changing climate and species' ecology drives macroevolutionary dynamics. Science, 332(6027), 349-351.

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic biology, 61(5), 717-726.

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. Methods in Ecology and Evolution, 8(9), 1103-1112.

Faircloth, B. C., Alda, F., Hoekzema, K., Burns, M. D., Oliveira, C., Albert, J. S., ... & Sidlauskas, B. L. (2020). A target enrichment bait set for studying relationships among ostariophysan fishes. Copeia, 108(1), 47-60.

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. Biological Conservation, 61(1), 1-10.

Faurby, S., & Antonelli, A. (2018). Evolutionary and ecological success is decoupled in mammals. Journal of Biogeography, 45(10), 2227-2237.

Faurby, S., Werdelin, L., & Antonelli, A. (2019). Dispersal ability predicts evolutionary success among mammalian carnivores. BioRxiv.

FitzJohn, R. G., Maddison, W. P., & Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Systematic Biology, 58(6), 595-611.

FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution, 3(6), 1084-1092.

Fiz-Palacios, O., Schneider, H., Heinrichs J. & Savolainen V. (2011). Diversification of land plants: insights from a family-level phylogenetic analysis. BMC Evolutionary Biology, 11, 341.

French, C. M., Deutsch, M. S., Chávez, G., Almora, C. E., & Brown, J. L. (2019). Speciation with introgression: phylogeography and systematics of the *Ameerega petersi* group (Dendrobatidae). Molecular Phylogenetics and Evolution, 138, 31-42.

Friedman, M., Feilich, K. L., Beckett, H. T., Alfaro, M. E., Faircloth, B. C., Černý, D., ... & Harrington, R. C. (2019). A phylogenomic framework for pelagiarian fishes (Acanthomorpha: Percomorpha) highlights mosaic radiation in the open ocean. Proceedings of the Royal Society B, 286(1910), 20191502.

Garzione, C. N., Hoke, G. D., Libarkin, J. C., Withers, S., MacFadden, B., Eiler, J., ... & A. Mulch (2008). Rise of the Andes. Science, 320(5881), 1304.

Garzione, C. N., Auerbach, D. J., Smith, J. J. S., Rosario, J. J., Passey, B. H., Jordan, T. E., & Eiler, J. M. (2014). Clumped isotope evidence for diachronous surface cooling of the Altiplano and pulsed surface uplift of the Central Andes. Earth and Planetary Science Letters, 393, 173-181.

Gross, P. (1993). Insect Behavioral and Morphological Defenses Against Parasitoids. Annual Review of Entomology, 38(1), 251-273.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk W. & Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307-321.

Hall, R. (2011). Australia–SE Asia collision: plate tectonics and crustal flow. Geological Society, London, Special Publications, 355(1), 75-109.

Hamilton, C. A., St Laurent, R. A., Dexter, K., Kitching, I. J., Breinholt, J. W., Zwick, A., ... & Kawahara, A. Y. (2019). Phylogenomics resolves major relationships and reveals significant diversification rate shifts in the evolution of silk moths and relatives. BMC Evolutionary Biology, 19(1), 1-13.

Hamilton, C. A., Winiger, N., Rubin, J. J., Breinholt, J., Rougerie, R., Kitching, I. J., ... & Kawahara, A. Y. (2020). Evolution of body size and wing shape trade-offs in arsenurine silkmoths. bioRxiv.

Hardy, N. B., & Otto, S. P. (2014). Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. Proceedings of the Royal Society B: Biological Sciences, 281(1795), 20132960.

Harris, R. S. (2007). Improved pairwise alignment of genomic DNA. PhD Thesis, The Pennsylvania State University.

Harrison, T., & Yumin, G. (1999). Taxonomy and phylogenetic relationships of early Miocene catarrhines from Sihong, China. Journal of Human Evolution, 37(2), 225-277.

Heinicke, M. P., Greenbaum, E., Jackman, T. R., & Bauer, A. M. (2011). Phylogeny of a trans-Wallacean radiation (Squamata, Gekkonidae, *Gehyra*) supports a single early colonization of Australia. Zoologica Scripta, 40(6), 584-602.

Hoelzer, G. A., & Meinick, D. J. (1994). Patterns of speciation and limits to phylogenetic resolution. Trends in Ecology & Evolution, 9(3), 104-107.

Hoorn, C., Wesselingh, F. P., Ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., ... & Jaramillo, C. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science, 330(6006), 927-931.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. Annual Review of Ecology, Evolution, and Systematics, 46, 523-549.

Huang, X. & Madan A. (1999). CAP3: A DNA Sequence Assembly Program. Genome Research, 9(9), 868-877.

Huchon, D., Chevret, P., Jordan, U., Kilpatrick, C. W., Ranwez, V., Jenkins, P. D., ... & Schmitz, J. (2007). Multiple molecular evidences for a living mammalian fossil. Proceedings of the National Academy of Sciences, 104(18), 7495-7499.

Hutchinson, G. E. & MacArthur R. H. (1959). A Theoretical Ecological Model of Size Distributions Among Species of Animals. The American Naturalist, 93(869), 117-125.

Iturralde-Vinent, M., & MacPhee, R. D. (1999). Paleogeography of the Caribbean region: implications for Cenozoic biogeography. Bulletin of the American Museum of Natural History, 238, 1–95.

Ivany, L. C., Patterson, W. P., & Lohmann, K. C. (2000). Cooler winters as a possible cause of mass extinctions at the Eocene/Oligocene boundary. Nature, 407(6806), 887-890.

Jahren, A. H. (2007). The Arctic Forest of the Middle Eocene. Annual Review of Earth and Planetary Sciences, 35(1), 509-540.

Janz, N. & Nylin S. (2008). The oscillation hypothesis of host-plant range and speciation. In The evolutionary biology of herbivorous insects: specialization, speciation and radiation. Tilman D. (eds.), University of California Press, 203-215.
Janzen, D. H. (1984). Two ways to be a tropical big moth: Santa Rosa saturniids and sphingids. Oxford Surveys in Evolutionary Biology, 1, 85-140.
Jervis, M. A., Ferns, P. N., & Boggs, C. L. (2007). A trade-off between female lifespan and larval diet breadth at the interspecific level in Lepidoptera. Evolutionary Ecology, 21(3), 307-323.
Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. Nature, 491(7424), 444-448.
Jiang, D., Klaus, S., Zhang, Y. P., Hillis, D. M., & Li, J. T. (2019). Asymmetric biotic interchange across the Bering land bridge between Eurasia and North America. National Science Review, 6(4), 739-745.
Jousselin, E. & Elias M. (2019). Testing host-plant driven speciation in phytophagous insects: a phylogenetic perspective. Preprints, 2019020215.
Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14(6), 587-589.
Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30(4), 772-780.
Kawahara, A. Y., & Barber, J. R. (2015). Tempo and mode of antibat ultrasound production and sonar jamming in the diverse hawkmoth radiation. Proceedings of the National Academy of Sciences, 112(20), 6407-6412.
Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F., Donath, A., ... & Barber, J. R. (2019). Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. Proceedings of the National Academy of Sciences, 116(45), 22657-22663.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647-1649.
Kitching, I. J., & Sadler, S. (2011). Lepidoptera, Insecta. In Paleontology and Geology of Laetoli: Human Evolution in Context (pp. 549-554). Springer, Dordrecht.
Kitching, I. J., Rougerie, R., Zwick, A., Hamilton, C. A., St Laurent, R. A., Naumann, S., ... & Kawahara, A. Y. (2018). A global checklist of the Bombycoidea (Insecta: Lepidoptera). Biodiversity Data Journal, 6.
Klaus, K. V., & Matzke, N. J. (2020). Statistical comparison of trait-dependent biogeographical models indicates that Podocarpaceae dispersal is influenced by both seed cone traits and geographical distance. Systematic Biology, 69(1), 61-75.
Kristensen, N. P. (2012). Molecular phylogenies, morphological homologies and the evolution of moth 'ears'. Systematic Entomology, 37(2), 237-239.
Lemaire, C. (2002). The Saturniidae of America. Les Saturniidae américains (= Attacidae). Hemileucinae. Goecke & Evers, Keltern, Germany, 1388 pp., 140 pls.
Lamarre, G. P. A., Mendoza, I., Rougerie, R., Decaëns, T., Hérault, B., & Beneluz, F. (2015). Stay out (almost) all night: contrasting responses in flight activity among tropical moth assemblages. Neotropical entomology, 44(2), 109-115.
Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution, 34(3), 772-773.
Le Roy, C., Debat, V., & Llaurens, V. (2019). Adaptive evolution of butterfly wing shape: from morphology to behaviour. Biological Reviews, 94(4), 1261-1281.
Lei, M. & Dong, D. (2016). Phylogenomic analyses of bat subordinal relationships based on transcriptome data. Scientific Reports, 6, 27726.
LePage, B. A., Yang, H., & Matsumoto, M. (2005). In The Geobiology and Ecology of Metasequoia. LePage, B. A., Williams, C. J., Yang, H. (Eds.). Springer, NewYork, pp. 4–81.

Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. Systematic Biology, 56(5), 701-710.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist, 207(2), 437-453.
Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 27(21), 2957-2963.
Mai, U., & Mirarab, S. (2018). TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC genomics, 19(5), 23-40.
Maliet, O., Hartig, F., & Morlon, H. (2019). A model with many small shifts for estimating species-specific diversification rates. Nature ecology & evolution, 3(7), 1086-1092.
Matos-Maraví, P., Clouse, R. M., Sarnat, E. M., Economo, E. P., LaPolla, J. S., Borovanska, M., ... & Janda, M. (2018). An ant genus-group (Prenolepis) illuminates the biogeography and drivers of insect diversification in the Indo-Pacific. Molecular Phylogenetics and Evolution, 123, 16-25.
Maurer, B. A. (1998). The evolution of body size in birds. I. Evidence for non-random diversification. Evolutionary Ecology, 12(8), 925.
Michener, C. D. (1952). The Saturniidae (Lepidoptera) of the Western Hemisphere: Morphology, phylogeny and classification. Bulletin of the American Museum of Natural History, 98, 338-501.
Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution, 37(5), 1530-1534.
Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics, 30(17), i541-i548.
Mirarab, S., & Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics, 31(12), i44-i52.
Misof, B. (2002). Diversity of Anisoptera (Odonata): Infering speciation processes from patterns of morphological diversity. Zoology, 105(4), 355-365.
Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., ... & Niehuis, O. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science, 346(6210), 763-767.
Modica, M. V., Gorson, J., Fedosov, A. E., Malcolm, G., Terryn, Y., Puillandre, N., & Holford, M. (2020). Macroevolutionary Analyses Suggest That Environmental Factors, Not Venom Apparatus, Play Key Role in Terebridae Marine Snail Diversification. Systematic Biology, 69(3), 413-430.
Moen, D., & Morlon, H. (2014). Why does diversification slow down? Trends in Ecology & Evolution, 29(4), 190-197.
Montes, C., Cardona, A., McFadden, R., Morón, S. E., Silva, C. A., Restrepo-Moreno, S., ... & Bayona, G. A. (2012). Evidence for middle Eocene and younger land emergence in central Panama: Implications for Isthmus closure. Bulletin, 124(5-6), 780-799.
Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J. C., Valencia, V., ... & Niño, H. (2015). Middle Miocene closure of the Central American seaway. Science, 348(6231), 226-229.
Morley, R. J. (2003). Interplate dispersal paths for megathermal angiosperms. Perspectives in Plant Ecology, Evolution and Systematics, 6(1-2), 5-20.
Morlon, H., Potts, M. D., & Plotkin, J. B. (2010). Inferring the dynamics of diversification: a coalescent approach. PLoS Biol, 8(9), e1000493.
Morlon, H. (2014). Phylogenetic approaches for studying diversification. Ecology letters, 17(4), 508-525.
Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403(6772), 853-858.
Nässig, W. A., Naumann, S., & Oberprieler, R. G. (2015). Notes on the Saturniidae of the Arabian Peninsula, with description of a new species (Lepidoptera: Saturniidae). Nachrichten des Entomologischen Vereins Apollo (NF), 36(1), 31-38.
Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32(1), 268-274.
Nilsen, T. H. (1978). Lower Tertiary laterite on the Iceland–Faeroe ridge and the Thulean land bridge. Nature, 274(5673), 786-788.

Nirmala, X., Mita, K., Vanisree, V., Žurovec, M., & Sehnal, F. (2001). Identification of four small molecular mass proteins in the silk of *Bombyx mori*. Insect Molecular Biology, 10(5), 437-445.
Oberprieler, R. G., & Nassig, W. A. (1994). Tarn-oder Warntrachten-ein Vergleich larvaler und imaginaler Strategien bei Saturniinen (Lepidoptera: Saturniidae). Nachr. Entomol. Ver. Apollo, 15, 267-303.
Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526-528.
Patriat, P., & Achache, J. (1984). India–Eurasia collision chronology has implications for crustal shortening and driving mechanism of plates. Nature, 311(5987), 615-621.
Peigler, R. S. & Oberprieler R. G. (2017). Ethnographic description of cocoons and silk of the moth families Saturniidae, Lasiocampidae and Psychidae. Nachrichten des Entomologischen Vereins Apollo, 38(2-3), 113-120.
Philippe, H., de Vienne, D. M., Ranwez, V., Roure, B., Baurain, D., & Delsuc, F. (2017). Pitfalls in supermatrix phylogenomics. European Journal of Taxonomy, (283).
Pindell, J. L., & Kennan, L. (2009). Tectonic evolution of the Gulf of Mexico, Caribbean and northern South America in the mantle reference frame: an update. Geological Society, London, Special Publications, 328(1), 1-55.
Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. R news, 6(1), 7-11.
Prebus, M. (2017). Insights into the evolution, biogeography and natural history of the acorn ants, genus *Temnothorax* Mayr (hymenoptera: Formicidae). BMC Evolutionary Biology, 17(1), 250.
Price, S. A., Hopkins, S. S., Smith, K. K., & Roth, V. L. (2012). Tempo of trophic evolution and its impact on mammalian diversification. Proceedings of the National Academy of Sciences, 109(18), 7008-7012.
Prinzing, A. (2003). Are generalists pressed for time? An interspecific test of the time-limited disperser model. Ecology, 84(7), 1744-1755.
R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature communications, 4(1), 1-8.
Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., ... & Larson, J. G. (2014). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods in Ecology and Evolution, 5(7), 701-707.
Rainford, J. L., Hofreiter, M., & Mayhew, P. J. (2016). Phylogenetic analyses suggest that diversification and body size evolution are independent in insects. BMC Evolutionary Biology, 16(1), 8.
Regier, J. C., Grant, M. C., Mitter, C., Cook, C. P., Peigler, R. S., & Rougerie, R. (2008). Phylogenetic relationships of wild silkmoths (Lepidoptera: Saturniidae) inferred from four protein-coding nuclear genes. Systematic Entomology, 33(2), 219-228.
Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution, 3(2), 217-223.
Romiguier, J. & Roux C. (2017). Analytical Biases Associated with GC-Content in Molecular Evolution. Frontiers in Genetics, 8, 16.
Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology, 61(3), 539-542.
Rougerie, R., Naumann, S., & Nässig, W. A. (2012). Morphology and molecules reveal unexpected cryptic diversity in the enigmatic genus Sinobirma Bryk, 1944 (Lepidoptera: Saturniidae). PLoS One, 7(9), e43920.
Rubin, J. J., Hamilton, C. A., McClure, C. J., Chadwell, B. A., Kawahara, A. Y., & Barber, J. R. (2018). The evolution of anti-bat sensory illusions in moths. Science advances, 4(7), eaar7428.
Rubinoff, D., & Doorenweerd, C. (2020). In and out of America: Ecological and species diversity in Holarctic giant silkmoths suggests unusual dispersal, defying the dogma of an Asian origin. Journal of Biogeography, 47(4), 903-914.

Sayyari, E., & Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution, 33(7), 1654-1668.
Ségalen, L., Lee-Thorp, J. A., & Cerling, T. (2007). Timing of C4 grass expansion across sub-Saharan Africa. Journal of Human Evolution, 53(5), 549-559.
Seiffert, E. R. (2007). Evolution and extinction of Afro-Arabian primates near the Eocene-Oligocene boundary. Folia Primatologica, 78(5-6), 314-327.
Sepulchre, P., Arsouze, T., Donnadieu, Y., Dutay, J. C., Jaramillo, C., Le Bras, J., ... & Waite, A. J. (2014). Consequences of shoaling of the Central American Seaway determined from modeling Nd isotopes. Paleoceanography, 29(3), 176-189.
Shi, J. J. & Rabosky D. L. (2015). Speciation dynamics during the global radiation of extant bats. Evolution, 69(6), 1528-1545.
Silvestro, D., Cascales-Miñana, B., Bacon, C. D., & Antonelli, A. (2015). Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. New Phytologist, 207(2), 425-436.
Singer, M. S. (2008). Evolutionary ecology of polyphagy. In Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects. Tilmon K. J. (Eds.). University of California Press, Berkeley, California, USA, 29-42.
Soubrier, J., Steel, M., Lee, M. S., Der Sarkissian, C., Guindon, S., Ho, S. Y., & Cooper, A. (2012). The influence of rate heterogeneity among sites on the time dependence of molecular rates. Molecular Biology and Evolution, 29(11), 3345-3358.
Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313.
Stephens, P. A., Boyd, I. L., McNamara, J. M., & Houston, A. I. (2009). Capital breeding and income breeding: their meaning, measurement, and worth. Ecology, 90(8), 2057-2067.
Struck, T. H. (2014). TreSpEx–-Detection of misleading signal in phylogenetic reconstructions based on tree information. Evolutionary Bioinformatics, 10, EBO-S14239.
Sukumaran, J., Economo, E. P., & Lacey Knowles, L. (2016). Machine learning biogeographic processes from biotic patterns: a new trait-dependent dispersal and diversification model with model choice by simulation-trained discriminant analysis. Systematic Biology, 65(3), 525-545.
Sukumaran, J., & Knowles, L. L. (2018). Trait-dependent biogeography:(re) integrating biology into probabilistic historical biogeographical models. Trends in ecology & evolution, 33(6), 390-398.
Sutherland, T. D., Young, J. H., Weisman, S., Hayashi, C. Y., & Merritt, D. J. (2010). Insect silk: one name, many materials. Annual review of entomology, 55.
Tagliacollo, V. A. & Lanfear R. (2018). Estimating Improved Partitioning Schemes for Ultraconserved Elements. Molecular Biology and Evolution, 35(7), 1798-1811.
Talavera, G. & Castresana J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Systematic Biology, 56(4), 564-577.
Tammaru, T. & Haukioja E. (1996). Capital breeders and income breeders among Lepidoptera: consequences to population dynamics. Oikos, 77(3), 561-564.
Thomas, G. H., Hartmann, K., Jetz, W., Joy, J. B., Mimoto, A., & Mooers, A. O. (2013). PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences. Methods in Ecology and Evolution, 4(11), 1011-1017.
Thorne, J. L., Kishino, H., & Painter, I. S. (1998). Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution, 15(12), 1647-1657.
Tiffney, B. H. (1985). The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. Journal of the Arnold Arboretum, 66(2), 243-273.
Vandenberghe, N., Hilgen, F. J., & Speijer, R. P. (2012). The Paleogene period. In The geologic time scale. Gradstein FM, Ogg JG, Schmitz M, Ogg G, (eds.). Amsterdam, Elsevier, 855-921.
Wahlberg, N., Wheat, C. W., & Peña, C. (2013). Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLOS one, 8(11), e80875.
Wang, H., Holloway, J. D., Janz, N., Braga, M. P., Wahlberg, N., Wang, M., & Nylin, S. (2017). Polyphagy and diversification in tussock moths: Support for the oscillation hypothesis from extreme generalists. Ecology and evolution, 7(19), 7975-7986.

Willis, C. G., Hall, J. C., Rubio de Casas, R., Wang, T. Y., & Donohue, K. (2014). Diversification and the evolution of dispersal ability in the tribe Brassiceae (Brassicaceae). Annals of Botany, 114(8), 1675-1686.
Wolfe, J. A. (1975). Some Aspects of Plant Geography of the Northern Hemisphere During the Late Cretaceous and Tertiary. Annals of the Missouri Botanical Garden, 62(2), 264-279.
Yang, Z. (1995). A space-time process model for the evolution of DNA sequences. Genetics, 139, 993-1005.
Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences, 13(5), 555-556.
Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. science, 292(5517), 686-693.
Zachos, J. C., Dickens, G. R., & Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature, 451(7176), 279-283.
Zhang J., Sun B. & Zhang X. (1994). Miocene insects and spiders from Shanwang, Shandong. Beijing, China: Science Press, 298 pp., 44 pls.
Zhang, Z., Ramstein, G., Schuster, M., Li, C., Contoux, C., & Yan, Q. (2014). Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene. Nature, 513(7518), 401-404.
Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC bioinformatics, 19(6), 153.

Supplementary Material

Historical biogeography analyses

Ancestral area reconstructions were performed using DECX (Beeravolu & Condamine 2016) with both adjacency and dispersal rate matrices as detailed here.



Areas considered
WP: Western Palearctic
EP : Eastern Palearctic
AF: Africa
WA: Malayan region
IN: India
AU: Australia + Pacific islands
MA: Madagascar
SA: South America
CA: Central America
WN: Western Nearctic
EN: Eastern Nearctic

Figure S8 – Biogeographical areas considered.

Symbols used
» : Adjacent areas, no barrier, p=1.
X : small water barrier, p=0.5.
X : large water barrier, p=0.1.
X : one terrestrial area, p=0.75.
X : two terrestrial areas, p=0. 5.

Present – 5Ma: Pliocene (+Pléistocène)

Dispersal rate matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	1	0.75	0.75	0.1	0.1	0.1	0.1	0.1	0.1	0.25
CA	»	-	1	1	0.1	0.1	0.1	0.25	0.1	0.25	0.375
WN	X	»	-	1	0.25	0.1	0.1	0.375	0.375	0.375	0.5
EN	X	»	»	-	0.1	0.1	0.1	0.25	0.25	0.25	0.375
AF	X	X	XX	X	-	0.1	0.5	0.1	0.75	1	0.75
AU	X	X	X	X	X	-	0.1	0.5	0.375	0.25	0.375
MD	X	X	X	X	X	X	-	0.1	0.25	0.375	0.25
WA	X	XX	XX	XX	X	X	X	-	1	0.75	1
IN	X	X	XX	XX	X	XX	XX	»	-	1	1
WP	X	XX	XX	XX	»	XX	XX	X	»	-	1
EP	XX	XX	X	XX	X	XX	XX	»	»	»	-

Adjacency matrix

	S	C	WN	EN	AF	AU	M	O	I	WP	EP
S	-	1	0	0	0	0	0	0	0	0	0
C	»	-	1	1	0	0	0	0	0	0	0
WN	X	»	-	1	0	0	0	0	0	0	1
EN	X	»	»	-	0	0	0	0	0	0	0
AF	X	X	XX	X	-	0	1	0	0	1	0
AU	X	X	X	X	X	-	0	1	0	0	0
M	X	X	X	X	X	X	-	0	0	0	0
O	X	XX	XX	XX	X	X	X	-	1	0	1
I	X	X	XX	XX	X	XX	XX	»	-	1	1
WP	X	XX	XX	XX	»	XX	XX	X	»	-	1
EP	XX	XX	X	XX	X	XX	XX	»	»	»	-

5Ma – 14Ma: late Miocene

Panama: Land bridge considered. Phylogenetic and fossils evidences: Bacon (2015) detected a major shift in the migration rates around 6Ma but showed that lineages started to disperse way before and detected a first migration rate shift around 20 Ma. Geologic evidences: Montes et al. (2015) argued the Central American Seaway was closed in the Middle Miocene using detrital zircon geochronology.

Bering strait: closed (Gladenkov et al. 2002, Gladenkov & Gladenkov 2004, Verhoeven et al. 2011) with temperate conditions.

North Atlantic Land Bridge (NALB): Considering land bridges EN↔WP. But according to Denk et al. (2011) who summarized geological data from different studies, the subsidence of different parts of the Greenland-Scotland Transverse Ridge (GSTR) do not overlapn, though organisms are known to have colonized Iceland from Groenland (7-5.5Ma for the lastest) and from Scotland (9-8Ma). We considered small water barrier for 5-34Ma, then terrestrial connection for 34-66Ma (see following tables).

Africa: Rosenbaum et al. (2002) considered a land bridge through Sicily from early Oligocene. In that easy dispersion matrix, we considered that land bridge. Land connection through Arabic Peninsula considered from Early Miocene (Early Langhian according to Meulenkamp & Sissingh 2003).

Dispersal rate matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	1	0.75	0.75	0.1	0.1	0.1	0.1	0.1	0.1	0.5
CA	»	-	1	1	0.1	0.1	0.1	0.5	0.5	0.5	0.75
WN	X	»	-	1	0.5	0.25	0.1	0.75	0.75	0.75	1
EN	X	»	»	-	0.25	0.1	0.1	0.5	0.5	0.5	0.75
AF	X	X	X	XX	-	0.1	0.5	0.1	0.75	1	0.75
AU	X	X	XX	X	X	-	0.1	0.5	0.375	0.25	0.375
MD	X	X	X	X	X	X	-	0.1	0.25	0.375	0.25
WA	X	X	X	X	X	X	X	-	1	0.75	1
IN	X	X	X	X	X	XX	XX	»	-	1	1
WP	X	X	X	X/X	»	XX	XX	X	»	-	1
EP	X	X	»	X	X	XX	XX	»	»	»	-

Adjacency matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	1	0	0	0	0	0	0	0	0	0
CA	»	-	1	1	0	0	0	0	0	0	0
WN	X	»	-	1	0	0	0	0	0	0	1
EN	X	»	»	-	0	0	0	0	0	1.	0
AF	X	X	X	XX	-	0	1	0	0	1	0
AU	X	X	XX	X	X	-	0	1	0	0	0
MD	X	X	X	X	X	X	-	0	0	0	0
WA	X	X	X	X	X	X	X	-	1	0	1
IN	X	X	X	X	X	XX	XX	»	-	1	1
WP	X	X	X	X/X	»	XX	XX	X	»	-	1
EP	X	X	»	X	X	XX	XX	»	»	»	-

14Ma – 23Ma: early Miocene

Panama: Panama isthmus did not exist yet. However, only about 200km segregate South and Central America (Montes 2012,2015).

India ↔Madagascar: possible dispersal events through archipelagos. Here we considered a small water barrier.

Dispersal rate matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	0.5	0.375	0.375	0.1	0.1	0.1	0.1	0.1	0.1	0.25
CA	X	-	1	1	0.1	0.1	0.1	0.5	0.5	0.5	0.75
WN	XX	»	-	1	0.5	0.25	0.1	0.75	0.75	0.75	1
EN	XX	»	»	-	0.25	0.1	0.1	0.5	0.5	0.5	0.75
AF	X	X	X	XX	-	0.1	0.5	0.1	0.75	1	0.75
AU	X	X	XX	X	X	-	0.1	0.5	0.375	0.25	0.375
MD	X	X	X	X	X	X	-	0.1	0.5	0.375	0.25
WA	X	X	X	X	X	X	X	-	1	0.75	1
IN	X	X	X	X	X	XX	X	»	-	1	1
WP	X	X	X	X/XX	»	XX	XX	X	»	-	1
EP	XX	X	»	X	X	XX	XX	»	»	»	-

Adjacency matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	1	0	0	0	0	0	0	0	0	0
CA	X	-	1	1	0	0	0	0	0	0	0
WN	XX	»	-	1	0	0	0	0	0	0	1
EN	XX	»	»	-	0	0	0	0	0	1	0
AF	X	X	X	XX	-	0	1	0	0	1	0
AU	X	X	XX	X	X	-	0	1	0	0	0
MD	X	X	X	X	X	X	-	0	1	0	0
WA	X	X	X	X	X	X	X	-	1	0	1
IN	X	X	X	X	X	XX	X	»	-	1	1
WP	X	X	X	X/X	»	XX	XX	X	»	-	1
EP	-	1	0	0	0	0	0	0	0	0	0

23Ma – 34Ma: Oligocene

See above comments about the Africa/Palearctic connections.

Dispersal rate matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	0.5	0.375	0.375	0.1	0.1	0.1	0.1	0.1	0.1	0.1
CA	X	-	1	1	0.1	0.1	0.1	0.5	0.5	0.5	0.75
WN	XX	»	-	1	0.25	0.1	0.1	0.75	0.75	0.75	1
EN	XX	»	»	-	0.1	0.1	0.1	0.5	0.5	0.5	0.75
AF	X	X	XX	X	-	0.1	0.5	0.1	0.75	1	0.75
AU	X	X	X	X	X	-	0.1	0.5	0.375	0.25	0.375
MD	X	X	X	X	X	X	-	0.1	0.5	0.1	0.1
WA	X	X	X	X	X	X	X	-	1	0.75	1
IN	X	X	X	X	XX	X	X	»	-	1	1
WP	X	X	X	X/X	»	X	X	X	»	-	1
EP	X	X	»	X	XX	X	X	»	»	»	-

Adjacency matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	1	0	0	0	0	0	0	0	0	0
CA	X	-	1	1	0	0	0	0	0	0	0
WN	XX	»	-	1	0	0	0	0	0	0	1
EN	XX	»	»	-	0	0	0	0	0	1	0
AF	X	X	XX	X	-	0	1	0	0	1	0
AU	X	X	X	X	X	-	0	1	0	0	0
MD	X	X	X	X	X	X	-	0	1	0	0
WA	X	X	X	X	X	X	X	-	1	0	1
IN	X	X	X	X	XX	X	X	»	-	1	1
WP	X	X	X	X/X	»	X	X	X	»	-	1
EP	X	X	»	X	XX	X	X	»	»	»	-

34Ma – 56Ma: Eocene

Africa: no land bridge WP↔Africa

Australia: see supplementary materials of de Bruyn et al. (2014) for detailed maps. Australian landmass was still far from Asia, only getting close enough to consider a small water barrier at 35Ma. Heinicke et al. (2011) estimated that *Gehyra* Gekko colonized Australia in Early Oligocene/Late Eocene (30-38Ma).

Dispersal rate matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	0.5	0.375	0.375	0.1	0.25	0.1	0.1	0.1	0.1	0.1
CA	X	-	1	1	0.1	0.1	0.1	0.5	0.5	0.5	0.75
WN	XX	»	-	1	0.25	0.1	0.1	0.75	0.75	0.75	1
EN	XX	»	»	-	0.1	0.1	0.1	0.5	0.75	1	0.75
AF	X	X	XX	X	-	0.1	0.5	0.1	0.375	0.5	0.375
AU	XX	X	X	X	X	-	0.1	0.1	0.1	0.1	0.1
MD	X	X	X	X	X	X	-	0.1	0.5	0.1	0.1
WA	X	XX	X	X	X	X	X	-	1	0.75	1
IN	X	X	X	X	XX	X	X	»	-	1	1
WP	X	X	X	»	X	X	X	X	»	-	1
EP	-	0.5	0.375	0.375	0.1	0.25	0.1	0.1	0.1	0.1	0.1

Adjacency matrix

	SA	CA	WN	EN	AF	AU	MD	WA	IN	WP	EP
SA	-	1	0	0	0	0	0	0	0	0	0
CA	X	-	1	1	0	0	0	0	0	0	0
WN	XX	»	-	1	0	0	0	0	0	0	1
EN	XX	»	»	-	0	0	0	0	0	1	0
AF	X	X	XX	X	-	0	1	0	0	1	0
AU	X	X	X	X	X	-	0	0	0	0	0
MD	X	X	X	X	X	X	-	0	1	0	0
WA	X	XX	X	X	X	X	X	-	1	0	1
IN	X	X	X	X	XX	X	X	»	-	1	1
WP	X	X	X	»	X	X	X	X	»	-	1
EP	X	X	»	X	XX	X	X	»	»	»	-

56Ma – 66Ma: Paleocene

Panama: We do not consider island arc anymore. SA was very geographically isolated.

Dispersal rate matrix

	S	C	WN	EN	AF	AU	M	O	I	WP	EP
S	-	0.1	0.1	0.1	0.1	0.25	0.1	0.1	0.1	0.1	0
C	X	-	1	1	0.25	0.1	0.1	0.5	0.5	0.75	0.75
WN	X	»	-	1	0.25	0.1	0.1	0.75	0.75	0.75	1
EN	X	»	»	-	0.375	0.1	0.1	0.5	0.75	1	0.75
AF	X	XX	XX	XX	-	0.1	0.5	0.1	0.375	0.5	0.375
AU	XX	X	X	X	X	-	0.1	0.1	0.1	0.1	0.1
M	X	X	X	X	X	X	-	0.1	0.5	0.1	0.1
O	X	X	X	X	X	X	X	-	0.5	0.75	1
I	X	X	X	X	XX	X	X	X	-	1	1
WP	X	X	X	»	X	X	X	X	»	-	1
EP	X	X	»	X	XX	X	X	»	»	»	-

Adjacency matrix

	S	C	WN	EN	AF	AU	M	O	I	WP	EP
S	-	0	0	0	0	0	0	0	0	0	0
C	X	-	1	1	0	0	0	0	0	0	0
WN	X	»	-	1	0	0	0	0	0	0	1
EN	X	»	»	-	0	0	0	0	0	1	0
AF	X	XX	XX	XX	-	0	1	0	0	0	0
AU	X	X	X	X	X	-	0	0	0	0	0
M	X	X	X	X	X	X	-	0	1	0	0
O	X	X	X	X	X	X	X	-	1	0	1
I	X	X	X	X	XX	X	X	»	-	1	1
WP	-	0	0	0	0	0	0	0	0	0	0
EP	X	-	1	1	0	0	0	0	0	0	0

Supplementary figures

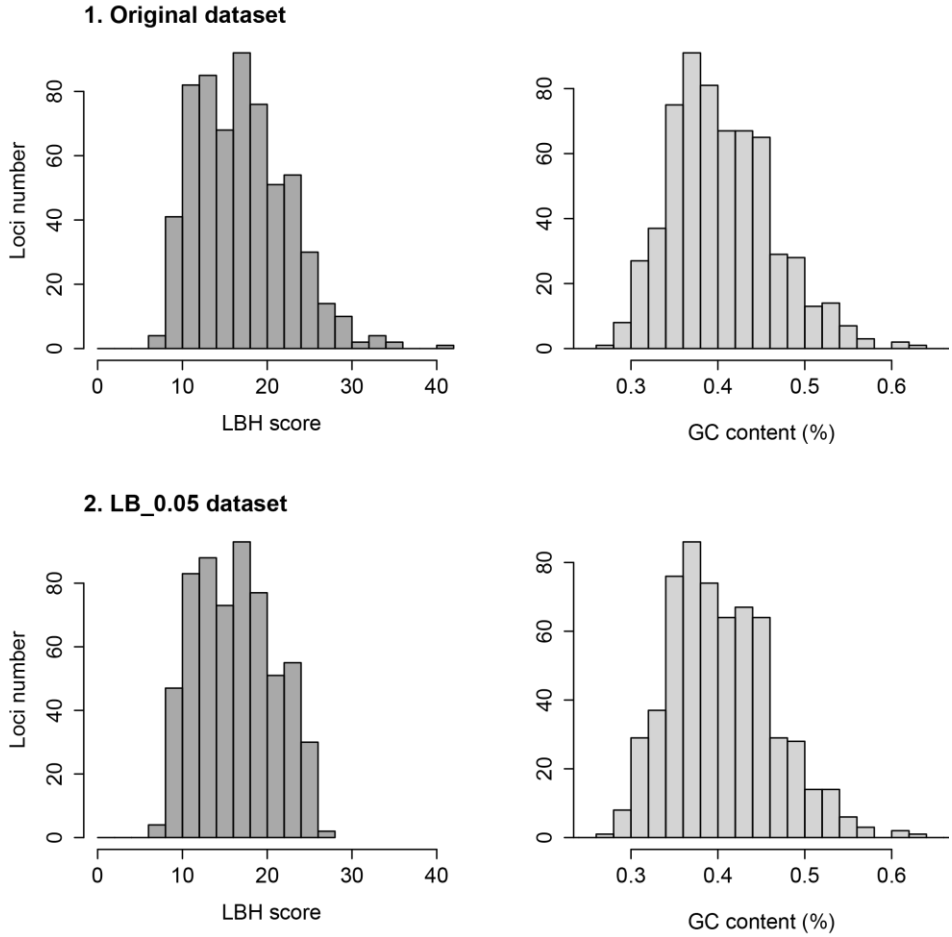


Figure S1 – Distribution of Long Branch Heterogeneity scores (left column) and GC contents (right column) among UCE *loci* of the original (top row) dataset and of the LB_0.05 (bottom row) dataset in which 5% of *loci* with the highest LB scores were discarded. Discarding these *loci* from the original genomic dataset corrected the negative skewness of the LBH score distribution.

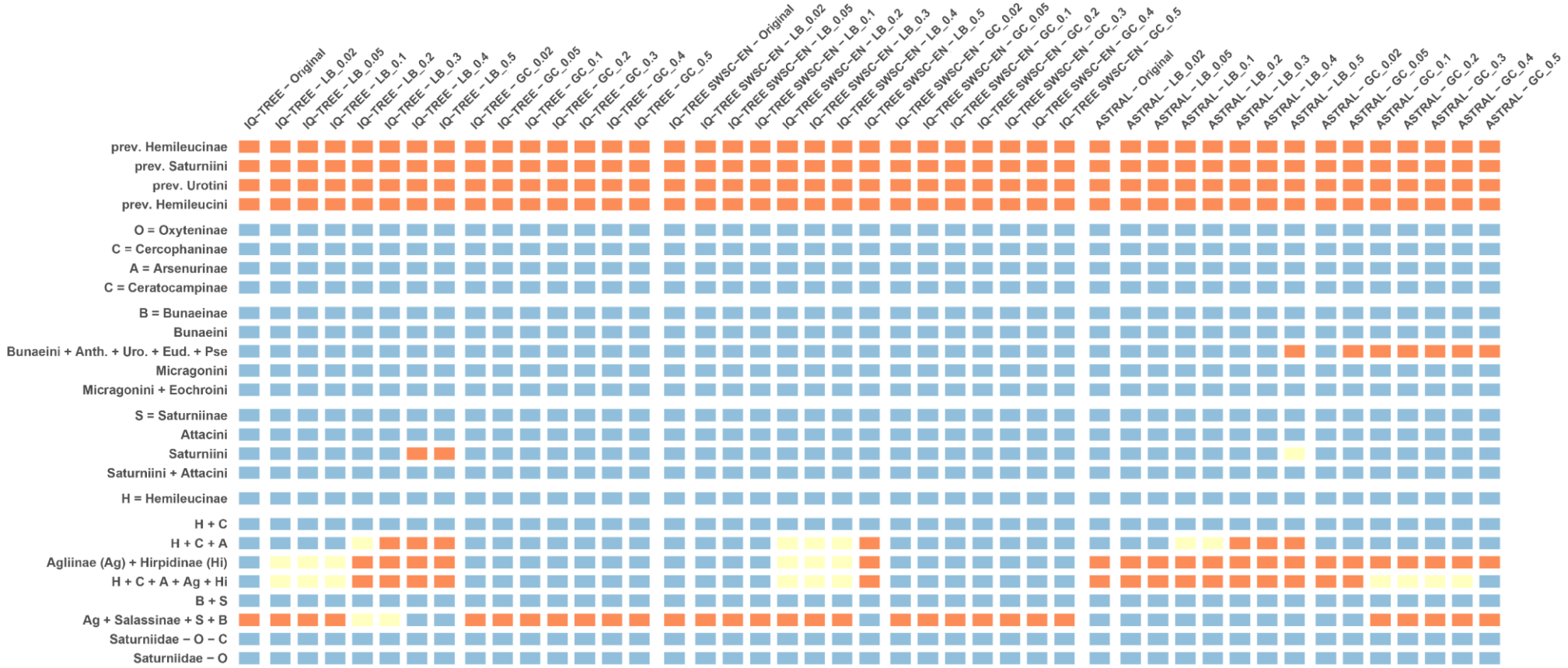


Figure S2 - Phylogenetic supports obtained with the different analyses (in column) for the main Saturniidae clades (in row). IQ-TREE or ASTRAL refer to the phylogenetic method used and SWSC-EN indicate when, for the IQ-TREE analyses, the Tagliacollo & Lanfear (2018) partitioning method was used. The results showed here were inferred from the original genomic dataset and the datasets in which 5% of *loci* were discarded based of their LB score (LB_X) or their GC composition (GC_X). Blue squares indicate that the clade monophyly was strongly supported, yellow weakly supported and orange non recovered.

Figure S3 (Online resource) - All inferred trees with IQ-TREE or ASTRAL from the different sub-datasets. The title of each tree indicates the software used and the sub-dataset considered. Subfamilies, species names, sample codes and number of loci are indicated at tips.

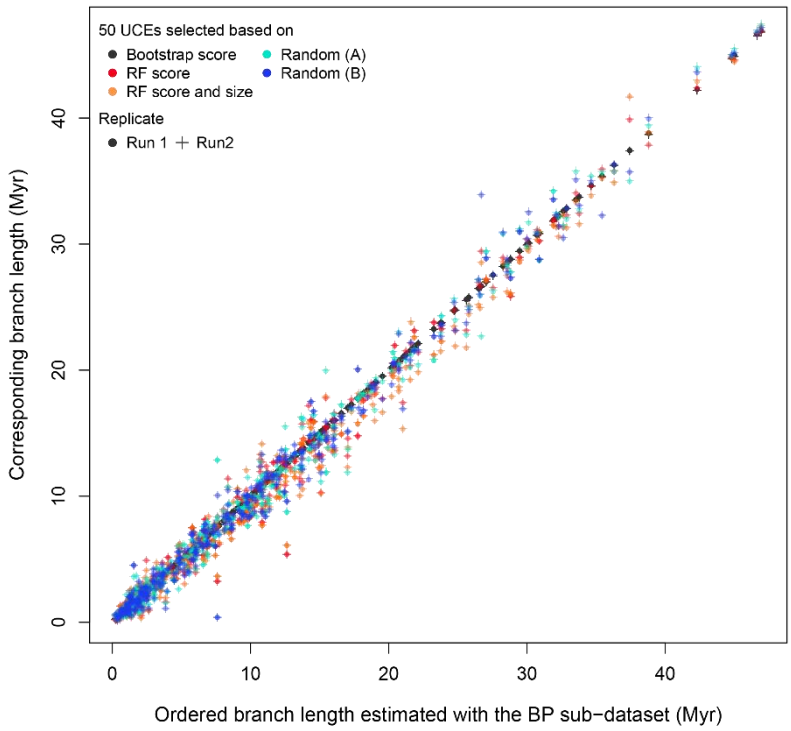
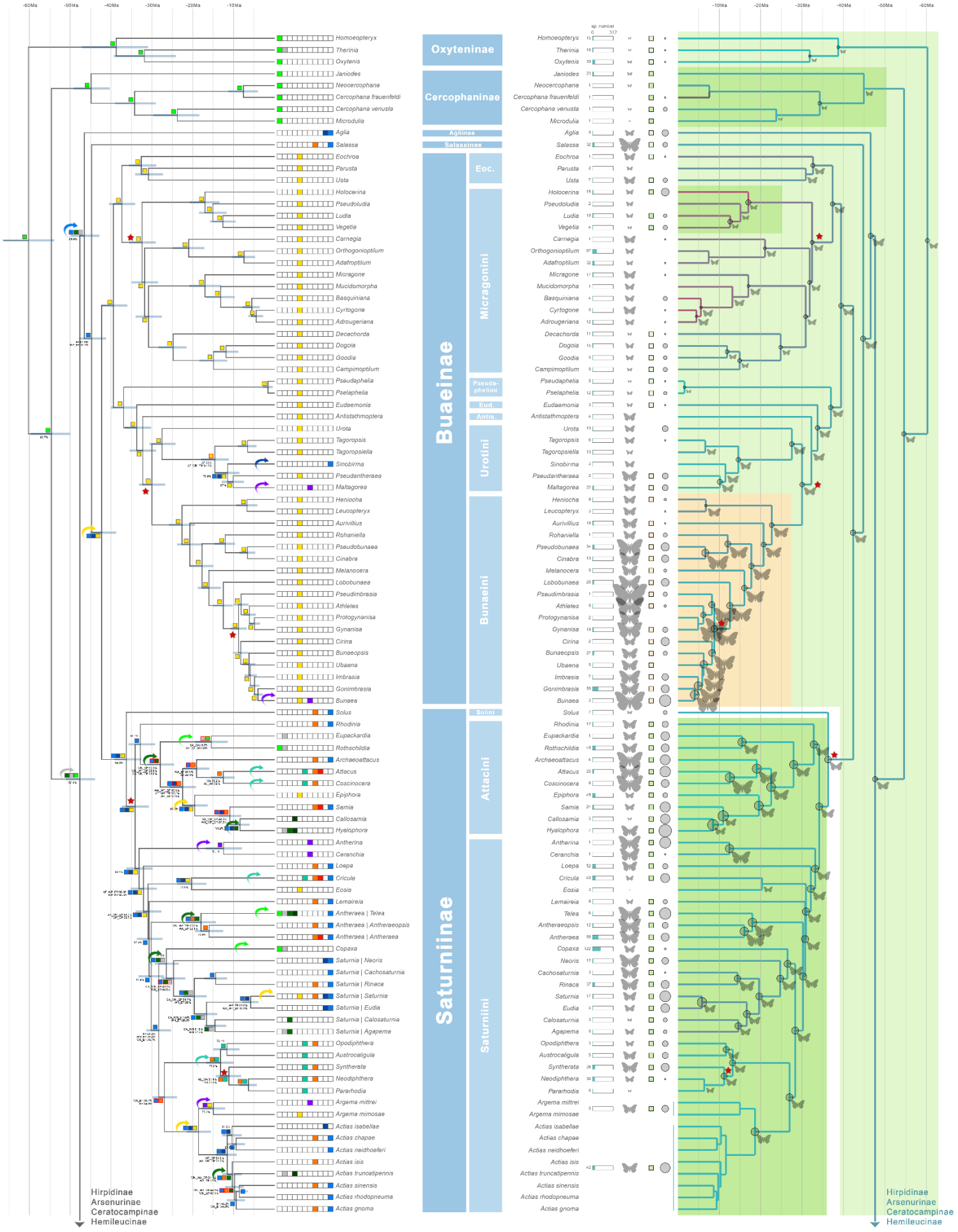
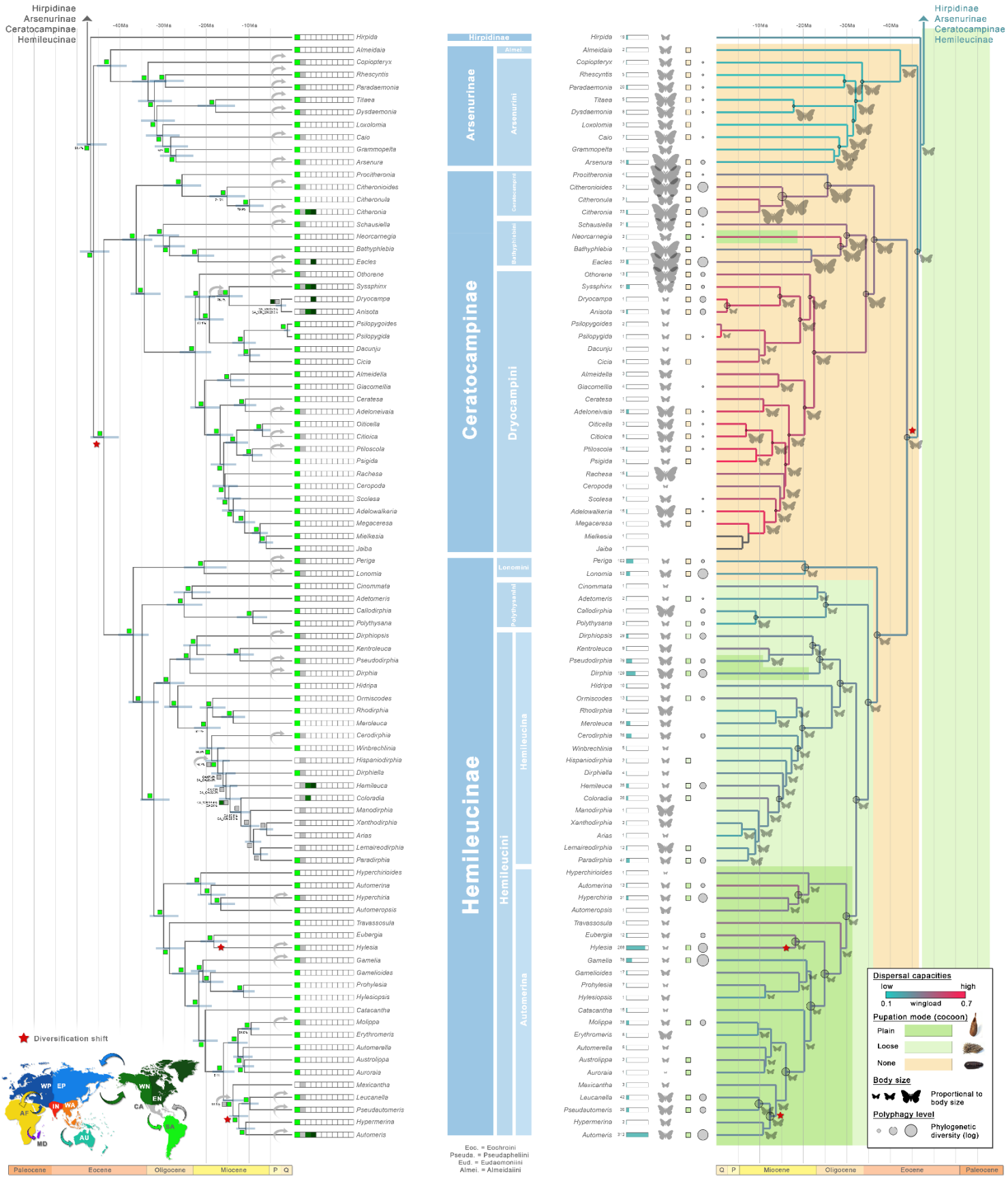


Figure S4 – Convergence of datation analyses. The divergence time estimates obtained with the different sub-dataset (see legend) are represented as a function of the estimates obtained with the BP sub-dataset (Bootstrap scores), for which 50 loci were sampled, because the generated gene trees were the most supported.

Figure S5 (next two pages, see Online resources) – Historical biogeography and traits evolution of Saturniidae. Red stars depict the diversification rate shifts as estimated with BAMM. Left panel: biogeographical history of Saturniidae as estimated with DECX from 1000 species-level trees. Ranges with the highest probabilities, as estimated with DECX, are represented by colored squares; red contour lines of squares indicate best ranges inferred with less than 50% confidence. When range estimate confidence values were below 90%, alternative estimates were depicted. Colored arrows highlight the main dispersal events as represented in the World map (bottom left corner). Blue bars indicate confidence intervals of node ages. Right panel: saturniid life-history trait evolution. Branches are colored according to wingload values. Background colors represent the pupation modes. The size of shaded silhouettes is proportional to body size values (log-normal). Polyphagy level is represented at tips and nodes with circles whose size is proportional to the estimated polyphagy score (log-normal). A .pdf file of this figure is also available as Online resource.


-60Ma
-50Ma
-40Ma
-30Ma
-20Ma
-10Ma
Oxyteninae
Cercophaninae
Agliinae
Salassinae
Buaeinae
Eoc.
Micragonini
Pseuda-pheliini
Eud.
Antis.
Urotini
Bunaeini
Solini
Saturniinae
Attacini
Saturniini
Homoeopteryx
Therinia
Oxytenis
Janiodes
Neocercophana
Cercophana frauenfeldi
Cercophana venusta
Microdulia
Aglia
Salassa
Eochroa
Parusta
Usta
Holocerina
Pseudoludia
Ludia
Vegetia
Carnegia
Orthogonioptilum
Adafroptilum
Micragone
Mucidomorpha
Basquiniana
Cyrtogone
Adrougeriana
Decachorda
Dogoia
Goodia
Campimoptilum
Pseudaphelia
Pselaphelia
Eudaemonia
Antistathmoptera
Urota
Tagoropsis
Tagoropsiella
Sinobirma
Pseudantheraea
Maltagorea
Heniocha
Leucopteryx
Aurivillius
Rohaniella
Pseudobunaea
Cinabra
Melanocera
Lobobunaea
Pseudimbrasia
Athletes
Protogynanisa
Gynanisa
Cirina
Bunaeopsis
Ubaena
Imbrasia
Gonimbrasia
Bunaea
Solus
Rhodinia
Eupackardia
Rothschildia
Archaeoattacus
Attacus
Coscinocera
Epiphora
Samia
Callosamia
Hyalophora
Antherina
Ceranchia
Loepa
Cricula
Eosia
Lemaireia
Antheraea | Telea
Antheraea | Antheraeopsis
Antheraea | Antheraea
Copaxa
Saturnia | Neoris
Saturnia | Cachosaturnia
Saturnia | Rinaca
Saturnia | Saturnia
Saturnia | Eudia
Saturnia | Calosaturnia
Saturnia | Agapema
Opodiphthera
Austrocaligula
Syntherata
Neodiphthera
Pararhodia
Argema mittrei
Argema mimosae
Actias isabellae
Actias chapae
Actias neidhoeferi
Actias isis
Actias truncatipennis
Actias sinensis
Actias rhodopneuma
Actias gnoma
sp. number
0
512
Hirpidinae
Arsenurinae
Ceratocampinae
Hemileucinae


Hirpidinae
Arsenurinae
Ceratocampinae
Hemileucinae
Diversification shift
Paleocene
Eocene
Oligocene
Miocene
Eoc. = Eochroini
Pseuda. = Pseudapheliini
Eud. = Eudaemoniini
Almei. = Almeidaiini
Dispersal capacities
low
high
0.1
wingload
0.7
Pupation mode (cocoon)
Plain
Loose
None
Body size
Proportional to body size
Polyphagy level
Phylogenetic diversity (log)

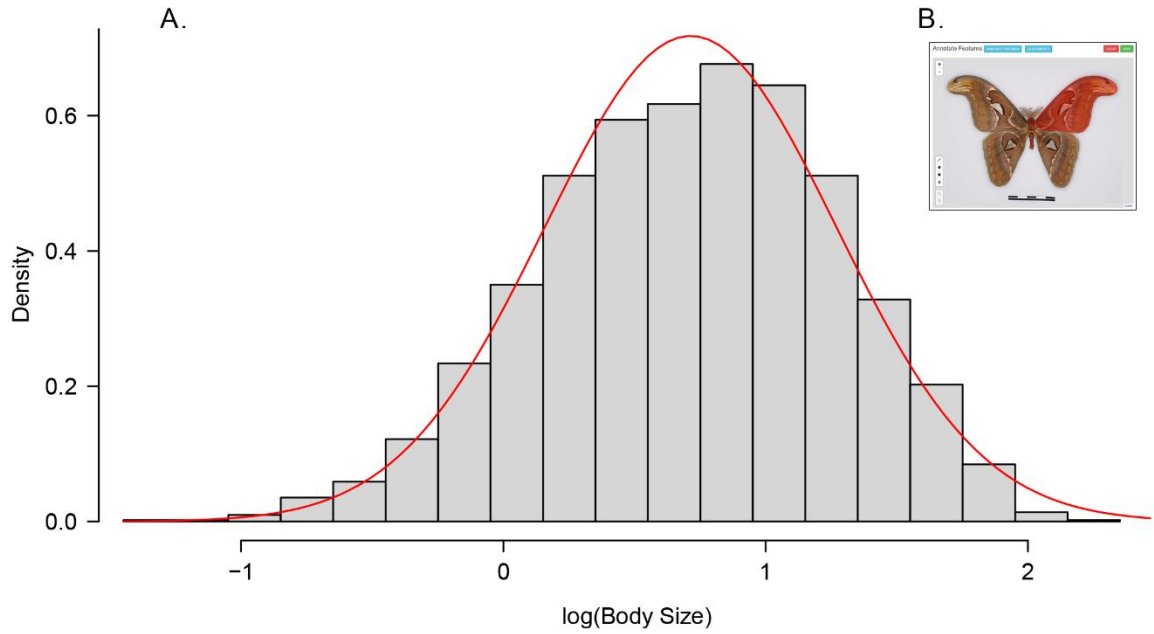


Figure S6 - (A) Histogram representing the log-transformed body size distribution of 2577 specimens, as estimated by the product of body length and the mean of the three measurements of thorax width. (B) Represents the tool used for measuring these features, as integrated in BOLD (www.boldsystems.org). The distribution is negatively skewed (one sided Agostino test, skew = -0.226, z = -4.594, p-value = 2.17e-06). Red line represents a normal distribution with similar mean and standard deviation.

Figure S7 (Online resource) - Polyphagy-dependent diversification models - subfamily level. Within each major subfamily (one page each) of Saturniidae, we categorized the genera into three categories, specialist (category 1), oligophagous (2) and polyphagous (3), considering quantiles p=0.33 and p=0.66. Genus polyphagy is indicated at the crown node of each genus. On the left of each page, we represented the species-level phylogenies of each subfamily as inferred with PASTIS; the genera missing food plant information were discarded. We did not consider intra-generic variation because of the way we estimated polyphagy level. On the right, histograms represented speciation, extinction, diversification and transition rates as estimated over 10,000 MCMC generations with the diversitree R package.

Figure S9 (following page) - Influence of priors on the posterior results in BAMM. We ran BAMM with several expectedNumberOfShifts values in order to understand how priors can affect the number and the position of recovered shifts. Here we represented the prior and posterior distributions and showed that priors can influence the number of recovered shifts, especially when expectedNumberOfShifts<3. The number of recovered shifts is not significantly influenced by the priors when expectedNumberOfShifts≥5. Analyses based on 20 millions of generations instead of 10 millions are highlighted by a '20M' blue box; analyses that failed to converge with a 'CON' red box.

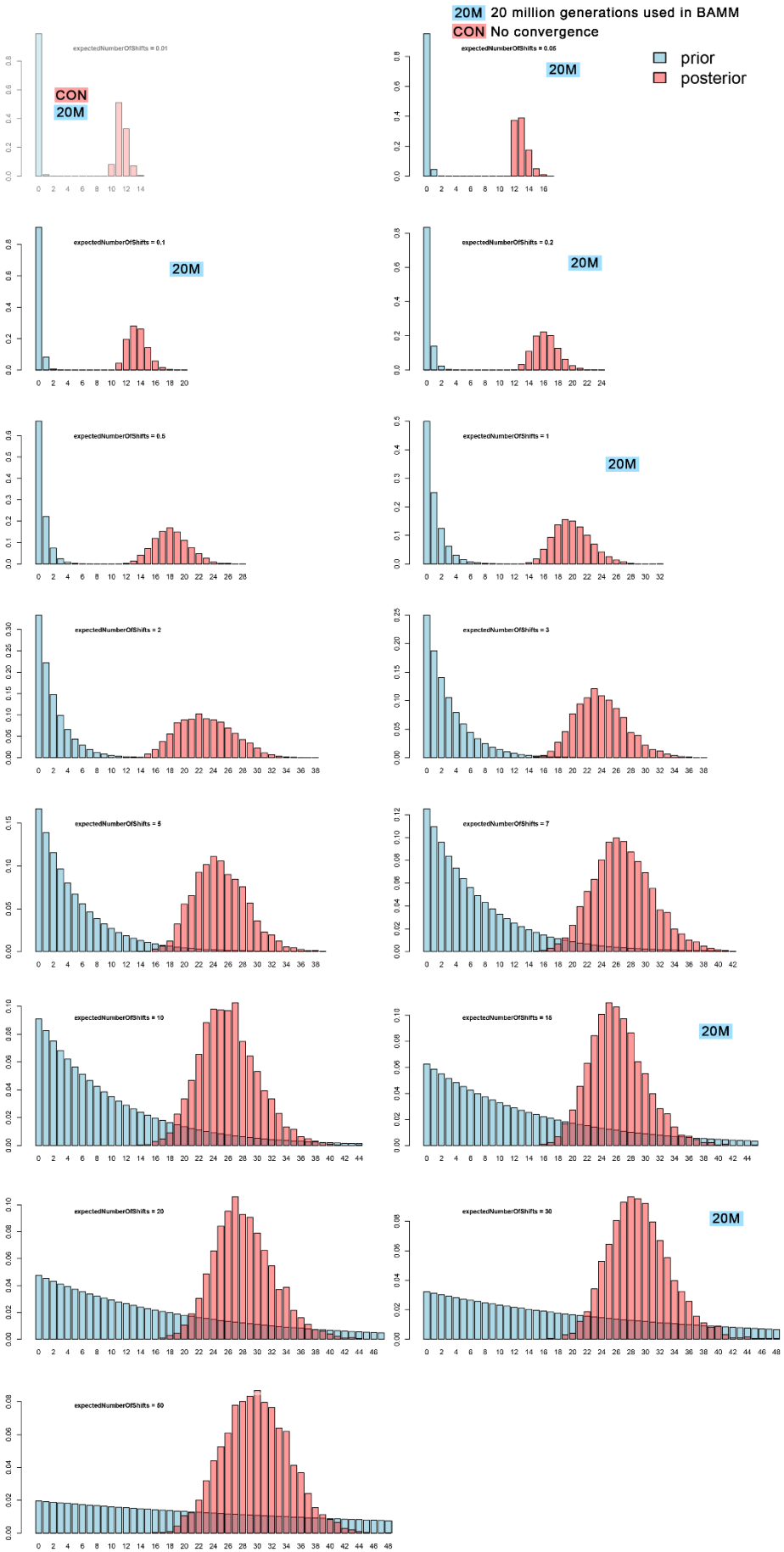

20M 20 million generations used in BAMM
CON No convergence
prior
posterior
expectedNumberOfShifts = 0.01
CON
20M
expectedNumberOfShifts = 0.05
20M
expectedNumberOfShifts = 0.1
20M
expectedNumberOfShifts = 0.2
20M
expectedNumberOfShifts = 0.5
expectedNumberOfShifts = 1
20M
expectedNumberOfShifts = 2
expectedNumberOfShifts = 3
expectedNumberOfShifts = 5
expectedNumberOfShifts = 7
expectedNumberOfShifts = 10
expectedNumberOfShifts = 15
20M
expectedNumberOfShifts = 20
expectedNumberOfShifts = 30
20M
expectedNumberOfShifts = 50

Figure S10 (Online resource) - Maximum shift credibility configuration obtained with BAMM for each of the expectedNumberOfShifts values. The position of the recovered shifts can slightly differ from one value to another. However, the positions of shift are very consistent and found on adjacent branches.

Supplementary tables

Table S1 - Genomic statistics. For every sample, we indicated the number of reads, number of contigs, total number of UCE loci, and number of UCE loci included in the genomic matrix used to infer the phylogenies.

Table S2 – Calibrations points used in the MCMCtree datation analyses.

Secondary calibrations		
Node	calibration (Ma)	Reference
Sphingidae / Saturniidae	[64.1;84.7]	Walhberg et al. 2013
Crown Sphingidae	[26.6;55.6]	Walhberg et al. 2013
Crown Saturniidae	[48.5;69.9]	Walhberg et al. 2013
Saturniinae/Cercophaninae	[43.0;64.1]	Walhberg et al. 2013
Ceratocampinae/Hemileucinae	[26.9;49.6]	Walhberg et al. 2013
Salassinae/Saturniinae	[23.9;45.5]	Walhberg et al. 2013

Fossils		
Node	calibration (Ma)	Reference
Crown Smerinthini	>15.2	Zhang & Zhang 1994
Crown Bunaeini	>3,66Ma	Kitching & Sadler 2011

Table S3 - Revised Saturniidae classification, as proposed in this study. See Figure 3.

Family	**Subfamily**	**Tribe**	**Subtribe**	**Genus**	**Subgenus**
Saturniidae	Agliinae			*Aglia*	
Saturniidae	Arsenurinae	Almeidaiini		*Almeidaia*	
Saturniidae	Arsenurinae	Arsenurini		*Arsenura*	
Saturniidae	Arsenurinae	Arsenurini		*Caio*	
Saturniidae	Arsenurinae	Arsenurini		*Copiopteryx*	
Saturniidae	Arsenurinae	Arsenurini		*Dysdaemonia*	
Saturniidae	Arsenurinae	Arsenurini		*Grammopelta*	
Saturniidae	Arsenurinae	Arsenurini		*Loxolomia*	
Saturniidae	Arsenurinae	Arsenurini		*Paradaemonia*	
Saturniidae	Arsenurinae	Arsenurini		*Rhescyntis*	
Saturniidae	Arsenurinae	Arsenurini		*Titaea*	
Saturniidae	Bunaeinae	Antistathmopterini		*Antistathmoptera*	
Saturniidae	Bunaeinae	Bunaeini		*Athletes*	
Saturniidae	Bunaeinae	Bunaeini		*Aurivillius*	
Saturniidae	Bunaeinae	Bunaeini		*Bunaea*	
Saturniidae	Bunaeinae	Bunaeini		*Bunaeopsis*	
Saturniidae	Bunaeinae	Bunaeini		*Cinabra*	
Saturniidae	Bunaeinae	Bunaeini		*Cirina*	
Saturniidae	Bunaeinae	Bunaeini		*Gonimbrasia*	*Gonimbrasia*
Saturniidae	Bunaeinae	Bunaeini		*Gonimbrasia*	*Nudaurelia*
Saturniidae	Bunaeinae	Bunaeini		*Gynanisa*	
Saturniidae	Bunaeinae	Bunaeini		*Heniocha*	
Saturniidae	Bunaeinae	Bunaeini		*Imbrasia*	
Saturniidae	Bunaeinae	Bunaeini		*Leucopteryx*	
Saturniidae	Bunaeinae	Bunaeini		*Lobobunaea*	
Saturniidae	Bunaeinae	Bunaeini		*Melanocera*	
Saturniidae	Bunaeinae	Bunaeini		*Protogynanisa*	
Saturniidae	Bunaeinae	Bunaeini		*Pseudimbrasia*	
Saturniidae	Bunaeinae	Bunaeini		*Pseudobunaea*	
Saturniidae	Bunaeinae	Bunaeini		*Rohaniella*	
Saturniidae	Bunaeinae	Bunaeini		*Ubaena*	
Saturniidae	Bunaeinae	Eochroini		*Eochroa*	
Saturniidae	Bunaeinae	Eochroini		*Parusta*	
Saturniidae	Bunaeinae	Eochroini		*Usta*	
Saturniidae	Bunaeinae	Eudaemoniini		*Eudaemonia*	
Saturniidae	Bunaeinae	Micragonini		*Adafroptilum*	
Saturniidae	Bunaeinae	Micragonini		*Adrougeriana*	
Saturniidae	Bunaeinae	Micragonini		*Basquiniana*	
Saturniidae	Bunaeinae	Micragonini		*Campimoptilum*	
Saturniidae	Bunaeinae	Micragonini		*Carnegia*	
Saturniidae	Bunaeinae	Micragonini		*Cyrtogone*	
Saturniidae	Bunaeinae	Micragonini		*Decachorda*	
Saturniidae	Bunaeinae	Micragonini		*Dogoia*	
Saturniidae	Bunaeinae	Micragonini		*Goodia*	
Saturniidae	Bunaeinae	Micragonini		*Holocerina*	

Saturniidae	Bunaeinae	Micragonini		*Ludia*	
Saturniidae	Bunaeinae	Micragonini		*Micragone*	
Saturniidae	Bunaeinae	Micragonini		*Mucidomorpha*	
Saturniidae	Bunaeinae	Micragonini		*Orthogonioptilum*	
Saturniidae	Bunaeinae	Micragonini		*Pseudoludia*	
Saturniidae	Bunaeinae	Micragonini		*Vegetia*	
Saturniidae	Bunaeinae	Pseudapheliini		*Pselaphelia*	
Saturniidae	Bunaeinae	Pseudapheliini		*Pseudaphelia*	
Saturniidae	Bunaeinae	Urotini		*Maltagorea*	
Saturniidae	Bunaeinae	Urotini		*Pseudantheraea*	
Saturniidae	Bunaeinae	Urotini		*Sinobirma*	
Saturniidae	Bunaeinae	Urotini		*Tagoropsiella*	
Saturniidae	Bunaeinae	Urotini		*Tagoropsis*	
Saturniidae	Bunaeinae	Urotini		*Urota*	
Saturniidae	Ceratocampinae	Bathyphlebiini		*Bathyphlebia*	
Saturniidae	Ceratocampinae	Bathyphlebiini		*Eacles*	
Saturniidae	Ceratocampinae	Bathyphlebiini		*Neorcarnegia*	
Saturniidae	Ceratocampinae	Bathyphlebiini		*Schausiella*	
Saturniidae	Ceratocampinae	Ceratocampini		*Citheronia*	
Saturniidae	Ceratocampinae	Ceratocampini		*Citheronioides*	
Saturniidae	Ceratocampinae	Ceratocampini		*Citheronula*	
Saturniidae	Ceratocampinae	Ceratocampini		*Procitheronia*	
Saturniidae	Ceratocampinae	Dryocampini		*Adeloneivaia*	
Saturniidae	Ceratocampinae	Dryocampini		*Adelowalkeria*	
Saturniidae	Ceratocampinae	Dryocampini		*Almeidella*	
Saturniidae	Ceratocampinae	Dryocampini		*Anisota*	
Saturniidae	Ceratocampinae	Dryocampini		*Ceratesa*	
Saturniidae	Ceratocampinae	Dryocampini		*Ceropoda*	
Saturniidae	Ceratocampinae	Dryocampini		*Cicia*	
Saturniidae	Ceratocampinae	Dryocampini		*Citioica*	
Saturniidae	Ceratocampinae	Dryocampini		*Dacunju*	
Saturniidae	Ceratocampinae	Dryocampini		*Dryocampa*	
Saturniidae	Ceratocampinae	Dryocampini		*Giacomellia*	
Saturniidae	Ceratocampinae	Dryocampini		*Jaiba*	
Saturniidae	Ceratocampinae	Dryocampini		*Megaceresa*	
Saturniidae	Ceratocampinae	Dryocampini		*Mielkesia*	
Saturniidae	Ceratocampinae	Dryocampini		*Oiticella*	
Saturniidae	Ceratocampinae	Dryocampini		*Othorene*	
Saturniidae	Ceratocampinae	Dryocampini		*Psigida*	
Saturniidae	Ceratocampinae	Dryocampini		*Psilopygida*	
Saturniidae	Ceratocampinae	Dryocampini		*Psilopygoides*	
Saturniidae	Ceratocampinae	Dryocampini		*Ptiloscola*	
Saturniidae	Ceratocampinae	Dryocampini		*Rachesa*	
Saturniidae	Ceratocampinae	Dryocampini		*Scolesa*	
Saturniidae	Ceratocampinae	Dryocampini		*Syssphinx*	
Saturniidae	Cercophaninae	Cercophanini		*Cercophana*	
Saturniidae	Cercophaninae	Cercophanini		*Microdulia*	
Saturniidae	Cercophaninae	Cercophanini		*Neocercophana*	
Saturniidae	Cercophaninae	Janiodini		*Janiodes*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Ancistrota*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Auroraia*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Austrolippa*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Automerella*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Automerina*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Automeris*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Automeropsis*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Catacantha*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Erythromeris*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Eubergia*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Gamelia*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Gamelioides*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hylesia*	*Bertrandylesia*
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hylesia*	*Extremylesia*
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hylesia*	*Gamylesia*
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hylesia*	*Hylesia*
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hylesia*	*Linylesia*
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hylesia*	*Micrattacus*
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hylesiopsis*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hyperchiria*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hyperchirioides*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Hypermerina*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Leucanella*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Mexicantha*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Molippa*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Prohylesia*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Pseudautomeris*	
Saturniidae	Hemileucinae	Hemileucini	Automeriina	*Travassosula*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Arias*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Cerodirphia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Coloradia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Dirphia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Dirphiella*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Dirphiopsis*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Eudyaria*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Heliconisa*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Hemileuca*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Hidripa*	

Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Hispaniodirphia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Ithomisa*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Kentroleuca*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Lemaireodirphia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Manodirphia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Meroleuca*	*Dihirpa*
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Meroleuca*	*Meroleuca*
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Meroleuca*	*Meroleucoides*
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Ormiscodes*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Paradirphia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Periphoba*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Pseudodirphia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Rhodirphia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Winbrechlinia*	
Saturniidae	Hemileucinae	Hemileucini	Hemileucina	*Xanthodirphia*	
Saturniidae	Hemileucinae	Lonomiini		*Lonomia*	
Saturniidae	Hemileucinae	Lonomiini		*Periga*	
Saturniidae	Hemileucinae	Polythysanini		*Adetomeris*	
Saturniidae	Hemileucinae	Polythysanini		*Callodirphia*	
Saturniidae	Hemileucinae	Polythysanini		*Cinommata*	
Saturniidae	Hemileucinae	Polythysanini		*Polythysana*	
Saturniidae	Hirpidinae			*Hirpida*	
Saturniidae	Oxyteninae			*Homoeopteryx*	
Saturniidae	Oxyteninae			*Oxytenis*	
Saturniidae	Oxyteninae			*Therinia*	
Saturniidae	Salassinae			*Salassa*	
Saturniidae	Saturniinae	Attacini		*Archaeoattacus*	
Saturniidae	Saturniinae	Attacini		*Attacus*	
Saturniidae	Saturniinae	Attacini		*Callosamia*	
Saturniidae	Saturniinae	Attacini		*Coscinocera*	
Saturniidae	Saturniinae	Attacini		*Epiphora*	
Saturniidae	Saturniinae	Attacini		*Eupackardia*	
Saturniidae	Saturniinae	Attacini		*Hyalophora*	
Saturniidae	Saturniinae	Attacini		*Rhodinia*	
Saturniidae	Saturniinae	Attacini		*Rothschildia*	
Saturniidae	Saturniinae	Attacini		*Samia*	
Saturniidae	Saturniinae	Saturniini		*Actias*	
Saturniidae	Saturniinae	Saturniini		*Antheraea*	*Antheraea*
Saturniidae	Saturniinae	Saturniini		*Antheraea*	*Antheraeopsis*
Saturniidae	Saturniinae	Saturniini		*Antheraea*	*Telea*
Saturniidae	Saturniinae	Saturniini		*Antherina*	
Saturniidae	Saturniinae	Saturniini		*Argema*	
Saturniidae	Saturniinae	Saturniini		*Austrocaligula*	
Saturniidae	Saturniinae	Saturniini		*Ceranchia*	
Saturniidae	Saturniinae	Saturniini		*Copaxa*	
Saturniidae	Saturniinae	Saturniini		*Cricula*	
Saturniidae	Saturniinae	Saturniini		*Eosia*	
Saturniidae	Saturniinae	Saturniini		*Lemaireia*	
Saturniidae	Saturniinae	Saturniini		*Loepa*	
Saturniidae	Saturniinae	Saturniini		*Neodiphthera*	
Saturniidae	Saturniinae	Saturniini		*Opodiphthera*	
Saturniidae	Saturniinae	Saturniini		*Pararhodia*	
Saturniidae	Saturniinae	Saturniini		*Saturnia*	*Agapema*
Saturniidae	Saturniinae	Saturniini		*Saturnia*	*Cachosaturnia*
Saturniidae	Saturniinae	Saturniini		*Saturnia*	*Calosaturnia*
Saturniidae	Saturniinae	Saturniini		*Saturnia*	*Eudia*
Saturniidae	Saturniinae	Saturniini		*Saturnia*	*Neoris*
Saturniidae	Saturniinae	Saturniini		*Saturnia*	*Perisomena*
Saturniidae	Saturniinae	Saturniini		*Saturnia*	*Rinaca*
Saturniidae	Saturniinae	Saturniini		*Saturnia*	*Saturnia*
Saturniidae	Saturniinae	Saturniini		*Syntherata*	
Saturniidae	Saturniinae	Solini		*Solus*	

The novel genus described by Brechlin (2019), *Hirpsinjaevia*, should be considered as a Hirpidinae (see Chapitre 3).

Table S4 (Online Resource) - Saturniidae morphology measurements. 12850 measurements were done for 2577 saturniid specimens. The measurements considered were: (i) body length, (ii) thorax width between the junction points of thorax and forewings, (iii) thorax width at the middle of the thorax, (iv) thorax width between the junction points of thorax and hindwings and (v) the forewing surface. Measurements were performed in BOLD (www.boldsystems.org) and rescaled thanks to a scale measurement.

Table S5 - Parameter values from best models of trait-dependent diversification analyses. The full model was recovered as the best model for all life-history traits considered (see Table 1 in main text). The values represent medians of each parameter (λ = speciation rate, μ = extinction rate and λ-μ = diversification rate) as estimated with 10,000 MCMC generations (10% of burnin).

MCMC estimates		λ_1	μ_1	$\lambda_1 - \mu_1$	λ_2	μ_2	$\lambda_2 - \mu_2$	λ_3	μ_3
Body Size (BS)	cat1: small BS cat2: medium BS cat3: large BS	0.142	3.54e-03	0.138	0.164	0.011	0.152	0.177	3.70e-03
Dispersal capacities - Wingload (WL)	cat1: low WL cat2: medium WL cat3: high WL	0.117	3.49e-03	0.113	0.171	6.96e-03	0.163	0.184	2.79e-03
Polyphagy level (POL)	cat1: low POL cat2: medium POL cat3: high POL	0.127	0.0302	0.0964	0.148	2.77e-03	0.145	0.214	1.62e-03
Pupation mode	cat1: No cocoon cat2: loose cocoon cat3: plain cocoon	0.172	0.0169	0.154	0.128	0.0293	0.0984	0.178	1.59e-03

Table S6 – Food plant dataset (Ballesteros-Mejia et al. submitted, see attached articles).

Table S7 (Online resource) - Genus-level pupation mode dataset.

To be submitted in PNAS

Life-history innovations drove spatial and temporal diversification in capital-breeding insects

Pierre Arnal[1*], Astrid Cruaud[2], Jean-Yves Rasplus[2], Marianne Elias[1], Liliana Ballesteros[1], Fabien Condamine[3], Thibaud Decaëns[4], Delphine Gey[5], Winnie Hallwachs[6], Daniel H. Janzen[6], Ian J. Kitching[7], Sébastien Lavergne[8], Carlos Lopez-Vaamonde[9,10], Jérôme Murienne[11], Sabine Nidelet[2], Sujeevan Ratnasingham[12], Rodolphe Rougerie[1*]

[1] Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, EPHE, Sorbonne Université, Université des Antilles; Paris, 75005, France.

[2] Centre de Biologie pour la Gestion des Populations (CBGP), INRAE, CIRAD, IRD, Montpellier SupAgro, Université de Montpellier; Montpellier, 34000, France.

[3] Institut des Sciences de l'Evolution de Montpellier (ISEM), CNRS, Université de Montpellier: Montpellier, 34095, France.

[4] Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), Université de Montpellier, CNRS, EPHE, IRD, Université Paul Valéry Montpellier 3: Montpellier, 34090, France.

[5] Service de Systématique moléculaire (SSM), CNRS, Muséum National d'Histoire Naturelle, Sorbonne Université: Paris, 75005, France.

[6] Department of Biology, University of Pennsylvania: Philadelphia, 19104, United States of America.

[7] Department of Life Sciences, Natural History Museum: London, United Kingdom.

[8] Laboratoire d'Écologie Alpine (LECA), Université Grenoble Alpes, CNRS, Université Savoie Mont Blanc: Grenoble, 38000, France.

[9] Institut de Recherche sur la Biologie de l'Insecte (IRBI), CNRS, Université de Tours, UFR Sciences et Techniques: Tours, 37200, France.

[10] Unité de Recherche de Zoologie Forestière, INRAE: Orléans, 45075, France.

[11] Evolution et Diversité Biologique (EDB), Université Toulouse 3 Paul Sabatier, CNRS, IRD: Toulouse, 31077, France.

[12] Centre for Biodiversity Genomics, University of Guelph: Guelph, Ontario, Canada.

[*] Correspondence: pierrearnal34@gmail.com & rodolphe.rougerie@mnhn.fr.

ABSTRACT

Background. Today's uneven geographical and phylogenetic distribution of biodiversity reflects complex underlying temporal and spatial dynamics. The development of large and robust phylogenies, the assembly of distribution and trait databases, and new analytical methods are shedding light on the respective roles of abiotic and biotic factors on patterns of diversification and emphasize the importance of adaptive traits. Yet, very few studies have investigated the role of these factors on insect diversification dynamics at a global scale, though it constitutes the most diverse group of organisms on Earth.

Results. Using a combination of phylogenomics, diversification and historical biogeography analyses we propose a comprehensive account of the spatial and temporal diversification dynamics in wild silkmoths (Saturniidae). These moths are typical capital-breeding insects in which non-feeding adults have short lifespan entirely devoted to reproduction. Considering several key life-history traits analyzed using trait-dependent models of diversification, we found that wild silkmoths evolved toward larger body-size and increased level of polyphagy of their larval stages, in contrast to most other insects. Polyphagy, pupation mode and dispersal capacities positively impacted diversification rates, but only the first two traits likely favored the mobility of lineages between distant biogeographical regions. Therefore, wild silkmoths challenge traditional expectations with poor dispersers being worldwide colonizers.

Conclusion. Our work demonstrates the importance of life-history innovations in driving the dynamics of diversification in a family of moths. In particular, high polyphagy level and protection against environmental hazards (plain cocoons) are proposed as key adaptations for coping with large size and thus for capital-breeders to colonize and thrive on all continents.

SIGNIFICANCE

In this study, we carried out a global scale analysis of spatial and temporal evolution of wild silkmoths, a diverse group of capital-breeding insects. Through the analysis of a set of life-history traits we show that – in contrast to most other insects – wild silkmoths evolved toward larger adult body-size and increased level of polyphagy of their caterpillars. Furthermore, polyphagy and strong cocoons likely favored the mobility of lineages between distant biogeographical regions. Through key adaptations driven by their reproductive strategy, wild silkmoths adopted a capital breeding strategy and challenge traditional expectations with poor dispersers being worldwide colonizers.

KEYWORDS

Adaptation, Insects, Life-history traits, Reproductive strategy, Phylogenomics, Wild Silkmoths, Capital breeding, Saturniidae, Ultraconserved elements.

INTRODUCTION

Large-scale diversification studies have mostly focused on vertebrates (Jetz, Thomas et al. 2012; Bonetti & Wiens 2014; Claramunt & Cracraft 2015; Cooney, Bright et al. 2017) or plants (Fiz-Palacio et al. 2011; Antonelli, Zizka et al. 2015; Silvestro, Cascales-Minana et al. 2015) and only a few have addressed the role of adaptive traits in spatial and temporal diversification dynamics of widespread taxa (Price, Hopkins et al. 2012; Modica, Gorson et al. 2020). In insects, studies are impeded by limited knowledge of their diversity, distribution, evolution, traits, or ecology (Hortal et al. 2015). In addition, studies are rarely global (but see Economo et al. 2018, 2019), and often focus on higher taxonomic ranks with limited insights into the drivers of diversification (but see Condamine et al. 2018; Matos-Maravi et al. 2018; Cozzarolo et al. 2019; Dorey et al. 2020).

We studied a family of moths, the wild silkmoths (Saturniidae), which are among the best documented insects, certainly because these emblematic moths and their caterpillars are spectacular in their size and forms, often abundant, and have attracted the attention of naturalists for centuries (Fig. 1). As most diverse clades of organisms, their diversity is unevenly distributed, both phylogenetically (*ca* 80% of the species belong to 2 of the 8 subfamilies) and geographically (>95% of species inhabit the intertropical region, 66% in the Neotropics alone; Fig. 1). Wild silkmoths exhibit a set of life-history traits associated with the way they allocate resources to their reproduction as capital breeders (Stephens et al. 2009): the adult moths do not feed, achieve a large size (females are pro-ovigenic, carrying eggs in large abdomens), have low mobility and short lifespan (<7 days); their caterpillars are often gregarious and generalists in their use of food-plants (Janzen 1984). In addition, saturniid moths adopted different strategies to protect their pupa – the most vulnerable stage to predators, parasitoids and pathogens – from underground pupation to multi-layered silk cocoons, well known for their use in production of natural silk textiles (Li et al. 2017). As one of the most diverse group of capital breeding organisms and as one of the best documented group of insects, wild silkmoths represent an releveant model to investigate the role of life-history traits in diversification dynamics.

Here, we inferred the backbone phylogeny of the family Saturniidae from 1,381 Ultra-Conserved Elements (UCEs) sampled from 236 species representing all 179 genera and most subgenera, as well as 15 outgroups. Then, using a dated species-level tree built by assigning each of the 3,451 described species to their (sub)genera, we investigated if and how key life-history traits (body size; dispersal capacity; larval diet breadth; pupation mode) and major geological and climatic events have shaped their current distribution and influenced their diversification dynamics.

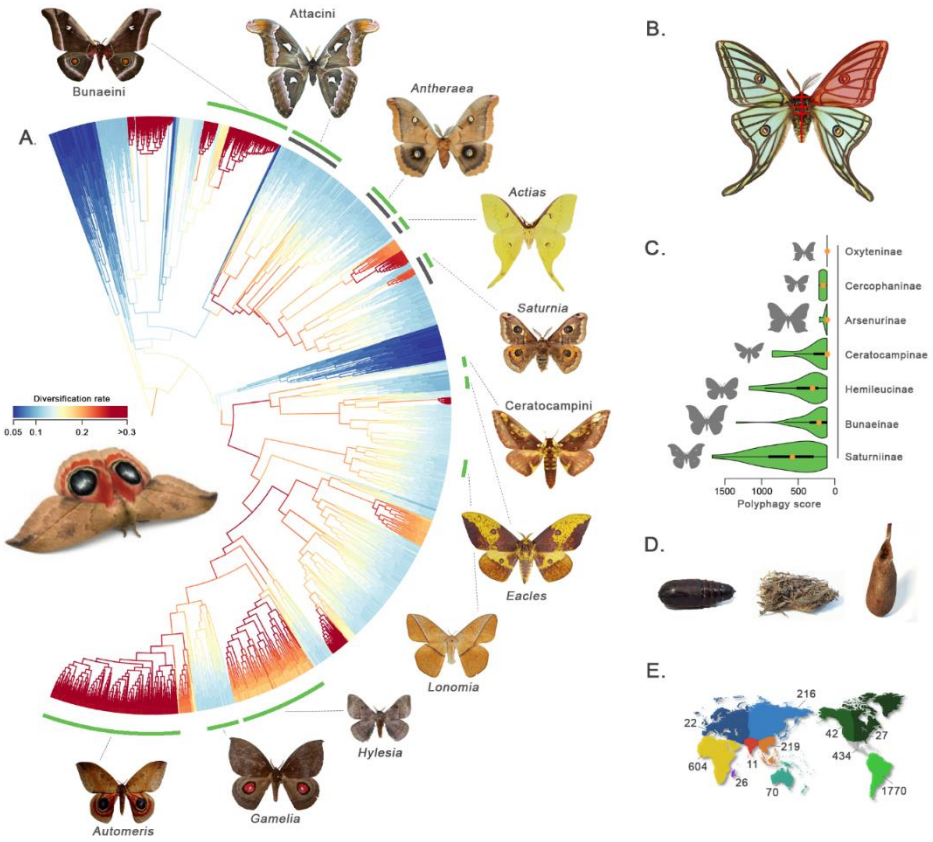


Figure 1 - Saturniidae life traits and distribution. (A) Saturniidae species-level dated phylogeny as estimated with ASTRAL and MCMCtree, in which species were grafted with PASTIS. Branch are colored according to branch-specific diversification rates as estimated with BAMM. External circle segments depict widespread clades (black) and polyphagous clades (green). (B) Morphological measurements considered in this study: thorax width (three measurements), body length and forewing surface. (C) Polyphagy score in the main Saturniidae subfamilies as estimated with the Phylogenetic Diversity (PD) metric. (D) The three modes of pupation in Saturniidae, *i.e.*, from left to right, pupation underground, loose cocoon and plain cocoon. (E) Number of saturniid species per biogeographical area.

RESULTS AND DISCUSSION

Phylogenomics illuminates the temporal and spatial evolution of wild silkmoths

Maximum likelihood (ML, IQ-TREE) and coalescence-based (ASTRAL) analyses of the UCE dataset (50%-complete matrix; 255 taxa; 1,171 markers; 788,018 bp with less than 6% gaps) produced similar topologies (Figs 2, S1, S2). The only notable exception is the position of Agliinae, recovered as sister to Salassinae + Saturniinae in the ASTRAL tree, while it is sister to *Hirpida* in the ML tree. Complementary analyses revealed that the position of Agliinae in the ML tree could be due to a long branch attraction artefact towards *Hirpida.* Agliinae was indeed recovered as sister to Salassinae + Saturniinae by IQ-TREE when the 20% most biased UCEs in terms of heterogeneity among branch lengths were removed (Fig. S2), a node also recovered in the phylogenetic analyses of Hamilton et al. (2019) from an independent genomic dataset. We therefore relied on the ASTRAL topology for the analyses of divergence ages, diversification and traits. In addition, GC content did not affect tree inference. On all trees, the monophyly of Saturniidae was well supported as was that of all previously

recognized subfamilies and tribes, except for subfamily Hemileucinae and the tribes Hemileucini, Urotini and Saturniini (Fig. S2). Our results revealed the necessary re-assignments of several genera and we propose a new classification for the family (Figs 1, S1 and S3) recognizing 10 subfamilies, 18 tribes, and 2 subtribes (see *Taxonomic treatment* in SI (see Chapitre 1 – Table S3)).

Ancestral range estimations unambiguously support a neotropical origin of saturniids (Figs 2; S4), shortly after the K-T extinction event (ca 60.3 Ma). From this origin, the biogeographical history of wild silkmoths remarkably mirrors the history of Earth through the Cenozoic period. Three lineages subsequently diverged and diversified in the Neotropics, with the earliest shift in diversification rates occurring ca. 44Ma in the branch leading to Ceratocampinae and Hemileucinae, in South-America. Two other shifts were identified within Hemileucinae (ca. 16Ma): i) along the branch leading to *Leucanella*, *Pseudautomeris*, *Hypermerina* and *Automeris*, and ii) within *Hylesia*. These shifts may be linked to the Middle Miocene Climatic Optimum and the rise of the Northern Andes that initiated ca 23Ma and accelerated by 15Ma (Garzione et al. 2008,2014; Hoorn et al. 2010). Saturniid moths also expanded their range northward, out of the Neotropics, possibly through the island arc (Iturralde-Vinent & MacPhee 1999) that connected Colombia to Yucatan from late Cretaceous to middle Eocene (Morley 2003; Pindell & Kennan 2009) when warm Eocene climates made the Nearctic region hospitable to megathermal lineages (Zachos et al. 2008). From there, a dispersal event to the Palearctic region occurred through Beringia in the early Eocene (*ca.* 47Ma). This dispersal might have been facilitated by deciduous forests extending beyond the Arctic Circle (Tiffney 1985; Jahren 2007; Baskin & Baskin 2016) and hosting food plants (e.g. oak, beech, etc.) on which early branching lineages still feed today (i.e. Agliinae, Salassinae, *Solus* and *Rhodinia*). In this clade occurring mostly in the Old World, early branching lineages are now mostly restricted to the Sino-Himalayan mountains where their host-plants found suitable refuges during the Eocene–Oligocene climatic deterioration (i.e. the "Grande Coupure"; Zachos et al. 2001). By ca. 44Ma, one lineage – the Bunaeinae – thrived in the Afrotropics (Africa + Madagascar), a region from where they never escaped with the noticeable exception of *Sinobirma* (Rougerie et al. 2012). Three shifts in diversification rates were detected within the Afrotropical Bunaeinae. The first two happened ca. 35Ma and 31Ma on the branches leading to Micragonini and Bunaeini+Urotini, respectively, and appear concomitant with cooling climatic conditions at the Eocene–Oligocene boundary (~33.5–26 Mya) that induced a dramatic faunal and floral turnover over the globe (Ivany et al. 2000, Seiffert 2007). The third shift in diversification rates occurred within Bunaeini, ca. 10 Ma, and may be linked to the aridification of Northern Africa (Zhang et al. 2014), the emergence of C4 plants and the fragmentation of forests in the Eastern side of Africa (Cerling et al. 1997, Ségalen et al. 2007). Another Old World lineage – the Saturniinae – experienced an early shift in diversification rate (ca. 35Ma) along the branch leading to tribes Attacini and Saturniini. Both tribes colonized all land masses; they returned to the New World, dispersed several times to Africa and Madagascar, and reached

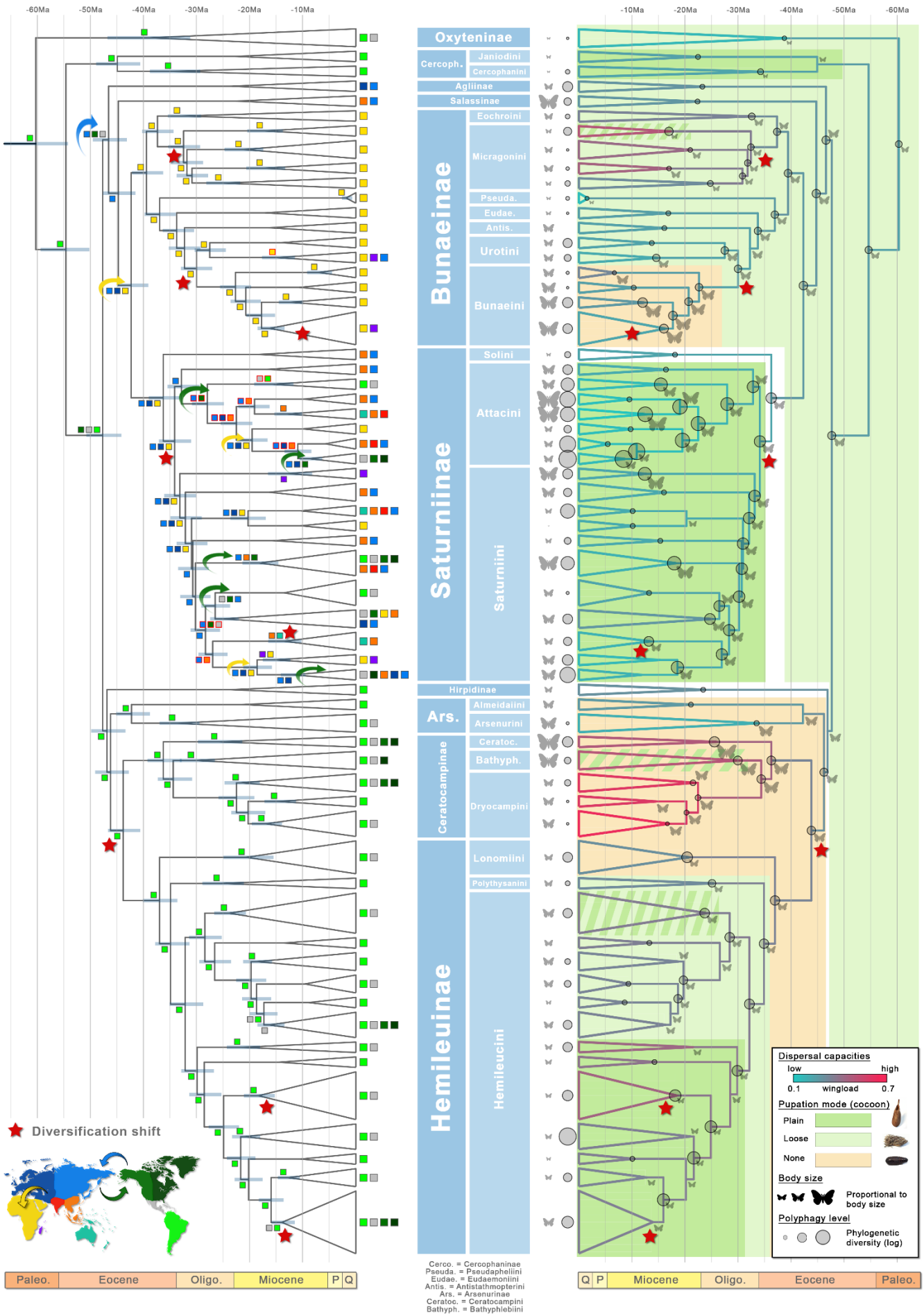
the Australian region, revealing a more recent shift in diversification rate ca 12Ma when the lineage leading to *Neodiphthera* and *Syntherata* emerged in the Oriental and Australian regions.

Life-history traits drove diversification dynamics of wild silkmoths

The body size conundrum of capital-breeding insects

The analyses of measurements from images of 1576 species in 172 genera (2577 specimens, and 12,850 measurements in total; see Table S2) reveal that the common ancestor to all wild silkmoths was small, falling into the bottom tier of the smallest extant saturniids. Body size carries phylogenetic signal (Bloomberg's K=0.84; p=0.001), but the evolution toward smaller or larger body sizes occurred several times throughout the evolution of the family. Salassinae, Bunaeini, Attacini, Bathyphlebiini and Citheroniini are striking examples of independent evolution of gigantism, whereas Pseudapheliini, Micragonini, Oxyteninae, and several genera in Dryocampini, Hemileucini and Saturniini are small moths that can be as much as 20 times smaller than the largest members of the family. Our results reveal an overall high heterogeneity in the evolution of this trait, with similar transition rate toward larger or smaller body size (Figure 3E, Table S3). Interestingly, this contrasts with the negatively skewed distribution of saturniid body size (log-transformed; Figure S5), showing that larger saturniids are more numerous than smaller ones. Furthermore, large body-sized moths (Figure 3A, Table S3) had higher diversification rates than small ones, whereas lineages of medium-sized moths have intermediate values. These results suggest that, for saturniids, evolution toward smaller body size can be an evolutionary dead-end, an observation contrasting with previously proposed models in other groups of animals (Hutchinson & MacArthur 1959; Maurer 1998), including insects (Rainford et al. 2016; but see Misof 2002 for a similar pattern in Anisoptera), and that may reflect the importance of being large when you are a capital-breeder. The capital-breeding reproductive strategy of wild silkmoths indeed implies that the larvae storestorage by the larvae of all the importantthe resources needed by the adults need for their reproduction and for the formation of a large amount of formed eggs (Tammaru & Haukioja 1996; Davis et al. 2016).

Figure 2 (following page) – Mirrored dated (scale at top/bottom) reconstructions of the biogeographical history (left) and of the evolution of life-history traits in wild silkmoths lineages. The newly adopted classification is given in the center of the figure. Several monophyletic lineages were grouped for the sake of visualization (no proportionality). Red stars depict the diversification rate shifts as estimated with BAMM. Left-hand tree: biogeographic history of Saturniidae as estimated with DECX on 1000 PASTIS, species-level trees. Only the best ranges are represented. At nodes, red border was used when the confidence about the estimation was <50%. Arrows highlight the main dispersal events as represented in the World map at the left bottom. Blue bars indicate node age 95% confidence intervals. Right-hand tree: saturniid life traits evolution. Branches were colored according to wingload values. Background colors represent the pupation modes. Shaded silhouette size are proportional to body size (log-normalized). Polyphagy level is depicted at tips and nodes with circles which size is proportional to the estimated Polyphagy Diversity score (log-normalized). The full genus-level phylogeny is represented in the Figure S4.


-60Ma
-50Ma
-40Ma
-30Ma
-20Ma
-10Ma
Oxyteninae
Cercoph.
Janiodini
Cercophanini
Agliinae
Salassinae
Bunaeinae
Eochroini
Micragonini
Pseuda.
Eudae.
Antis.
Urotini
Bunaeini
Saturniinae
Solini
Attacini
Saturniini
Hirpidinae
Ars.
Almeidaiini
Arsenurini
Ceratocampinae
Ceratoc.
Bathyph.
Dryocampini
Hemileuinae
Lonomiini
Polythysanini
Hemileucini
Diversification shift
Paleo.
Eocene
Oligo.
Miocene
P
Q
Cerco. = Cercophaninae
Pseuda. = Pseudapheliini
Eudae. = Eudaemoniini
Antis. = Antistathmopterini
Ars. = Arsenurinae
Ceratoc. = Ceratocampini
Bathyph. = Bathyphlebiini
Dispersal capacities
low
high
0.1
wingload
0.7
Pupation mode (cocoon)
Plain
Loose
None
Body size
Proportional to body size
Polyphagy level
Phylogenetic diversity (log)

Dispersal capacity promotes diversification, but not biogeographical interchanges

Analyses were based on the calculation of wingload ratio derived from the previously mentioned measurements (Table S2). Most lineages are poor dispersers with low wingload ratio and the ancestor of all Saturniidae was likely also a poor disperser. Phylogenetic signal of wingload is strong (Bloomberg's K=1.04; p=0.001). Diversification rates in lineages with high or medium dispersal capacity are higher than in those with low capacities (Fig. 3, Table S3) as already observed in other organisms (Willis et al. 2014, Faurby & Antonelli 2018). Indeed, high dispersal ability increases the probability of reproductive isolation in novel habitats (e.g. to escape interspecific competition). A significant increase in dispersal capacity is observed in Ceratocampinae and, to a lesser extent, in Micragonini (Figs 2, S4). However, dispersal capacity does not seem to have played an important role in long distance colonisation. Indeed, these two groups never dispersed outside the New World and the African continent, respectively.

Polyphagy – the engine driving wild silkmoths' diversification

Food plants were compiled for 586 species in 121 genera, which represents 2,568 records of associations between genera and plant families (Table S4). Diet breadth appears variable (Figs 2, S4) and ranges from specialized genera (e.g. *Tagoropsis* with all documented species feeding on Sapindaceae) to generalists feeding on many plant families (e.g. *Automeris* on >14 plant families). We found a strong phylogenetic signal in diet breadth (Bloomberg's K= 0.547; p=8.0e-04) and inferred that the ancestor of extant Saturniidae was oligophagous. Diet breath increased several times independently and polyphagous lineages (e.g. Saturniinae; Bunaeini; *Hylesia*, *Automeris*) diversified 66% and 46% faster than specialized and oligophagous lineages (e.g. most Arsenurinae and Ceratocampinae) respectively. Transition rates toward specialization are significantly lower than those toward generalism (Fig. 3), which strongly advocates an evolutionary trend toward an increasing polyphagy. However, while an increase in diet breath seems to correlate with higher diversification rates in Hemileucinae, Bunaeinae and Saturniinae, (Fig. S6, Table S5) this is not the case for Ceratocampinae, whose diversification is better explained by higher dispersal capacities. While most herbivorous insects tend to evolve toward specialization over evolutionary time, thus gaining greater fitness (Singer 2008), wild silkmoths followed the opposite path toward generalism. A similar pattern is observed in other capital-breeding insects like fruit flies (Clarke 2017). Finally, although not all highly polyphagous lineages of Saturniidae successfully colonized new areas (Figs 2, S4), Attacini and Saturniini which display the highest frequency of long-distance dispersal events in the family, are also those comprising the most polyphagous genera. Consequently, polyphagy may have facilitated the establishment and adaptation of populations in new environments with different plant communities.

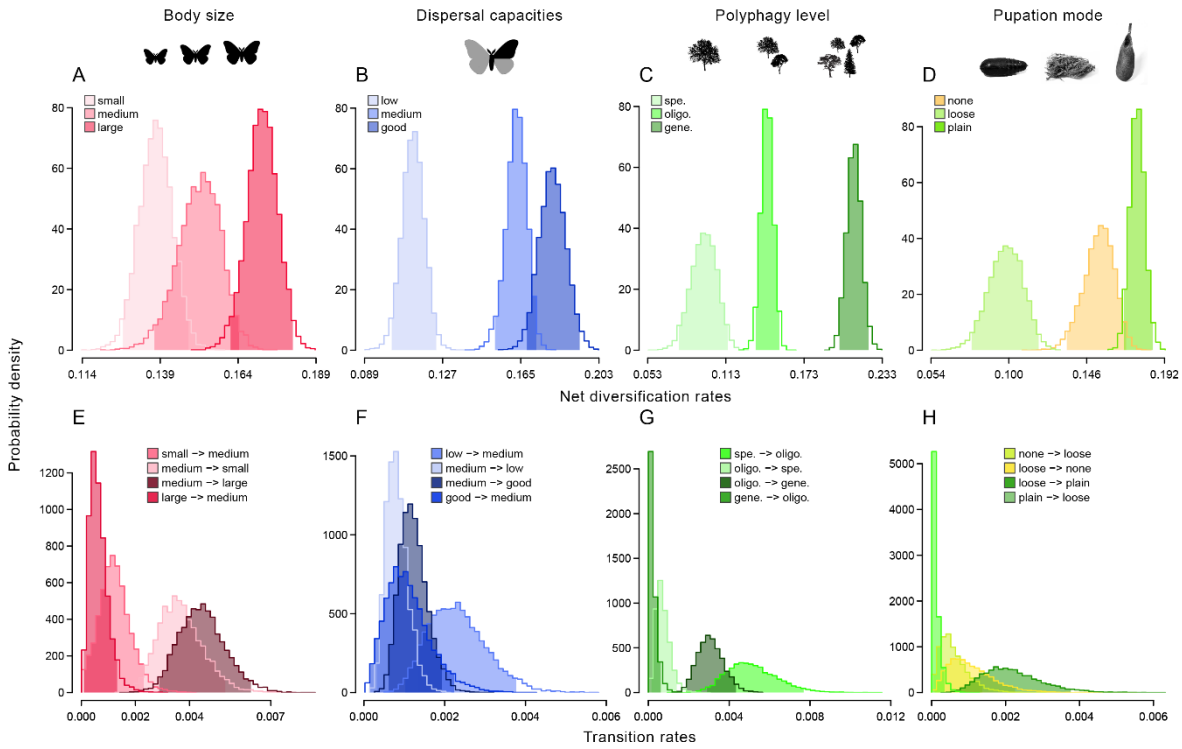


Figure 3 - Posterior probability distributions of parameters obtained from the MuSSE analysis of trait dependent diversification models. Upper row (A, B, C, D): net diversification rate (λ-μ); lower row (E, F, G, H): transition rates. Shaded areas correspond to the 95% credibility intervals.

Protection of pupae promoted diversification

Pupation modes were compiled for 139 (sub)genera representing all higher taxa but three (Table S6). One third of the documented taxa pupate underground, without spinning a cocoon, while the other two thirds pupate inside a cocoon. Cocoons showed different degrees of elaborateness, from a loose net of silk threads to plain cocoons that offer a better protection to the pupa. Phylogenetic signal in pupation mode is strong (Figs 2, S4) and the ancestor to all extant wild silkmoths may have pupated in a loose cocoon. Evolution of a plain cocoon occurred several times independently in the ancestors to Cercophaninae, some Micragonini, Saturniinae, some Hemileucinae, and in a single genus of Ceratocampinae (*Neorcarnegia*). A shift to underground pupation, occurred in the ancestor to Arsenurinae, Ceratocampinae and Hemileucinae and in the ancestor of Bunaeini. The causes triggering these shifts remain intriguing. However, the lineages branching just after the shifts have specialized caterpillar (Figs 2, S4). Low plant diversity and frequent exposure to harsh environmental conditions (drought, fires), may be an explanation. The capacity to spin a cocoon, either loose or plain, re-evolved twice independently within the New World lineages that had previously shifted to underground pupation: *Neorcarnegia* (Ceratocampinae), and in all Hemileucinae but Lonomiini. Diversification rates were significantly slower in lineages spinning loose cocoons than in those pupating underground or spinning plain cocoon (Fig. S3, Table S3). The latter two conditions are efficient strategies to escape predators and parasites (Gross 1993) and survive unfavourable (drought, cold) or even hostile (fire,

flooding) environmental conditions (Danks 2004). Interestingly, the positive diversification shift identified along the branch leading to Saturniinae (Saturniini, Attacini and Solini), is congruent with a transition from loose to plain cocoons, conferring higher level of protection. Similar transitions however happened in several other clades (e.g. Hemileucinae) (Figs 2, S4) without causing an apparent shift in diversification rate, suggesting that other factors are acting on major changes of evolutionary dynamics. However, we used the same character state ("plain cocoons") to, for example, describe thin cocoons of many Hemileucinae, and the strong cocoons of Saturniini and Attacini, which may confuse the results. A reasonable hypothesis is that strong cocoons might have been key to surviving cold seasons and to successfully colonize all land masses from Eastern Palearctic to Neotropical mountains through the Nearctic region as only Saturniinae managed to do.

CONCLUSION

Our study offers a comprehensive overview of the temporal and spatial dynamics of diversification in a globally distributed diverse group of phytophagous insects, unveiling the main events that shaped the distribution of extant diversity in saturniid moths. Throughout their evolutionary history, wild silkmoths seized multiple opportunities linked to changes in environmental conditions (climate, topology) to expand geographically, from a Neotropical origin during the Paleocene to a global distribution today, proving capable of dispersing across landmasses repeatedly. We show however that not all lineages experienced similar "success" in terms of diversification and range expansion. In particular, our results emphasize the importance of life-history traits in shaping the evolution of these moths, both in terms of diversification and geographical mobility. Neither dispersal capacity, nor body size appear to be good predictors of long-distance dispersal; instead the features that seem to be the most critical in successfully colonizing new biogeographical regions are both the ability to exploit a diversity of food plants (polyphagy) and to spin plain cocoons.

Janzen (Janzen 1984) hypothesized that the shift to capital-breeding in saturniids, as opposed to the income-breeding strategy of their sister group – the Sphingidae family - was an adaptive response in environments where adult moths are exposed to high predation pressure. Sphingid moths, in need of feeding to sustain egg production, are engaged into an "arm race" with bats through an arsenal of morphological and behavioral adaptations (Kristensen 2012, Kawahara & Barber 2015, Rubin, Hamilton et al. 2018). The split between the two lineages (ca. 60Ma; see also Kawahara et al. 2019) coincides with the origin and diversification of bats (Shi & Rabosky 2015, Lei & Dong 2016), and Saturniidae have reduced their exposure to predation through short adult activity and capitalization on resources accumulated during larval stages. Our results emphasize the importance of polyphagy and pupation mode in driving the diversification dynamics of saturniid moths over macro-evolutionary scale, and thus reveal the major role of key adaptive traits selected to cope with their innovative reproductive

strategy. Within Saturniidae, members of subfamily Saturniinae, which could be considered as super capital breeders because they combine large body size with the ability to exploit multiple food plants and to spin strong cocoons, are the only lineage that managed to colonize the world despite their poor flying ability. Thus, evolution towards optimized capital breeding transformed poor dispersers into successful colonizers, challenging all expectations.

MATERIAL AND METHODS

Sampling, library preparation and data processing. DNA was extracted from legs or thorax muscles of 254 ethanol preserved, dried in envelopes, or pinned specimens (Table S1) using the Qiagen DNeasy Blood and Tissue kit. Library preparation and raw data processing followed (Cruaud et al. 2019).

Phylogenomic analyses and Exploration of systematic bias. Only the 1,171 loci obtained in at least 50% of the samples were analysed. We used TreeShrink v1.3.1 (Mai & Mirarab 2018) to remove potential paralogs. Phylogenetic trees were inferred from a concatenated matrix using IQTREE v1.6.3 (Nguyen et al. 2014) and different partitioning strategies as well as a coalescence-based approach using ASTRAL-III v5.5.6 (Zhang et al. 2018). Data subsets were built to test for possible inference bias induced by high GC content or heterogeneity in evolutionary rates among taxa.

Divergence time estimates. We generated a time-calibrated tree with MCMCTree (Yang 1997), using 8 calibration points (Table S7) and the ASTRAL tree as input. To make computation tractable we compared the results obtained with 5 loci sets: i) 50 loci that produced the most supported gene trees ii) 50 loci that produced trees with the lowest Robinson-Fould (RF) distance as compared to the ASTRAL tree iii) 500bp-length loci with the lowest RF distance as compared to the ASTRAL tree iv and v) 2 sets of 50 random loci.

Species-level phylogenies inference. We generated a species level phylogeny using PASTIS (Thomas, Hartmann et al. 2013) and MrBayes (Ronquist et al. 2012) that includes all the 3,451 Saturniid species listed in (Kitching, Rougerie et al. 2018) by fixing every node of the phylogeny to match median ages estimated by MCMCtree and by constraining genera and subgenera to be monophyletic.

Biogeographical analyses. Species distributions were categorized into 11 major biogeographical areas: South America (SA); Central America (CA); Eastern Neartic (EN); Western Neartic (WN); Eastern Palaearctic (EP); Western Palaearctic (WN); Madagascar (MD); Africa (AF); India (IN); Malayan region (WA) and Australian region (AU). Cenozoic was divided into: Paleocene (66-56Ma), Eocene (56-34Ma), Oligocene (34-23Ma), early Miocene and Langhian (23-14Ma), Serravallian and late Miocene (14-5Ma) and Pliocene and Quaternary (5Ma to present). Adjacency and dispersal rate matrices are provided in the *SI*. Ancestral area reconstruction analyses were performed using DECX (Beeravolu & Condamine 2016) on 1000 trees sampled in the posterior distribution of the species-level analysis. Results obtained at each node of interest were averaged. Because the intra-generic phylogenetic

relationships were unknown, we randomized the positions of the species within genera in each of the species-level phylogeny replicate.

Diversification analyses. Analyses were performed with (i) BAMM v2.5 (Rabosky, Santini et al. 2013) to identify diversification rate shifts without a priori and (ii) the R-package *diversitree* v.0.9-8 (FitzJohn 2012) to simultaneously model trait evolution and its impact on diversification. We used the species-level phylogenetic tree from which genera without trait data were pruned and a few subfamily level trees for a deeper exploration of the influence of larval polyphagy on diversification rates (Bunaeinae, Ceratocampinae, Hemileucinae and Saturniinae).

Traits measurement, compilation and evolutionary analyses. Morphological traits (*body length*; *forewing length; thorax width; forewing surface)* were measured from images representing all existing genera but two using an image annotation tool in the Barcode of Life Datasystems (www.boldsystem.org; (Ratnasingham & Hebert 2007)). *body size* was defined as *thorax width* x *body length* and *wingload* was defined as *body size* / *forewing surface*. The phylogenetic diversity (Faith, 1992) of hostplants families consumed by congeneric species (calculated on the angiosperm phylogeny by Magallón et al. (2015) was used to described the level of larval polyphagy of the genus. Pupation mode was compiled from the literature. Ancestral values of *body size*; *wingload*; larval polyphagy and pupation modes were estimated using the *ace* function in the R package *ape* (Paradis & Schliep 2019).

Acknowledgment

This study was supported by funding from ANR to project SPHINX (ANR-16-CE02-0011-01) and from the French Foundation for Research on Biodiversity (FRB) and the synthesis center CESAB to project ACTIAS. We thank Leila Zekraoui and Cedric Mariac for the access the Bioruptor © (UMR DIADE - IRD), Helene Vigne and Audrey Weber (UMR AGAP - CIRAD) for the sequencing facilities and the genotoul bioinformatics platform Toulouse Occitanie for providing help and assistance for computing. We express all our gratitude to Carlos Mielke, Stefan Naumann, Ron Brechlin, Thierry Bouyer, Tomas Melichar, Hermann Staude, Alan Gardiner, Axel Hausmann, Frank Meister, Eric vanSchayck, Ulrich Paukstadt, Jurgen Vanhoudt, Patrick Basquin, Philippe Darge, Yves Estradel, Daniel Herbin, Frédéric Bénéluz and Jeremy Dickens and to all collaborators who sent us sample legs of the species studied here.

Data accessibility

Demultiplexed reads are available as a NCBI Sequence Read Archive (ID#XXX).

Supplementary Material

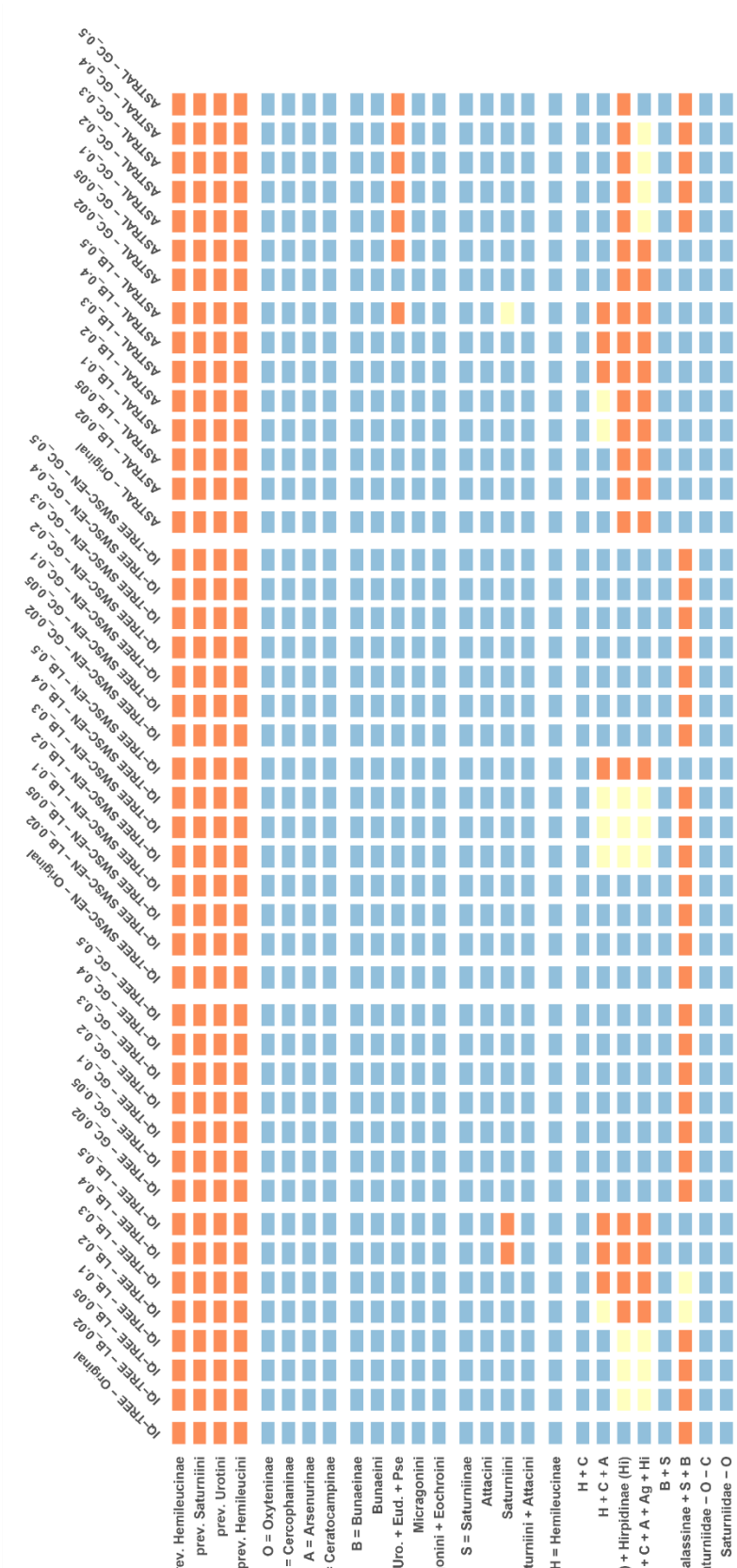
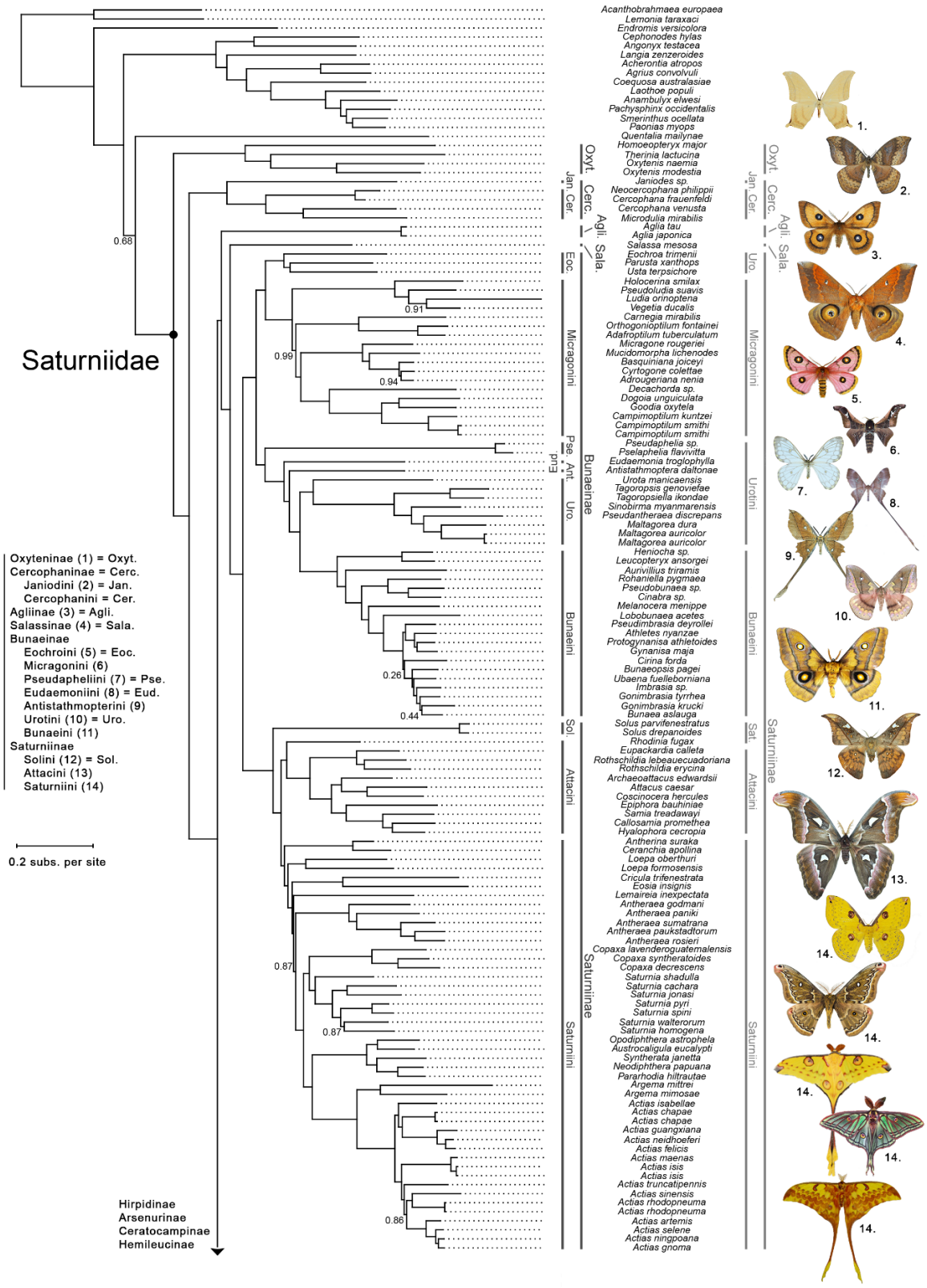


Figure S1 - Phylogenetic supports obtained with the different analyses (in column) for the main Saturniidae clades (in row). IQ-TREE or ASTRAL refer to the phylogenetic software used and SWSC-EN indicate when, for the IQ-TREE analyses, the Tagliacollo & Lanfear (2018) partitioning method was used. The results showed here were inferred from the original genomic dataset and the datasets in which 5% of *loci* were discarded based of their LB score (LB_X) or their GC composition (GC_X). Blue square indicate that the clade monophyly were strongly supported, yellow weakly supported and orange non recovered.

Figure S2 (Online resource – Figure S3 in Chapitre 1) - All inferred trees with IQ-TREE or ASTRAL from the different sub-datasets. Titles depicted the software used and the sub-dataset considered. Subfamilies, species names, sample codes and number of loci are indicated at tips.

Figure S6 (Figure 2 in Chapitre 1) – Genus-level phylogeny of the Saturniidae family inferred from the LB_0.05 dataset with IQTREE, constraining the topology to match the one obtained with ASTRAL. The presented supports are localPP, estimated with ASTRAL. localPP equal to 1 (resolved clade) are not represented. The scale bar (left side) represent the mean number of substitutions per site. Species names are indicated on the right side of the phylogeny, bordered by the new (left) and the old (right) Saturniidae classifications. Numbered photos refer to the different saturniid clades as indicated on the left side of the phylogeny.



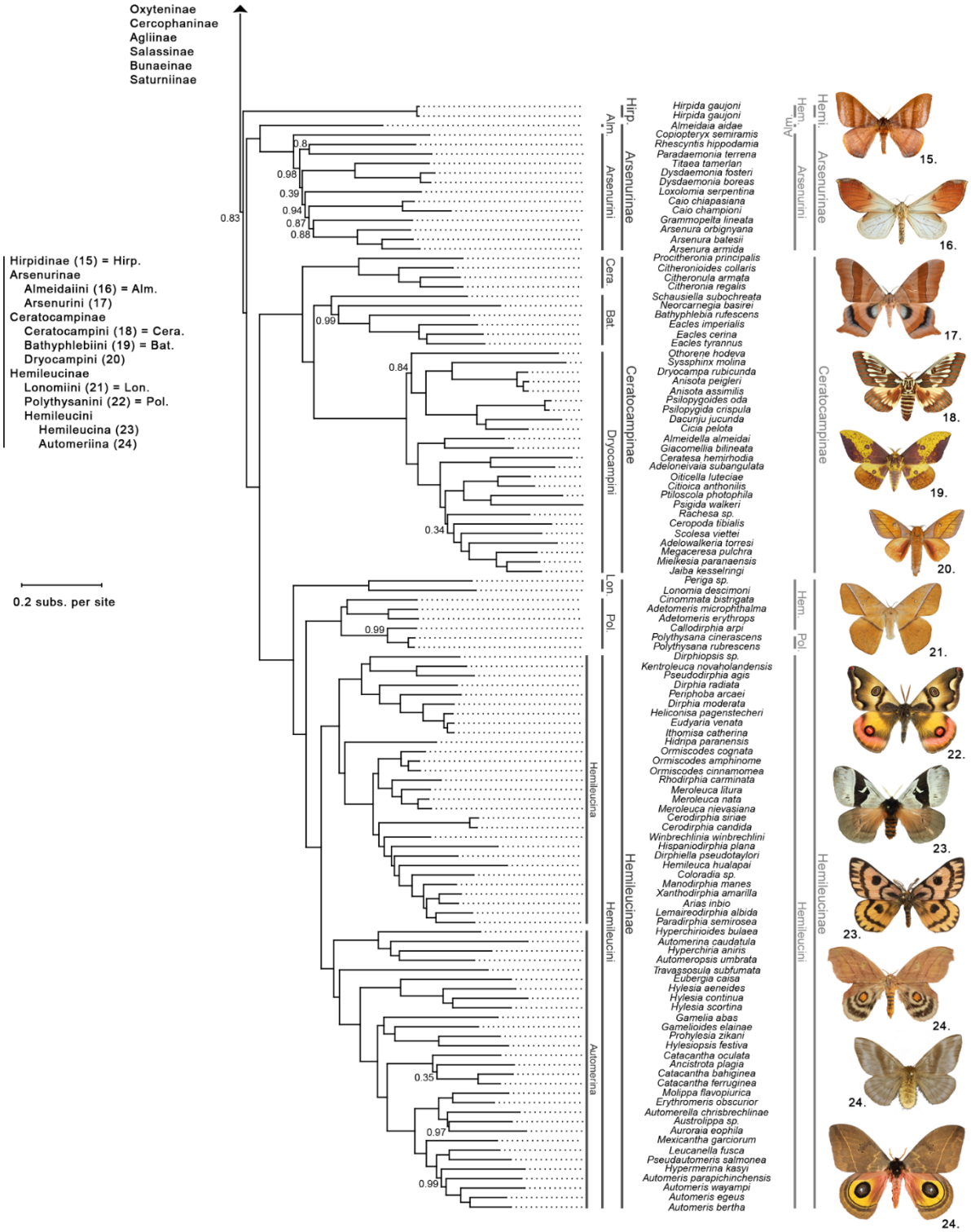
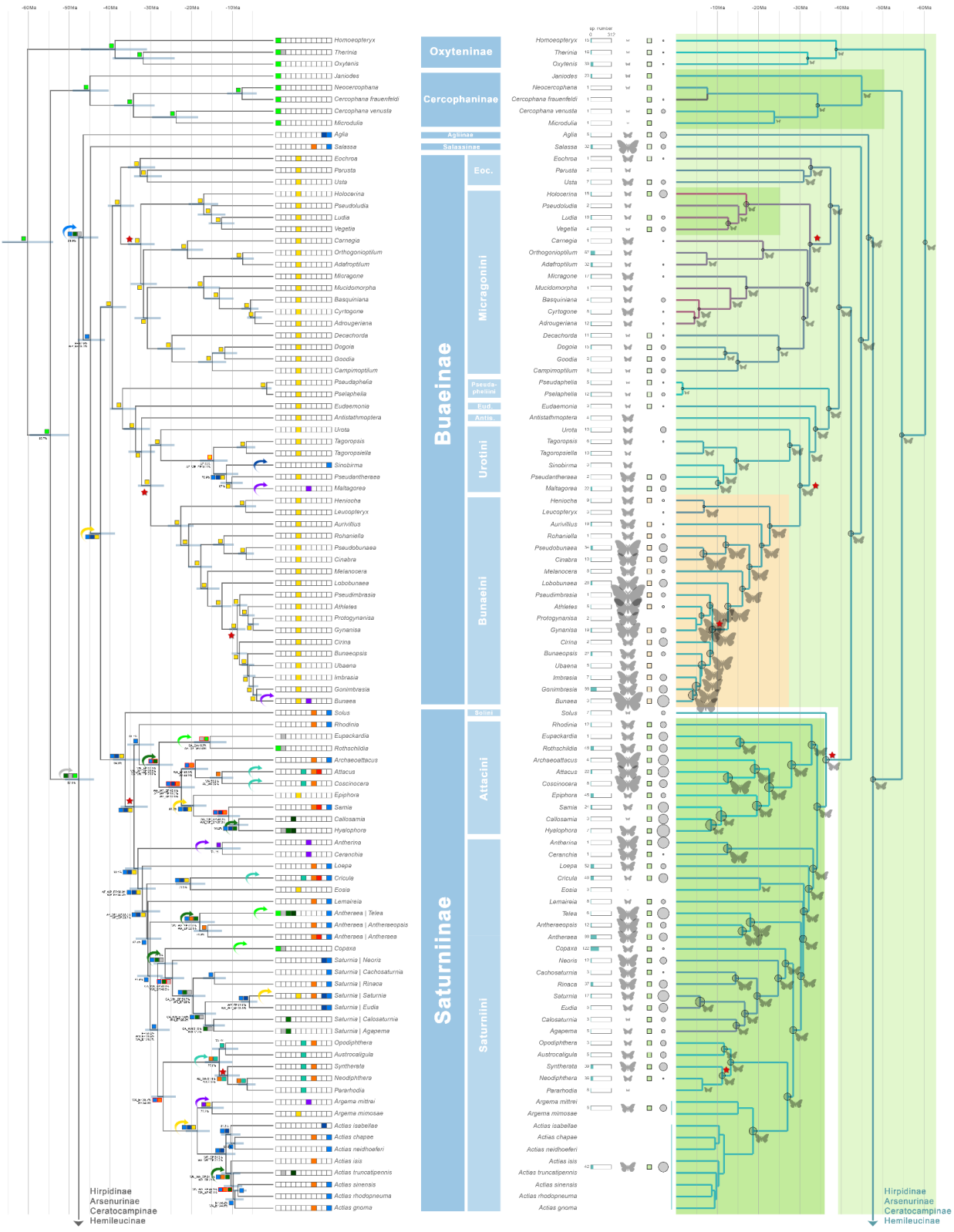
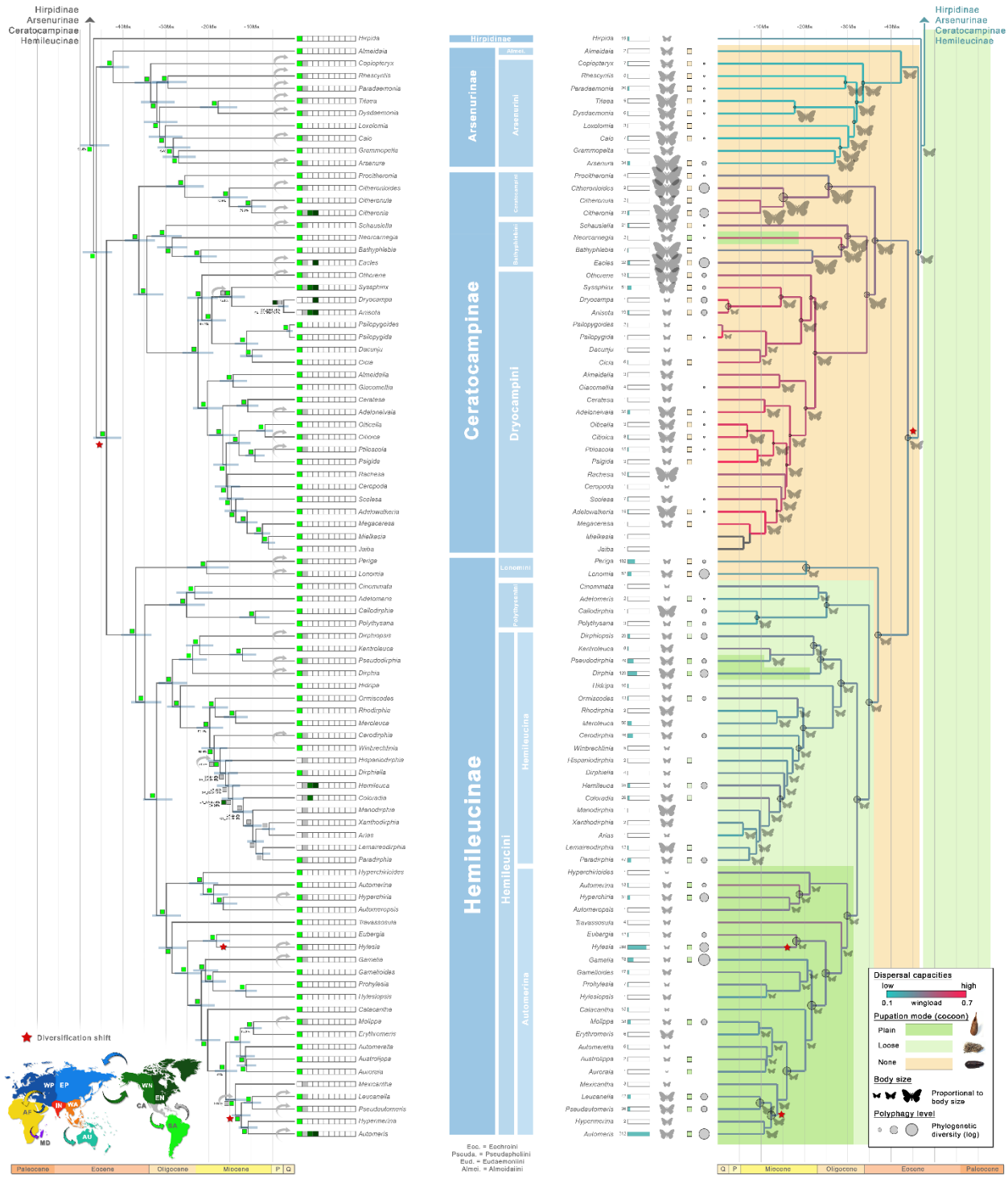

Oxyteninae
Cercophaninae
Agliinae
Salassinae
Bunaeinae
Saturniinae
Hirpidinae (15) = Hirp.
Arsenurinae
Almeidaiini (16) = Alm.
Arsenurini (17)
Ceratocampinae
Ceratocampini (18) = Cera.
Bathyphlebiini (19) = Bat.
Dryocampini (20)
Hemileucinae
Lonomiini (21) = Lon.
Polythysanini (22) = Pol.
Hemileucini
Hemileucina (23)
Automeriina (24)
0.2 subs. per site
0.83
0.8
0.98
0.39
0.94
0.87
0.88
0.99
0.84
0.34
0.99
0.35
0.97
0.99
Hirp.
Alm.
Arsenurini
Cera.
Bat.
Dryocampini
Lon.
Pol.
Hemileucina
Hemileucini
Automeriina
Arsenurinae
Ceratocampinae
Hemileucinae
Hemi.
Hirpida gaujoni
Hirpida gaujoni
Almeidaia aidae
Copiopteryx semiramis
Rhescyntis hippodamia
Paradaemonia terrena
Titaea tamerlan
Dysdaemonia fosteri
Dysdaemonia boreas
Loxolomia serpentina
Caio chiapasiana
Caio championi
Grammopelta lineata
Arsenura orbignyana
Arsenura batesii
Arsenura armida
Procitheronia principalis
Citheronioides collaris
Citheronula armata
Citheronia regalis
Schausiella subochreata
Neocarnegia basirei
Bathyphlebia rufescens
Eacles imperialis
Eacles cerina
Eacles tyrannus
Othorene hodeva
Sysphinx molina
Dryocampa rubicunda
Anisota peigleri
Anisota assimilis
Psilopygoides oda
Psilopygida crispula
Dacunju jucunda
Cicia pelota
Almeidella almeidai
Giacomellia bilineata
Ceratesa hemirhodia
Adelonelvaia subangulata
Oiticella luteciae
Citioica anthonilis
Ptiloscola photophila
Psigida walkeri
Rachesa sp.
Ceropoda tibialis
Scolesa viettei
Adelowalkeria torresi
Megaceresa pulchra
Mielkesia paranaensis
Jaiba kesselringi
Periga sp.
Lonomia descimoni
Cinommata bistrigata
Adetomeris microphthalma
Adetomeris erythrops
Callodirphia arpi
Polythysana cinerascens
Polythysana rubrescens
Dirphiopsis sp.
Kentroleuca novaholandensis
Pseudodirphia agis
Dirphia radiata
Periphoba arcaei
Dirphia moderata
Heliconisa pagenstecheri
Eudyaria venata
Ithomisa catherina
Hidripa paranensis
Ormiscodes cognata
Ormiscodes amphinome
Ormiscodes cinnamomea
Rhodirphia carminata
Meroleuca litura
Meroleuca nata
Meroleuca nievasiana
Cerodirphia siriae
Cerodirphia candida
Winbrechlinia winbrechlini
Hispaniodirphia plana
Dirphiella pseudotaylori
Hemileuca hualapai
Coloradia sp.
Manodirphia manes
Xanthodirphia amarilla
Arias inbio
Lemaireodirphia albida
Paradirphia semirosea
Hyperchirioides bulaea
Automerina caudatula
Hyperchiria aniris
Automeropsis umbrata
Travassosula subfumata
Eubergia caisa
Hylesia aeneides
Hylesia continua
Hylesia scortina
Gamelia abas
Gamelioides elainae
Prohylesia zikani
Hylesiopsis festiva
Catacantha oculata
Ancistrota plagia
Catacantha bahiginea
Catacantha ferruginea
Molippa flavopiurica
Erythromeris obscurior
Automerella chrisbrechlinae
Austrolippa sp.
Auroraia eophila
Mexicantha garciorum
Leucanella fusca
Pseudautomeris salmonea
Hypermerina kasyi
Automeris parapichinchensis
Automeris wayampi
Automeris egeus
Automeris bertha
15.
16.
17.
18.
19.
20.
21.
22.
23.
23.
24.
24.
24.

Figure S4 (two following page - Figure S5 in Chapitre 1) – Historical biogeography and traits evolution of Saturniidae. Red stars depict the diversification rate shifts as estimated with BAMM. Left panel: biogeographic history of Saturniidae as estimated with DECX on 1000 PASTIS, species-level trees. The best ranges, as estimated with DECX, are represented by colored squares. At nodes, red border was used when the confidence about the estimation was <50%. When confidence values were below 90%, alternative estimates were depicted. Arrows highlight the main dispersal events as represented in the World map at the left bottom. Blue bars indicate node age confidence intervals. Right panel: saturniid life traits evolution. Branches were colored according to wingload values. Background colors represent the pupation modes. Shaded silhouette size are proportional to body size (log-normalized). Polyphagy level is depicted at tips and nodes with circles which size is proportional to the estimated Polyphagy Diversity score (log-normalized). A .pdf file is available as Online resource.




Hirpidinae
Arsenurinae
Ceratocampinae
Hemileucinae
Diversification shift
Paleocene
Eocene
Oligocene
Miocene
Arsenurinae
Ceratocampinae
Hemileucinae
Dryocampini
Hemileucini
Automerina
Dispersal capacities
low
high
0.1
wingload
0.7
Pupation mode (cocoon)
Plain
Loose
None
Body size
Proportional to body size
Polyphagy level
Phylogenetic diversity (log)

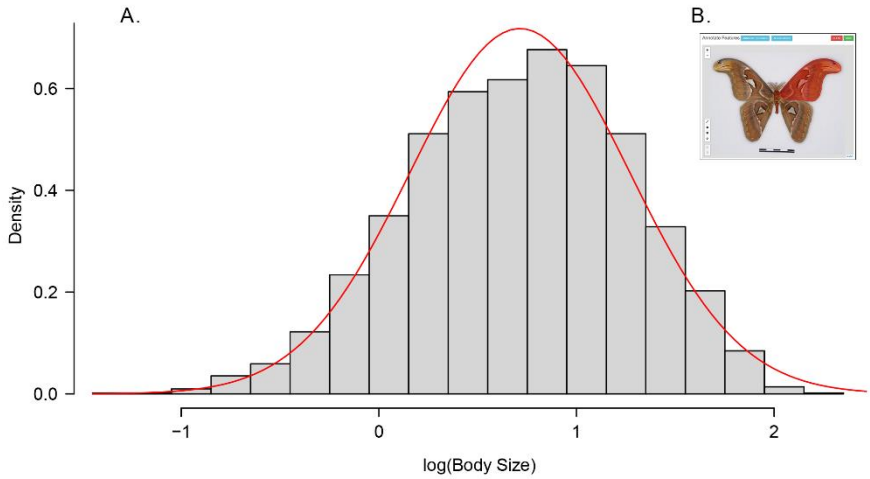


Figure S5 (Figure S6 in Chapitre 1) - (A) Histogram representing the log-transformed body size distribution of 2577 specimens, as estimated by the product of body length and the mean of the three measurements of thorax width. (B) Represents the tool used for measuring these features, as integrated in BOLD (www.boldsystems.org). The distribution is negatively skewed (one sided Agostino test, skew = -0.226, z = -4.594, p-value = 2.17e-06). Red line represents a normal distribution with similar mean and standard deviation.

Figure S6 (Online resource – Figure S7 in Chapitre 1) - Polyphagy-dependent diversification models - subfamily level. Within each major subfamily (one page each) of Saturniidae, we categorized the genera into three categories, specialist (category 1), oligophagous (2) and polyphagous (3), considering quantiles p=0.33 and p=0.66. Genus polyphagy is indicated at the crown node of each genus. On the left of each page, we represented the species-level phylogenies of each subfamily as inferred with PASTIS; the genera missing food plant information were discarded. We did not consider intra-generic variation because of the way we estimated polyphagy level. On the right, histograms represented speciation, extinction, diversification and transition rates as estimated over 10,000 MCMC generations with the diversitree R package.

Table S1 – Voucher information. For every sample, we indicated the number of reads, contigs, UCEs in total and UCEs in the genomic matrix used to infer the phylogenies.

Table S2 (Table S4 in Chapitre 1) - Saturniidae morphology measurements. Online Resource. 12850 measurements were done for 2577 saturniid specimens. The variables considered were (i) body length, (ii) thorax width between the junction points of thorax and forewings, (iii) thorax width at the middle of the thorax, (iv) thorax width between the junction points of thorax and hindwings and (v) the forewing surface. Measurements were performed on the BOLD platform (www.boldsystems.org) and rescaled thanks to a scale measurement.

Table S3 (Table S5 in Chapitre 1) - Parameter values from best models of trait-dependent diversification analyses. The full model was recovered as the best model for all life-history traits considered (see Table 1 in main text). The values represent medians of each parameter (λ = speciation rate, μ = extinction rate and λ-μ = diversification rate) as estimated with 10,000 MCMC generations (10% of burnin).

MCMC estimates		λ_1	μ_1	$\lambda_1 - \mu_1$	λ_2	μ_2	$\lambda_2 - \mu_2$	λ_3	μ_3
Body Size (BS)	cat1: small BS cat2: medium BS cat3: large BS	0.142	3.54e-03	0.138	0.164	0.011	0.152	0.177	3.70e-03
Dispersal capacities - Wingload (WL)	cat1: low WL cat2: medium WL cat3: high WL	0.117	3.49e-03	0.113	0.171	6.96e-03	0.163	0.184	2.79e-03
Polyphagy level (POL)	cat1: low POL cat2: medium POL cat3: high POL	0.127	0.0302	0.0964	0.148	2.77e-03	0.145	0.214	1.62e-03
Pupation mode	cat1: No cocoon cat2: loose cocoon cat3: plain cocoon	0.172	0.0169	0.154	0.128	0.0293	0.0984	0.178	1.59e-03

Table S4 (Online resource – Table S6 in Chapitre 1) – Food plant dataset (Ballesteros-Mejia et al. submitted, see attached articles).

Table S5 (Table 2 in Chapitre 1) – Results of the MuSSE analyses of polyphagy-dependent diversification models fitted to the phylogeny of the four most diverse subfamilies of wild silkmoths. For each subfamily, we compared three models: (i) a null model in which the diversification rates were homogenous across the phylogeny, (ii) a qfree model in which the transition rates between the different polyphagy levels were not fixed as equal (but $q_{13} \sim 0$, $q_{31} \sim 0$), (iii) a full model in which we estimated distinct diversification rates for the three polyphagy categories as well as distinct transition rates. In subfamilies Bunaeinae, Hemileucinae, and Saturniinae, the full model was ranked first according to AIC scores. In Ceratocampinae (Cerato.), polyphagy level did not better explain species diversity. ΔAIC indicates the AIC difference to the best model. Df is the number of degrees of freedom for each model. The evolution of traits and diversification rates throughout the saturniid evolution can be seen in Figure S6.

Polyphagy-dependent diversification models		Df	lnLik	AIC	ΔAIC
Bunaeinae	null model	3	-1203.7	2413	17
	qfree model	6	-1199.5	2411	15
	full model	10	-1183.1	2396	0
Cerato.	null model	3	-709.94	1426	1
	qfree model	6	-706.64	1245	0
	full model	10	-704.39	1429	3
Hemileuc.	null model	3	-3605.1	7216	26
	qfree model	6	-3596.7	7205	15
	full model	10	-3585.2	7190	0
Saturniinae	null model	3	-1897.7	3801	4
	qfree model	6	-1896.6	3805	8
	full model	10	-1888.5	3797	0

Table S6 (Online resource – Table S7 in Chapitre 1) - Genus-level pupation mode dataset.

Table S7 – Calibrations points used in the MCMCtree datation analyses.

Secondary calibrations		
Node	calibration (Ma)	Reference
Sphingidae / Saturniidae	[64.1;84.7]	Walhberg et al. 2013
Crown Sphingidae	[26.6;55.6]	Walhberg et al. 2013
Crown Saturniidae	[48.5;69.9]	Walhberg et al. 2013
Saturniinae/Cercophaninae	[43.0;64.1]	Walhberg et al. 2013
Ceratocampinae/Hemileucinae	[26.9;49.6]	Walhberg et al. 2013
Salassinae/Saturniinae	[23.9;45.5]	Walhberg et al. 2013

Fossils		
Node	calibration (Ma)	Reference
Crown Smerinthini	>15.2	Zhang & Zhang 1994
Crown Bunaeini	>3,66Ma	Kitching & Sadler 2011

REFERENCES

Antonelli, A., A. Zizka, D. Silvestro, R. Scharn, B. Cascales-Minana and C. D. Bacon (2015). "An engine for global plant diversit y: highest evolutionary turnover and emigration in the American tropics." Front Genet **6**: 130.

Baskin, J. M. and C. C. Baskin (2016). "Origins and Relationships of the Mixed Mesophytic Forest of Oregon–Idaho, China, and Kentucky: Review and Synthesis1." Annals of the Missouri Botanical Garden **101**(3): 525-552.

Beeravolu, C. R. and F. L. Condamine (2016). "An Extended Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction and Cladogenesis." bioRxiv: 038695.

Bonetti, M. F. and J. J. Wiens (2014). "Evolution of climatic niche specialization: a phylogenetic analysis in amphibians." Proceedings of the Royal Society B: Biological Sciences **281**(1795).

Cerling, T. E., J. M. Harris, B. J. MacFadden, M. G. Leakey, J. Quade, V. Eisenmann and J. R. Ehleringer (1997). "Global vegetation change through the Miocene/Pliocene boundary." Nature **389**(6647): 153-158.

Claramunt, S. and J. Cracraft (2015). "A new time tree reveals Earth history's imprint on the evolution of modern birds." Science advances. **1**(11): e1501005.

Clarke, A. R. (2017). "Why so many polyphagous fruit flies (Diptera: Tephritidae)? A further contribution to the 'generalism' debate." Biological Journal of the Linnean Society **120**: 245-257.

Condamine, F. L., J. Rolland, S. Hohna, F. A. H. Sperling and I. Sanmartin (2018). "Testing the Role of the Red Queen and Court Jester as Drivers of the Macroevolution of Apollo Butterflies." Syst Biol **67**(6): 940-964.

Cooney, C. R., J. A. Bright, E. J. Capp, A. M. Chira, E. C. Hughes, C. J. Moody, L. O. Nouri, Z. K. Varley and G. H. Thomas (2017). "Mega-evolutionary dynamics of the adaptive radiation of birds." Nature.

Cozzarolo, C.-S., M. Balke, S. Buerki, N. Arrigo, C. Pitteloud, M. Gueuning, N. Salamin, M. Sartori and N. Alvarez (2019). "Biogeography and Ecological Diversification of a Mayfly Clade in New Guinea." Frontiers in Ecology and Evolution **7**: 223.

Cruaud, A., S. Nidelet, P. Arnal, A. Weber, L. Fusu, A. Gumovsky, J. T. Huber, A. Polaszek and J. Y. Rasplus (2019). "Optimized DNA extraction and library preparation for minute arthropods: Application to target enrichment in chalcid wasps used for biocontrol." Molecular Ecology Resources **19**(3): 702-710.

Danks, H. V. (2004). "The roles of insect cocoons in cold conditions." European Journal of Entomology **101**(3): 433-437.

Davis, R. B., J. Javoiš, A. Kaasik, E. Õunap, and T. Tammaru (2016). "An ordination of life histories using morphological proxies: capital vs. income breeding in insects." Ecology **97**(8): 2112-2124.

Dorey, J. B., S. V. C. Groom, E. H. Freedman, C. S. Matthews, O. K. Davies, E. J. Deans, C. Rebola, M. I.

Economo, E. P., N. Narula, N. R. Friedman, M. D. Weiser and B. Guenard (2018). "Macroecology and macroevolution of the latitudinal diversity gradient in ants." Nat Commun **9**(1): 1778.

Economo, E. P., J.-P. Huang, G. Fischer, E. M. Sarnat, N. Narula, M. Janda, B. Guénard, J. T. Longino, L. L. Knowles and I. Simova (2019). "Evolution of the latitudinal diversity gradient in the hyperdiverse ant genus Pheidole." Global Ecology and Biogeography **28**(4): 456-470.

Faurby, S. and A. Antonelli (2018). "Evolutionary and ecological success is decoupled in mammals." Journal of Biogeography **45**(10): 2227-2237.

FitzJohn, R. G. (2012). "Diversitree: comparative phylogenetic analyses of diversification in R." Methods in Ecology and Evolution **3**(6): 1084-1092.

Fiz-Palacios, O., H. Schneider, J. Heinrichs and V. Savolainen (2011). "Diversification of land plants: insights from a family-level phylogenetic analysis." BMC Evolutionary Biology **11**: 341.

Garzione, C. N., D. J. Auerbach, J. Jin-Sook Smith, J. J. Rosario, B. H. Passey, T. E. Jordan and J. M. Eiler (2014). "Clumped isotope evidence for diachronous surface cooling of the Altiplano and pulsed surface uplift of the Central Andes." Earth and Planetary Science Letters **393**: 173-181.

Garzione, C. N., G. D. Hoke, J. C. Libarkin, S. Withers, B. MacFadden, J. Eiler, P. Ghosh and A. Mulch (2008). "Rise of the Andes." Science **320**(5881): 1304.

Gross, P. (1993). "Insect Behavioral and Morphological Defenses Against Parasitoids." Annual Review of Entomology **38**(1): 251-273.
Hoorn, C., F. P. Wesselingh, H. ter Steege, M. a. Bermudez, a. Mora, J. Sevink, I. Sanmartín, a. Sanchez-Meseguer, C. L. Anderson, J. P. Figueiredo, C. Jaramillo, D. Riff, F. R. Negri, H. Hooghiemstra, J. Lundberg, T. Stadler, T. Särkinen and a. Antonelli (2010). "Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity." Science **330**(6006): 927-931.
Hortal, J., F. de Bello, J. A. F. Diniz-Filho, T. M. Lewinsohn, J. M. Lobo and R. J. Ladle (2015). "Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity." Annual Review of Ecology, Evolution, and Systematics **46**(1): 523-549.
Hutchinson, G. E., and R. H. MacArthur (1959). "A theoretical ecological model of size distributions among species of animals." The American Naturalist **93**(869): 117-125.
Iturralde-Vinent, M. A. and R. D. MacPhee (1999). "Paleogeography of the Caribbean Region: Implications for Cenozoic biogeography." Bulletin of the American Museum of Natural History **238**: 1-95.
Ivany, L. C., W. P. Patterson and K. C. Lohmann (2000). "Cooler winters as a possible cause of mass extinctions at the Eocene/Oligocene boundary." Nature **407**(6806): 887-890.
Jahren, A. H. (2007). "The Arctic Forest of the Middle Eocene." Annual Review of Earth and Planetary Sciences **35**(1): 509-540.
Janzen, D. H. (1984). "Two ways to be a tropical big moth: Santa Rosa saturniids and sphingids." Oxford Surveys in Evolutionary Biology **1**: 85-140.
Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann and A. O. Mooers (2012). "The global diversity of birds in space and time." Nature **491**(7424): 444-448.
Kawahara, A. Y. and J. R. Barber (2015). "Tempo and mode of antibat ultrasound production and sonar jamming in the diverse hawkmoth radiation." Proceedings of the National Academy of Sciences **112**(20): 6407-6412.
Kawahara, A. Y., D. Plotkin, M. Espeland, K. Meusemann, E. F. A. Toussaint, A. Donath, F. Gimnich, P. B. Frandsen, A. Zwick, M. Dos Reis, J. R. Barber, R. S. Peters, S. Liu, X. Zhou, C. Mayer, L. Podsiadlowski, C. Storer, J. E. Yack, B. Misof and J. W. Breinholt (2019). "Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths." Proc Natl Acad Sci U S A **116**(45): 22657-22663.
Kitching, I., R. Rougerie, A. Zwick, C. Hamilton, R. St Laurent, S. Naumann, L. Ballesteros Mejia and A. Kawahara (2018). "A global checklist of the Bombycoidea (Insecta: Lepidoptera)." Biodiversity Data Journal **6**: e22236.
Kristensen, N. P. (2012). "Molecular phylogenies, morphological homologies and the evolution of moth 'ears'." Systematic Entomology **37**(2): 237-239.
Lei, M. and D. Dong (2016). "Phylogenomic analyses of bat subordinal relationships based on transcriptome data." Scientific reports **6**: 27726-27726.
Li, W., Z. Zhang, L. Lin and O. Terenius (2017). "Antheraea pernyi (Lepidoptera: Saturniidae) and its importance in sericulture, food consumption, and traditional Chinese medicine. " Journal of Economic Entomology **110**(4): 1404-1411.
Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes and T. Hernández-Hernández (2015). "A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity." New Phytologist **207**(2): 437-453.
Mai, U. and S. Mirarab (2018). "TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees." BMC Genomics **19**(5): 272.
Matos-Maravi, P., R. M. Clouse, E. M. Sarnat, E. P. Economo, J. S. LaPolla, M. Borovanska, C. Rabeling, J. Czekanski-Moir, F. Latumahina, E. O. Wilson and M. Janda (2018). "An ant genus-group (Prenolepis) illuminates the biogeography and drivers of insect diversification in the Indo-Pacific." Mol Phylogenet Evol **123**: 16-25.
Maurer, B. A. (1998). "The evolution of body size in birds. I. Evidence for non-random diversification". Evolutionary Ecology **12**(8): 925.
Misof, B. (2002). "Diversity of Anisoptera (Odonata): Infering speciation processes from patterns of morphological diversity." Zoology **105**(4): 355-365.

Modica, M. V., J. Gorson, A. E. Fedosov, G. Malcolm, Y. Terryn, N. Puillandre and M. Holford (2020). "Macroevolutionary Analyses Suggest That Environmental Factors, Not Venom Apparatus, Play Key Role in Terebridae Marine Snail Diversification." Systematic Biology **69**(3): 413-430.
Morley, R. J. (2003). "Interplate dispersal paths for megathermal angiosperms." Perspectives in Plant Ecology, Evolution and Systematics **6**(1-2): 5-20.
Nguyen, L.-T., H. A. Schmidt, A. von Haeseler and B. Q. Minh (2014). "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." Molecular Biology and Evolution **32**(1): 268-274.
Paradis, E. and K. Schliep (2019). "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R." Bioinformatics **35**(3): 526-528.
Pindell, J. L. and L. Kennan (2009). "Tectonic evolution of the Gulf of Mexico, Caribbean and northern South America in the mantle reference frame: an update." Geological Society, London, Special Publications **328**(1): 1-55.
Price, S. A., S. S. B. Hopkins, K. K. Smith and V. L. Roth (2012). "Tempo of trophic evolution and its impact on mammalian diversification." Proceedings of the National Academy of Sciences **109**(18): 7008-7012.
Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang and M. E. Alfaro (2013). "Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation." Nature communications **4**(1): 1-8.
Rainford, J. L., M. Hofreiter and P. J. Mayhew (2016). "Phylogenetic analyses suggest that diversification and body size evolution are independent in insects." BMC Evolutionary Biology **16**(1): 8.
Ratnasingham, S. and P. D. N. Hebert (2007). "BOLD: The Barcode of Life Data System (http://www.barcodinglife.org)." Molecular Ecology Notes **7**(3): 355-364.
Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck (2012). "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space." Systematic biology **61**(3): 539-542.
Rougerie, R., S. Naumann and W. A. Nässig (2012). "Morphology and molecules reveal unexpected cryptic diversity in the enigmatic genus *Sinobirma* Bryk, 1944 (Lepidoptera: Saturniidae)." PLoS ONE **7**(9): e43920.
Rubin, J. J., C. A. Hamilton, C. J. W. McClure, B. A. Chadwell, A. Y. Kawahara and J. R. Barber (2018). "The evolution of anti-bat sensory illusions in moths." Science Advances **4**: eaar7428.
Ségalen, L., J. A. Lee-Thorp and T. Cerling (2007). "Timing of C4 grass expansion across sub-Saharan Africa." Journal of Human Evolution **53**(5): 549-559.
Seiffert, E. R. (2007). "Evolution and Extinction of Afro-Arabian Primates Near the Eocene-Oligocene Boundary." Folia Primatologica **78**(5-6): 314-327.
Shi, J. J. and D. L. Rabosky (2015). "Speciation dynamics during the global radiation of extant bats." Evolution **69**(6): 1528-1545.
Silvestro, D., B. Cascales-Minana, C. D. Bacon and A. Antonelli (2015). "Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record." New Phytol **207**(2): 425-436.
Singer, M. S. (2008). "Evolutionary ecology of polyphagy", p. 29-42. In: K.J. TILMON (Ed). Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects. Berkeley, University of California Press.
Smith, F. A., J. L. Payne, N. A. Heim, M. A. Balk, S. Finnegan, M. Kowalewski, S. K. Lyons, C. R. McClain, D. W. McShea, P. M. Novack-Gottshall, P. S. Anich and S. C. Wang (2016). "Body Size Evolution Across the Geozoic." Annual Review of Earth and Planetary Sciences **44**(1): 523-553.
Stephens, P. A., I. L. Boyd, J. M. McNamara and A. I. Houston (2009). "Capital breeding and income breeding: their meaning, measurement, and worth." Ecology **90**(8): 2057-2067.
Tammaru, T., and E. Haukioja (1996). "Capital breeders and income breeders among Lepidoptera: consequences to population dynamics." Oikos **77**: 561-564.

Thomas, G. H., K. Hartmann, W. Jetz, J. B. Joy, A. Mimoto and A. O. Mooers (2013). "PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences." Methods in Ecology and Evolution **4**(11): 1011-1017.
Tiffney, B. H. (1985). "Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America." Journal of the Arnold Arboretum **66**(1): 73-94.
Willis, C. G., J. C. Hall, R. Rubio de Casas, T. Y. Wang and K. Donohue (2014). "Diversification and the evolution of dispersal ability in the tribe Brassiceae (Brassicaceae)." Annals of Botany **114**(8): 1675-1686.
Yang, Z. (1997). "PAML: a program for package for phylogenetic analysis by maximum likelihood." CABIOS **15**: 555-556.
Zachos, J., M. Pagani, L. Sloan, E. Thomas and K. Billups (2001). "Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present." Science **292**: 686-693.
Zachos, J. C., G. R. Dickens and R. E. Zeebe (2008). "An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics." Nature **451**: 279-283.
Zhang, C., M. Rabiee, E. Sayyari and S. Mirarab (2018). "ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees." BMC Bioinformatics **19**(6): 153.
Zhang, Z., G. Ramstein, M. Schuster, C. Li, C. Contoux and Q. Yan (2014). "Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene." Nature **513**(7518): 401-404.

Submitted in Biodiversity Data Journal

Biodiversity Data Journal : Data Paper (Biosciences)

Print

A global food plant dataset for wild silkmoths and hawkmoths, and its use in documenting polyphagy of their caterpillars (Lepidoptera: Bombycoidea: Saturniidae, Sphingidae)

Liliana Ballesteros Mejia[‡,§,|], Pierre Arnal[‡], Winnie Hallwachs[¶], Jean Haxaire[#,¤], Daniel Janzen[«], Ian J. Kitching[»], Rodolphe Rougerie[‡]

‡ Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS,Sorbonne Université, EPHE, Université des Antilles (1st authorship shared between the first two authors), Paris, France

§ CESAB, Centre de Synthèse et d'Analyse sur la Biodiversité, Montpellier, France

| Ecologie, Systématique and Evolution, Université Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France

¶ Department of Biology, University of Pennsylvania, Philadelphia, United States of America

Correspondent of Muséum national d'Histoire naturelle, Paris, France

¤ Associate researcher of Insectarium de Montréal, Quebec, Canada

« University of Pennsylvania, Philadelphia, United States of America

» Department of Life Sciences, Natural History Museum, London, United Kingdom

Corresponding author: Rodolphe Rougerie (rodolphe.rougerie@mnhn.fr)

© Liliana Ballesteros Mejia, Pierre Arnal, Winnie Hallwachs, Jean Haxaire, Daniel Janzen, Ian Kitching, Rodolphe Rougerie

OPEN ACCESS

Citation:

Abstract

Background

Herbivorous insects represent a major fraction of global biodiversity and the relationships they have established with their food plants range from strict specialists to broad generalists. Our knowledge of these relationships is of primary importance to basic (e.g. the study of insect ecology and evolution) and applied biology (e.g. monitoring of pest or invasive species), and yet remains very fragmentary and understudied. In Lepidoptera, caterpillars of families Saturniidae and Sphingidae are rather well known and considered to have adopted contrasting preferences in their use of food plants. The former are regarded as being rather generalist feeders, whereas the latter are more specialist.

New information

To assemble and synthesize the vast amount of existing data on food plants of Lepidoptera families Saturniidae and Sphingidae, we combined three major existing databases to produce a dataset collating more than 26,000 records for 1256 species (25% of all species) in 121 (67%) and 167 (81%) genera of Saturniidae and Sphingidae, respectively. This dataset is used here to document the level of polyphagy of each of these genera using summary statistics as well as the calculation of a polyphagy score derived from the analysis of Phylogenetic Diversity of the food plants used by the species in each genus.

Keywords

Lepidoptera, food plant, ecology, life-history traits, caterpillar, polyphagy

Introduction

Herbivorous insects represent a major fraction of global biodiversity (Fiedler 1998) and are central to studies of numerous and diverse ecological and evolutionary processes, such as resource specialization (Devictor et al. 2008), coevolution (Thompson 1988) and food web dynamics (Vidal and Murphy 2017). Elucidating the degree of food plant-insect specificty helps understand community assembly, ecosystem dynamics and latitudinal gradients of species richness (Ødegaard 2006). Moreover, insect-plant interactions are central to the understanding of niche breadth and they play a key role in mediating competition that structures communities, and backdrop the human view of entire networks of interacting species (Devictor et al. 2008, Forister et al. 2014). The different levels of specialization observed in phytophagous insects, from strict specialists to highly generalist species, are traits that are also considered as possibly important drivers of speciation or adaptive radiation (Janz and Nylin 2008, Jousselin and Elias 2019, Wang et al. 2017).

The Lepidoptera families Saturniidae (wild silkmoths) and Sphingidae (hawkmoths, sphinx moths) are among the best-known insect families worldwide, both taxonomically and biologically, and they are generally characterized by being large-bodied moths (Janzen 1984a). A recently published taxonomic checklist (Kitching et al. 2018) revealed a combined species richness of around 5000 species globally. These two families exhibit contrasting life-history strategies both as adults - Sphingidae (feeding, long-lived adults) and Saturniidae (non-feeding, short-lived adults) (Janzen 1984a) - and as caterpillars - Sphingidae (fast growing, many toxic plant specialists) and Saturniidae (slow-growing, many tannin and resin-rich plant specialists) (Janzen 1981, Janzen 1984a). In the Neotropics sphingid caterpillars seem to specialize on only a relatively small number of plant families, feeding on both young and old, relatively tender leaves that contain low molecular weight toxic compounds, whereas saturniid caterpillars feed on tougher, as well as younger, leaves of an often wide range of plant families that contain high levels of large polymeric molecules (tannins, resins) that interfere with digestion (Janzen 1981). Consequently, sphingid caterpillars digest more nutrients per bite and need less time to reach a given full size than do saturniids (Bernays and Janzen 1988).

A massive amount of data is available on the larval food plants in the wild of the two families, both in the literature and in institutional and personal databases. For the Lepidoptera as a whole, the HOSTS database (Robinson et al. 2010b) comprises the most comprehensive collation of information about what caterpillars overall are believed to eat. It contains some 180,000 records for about 22,000 Lepidoptera species extracted from 1600 documents (Robinson et al. 2010b). Although HOSTS has

not been updated for almost a decade, the subset of records for the superfamily Bombycoidea has been independently maintained and added to by IJK and this updated version is used here. Another spectacular effort towards gathering food plant data for Lepidoptera is the inventory of caterpillars in the Area de Conservacion Guanascate (ACG) in northwestern Costa Rica (Janzen and Hallwachs 2016, Janzen and Hallwachs 2020). It comprises ~70,000 records of reared wild-caught larvae of Saturniidae and Sphingidae linked to their DNA barcodes. Besides these two main public data repositories, one of the authors (JH) has built his own personal database for Sphingidae over 20 years, compiling records from the literature, web resources, personal field observations and communications from collaborators. In addition, food plant information is also scattered across the published literature, including a few more recent food plant catalogues, such as in Stone (1994), Santin (2004), Meister (2011), but also webpages and personal databases, all of which makes the process of collating and resolving the information very difficult and time consuming.

All three databases cited above are and remain independently maintained and updated. Here we publish a single dataset resulting from their combination. Our aim is to make this massive amount of information available as a single dataset that allows its use for ecological and evolutionary analyses. In particular, we want to investigate the role of food plant use in the evolution of the two families (Arnal et al., in prep.), especially with respect to the degree of polyphagy, defined as the plasticity in the use of different food plants for caterpillars to complete their development. We provide further details about the contents of this dataset in the following sections, as well as a number of caveats to avoid incorrect interpretation and use of these data. In addition to variables summarizing the level of polyphagy of the caterpillars of sphingid and saturniid moths, we also provide a polyphagy score based on a calculation of Phylogenetic Diversity (Faith 1992) of the food plant families used by the species included in the database.

General description

Purpose: The food plant dataset

This dataset (Suppl. material 1) is a synthesis of current knowledge regarding the food plants eaten by the caterpillars of two families of Lepidoptera (Saturniidae and Sphingidae). It aims to capture the state of knowledge at the time of assembly of the dataset so that it can be used to investigate the role of food plants use breadth in the spatial and temporal evolution of both families (Arnal et al., in prep.).

Table 1. Download as CSV XLSX

Table 1. General overview of the contribution of each database to our dataset Suppl. material 1

Database	Family	Number of records	Geographical coverage
HOSTS	Saturniidae	10586	Worldwide
	Sphingidae	10528	Worldwide
DHJ	Saturniidae	2297	Local in three adjoining ecosystems
	Sphingidae	1322	Local in three adjoining

			ecosystems
JH	Sphingidae	2401	Worldwide

This dataset of larval food plant records for sphingids and saturniids worldwide is the result of the integration, with significant data reconciliation and standardization, of these three largely independent data sources:

1) Information for Sphingidae and Saturniidae embedded in the HOSTS database (Robinson et al. 2010b); as further added to and refined by IJK, downloaded on March 2nd, 2018) (hereafter HOSTS);

2) An inventory of the caterpillars, their food plants and parasitoids of Area de Conservacion Guanacaste (ACG, Janzen DH, downloaded on July 16th 2018 for Saturniidae and July 18th 2018 for Sphingidae) (hereafter DHJ);

3) The personal database of Jean Haxaire (Associate Researcher to MNHN, imported on July 17th 2018) (hereafter JH).

A "record" refers to a unique combination of caterpillar species, plant species and source. Records in the dataset resulting from rearing experiments in captivity or from introduced plant species are listed separately as they often do not represent natural insect-plant associations. Redundancy (duplication) of records among the three databases following their combination was not a concern for our research objectives; the dataset should be treated as qualitative, and the frequency of records ignored (see list of points in next section).

A total of 25,937 records was compiled from the three databases in a single dataset given as Suppl. material 1. Table 1 below provides details of the number of records contributed by each of the independent databases. We followed the plant taxonomy of the International Plant Names Index (IPNI; https://www.ipni.org) and the latest moth taxonomy (Kitching et al. 2018), though both do not coincide with some of the names used by all three sources (see 'call for caution' below).

This compilation provides information for 137 genera and 757 species of Saturniidae and 166 genera and 725 species of Sphingidae.

As an example of the uses of this dataset, we report basic polyphagy variables as well as a polyphagy score based on Phylogenetic Diversity (PD, Faith 1992) of the food plants used by the caterpillars of saturniid and sphingid moths. Using a recent dated angiosperm phylogeny (Magallón et al. 2015) we measured the PD score, i.e. the total length of all phylogenetic tree branches connecting the different families of plants eaten by a given moth species *in natura*, using the *pd* function of the *picante* R package (Kembel et al. 2010). The species scores were then averaged within each genus to get genus scores in Suppl. material 2. Note that gymnosperm records were excluded from our calculations of PD scores to avoid bias caused by the considerable phylogenetic distance between angiosperms and gymnosperms.

The genus-level polyphagy variables and the polyphagy scores of Saturniidae and Sphingidae genera are provided as Suppl. material 2.

Additional information: Calls for caution:

1. The correctness of food plant identifications in databases and in the literature should be treated with considerable caution, as they were largely made by non-botanists; food plant names used

are also subject to taxonomic and nomenclatural uncertainty, and their correctness and validity may be considered equivocal in some cases.

2. The previous point also applies to moth names, especially when considering species-level identifications. These may be incorrect or outdated. For example, more than 1500 new species have been described within family Saturniidae in the past decade (Kitching et al. 2018), largely with the support of DNA barcoding analyses. Thus, food plant records may not account for recently split complexes of cryptic species, members of which may have quite different natural histories (e.g. Janzen (2012)).
3. The food plant dataset is derived from known food plant records at the time of its compilation; as such it represents a snapshot of the knowledge at that time and it may differ from the data compiled in the original sources and then updated independently (e.g. new records and/or corrections (e.g. identification errors or synonymies of the moth/caterpillar, or the plant, or both)).
4. All records are meant to represent actual instances in which caterpillars were found feeding and developing on the food plant. Records in the DHJ database all result from rearing trials of caterpillars found in the field on the food plant in question, and in many cases, identification of the caterpillar was confirmed through DNA barcoding of the resultant adult moths. A few records recognized as questionable (e.g. inconsistent locality/identification data) in the HOSTS and JH databases were filtered out and are not included in the present combined dataset.
5. The food plant dataset does not account for the frequency of use of a given food plant among other plants also listed for the same species of moth. The DHJ database, because it is based on individual specimen records, does include quantitative data; however, this information is not incorporated into the combined dataset, although it could bring additional information on local food plant preferences of species and populations. We note that this information would nevertheless be very difficult to analyze and interpret as it is conditional upon the local availability of food plants, as well as possibly seasonal conditions, local variations through time and difficulty of collecting.
6. The previous point also brings a note of caution in that polyphagy, as calculated here from the data available for a given species, may not be translatable to the population or site level, and vice versa. A species may have populations in which some caterpillars have a lower level of polyphagy than others, at least in part because the food plants that could be eaten do not occur in that ecosystem and because many species arrive by ecological fitting rather than *in situ* evolution (Janzen 1985a). This is especially the case with species following expanding frontier agriculture into new ecosystems, or following contemporary climate changes.
7. Strictly speaking, we define polyphagy as the capacity of a given individual caterpillar to feed and develop (through its complete life cycle) on different food plants. This can only be approximated by considering sibling individuals (as is sometimes the case in DHJ database), individuals from the same population, or, ultimately from the same species or higher taxonomic categories. We thus acknowledge that the scores of polyphagy at species and genus level should be recognized as human abstractions.
8. Polyphagy is constrained *in situ* by the local availability of food plants - an individual caterpillar cannot be polyphagous on species of plants that are not present.
9. Here we approximated polyphagy scores at species level for saturniid and sphingid moths, and we assume that they represent valuable information about the level of plasticity of individuals of the populations of a species to use different food plants. These scores were then used to calculate polyphagy scores at the genus level. Generic level of polyphagy is a human

abstraction, but it is seen as relevant information to understand the past diversification dynamics. Plasticity in the use of food plants may have favored or impeded geographical dispersal, and may have mitigated speciation or extinction processes, or influenced species natural histories in many other ways (Janzen 1985b).

10. We acknowledge that the polyphagy level derived from caterpillar plant feeding records approximates, but may not reflect precisely, the plasticity in oviposition site selection by female moths (see for instance Janzen 1984b). Indeed, caterpillars may be driven by starvation to feed on a different plant after consuming all leaves of the plant they started to develop on and which had been selected for oviposition by the female.

Geographic coverage

Description: The present dataset combines food plant records for saturniid and sphingid species worldwide.

Taxonomic coverage

Taxa included:

Rank	Scientific Name	Common Name
family	Saturniidae	Wild silkmoths
family	Sphingidae	Hawkmoths

Usage rights

Use license: Open Data Commons Attribution License

Data resources

Data package title: Global food plant dataset and polyphagy scores for Sphingidae and Saturniidae

Number of data sets: 2

Data set name: Global food plant dataset for Saturniidae and Sphingidae species

Data format: Excel data spreadsheet

Column label	Column description
Family	Taxonomic family of the moth genus/species
Subfamily	Taxonomic subfamily of the moth genus/species
Tribe	Taxonomic tribe of the moth genus/species
Moth_Genus_name	Genus name

Moth_Species_Name	Species name
Number_PlantGenus	Total number of plant genera known to be eaten in natural environment by caterpillars of this species of moth
Plant_GenusNames	Names of plant genera known to be eaten in natural environment by caterpillars of this species of moth
Number_PlantSpecies	Total number of plant species known to be eaten in natural environments by caterpillars of this species of moth
Plant_SpeciesNames	Names of plant species known to be eaten in natural environments by caterpillars of this species of moth
Number_PlantFamily	Total number of plant families known to be eaten in natural environments by caterpillars of this species of moth
Plant_FamilyNames	Names of plant families known to be eaten in natural environments by caterpillars of this species of moth
Number_PlantOrders	Total number of plant orders known to be eaten in natural environments by caterpillars of this species of moth
Plant_OrderNames	Names of plant orders known to be eaten in natural environments by caterpillars of this species of moth
Number_PlantGenus_Capt	Total number of plant genera known to be eaten in captivity by caterpillars of this species of moth
Plant_GenusNames_Capt	Names of plant genera known to be eaten in captivity by caterpillars of this species of moth
Number_PlantSpecies_Capt	Total number of plant species known to be eaten in captivity by caterpillars of this species of moth
Plant_SpeciesNames_Capt	Names of plant species known to be eaten in captivity by caterpillars of this species of moth
Number_PlantFamily_Capt	Total number of plant families known to be eaten in captivity by caterpillars of this species of moth
Plant_FamilyNames_Capt	Names of plant families known to be eaten in captivity by caterpillars of this species of moth
Number_PlantOrders_Capt	Total number of plant orders known to be eaten in captivity by caterpillars of this species of moth
Plant_OrderNames_Capt	Names of plant orders known to be eaten in captivity by caterpillars of this species of moth

Data set name: Polyphagy variables and score for Saturniidae and Sphingidae

Data format: Excel data spreadsheet

Column label	Column description
Family	Taxonomic family of the moth genus
Subfamily	Taxonomic subfamily of the moth genus
Tribe	Taxonomic tribe of the moth genus
Moth_Genus_Name	Genus name
NumberSampledMothSpecies	Number of moth species within the genus that have food plant information available
TotalMothSpecies	Total number of moth species within the genus
TotalNumberGenus	Total number of plant genera known to be eaten in natural environment by caterpillars of this genus of moth
AverageNumberGenus	The average number of plant genera known to be eaten in natural environments by species within this genus of moth
TotalNumberFamilies	Total number of plant families known to be eaten in natural environments by caterpillars of this genus of moth
AverageNumberFamilies	The average number of plant families known to be eaten in natural environments by species within this genus of moth
TotalNumberOrders	Total number of plant orders known to be eaten in natural environments by caterpillars of this genus of moth
AverageNumberOrders	The average number of plant orders known to be eaten in natural environments by species within this genus of moth
PD_score	Phylogenetic score based on the total branch length in the phylogenetic tree connecting the different families of plant eaten by a given moth species and summarized by genus. NOTE: Only computed from records of plant species eaten in natural environments; records on Gymnosperms were dicarded prior calculation (see text).

Acknowledgements

- This work is a product of the ACTIAS group funded by the synthesis center CESAB of the French Foundation for Research on Biodiversity (FRB; www.fondationbiodiversité.fr).
- PA and RR are supported by ANR grant SPHINX ANR-16-CE02-0011-01.
- We thank the entire ACG parataxonomist team (see Janzen and Hallwachs, 2020) for finding, rearing and processing the specimens described and discussed here in database DHJ, and we thank the Centre for Biodiversity Genomics at the University of Guelph for DNA barcoding them and otherwise managing their molecular data; Area de Conservacion Guanacaste/MINAE and donors to the Guanacaste Dry Forest Conservation Fund for preserving the ACG rain forest in

which they live; the Systematic Entomology Laboratory of the USDA and US National Museum/Smithsonian Institution for receiving and permanently housing the specimens and providing the human resources to carry out this taxonomic work, and finally, many dozens of members of the taxasphere for applying names to both the moths and their food plants.

- IJK would like to thank the Trustees of the Loke Wan Tho Memorial Foundation for their generous support of the HOSTS project and his colleagues at the NHM who undertook the original HOSTS project: Phillip Ackery, George Beccaloni, Luis Hernández, Adrian Hine, Sven Loburg, Mike Lowndes and most of all, the late Gaden Robinson, whose dedication saw the project to completion. He is also extremely grateful to the many people who contributed their own rearing records of Lepidoptera or personal accumulations of data for inclusion in the HOSTS database, particularly Mike Bigger (UK), John W. Brown (USA), Chris Conlan (USA), Rob Ferber (USA), Konrad Fiedler (Germany), Jeremy Holloway (UK), Frank Hsu (USA), Jurie Intachat (Malaysia), Alec McClay (Canada), Bill Palmer (Australia), Pierre Plauzoles (USA) and the generous individuals who contributed rearing records through the WorldWideWeb and who are known to us only as an email address. IJK is particularly grateful to Julian Donahue and the Los Angeles County Museum of Natural History for allowing us to include data into HOSTS on Microlepidoptera from the card catalogue prepared by the late J.A. Comstock and C. Henne, and for access to manuscript records by Noel McFarland. Full acknowledgements for the HOSTS database can be found at https://www.nhm.ac.uk/our-science/data/hostplants/#9.

Author contributions

PA, LBM, IJK and RR designed the study and organized the assembly of the dataset. JH, IJK, WH and DHJ compiled the three databases; LBM carried out their combination and computed summary statistics of polyphagy levels. PA computed the calculation of polyphagy scores.

LBM wrote the frist draft of the manuscript, then all authors contributed to its redaction and to the edition of its final version.

References

- Bernays EA, Janzen DH (1988) Saturniid and Sphingid Caterpillars: Two Ways to Eat Leaves. Ecology 69 (4): 1153-1160. https://doi.org/10.2307/1941269
- Devictor V, Julliard R, Jiguet F (2008) Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. Oikos 117 (4): 507-514. https://doi.org/10.1111/j.0030-1299.2008.16215.x
- Faith D (1992) Conservation evaluation and phylogenetic diversity. Biological Conservation 61 (1): 1-10. https://doi.org/10.1016/0006-3207(92)91201-3
- Fiedler K (1998) Diet breadth and host plant diversity of tropical- vs. temperate-zone herbivores: South-East Asian and West Palaearctic butterflies as a case study. Ecological Entomology 23 (3): 285-297. https://doi.org/10.1046/j.1365-2311.1998.00132.x
- Forister M, Novotny V, Panorska A, Baje L, Basset Y, Butterill P, Cizek L, Coley P, Dem F, Diniz I, Drozd P, Fox M, Glassmire A, Hazen R, Hrcek J, Jahner J, Kaman O, Kozubowski T, Kursar T, Lewis O, Lill J, Marquis R, Miller S, Morais H, Murakami M, Nickel H, Pardikes N, Ricklefs R, Singer M,

Smilanich A, Stireman J, Villamarín-Cortez S, Vodka S, Volf M, Wagner D, Walla T, Weiblen G, Dyer L (2014) The global distribution of diet breadth in insect herbivores. Proceedings of the National Academy of Sciences 112 (2): 442-447. https://doi.org/10.1073/pnas.1423042112

- Janzen DH (1981) Patterns of Herbivory in a Tropical Deciduous Forest. Biotropica 13 (4). https://doi.org/10.2307/2387805
- Janzen DH (1984a) Two ways to be a tropical big moth: Santa Rosa saturniids and sphingids. Oxford Surveys in Evolutionary Biology 1: 85-140.
- Janzen DH (1984b) Natural history of *Hylesia lineata* (Saturniidae: Hemileucinae) in Santa Rosa National Park, Costa Rica. Journal of the Kansas Entomological Society 57: 490-514.
- Janzen DH (1985a) On ecological fitting. Oikos 45: 308-310.
- Janzen DH (1985b) A host plant is more than its chemistry. Illinois Natural History Bulletin 33: 141-174.
- Janzen DH, et al. (2012) What happens to the traditional taxonomy when a well-known tropical saturniid moth fauna is DNA barcoded? Invertebrate Systematics 26 (6): 478-505. https://doi.org/https://doi.org/10.1071/IS12038
- Janzen DH, Hallwachs W (2016) DNA barcoding the Lepidoptera inventory of a large complex tropical conserved wildland, Area de Conservacion Guanacaste, northwestern Costa Rica. Genome 59 (9): 641-660. https://doi.org/10.1139/gen-2016-0005
- Janzen DH, Hallwachs W (2020) Caterpillars, pupae, butterflies & moths of ACG, Guanacaste, Costa Rica. http://Janzen.sas.upenn.edu/caterpillars/database.lasso. Accessed on: 2020-9-18.
- Janz N, Nylin S (2008) The oscillation hypothesis of host-plant range and speciation. In: Tilman D (Ed.) The evolutionary biology of herbivorous insects: specialization, speciation and radiation. University of California Press, 203-215 pp.
- Jousselin E, Elias M (2019) Testing Host-Plant Driven Speciation in Phytophagous Insects: A Phylogenetic Perspective. BioArXiv. 1910.09510 https://doi.org/10.20944/preprints201902.0215.v1
- Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO (2010) Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26 (11): 1463-1464.
- Kitching I, Rougerie R, Zwick A, Hamilton C, St Laurent R, Naumann S, Ballesteros-Mejia L, Kawahara A (2018) A global checklist of the Bombycoidea (Insecta: Lepidoptera). Biodiversity Data Journal 6 https://doi.org/10.3897/bdj.6.e22236
- Magallón S, Gómez-Acevedo S, Sánchez-Reyes L, Hernández-Hernández T (2015) A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist 207 (2): 437-453. https://doi.org/10.1111/nph.13264
- Meister F (2011) *Die Zucht von tropischen Wilden Seidenspinnern (Lepidoptera: Saturniidae)*. [*A guide to the breeding of tropical silk moths*]. Verlag Dr. Friedrich Pfeil, Munchen, 220 pp. [In English].
- Ødegaard F (2006) Host Specificity, Alpha- and Beta-Diversity of Phytophagous Beetles in Two Tropical Forests in Panama. Biodiversity and Conservation 15 (1): 83-105. https://doi.org/10.1007/s10531-004-3106-5
- Robinson GS, Ackery PR, Kitching IJ, Beccaloni GW, Hernandez LM (2010) A Database of the World's Lepidopteran Hostplants. Natural History Museum, London. http://www.nhm.ac.uk/hosts. Accessed on: 2020-10-23.

- Santin A (2004) Répertoire des Plantes-Hôtes de Substitution des chenilles du monde. Éléments recueillis dans des publications entomologiques et résultats d'expérimentations personnelles. Deuxième edition. Office pour l'information éco-entomologique, Guyancourt, 1228 pp. [In French].
- Stone SE (1994) Foodplants of World Saturniidae. Memoirs of the Lepidopterists' Society 4: 1-186. [In English]. URL: https://www.lepsoc.org/catalog/memoirs
- Thompson J (1988) Coevolution and Alternative Hypotheses on Insect/Plant Interactions. Ecology 69 (4): 893-895. https://doi.org/10.2307/1941238
- Vidal M, Murphy S (2017) Bottom-up vs. top-down effects on terrestrial insect herbivores: a meta-analysis. Ecology Letters 21 (1): 138-150. https://doi.org/10.1111/ele.12874
- Wang H, Holloway J, Janz N, Braga M, Wahlberg N, Wang M, Nylin S (2017) Polyphagy and diversification in tussock moths: Support for the oscillation hypothesis from extreme generalists. Ecology and Evolution 7 (19): 7975-7986. https://doi.org/10.1002/ece3.3350

Supplementary materials

Suppl. material 1: Food plant dataset for worldwide saturniid and sphingid moths (Lepidoptera, Saturniidae, Sphingidae)

Authors: Kitching, I.J, Janzen, D.H.J., Hallwachs, W, Haxaire, J

Data type: food plant and moths association records

Brief description: This dataset lists food plants known to be fed on by caterpillars of saturniid and sphingid moths, worldwide. It includes both wild and captive records (listed separately).

Download file **(320.46 kb)**

Suppl. material 2: Polyphagy variables and score for Saturniidae and Sphingidae

Authors: Ballesteros-Mejia, L., Arnal, P., Kitching, I.J., Rougerie, R.

Data type: Summary variables, polyphagy score

Brief description: This table reports summary variables and score of polyphagy for Saturniidae and Sphingidae records of our dataset. Note that gymnosperm records were excluded from our calculations of PD scores to avoid bias caused by the considerable phylogenetic distance between angiosperms and gymnosperms. Polyphagy scores were not calculated for genera feeding on Gymnosperms only .

Download file **(31.44 kb)**

Chapitre 2

Le conservatisme de niche et des évènements de dispersification façonnèrent la diversification spatiale et temporelle des saturniidés du genre *Copaxa*

Dispersification events and niche conservatism shaped the spatial and temporal diversity dynamics of *Copaxa* moths into the Neotropics (Lepidoptera: Saturniidae)

Préambule

Étudier la diversification des Saturniidae à partir de phylogénies *higher-level* m'a permis de m'intéresser à l'évolution de traits relativement conservés, à l'histoire biogéographique de la famille depuis ses origines et sur tous les continents, et de mettre en évidence des évènements majeurs et des facteurs clés dans leur dynamique de diversification à grande échelle. De manière générale, ces phylogénies ne permettent cependant pas de s'intéresser au rôle de traits plus labiles, dont la variation intra-générique peut être importante. C'est pourquoi je me suis intéressé à l'évolution spatiale et temporelle d'un clade plus restreint, celui du genre *Copaxa*. Dans ce second Chapitre, j'ai ainsi inféré la première phylogénie moléculaire pour tous les représentants du genre *Copaxa* à l'aide d'une combinaison de marqueurs génomiques et génétiques. J'ai ensuite daté cette phylogénie, inféré l'histoire biogéographique et mesuré les taux de diversification au sein du groupe. Dans leur ensemble, les résultats détaillés dans ce Chapitre révèlent une forte influence du conservatisme de niche dans la diversification spatiale des *Copaxa*. Nous avons également mis en évidence deux évènements « clés » de leur diversification, qualifiés de « dispersification », *i.e.* la colonisation d'une nouvelle région biogéographique – l'Amérique du Sud – associée avec un shift positif de diversification.

Les travaux effectués dans ce Chapitre seront prochainement soumis dans *Journal of Biogeography* après avoir reçu les retours de l'ensemble des co-auteurs du manuscrit présenté dans les pages suivantes.

Dispersification events and niche conservatism shaped the spatial and temporal diversity dynamics of *Copaxa* moths into the Neotropics (Lepidoptera: Saturniidae)

Running title: Moth dispersification into the Neotropics

Pierre Arnal[1], Astrid Cruaud[2], Marianne Elias[1], Liliana Ballesteros-Mejia[1], Fabien Condamine[3], Thibaud Decaëns[4], Delphine Gey[5], Paul D. N. Hebert[6], Ian J. Kitching[7], Jean-Yves Rasplus[2], Sujeevan Ratnasingham[6], Rodolphe Rougerie[1]*

[1] Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, EPHE, Sorbonne Université, Université des Antilles; Paris, 75005, France.

[2] Centre de Biologie pour la Gestion des Populations (CBGP), INRAE, CIRAD, IRD, Montpellier SupAgro, Université de Montpellier; Montpellier, 34000, France.

[3] Institut des Sciences de l'Évolution de Montpellier (ISEM), CNRS, Université de Montpellier: Montpellier, 34095, France.

[4] Centre d'Écologie Fonctionnelle et Évolutive (CEFE), Université de Montpellier, CNRS, EPHE, IRD, Université Paul Valéry Montpellier 3: Montpellier, 34090, France.

[5] Service de Systématique moléculaire (SSM), CNRS, Muséum National d'Histoire Naturelle, Sorbonne Université: Paris, 75005, France.

[6] Centre for Biodiversity Genomics, University of Guelph: Guelph, Ontario, Canada.

[7] Department of Life Sciences, Natural History Museum: London, United Kingdom.

* Corresponding author: rodolphe.rougerie@mnhn.fr

Abstract

Aim

Neotropical mountains are home to a prodigious biodiversity, part of which was documented as originating from Holarctic lineages dispersing 'into the tropics' during the Cenozoic. Such pattern, well-known in plants, has rarely been documented for insects. Here we aim to unravel the diversification dynamics of a Holarctic saturniid lineage that dispersed into the New-World during the Oligocene and is now widely distributed throughout the Neotropics.

Location

Neotropics: Central and South-America.

Taxon

Lepidoptera, Saturniidae, genus *Copaxa*.

Methods

We used integrative taxonomy to redefine extant *Copaxa* species. We then assembled and analyzed a supermatrix combining genomic (Ultra-Conserved Elements and RADseq) and genetic (DNA barcodes) data to infer a dated phylogeny for all extant species of *Copaxa*. Using that analytical support, we performed historical biogeography and diversification analyses as well as ancestral state reconstruction for altitudinal preferences to elucidate the evolution of that diverse genus.

Results

We revised the diversity of genus *Copaxa* as comprising 130 described species and 19 additional Operational Taxonomic Units (OTUs). Our phylogenomic analyses resulted in a robust dated phylogenetic hypothesis for all extant species that was used to produce a detailed account of the spatial and temporal diversification of the genus. The common ancestor of extant *Copaxa* originated *ca*. 13 Ma in Central-America and two lineages experienced late Miocene dispersal events into South-America associated with diversification rate shifts ("dispersification"). The diversity of *Copaxa* is dominated by mountain species that diversified within restricted biogeographical ranges; adaptation to lowlands remains uncommon, but appeared several times independently within the genus.

Main conclusions

Phylogenetic niche conservatism, in conjunction with two independent dispersal events, played a major role in driving the diversification of *Copaxa* moths into the Neotropics. Colonizations of lowlands did not induce adaptive radiations, but may have enhanced dispersal across biogeographical regions.

Biosketch

Pierre Arnal is broadly interested in the spatial and temporal diversification of insects. This work is part of his PhD project at the Muséum national d'Histoire naturelle (Paris) focusing on the evolution of wild silkmoths. All authors are members of the SPHINX and ACTIAS consortia (PI: Rodolphe Rougerie) that investigate the macroevolution and macroecology of Saturniidae and Sphingidae moths.

Author contributions

PA, ME and RR designed the study. PA, AC, DG and RR did the lab work, PA, LBM, PDNH, IJK, RR and SR generated and analyzed DNA barcodes and revised the species diversity in the genus, PA conceived (with input from AC, ME, JYR, FC, TD and RR) and performed the analyses. Finally, PA wrote the manuscript, with assistance by ME and RR and receiving feedback from all co-authors.

Introduction

Neotropics contain 7 of the 25 biodiversity hotspots (Myers et al. 2000) and are the most speciose region on Earth (Jetz et al. 2012; Roll et al. 2017). As currently defined (Schultz, 2005; Antonelli & Sanmartín 2011), the Neotropical region, extending from central Mexico to southern Brazil, comprises two main geological entities, *i.e.* Central and South America, that have long, complex and distinct geological histories. The former was always connected to the Nearctic region within the Laurasia landmass, while South America was isolated for millions of years after the Gondwanan breakup (Torsvik & Cocks 2004; Jokat et al. 2003), except for the emergence of transitory volcanic bridges with the North American continent (Iturralde-Vinent 2006). As a result, the South American biota evolved in relative isolation (e.g. Aves: Arini, Tavares et al. 2006; Mammalia: Platyrrhini, Upham et al. 2019) and its diversity received limited input from other regions (Bacon et al. 2015; Antonelli et al. 2018). Nonetheless, several South American lineages are known to result from southward migrations, sometimes termed as 'into the tropics' dispersal events. Some predate (e.g. the bird family Furnariidae, *ca.* 30 Ma; Oliveros et al. 2019), but most are contemporaneous (e.g. the adaptive radiation in the Sigmodontinae rodent sub-family, *ca*. 12 Ma; Hershkovitz 1969; Parada et al. 2013) or subsequent to the closure of the Central America Seaway (CAS) about 10Ma (Sepulchre et al. 2014; Montes et al. 2012, 2015). This closure of the CAS and the subsequent establishment of a durable land bridge between South and Central America (Coates & Stallard 2013) had substantial climatic (Wolfe 1994; Molnar 2008) and biogeographical consequences (the Great American Biotic Interchange; see Stehli & Webb 2013). Bacon et al. (2015), who carried out a meta-analysis of 169 dated phylogenies, identified positive migration rate shifts *ca.* 8.8 and 5.2Ma, both southward and northward. Some Andean lineages of plants proved to result from the southward dispersion and subsequent radiations of lineages originating from the Holarctic region: Hughes & Eastwood (2006) estimated that the genus *Lupinus* (Fabaceae) colonized the Andes from Central America about 1.5Ma and subsequently experienced a prodigious radiation (2.49–3.72 species per Myr per lineage in the Andean clade); in *Hypericum* (Hypericaceae), a more ancient South American colonization, *ca.* 8Ma (Meseguer et al. 2013, 2015), also led to a significant Andean diversification, especially in the páramos (Nürk et al. 2013). Such examples of speciation following dispersal or vicariance, without any major shift in ecological niche (Phylogenetic Niche Conservatism (PNC), see Wiens 2004; Pyron et al. 2015), are in line with the hypothesis by Donoghue (2008) that lineages originating from cold and dry temperate environments of the Nearctic region were pre-adapted to tropical mountainous habitats and have first colonized and thrived in these habitats, before species originating from lowland tropical lineages could adapt to them.

However, the importance of PNC in governing the dispersal and diversification into the Neotropics of lineages originating from higher latitudes remains poorly understood, especially in insects that lack thoroughly documented analyses of their spatial and temporal evolutionary dynamics in tropical regions (Diniz-Filho et al. 2010, 2013). This is regrettable because insects are relevant models to address the

role of PNC in governing diversification in Neotropical mountains: they are highly diverse ectotherm organisms whose climatic niche is easier to track than in endotherms, because it evolves at a slower pace (Rolland et al. 2018).

Here we focus on *Copaxa* moths (Figure 1), a diverse genus of Lepidoptera from the family Saturniidae (wild silkmoths). *Copaxa* is known to be sister to the Holarctic genus *Saturnia* (Arnal et al. in prep.; see Chapitre 1; Rubinoff & Doorenweerd, 2019). The common ancestor to both genera was shown to have originated in Eastern Palearctic from where it colonized Western Nearctic and Central America *ca.* 27Ma (30.8-23.5Ma) through Beringia (Arnal et al. in prep.; Chapitre 1). The extant distribution of the genus *Copaxa* ranges across the entire Neotropics, from the North of the Sierra Madre cordilleras in Mexico to the South of the Atlantic Forest in Uruguay. It is found from lowland tropical rainforests of Amazonia (e.g. *C. marona* Schaus, 1906) to high elevation Andean páramos (e.g. *C. medea* (Maassen, 1890) up to 4,500m asl.). Its species diversity has been extensively characterized recently, thanks to vast DNA barcoding campaigns (Hebert et al. 2003; Hebert & Gregory 2005) and subsequent species descriptions that raised the number of valid described species from a few dozens in the early 2000s' (Wolfe & Conlan, 2002) to 122 (and 6 subspecies; Kitching et al. 2018; Figure 1). The affinities of most species for mountain habitats as well as known preferences of some *Copaxa* caterpillars for food plants in genera of temperate Fagales trees (*Fagus*, *Alnus*, *Quercus*; Wolfe 1993; Wolfe et al. 2003b) or more marginally on Gymnosperms (Wolfe 1993, 2003a) are suggestive of a role of PNC in the diversification of these moths into the Neotropics, and the genus thus represents an adequate model to investigate the dynamics of diversification "into the Neotropics" of a lineage originating from temperate regions.

In this study, we use an innovative approach combining genomic (Ultra Conserved Elements (UCEs) and Restriction-site Associated DNA (RAD) markers) and genetic data (COI DNA barcodes) to infer a dated phylogenetic hypothesis for all known species of *Copaxa*. This phylogeny then serves as the ground to reconstruct the spatial and temporal dynamics of diversification of the genus, and to investigate how phylogenetic niche conservatism, represented by lineage preferences for lowland, mid- or high-elevation habitats, may have governed these dynamics.

Materials & Methods

Species diversity in genus Copaxa

Copaxa wild silkmoths are medium-sized to large-sized moths belonging to the Saturniini tribe, within subfamily Saturniinae of the Saturniidae family. In its revision of New World Saturniinae, Lemaire (1978) recognized 30 valid species in the genus, but the integration over the past 10 years of DNA barcode data into the study of the diversity of these moths led to a spectacular increase in the number of species and subspecies recognized and described (Figure 1; Kitching et al. 2018). Nearly 80 new taxa

were described during this period and the status and validity of several older names were revisited, leading to a recent account (Kitching et al., 2018) of 128 described species and subspecies.

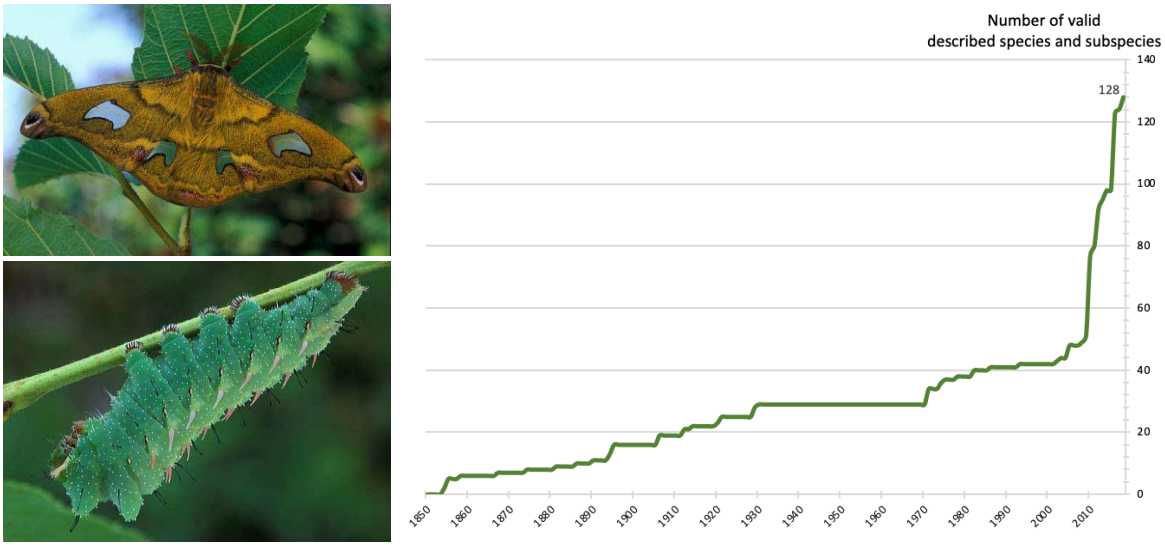


Figure 1 - Left panel: Adult male (up) and caterpillar (down) of *Copaxa sapatoza*, endemic of Eastern Cordillera in Colombia (photographs by Kirby Wolfe, ©); right panel: number of species described and considered valid in genus *Copaxa* after Kitching et al. (2018).

Lemaire (1978) proposed five species-groups to classify *Copaxa* species (Table S1). This classification was modified by Wolfe et al. (2003a), Wolfe (2005b), Brechlin & Meister (2010b) and lastly by Brechlin et al. (2016), recognizing three additional species groups as well as several sub-groups (see Suppl. Table 1) reflecting mostly affinities derived from morphological similitudes. However, they never were tested using a robust and objective phylogenetic analysis of characters, neither morphological nor molecular. In the present work, we aimed at considering the entire diversity of genus *Copaxa*; to do so, we used a comprehensive DNA barcode library for these moths as the ground to define the taxonomic units included in our analyses. This library was assembled during the course of the global DNA barcoding campaign for saturniid moths, initiated 15 years ago and built on the contributions of many taxonomists and several natural history institutions with large collections of these moths. The vast majority of DNA barcodes were produced at the Canadian Centre for DNA Barcoding (CCDB), hosted by the Centre for Biodiversity Genomics at the University of Guelph, Canada, following standard protocols (deWaard et al. 2008) developed as part of the international Barcode of Life (iBOL) program. As mentioned above, the integration of DNA barcodes triggered a significant increase in the number of species recognized and described. Because these efforts included material from the entire distribution range of the genus, its current taxonomic treatment can be considered fairly mature and nearly completed; nevertheless, we apply here an integrative approach considering DNA barcode results, morphology, distribution and ecological data to further refine species boundaries and delimit the operational Taxonomic Units (OTUs) used in our analyses. These OTUs may not always reflect the currently accepted species names, in which case we use provisional taxon names that can be derived, when relevant, from Barcode Index Numbers (BINs) automatically assigned by the Barcode of Life Datasystems (BOLD; www.boldsystems.org; see

Ratnasingham & Hebert, 2007; 2013). DNA barcode data (including sequence and specimen information, as well as images for most records) are publicly accessible in BOLD dataset DS-COPAX1; we used standard analytical tools included in BOLD to explore, analyze and curate DNA barcode records.

Molecular data acquisition

We used three types of DNA markers to infer a phylogeny of the genus *Copaxa*: Ultra Conserved Elements (UCEs), Restriction-site associated DNA (RADSeq), and the standard DNA barcode sequence (part of mitochondrial COI gene). Because the number of shared orthologous restriction sites decreases with phylogenetic divergence (Rubin et al. 2012; McCormack, et al. 2012), we expected limited phylogenetic information from RADSeq loci for early branching lineages within the genus, whose stem age was estimated about 26.5Ma ([23.5-29.0] 95% confidence interval; Arnal et al. in prep.; Chapitre 1). This is why we also used UCEs, which were shown to be informative at intra-generic level for various organisms (Blaimer et al. 2016; Prebus et al. 2017; Andersen et al. 2019). Because of costs and accessibility of suitable material, genomic markers could not be obtained for all species within the genus. Our sampling strategy thus consisted in targeting one or two representatives per recognized species-group for UCE markers, 25 to 50% of all species in these groups for RADSeq, but also to use DNA barcode results (from the exploration of Neighbor Joining trees) to select species that appeared strongly divergent for that marker from other species in their species group (e.g. *C. sophronia* and *C. syntheratoides*, placed in the *multifenestrata* group before present work, see results). These genomic markers were combined with the DNA barcode sequences for all species to take advantage of the phylogenetic signal of this mitochondrial marker for shallow nodes (Monteiro & Pierce 2000; Wilson 2010; Fang et al. 2017), expecting no or little negative influence of missing data on phylogenetic inferences (Wiens & Morrill 2011).

We processed 11 new *Copaxa* samples through the same protocol and data processing pipeline used by Arnal et al. (in prep.; Chapitre 1; see also Cruaud et al. 2019) to generate UCE sequences. By combining these with sequences of three additional *Copaxa* samples and the outgroup *Saturnia pyri* used in Arnal et al. (in prep.), we assembled a UCE dataset for 14 *Copaxa* samples representing all the recognized species groups, as well as for one outgroup.

Library preparation for RADSeq markers followed the protocol of Etter et al. (2011), and sequencing was performed on an Illumina HiSeq 2500 flowcell. A first library was built with 48 samples representing an initial selection of 43 *Copaxa* and 5 outgroup samples for which DNA was extracted from fresh tissues (legs) preserved dry or in ethanol at the Muséum national d'Histoire naturelle using either Qiagen QIAamp Mini Kit or Qiagen DNeasy Blood & Tissue Kit. A second library was assembled with 48 additional *Copaxa* samples, of which 10 were DNA extracts obtained in a similar way as for the first library, and 38 were DNA extracts produced at the CCDB following standard high-throughput

protocols (deWaard et al. 2008) to be used for sequencing of DNA barcodes in BOLD. These 38 samples were part of a larger set of 59 cryopreserved DNA extracts that contained very low initial DNA quantities (2 to 62 ng; average concentration=12.6 ng/µL, sd=11.6); they were selected after DNA was successfully amplified to a minimum value of 100ng using a Genomiphi V2 DNA Amplification Kit following manufacturer instructions. The RADIS pipeline (Cruaud et al. 2016) was used to generate nucleotide matrices from raw data. RADIS relies on Stacks (Catchen et al. 2013) to group reads into individual loci (*i.e.* stacks) and to identify homologous loci. We allowed a maximum of 3 differences between the reads (*M* parameter) and assigned secondary reads to primary stacks considering a maximum distance of 5 (N parameter). We subsequently discarded stacks with less than three reads. Then Stacks pools the consensus sequences of every stack of every sample and group similar sequences into homologous loci. Because we did not have any a priori on the divergence between the different *Copaxa* species for RADSeq markers, we tested different values of *n* (maximum number of differences between consensus of the different stacks) during the loci catalog building: n6, n8 and n10. During its last step, RADIS builds the nucleotide matrices. It only considers the loci for which the number of samples is higher than the parameter *S*. In order to test the effect of the amount of missing data on tree inference, we used different *S* parameter values: S26, S34 and S43, representing respectively 35%, 46% and 58% of the samples. The three *n* and *S* values imply the generation of 9 matrices with RADIS.

Finally, we also retrieved DNA barcode sequences from the BOLD DNA barcode library, selecting one sequence (full length (658bp) as far as possible) for each sample.

Phylogenetic inferences

We first analyzed the 9 RADSeq matrices produced by RADIS with IQTREE 1.6.3 (Nguyen et al. 2015) to assess the influence of parameters *n* and *S* and select the matrix to be used in the concatenated dataset. Phylogenetic supports were assessed with 1000 Ultrafast bootstraps (Hoang et al. 2018) and 1000 replicates of a Shimodaira-Hasegawa-like procedure (SH-aLRT; Guindon et al. 2010). We used three partitioning strategies: (i) a unique partition, (ii) one partition per locus, (iii) partitions as defined by PartitionFinder 2.1.1 using the *rclusterf* algorithm (data pre-partitioned per locus; Lanfear et al. 2017). Partition models were estimated with ModelFinder (Kalyaanamoorthy et al. 2017), implemented in IQTREE, with the TESTNEW option. This resulted in 27 trees corresponding to the 9 matrices and 3 partitioning strategies.

We built the UCE matrix using all loci for which at least 50% of the samples were available. Alignments were performed with MAFFT v7.245 (Katoh & Standley, 2013) for each individual locus and ambiguously aligned blocks were removed using Gblock_0.91b with relaxed constrains (-t=d -b2=b1 -b3=10 -b4=2 -b5=h) (Talavera & Castresana 2007). We then generated gene trees with ModelFinder and IQTREE 1.6.3 and used PMCOA (de Vienne et al. 2012) to identify and remove complete outliers (a complete locus or a complete sample) or cell outliers (an individual locus) from the DNA matrix. We

defined UCE partitions using the method of Tagliacollo & Lanfear (2018): we first split each UCE locus into three regions (two flanks and one core) using the Sliding-Window Site Characteristics Entropy (SWSC-EN) and subsequently grouped regions evolving at a similar pace with PartitionFinder 2.1.1 (Lanfear et al. 2016) using the *rclusterf* algorithm. This matrix was then combined with the RADSeq matrix generated in RADIS using n=8 and S=34 (see results), itself partitioned with PartitionFinder 2.1.1. Finally, the UCE+RADSeq matrix was combined with the DNA barcode dataset, as an additional single partition. ModelFinder was used to fit the best model of nucleotide evolution for each partition of our supermatrix.

To account for potential gene-specific effects on phylogenetic inference, we measured branch supports with a two steps bootstrap procedure (Seo et al. 2005) that consists in a first independent resampling of loci within the UCE and RADSeq matrices and then a resampling of all nucleotide positions within each marker, including the DNA barcode locus. We used a R script to generate 100 bootstrap matrix replicates and their corresponding partition scheme, and analyzed them with IQTREE. Because branch supports could be biased by the large amount of missing data in a few clades for which we did not have any genomic markers, we also computed a secondary set of bootstrap values after discarding samples for which only the DNA barcode was available.

Finally, we sought to identify rogue taxa with RogueNaRok (Aberer et al. 2013) using the bootstrap trees inferred from the two steps bootstrap procedure as input, considering a maximum dropset size of 1. Highly labile samples (rawImprovment score >1) were removed from subsequent divergence time and macroevolutionary analyses.

Divergence time estimations

To generate calibration points within genus *Copaxa*, we first re-analyzed the UCE dataset of Arnal et al. (in prep.; Chapitre 1) after including the three *Copaxa* species sequenced by the authors (*C. lavenderoguatemalensis, C. syntheratoides* and *C. decrescens*) and excluding all representatives of subfamilies Hirpidinae, Hemileucinae, Arsenurinae, Ceratocampinae, and Agliinae, to reduce computation time. We followed the same procedure as described in Arnal et al. (in prep.) to reduce the UCE matrix to the 50 loci that produced the best supported gene trees, which were analyzed with MCMCTREE in PALM v4.8 (Yang 1997) to estimate divergence times after constraining the topology to that inferred by Arnal et al. (in prep.; Chapitre 1) and using the same calibration points they did.

Because our *Copaxa* supermatrix included more than 120M nucleotides (see results), we reduced both the number of terminal taxa to 148, considering only one sample per species or OTU, and the number of loci by randomly sampling 25 UCE and 300 RADSeq loci along with the COI gene. We then used MCMCtree to estimate the divergence time, constraining the topology to that inferred with IQTREE. Branch lengths were estimated in PAML v4.5 with a maximum likelihood method introduced by Thorne

et al (1998); we used the two *Copaxa* calibration points estimated from our initial step, applying uniform priors (with soft bounds) that spanned the 95% confidence interval we estimated. We ran two analyses with 3 million generations in MCMCTREE (0.5M generations as burnin) for each dataset and combined them with LogCombiner 1.5.3 after checking for convergence.

Historical biogeography

The *Copaxa* genus originated from a lineage that crossed Beringia and dispersed toward Central and South America during the Oligocene and early Miocene (Arnal et al. in prep.; Chapitre 1); the extant distribution of its species ranges from Northern Mexico to Uruguay. To gain a better understanding of the timing of colonization through Central America, we considered three areas north of the Panama Isthmus (see Fig. 4): (from North to South) (i) the 'Transitional zone', comprising the Sierra Madre Oriental, Sierra Madre Occidental, Sierra Madre del Sur and the isolated Mexican Sierra de Santa Martha; (ii) the 'Nuclear zone', comprising the Sierra Madre de Chiapas; (iii) the 'Talamanca zone', comprising both the Talamanca and the Guanacaste cordilleras. From the Isthmus of Panama southward, we considered a fourth area, 'Choco', formed by a lowland zone spanning to the West to the Andes, from the Magdalena valley to the Rio Guayas estuary, here extended to include the Caribbean region of Colombia. We also divided the Andes into three areas: (i) North Western Andes spanning the Western and Central Cordilleras down to the Cero Mishahuanga; (ii) North Eastern Andes defined by the Eastern Colombian cordillera, the Serranía del Perijá, the Sierra Nevada de Santa Marta and by the Merida and la Costa cordilleras; (iii) the Central Andes spanning most of the Peruvian and Bolivian Andes, including the Yungas. Finally, we respectively considered the Amazonian Basin and the Atlantic Forest.

Ancestral area reconstruction analyses were performed with the BioGeoBEARS R package (Matzke 2013) using the DEC model. We manually set dispersal multipliers to 0.5 between two areas separated by a geographical (and/or climatic) barrier or by another area. The coefficient was multiplied by the number of barriers or areas to cross or when the barrier was particularly large. We restricted the maximum range size to three areas, except for the period between 0.1Ma and the present (see below). In addition, we did not allow any combination of areas including both Amazonia and Choco, because we considered the Andes to form a strong barrier impeding gene flow between these areas. Four time periods were defined. The first period of time spans the root of the tree to 10 Ma. During that period, the Central American Seaway (CAS) was still open and deep-sea current could go through the strait (Sepulchre et al. 2014; Montes 2012, 2015; Jaramillo et al. 2017). We considered that it acted as an efficient barrier against terrestrial dispersion even though the Central and South American land masses were very close. In our model, we thus assigned a 0.25 coefficient to dispersal between the Talamanca and Choco areas, and we forbid ancestral ranges to span over these two regions during that period of time. The second time period is 10-4 Ma. We allowed ranges to span over both the Talamanca and Choco regions and we increased the dispersal coefficient to 0.5, considering closure of the CAS (Montes

et al. 2012; Coaster & Stallard 2013). The beginning of the third period (4-0.1 Ma) was chosen because Cerrado lineages diversified mainly around 4Ma, suggesting an expansion of that dry biome and a wider separation between the Atlantic and Amazonian forests (Simon 2009), we thus reduced the manual dispersal multiplier from 1 to 0.5 between the Atlantic forest and its two adjacent areas, Central Andes and Amazonian forest. By contrast, during that period the Panama Isthmus emerged and we considered a terrestrial connection between the Choco and Talamanca areas (Coates & Stallard 2013). Finally, we considered a very short period of time (0.1Ma to present) during which we allowed ranges spanning over 4 areas. We considered that such wide ranges are only possible because of dispersal events that are too recent to have led to speciation. Only one species is distributed in 4 areas: *Copaxa bireni*.

Altitudinal preferences

Copaxa species occur in habitats spanning a wide range of altitudinal conditions. Some inhabit lowland forests (*e.g. C. marona* in Amazonia) whereas others can only be observed at very high elevation (*e.g. C. medea* above 4000m in the Central Andes). Adaptation to altitude is a key ecological trait that could help us understand the biogeographical and diversification dynamics of *Copaxa* species, both with respect to the role of Phylogenetic Niche Conservatism, but also because we hypothesize that variations in this trait may have favored dispersal and range expansion of the species capable of inhabiting low elevation environments, thus "escaping" isolation on mountains in a similar way as insular species in the context of the taxon cycle theory (Wilson, 1959, 1961). We retrieved all the altitudinal information available from records in the DNA barcode dataset and calculated median values for all species and OTUs. Median elevation values under 800m were considered as 'lowland', between 800 and 2000m as 'mid-mountain', and over 2000m as 'high-mountain'. We estimated the ancestral state at each node of the phylogeny applying a MuSSE model of evolution and using the asr.marginal function with the *diversitree* R package (FitzJohn 2012). All parameters of the model were free to vary, except for transition parameters between 'lowland' and 'high-mountain' (q_{13} and q_{31}) that were set to 0.

Diversification rates

To estimate diversification rates and their variations across the evolutionary history of genus *Copaxa*, we used two complementary methods. We first fitted the RevBayes (Höhna et al. 2016) model designed by Höhna et al. (2019) that considers extinct lineages when inferring branch-specific diversification rates. To facilitate probability computation, the model discretizes the prior speciation and extinction rates distributions into k categories. Here we used 8 rate categories and ran MCMC analyses for 5000 generations. We assessed convergence using Tracer (ESS>200; Rambaut et al. 2018) and summarized the posterior distribution with the *plot_branch_rates_tree* function of the RevGadgets R package (https://github.com/revbayes/RevGadgets). Because the biogeographical analyses described above brought evidence of independent and ancient colonization events of the South American continent in two major lineages (named A and B) of the genus, we hypothesized that these events induced shifts in

diversification rates. We tested this hypothesis using the birth-death method developed by Morlon et al. (2011) and implemented in the RPANDA R package (Morlon et al. 2016), an approach that extends previous birth-death methods in considering that speciation and/or extinction rates may change exponentially through time. We compared four different diversification regimes: (i) no diversification shifts, (ii) one diversification shift in the A clade, (iii) one diversification shift in the B clade and (iv) the two shifts combined. For each of these hypotheses, we split the tree at the shifts location and compared different diversification models for each subtree and the backbone: (i) a pure birth model, BCST, (ii) a birth-death model, BCST-DCST, (iii) a time-dependent birth model, BVAR, (iv) a time-dependent birth model with a constant extinction rate, BVAR-DCST, (v) a model with a constant birth rate but a time-dependent extinction rate, BCST-DVAR, and (vi) a time-dependent birth-death model, BVAR-DVAR. For each subtree we calculated the AICc score of the best model. For the different shift configurations, we summed AICc of the subtrees and the backbone and compared the obtained values to determine the best diversification shifts regime during the evolution of *Copaxa*. This procedure was applied to the MCMCtree consensus tree and to 100 trees randomly sampled in the posterior distribution of the MCMCtree analysis.

Results

Species diversity in genus *Copaxa*

To uniformize the taxonomic treatment of the genus, we raised to species level all six subspecies previously recognized (see Table S1) and considered lacking objective justification for the use of this rank. The DNA barcode reference library assembled for genus *Copaxa* comprises 1455 records representing all 130 named species here recognized as valid within the genus (see Table S1 and the checklist CL-COPAX in BOLD). A DNA barcode was obtained for the holotype specimen of 92 species (71% of all valid named taxa) and 247 paratypes were also sequenced for this marker. Our coverage is complete, but our taxonomic account differs from that of Kitching et al. (2018) in including several species described since this publication (Brechlin, 2018, 2019a, b), in considering *C. brunnea* Bouvier, 1929 as a *nomen nudum* resulting from improper formatting of the name *C. flavobrunnea* by Bouvier (1929), and in the treatment of *C. peggyae* Brechlin & Meister, 2013 syn. nov. and *C. rudloffi* Brechlin & Meister, 2010 syn. nov. as junior synonyms of *C. escalantei* Lemaire, 1971 and *C. moinieri* Lemaire, 1974, respectively (see notes under these two species in Table S1).

All 1455 records, including specimen and sequence data, as well as images, are publicly available in BOLD dataset DS-COPAX1; sequences are also deposited in GenBank (accession numbers available from DS-COPAX1). A neighbor joining tree built with BOLD tools using uncorrected pairwise distances, BOLD aligner and pairwise deletion options is provided in Figure S1 (Online resource). In addition to the 130 named species, we also considered 19 additional OTUs that may represent

undescribed cryptic species and all represent distinct BINs from the closely related species; they are all designated in our dataset by the name of this related species followed by the BIN code, except in the case of *C. curvilinea*DHJ01, provisionally named as part of the Lepidopteran inventory of Area de Conservacion Guanacaste in Costa Rica (Janzen et al. 2012).

Sequencing of phylogenomic data

Voucher information is compiled in Table S2. We successfully sequenced RADSeq markers for 69 *Copaxa* samples and 5 outgroups. As expected, the matrix sizes were significantly influenced by both the *S* and *n* parameters (Figure 2), and the number of loci was very heterogeneous among samples (Table S3; e.g. 38 to 1716 (median=1155) in matrix M3n6S34). The later was inversely proportional to the minimum number of samples per loci (*S*), but increased with the maximum number of differences allowed between stacks when building loci (*n*). Considering lower *S* values implied higher percentage of missing data but did not increase the percentage of informative positions. In contrast, higher values of *n*, leading to the grouping of more variable stacks, resulted in a larger number of informative sites, but the downside of considering higher *n* values was the risk of increasing paralog numbers, as expected.

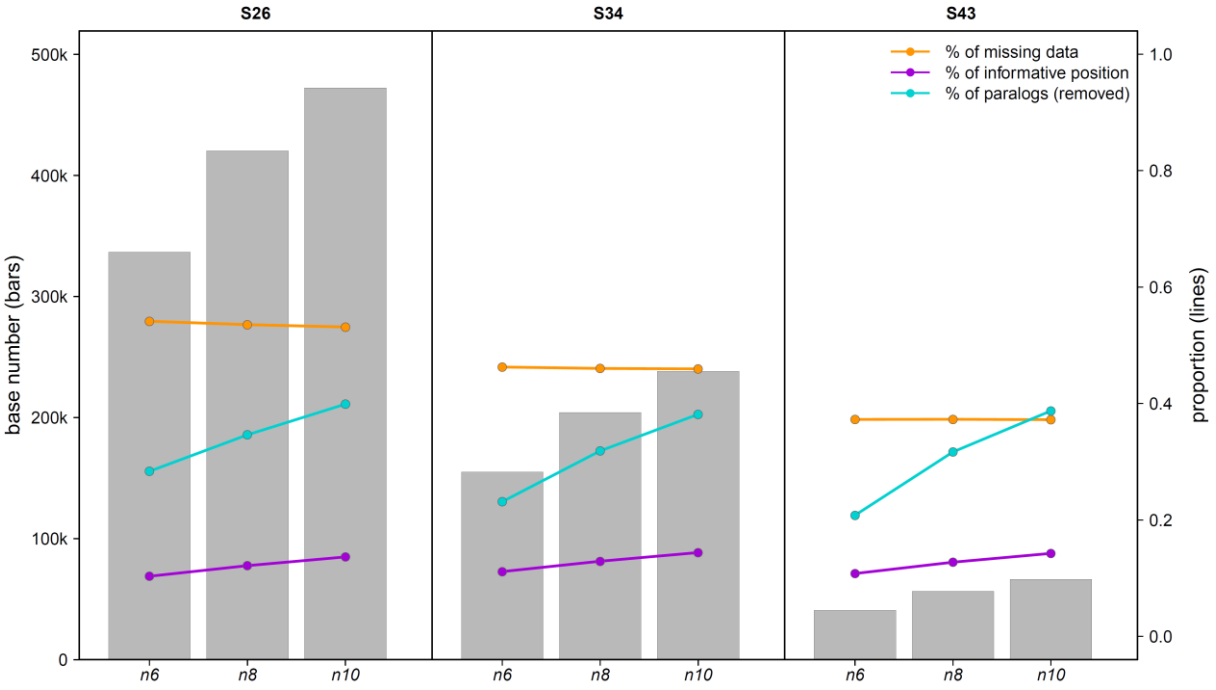


Figure 2 - Graphical representations of the influence of S (panels) and n (bars) values on RADSeq matrix size (bars) and proportions (lines) of missing data (orange), informative positions (purple), and paralogs removed (blue).

The UCE loci were sequenced for 11 *Copaxa* samples and combined to those of 3 samples from Arnal et al. (in prep.). Overall, the number of UCE loci per sample ranged from 107 to 1163. After discarding loci for which less than 7 samples were available, the UCE matrix included 856 loci (each 650bp long on average).

Phylogenetic inferences

The results obtained with the 27 different RADSeq matrices were highly congruent, independently of the partitioning strategies used (Figures S2, S3). The *n* and *S* parameters did not influence substantially the topologies but impacted the total branch lengths. Independently of the value of these parameters, we cannot firmly conclude from the RADSeq results about the position of the lineage grouping *C. mannana, C. muellerana,* and members of the *lavendera* group. To combine RADSeq with UCE and DNA barcode sequences into a supermatrix, we selected the matrix built with *n=8 and* S=*34*. This matrix is of reasonable size and contains a fair amount of informative loci, while avoiding large volume of missing data and a high risk of including paralogs.

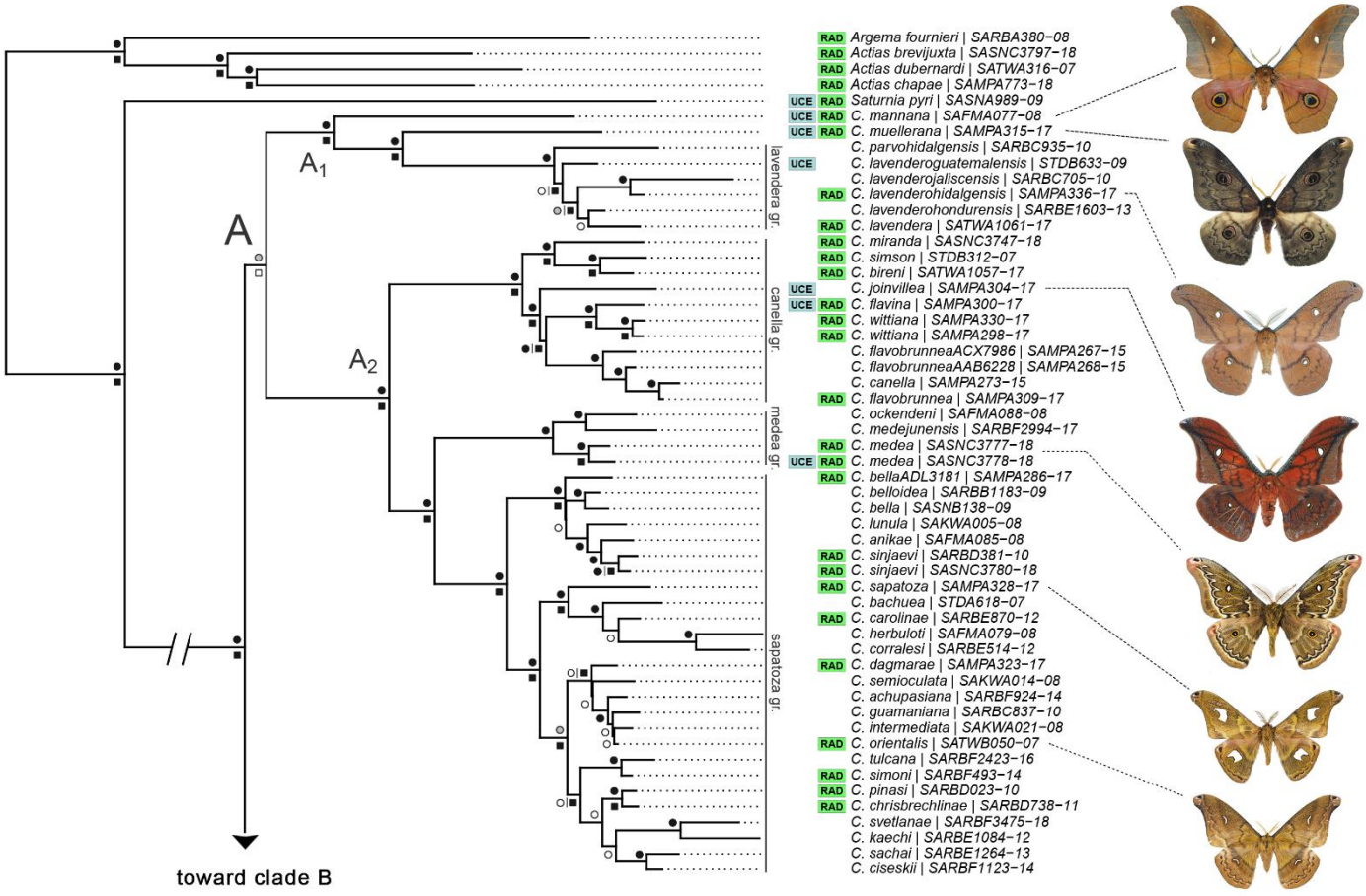
The phylogeny (Figure 3) inferred from the combination of UCE, RAD and COI markers included 158 samples representing all 130 valid *Copaxa* species and 18 OTUs (after removal of one rogue taxa identified with RogueNaRok: *Copaxa rufinans*ADF5627; code: SAMPA287-17), as well as 5 outgroups. It is very congruent with the topologies estimated with the RADSeq markers (Figure S3). While there remains some uncertainties about shallow or intermediate nodes, the phylogeny of *Copaxa*, which comprises two main lineages (hereafter referred as the A and B clades), is overall well supported and the backbone is nearly fully resolved. Again, the clade (named A_1) grouping *C. mannana, C. muellerana,* and the members of the *lavendera* group branches either with the A_2 sub-clade (*ca.* 2/3 of the bootstrap trees) or with the B clade (see Figure 3). Because our two-step bootstrap procedure is particularly stringent compared to classic bootstrap procedure (*e.g.* the SH-aLRT procedure or Ultrafast Bootstraps), we nonetheless consider the position of clade A_1 within lineage A as the most likely, pending further phylogenomic investigations. Beside this moot point, the position and monophyly of the different species groups are strongly supported by the two steps bootstrap procedure (Figure 3). The use of DNA barcodes as the only genetic marker for many samples had diverse degrees of success. Our results show that it is informative enough to place the different species in their respective species groups and we successfully resolved several shallow nodes with the sole COI marker. For example, in the *medea* species group, the positions of *C. ockendeni* and *C. medejunensis* are well supported even though we did not get any genomic markers for these two species (Figure 3). In contrast, our analyses failed to resolve some shallow nodes when the speciation rates were higher (*e.g. C. troetschi* and *C. andescens* in the *decrescens* group; Figure 3). The use of the sole DNA barcodes was even less successful in inferring the position of phylogenetically isolated species or clades. This is the case for *C. denda*, whom position has to be further investigated, that could not be assigned to any species group, as well as for the clade comprising *C. rufijaliscensis, C. pararufinans* and *C. rufimichoacanensis* (Figure 3). The lability of these clades dramatically decreased the bootstrap values of several deep nodes, as attested by secondary bootstrap values computed after excluding samples with DNA barcodes only.

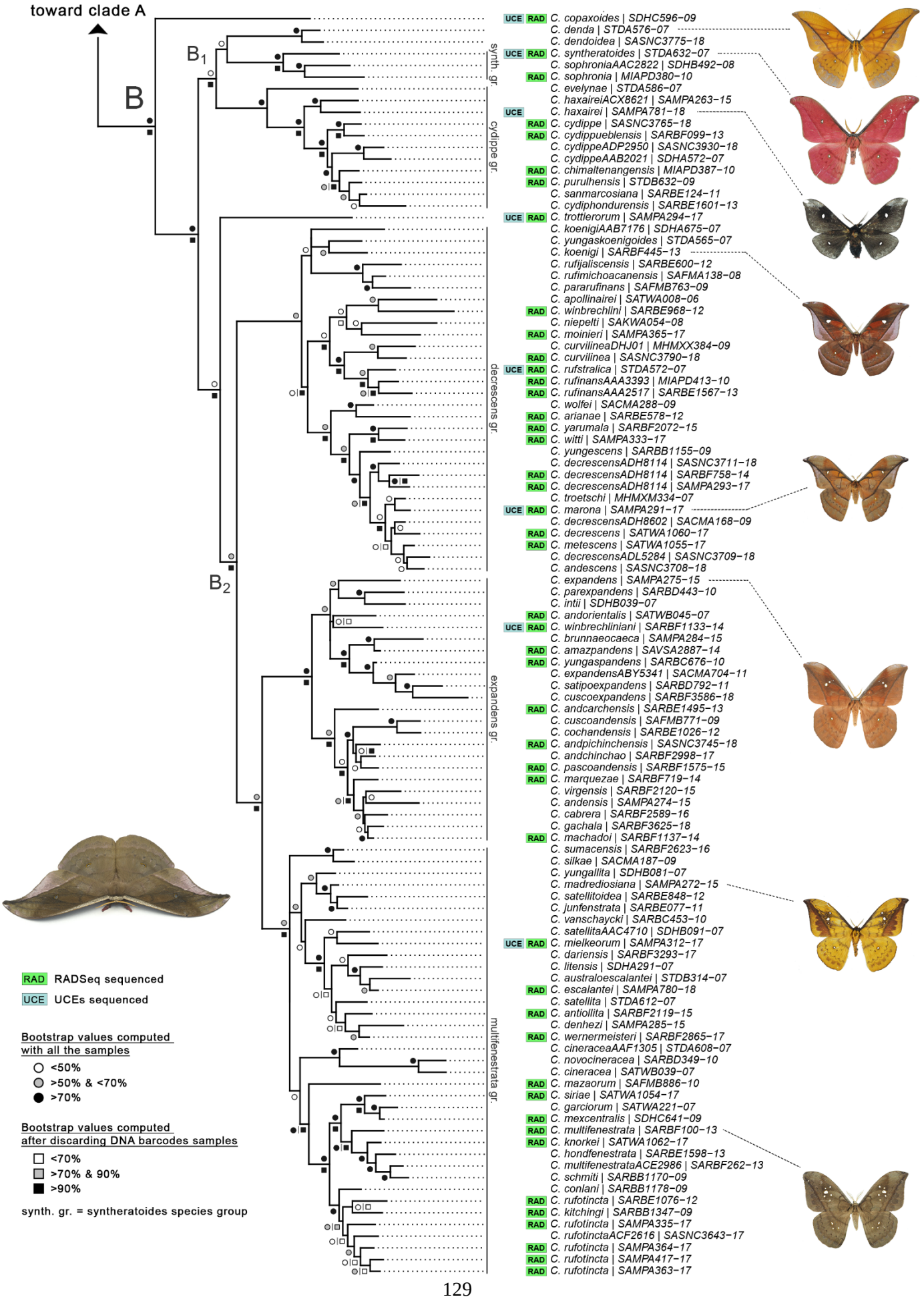
Notwithstanding those limits, our results support most of the morphological species group defined by Brechlin et al. (2016): the *cydippe*, *decrescens*, *expandens*, *medea* and *sapatoza* species groups are monophyletic and well supported (Figure 3). However, the *canella* and *multifenestrata* species groups are paraphyletic. Within the *canella* species group, Brechlin et al. (2016) defined the *lavendera* and *canella* subgroups that are well supported as distinct clades in our analyses, but not forming a monophyletic lineage. They are thus treated here as distinct species groups. We also found that *C. trottierorum*, *C. sophronia* and *C. syntheratoides* do not belong to the *multifenestrata* species group, which is however monophyletic when these species are excluded (Figure 3). *C. trottierorum* forms a very isolated lineage, with no extant relative, but *C. sophronia*, *C. sophronia*AAC2822 and *C. syntheratoides* are closely related and are here considered as part of a new species group: the *syntheratoides* species group. The subgroups defined by Brechlin et al. (2016) within the *decrescens* (*decrescens* and *rufinans* subgroups) and *expandens* (*expandens* and *andensis* subgroups) groups are paraphyletic and are abandoned here. Though previously grouped together in the *copaxoides* species group (Lemaire 1978, Brechlin et al. 2016), *C. mannana*, *C. muellerana* and *C. copaxoides* do not constitute a monophyletic group (Figure 3); they all are isolated lineages with no extant close relative and are then not assigned to any species group. As suggested by Lemaire (1978) on the basis of morphological characters, our analyses confirmed that the *medea* and *sapatoza* groups are sister groups. They form, together with the *lavendera* group what is here named clade A_2, itself sister to clade A_1, as defined above.

Within the B lineage, we define sub-clade B_1 as the clade comprising *C. denda*, *C. dendoidea* and the *syntheratoides* and *cydippe* species groups (Figure 3). We note that the position of *C. denda* and *C. dendoidea* in sub-clade B_1 remains poorly supported and we refrained from including both species within the *syntheratoides* group, although this relationship is to be considered the most likely according to our results. Subclade B_2 (Figure 3) includes the *decrescens*, *expandens* and *multifenestrata* species groups, which together include 2/3rd of species diversity within the genus (85 species and OTUs). Early divergences within B_2 are well supported, with the *decrescens* group branching first, then the *expandens* and *multifenestrata* groups being sister to each other. Shallower nodes, often only inferred from DNA barcode sequences remain only partially resolved within these three groups. Interestingly, the B lineage also includes two species representing long isolated lineages with no other extant representatives: *C. copaxoides*, sister to all other taxa in B clade, and *C. trottierorum*, sister to sub-group B_2. Together with *C. mannana* and *C. muellerana*, they form four monospecific lineages that were no assigned to any species groups.

The five species for which we sequenced RADSeq markers for multiple individuals were monophyletic, except for *C. rufotincta*. The paraphyly of this species is poorly supported and few RAD loci were available for most of the *C. rufotincta* samples (Table S3). In the following analyses we ignored that paraphyly and considered a unique representative for *C. rufotincta*.

Figure 3 (two following pages) - Maximum likelihood phylogenetic hypothesis for genus *Copaxa* built with IQTREE for 5 outgroups and 157 samples of *Copaxa* representing all 130 valid species within the genus and 18 OTUs; taxon names are preceded by symbols representing availability of genomic markers (RADSeq (green) and UCE (blue)); they are followed by the ProcessID code of the DNA barcode record used in the analysis (available for all terminals in BOLD dataset DS-COPAX1). The six main lineages discussed in the text are characterized as clades/sub-clades A, A_1, A_2, B, B_1, B_2 on branches, and species groups are indicated as vertical lines on the right side of the tree. Two-steps bootstrap support (BS) values are represented as symbols at nodes: circle represent regular BS values for the complete combined matrix (RADSeq + UCE + DNA barcodes, symbol colors as follow: black for BS>70%, grey for 50%<BS<70%, white for BS<50%); squares are secondary bootstrap supports after exclusion of records only represented by DNA barcodes (black for BS>90%, grey for 70%<BS<90%, white for BS<70%). Detailed BS values are given in Figure S4. Bottom left photograph by Armin Dett ©.




toward clade A
B
B1
B2
synth. gr.
cydippe gr.
decrescens gr.
expandens gr.
multifenestrata gr.
UCE RAD C. copaxoides | SDHC596-09
C. denda | STDA576-07
C. dendoidea | SASNC3775-18
UCE RAD C. syntheratoides | STDA632-07
C. sophroniaAAC2822 | SDHB492-08
RAD C. sophronia | MIAPD380-10
C. evelynae | STDA586-07
C. haxaireiACX8621 | SAMPA263-15
UCE C. haxairei | SAMPA781-18
RAD C. cydippe | SASNC3765-18
RAD C. cydippueblensis | SARBF099-13
C. cydippeADP2950 | SASNC3930-18
C. cydippeAAB2021 | SDHA572-07
RAD C. chimaltenangensis | MIAPD387-10
RAD C. purulhensis | STDB632-09
C. sanmarcosiana | SARBE124-11
C. cydiphondurensis | SARBE1601-13
UCE RAD C. trottierorum | SAMPA294-17
C. koenigiAAB7176 | SDHA675-07
C. yungaskoenigoides | STDA565-07
C. koenigi | SARBF445-13
C. rufijaliscensis | SARBE600-12
C. rufimichoacanensis | SAFMA138-08
C. pararufinans | SAFMB763-09
C. apollinairei | SATWA008-06
RAD C. winbrechlini | SARBE968-12
C. niepelti | SAKWA054-08
RAD C. moinieri | SAMPA365-17
C. curvilineaDHJ01 | MHMXX384-09
RAD C. curvilinea | SASNC3790-18
UCE RAD C. rufstralica | STDA572-07
RAD C. rufinansAAA3393 | MIAPD413-10
RAD C. rufinansAAA2517 | SARBE1567-13
C. wolfei | SACMA288-09
RAD C. arianae | SARBE578-12
RAD C. yarumala | SARBF2072-15
RAD C. witti | SAMPA333-17
C. yungescens | SARBB1155-09
C. decrescensADH8114 | SASNC3711-18
RAD C. decrescensADH8114 | SARBF758-14
RAD C. decrescensADH8114 | SAMPA293-17
C. troetschi | MHMXM334-07
UCE RAD C. marona | SAMPA291-17
C. decrescensADH8602 | SACMA168-09
RAD C. decrescens | SATWA1060-17
RAD C. metescens | SATWA1055-17
C. decrescensADL5284 | SASNC3709-18
C. andescens | SASNC3708-18
C. expandens | SAMPA275-15
C. parexpandens | SARBD443-10
C. intii | SDHB039-07
RAD C. andorientalis | SATWB045-07
UCE RAD C. winbrechliniani | SARBF1133-14
C. brunnaeocaeca | SAMPA284-15
RAD C. amazpandens | SAVSA2887-14
RAD C. yungaspandens | SARBC676-10
C. expandensABY5341 | SACMA704-11
C. satipoexpandens | SARBD792-11
C. cuscoexpandens | SARBF3586-18
RAD C. andcarchensis | SARBE1495-13
C. cuscoandensis | SAFMB771-09
C. cochandensis | SARBE1026-12
RAD C. andpichinchensis | SASNC3745-18
C. andchinchao | SARBF2998-17
RAD C. pascoandensis | SARBF1575-15
RAD C. marquezae | SARBF719-14
C. virgensis | SARBF2120-15
C. andensis | SAMPA274-15
C. cabrera | SARBF2589-16
C. gachala | SARBF3625-18
RAD C. machadoi | SARBF1137-14
C. sumacensis | SARBF2623-16
C. silkae | SACMA187-09
C. yungallita | SDHB081-07
C. madrediosiana | SAMPA272-15
C. satellitoidea | SARBE848-12
C. junfenstrata | SARBE077-11
C. vanschaycki | SARBC453-10
C. satellitaAAC4710 | SDHB091-07
UCE RAD C. mielkeorum | SAMPA312-17
C. dariensis | SARBF3293-17
C. litensis | SDHA291-07
C. australoescalantei | STDB314-07
RAD C. escalantei | SAMPA780-18
C. satellita | STDA612-07
RAD C. antiollita | SARBF2119-15
C. denhezi | SAMPA285-15
RAD C. wernermeisteri | SARBF2865-17
C. cineraceaAAF1305 | STDA608-07
C. novocineracea | SARBD349-10
C. cineracea | SATWB039-07
RAD C. mazaorum | SAFMB886-10
RAD C. siriae | SATWA1054-17
C. garciorum | SATWA221-07
RAD C. mexcentralis | SDHC641-09
RAD C. multifenestrata | SARBF100-13
RAD C. knorkei | SATWA1062-17
C. hondfenestrata | SARBE1598-13
C. multifenestrataACE2986 | SARBF262-13
C. schmiti | SARBB1170-09
C. conlani | SARBB1178-09
RAD C. rufotincta | SARBE1076-12
RAD C. kitchingi | SARBB1347-09
RAD C. rufotincta | SAMPA335-17
C. rufotinctaACF2616 | SASNC3643-17
RAD C. rufotincta | SAMPA364-17
RAD C. rufotincta | SAMPA417-17
RAD C. rufotincta | SAMPA363-17
RAD RADSeq sequenced
UCE UCEs sequenced
Bootstrap values computed with all the samples
<50%
>50% & <70%
>70%
Bootstrap values computed after discarding DNA barcodes samples
<70%
>70% & 90%
>90%
synth. gr. = syntheratoides species group

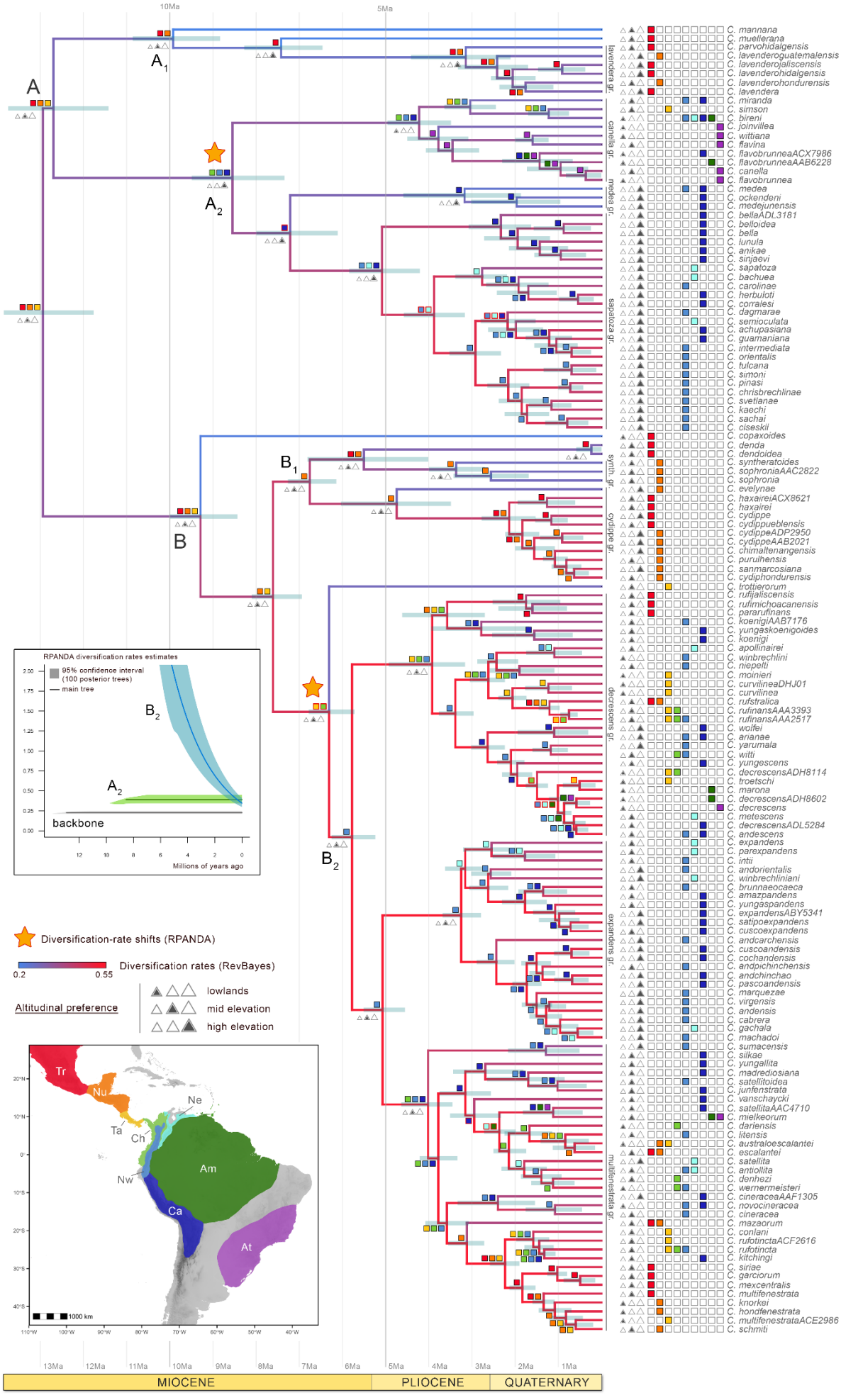
Estimates of divergence times

We re-estimated the divergence times of the Saturniidae family with a reduced dataset from Arnal et al. (in prep.; Chapitre 1) including three *Copaxa* samples. All MCMCtree analyses converged: effective sample sizes (ESS) were superior to 200 and the two independent runs led to identical estimates (Figure S5) that are very congruent with those in Arnal et al. (in prep.; Chapitre 1). We thus obtained two internal calibration points within the *Copaxa* phylogeny: between *C. decrescens* and *C. syntheratoides* (4.0-7.8 Ma) and between *C. decrescens/C. syntheratoides* and *C. lavenderoguatemalensis* (7.7-13.6 Ma). We then estimated the divergence times of the *Copaxa* phylogeny. The two independent runs launched on two randomly sampled sets of loci converged (ESS>200) and led to identical estimates (Figure S6). The results of divergence time estimation are further presented in the following paragraph along with inferred spatial and altitudinal diversification dynamics in the genus (Figure 4).

Spatial, temporal and altitudinal diversification dynamics of *Copaxa*

Our biogeographical model recovered with strong support that the most recent common ancestor of *Copaxa* was widely distributed in Central America, in the Transitional, Nuclear and Talamanca areas, *ca.* 13Ma ([11.59-13.75], Figure 4). Reconstruction of ancestral states revealed that the genus ancestors were most likely adapted to mid-elevation ranges. Most of the early diverging lineages, in both A and B clades, remained in Central America. This is the case of all the "relictual" taxa that do not belong to any recognized species group and have no extant relatives: *C. copaxoides*, *C. mannana*, *C. muellerana* and *C. trottierorum*. It is also the case of the *lavendera*, *syntheratoides* and *cydippe* species groups. Interestingly, these lineages, except for the *cydippe* species group (see discussion), poorly diversified and they all occur at intermediate or high altitudes (except *C. copaxoides* sampled over a broad altitudinal range).

Figure 4 (following page) – Historical biogeography dynamics of the *Copaxa* lineages. Blue bars at nodes represent the 95% confidence intervals of clade ages as estimated with MCMCtree (time-scale and main geological epochs are given at the bottom of the tree). The six main lineages discussed in the text are named as clades/sub-clades A, A_1, A_2, B, B_1, B_2; species groups are indicated as vertical lines on the right side of the tree. Biogeographical areas considered in this study are given as colored ranges in the bottom left corner map; the species ranges are represented at leaves using corresponding colors, as are the results of the ancestral reconstruction of biogeographical ranges (inferred with BioGeoBEARS) at nodes; red border indicates low confidence ($p<0.33$) in the estimates. Branches are colored according to the diversification rates estimated with RevBayes. Altitudinal preferences of extant species are depicted at leaves, following the legend on the left hand side of the tree; ancestral states, as estimated with *diversitree*, are indicated at nodes of the phylogeny backbone (see Figure S7 for a complete visualization of these estimates on the tree). Yellow stars represent the positive diversification rate shifts recovered with RPANDA. The graph on the left shows the diversification rate regimes estimated with RPANDA for clades A_2 and B_2, as well as for the backbone.


10Ma
5Ma
A
A1
A2
B
B1
B2
RPANDA diversification rates estimates
95% confidence interval (100 posterior trees)
main tree
backbone
Millions of years ago
Diversification-rate shifts (RPANDA)
Diversification rates (RevBayes)
0.2
0.55
Altitudinal preference
lowlands
mid elevation
high elevation
Tr
Nu
Ne
Ta
Ch
Am
Nw
Ca
At
1000 km
13Ma
12Ma
11Ma
10Ma
9Ma
8Ma
7Ma
6Ma
5Ma
4Ma
3Ma
2Ma
1Ma
MIOCENE
PLIOCENE
QUATERNARY
lavendera gr.
canella gr.
medea gr.
espadaza gr.
synth. gr.
cydippe gr.
decrescens gr.
expandens gr.
multifenestrata gr.
C. mannana
C. muellerana
C. parvohidalgensis
C. lavenderoguatemalensis
C. lavenderojaliscensis
C. lavenderohidalgensis
C. lavenderohondurensis
C. lavendera
C. miranda
C. simson
C. bireni
C. joinvillea
C. wittiana
C. flavina
C. flavobrunneaACX7986
C. flavobrunneaAAB6228
C. canella
C. flavobrunnea
C. medea
C. ockendeni
C. medejunensis
C. bellaADL3181
C. belloidea
C. bella
C. lunula
C. anikae
C. sinjaevi
C. sapatoza
C. bachuea
C. carolinae
C. herbuloti
C. corralesi
C. dagmarae
C. semioculata
C. achupasiana
C. guamaniana
C. intermediata
C. orientalis
C. fulcana
C. simoni
C. pinasi
C. chrisbrechlinae
C. svetlanae
C. kaechi
C. sachai
C. ciseskii
C. copaxoides
C. denda
C. dendoidea
C. syntheratoides
C. sophroniaAAC2822
C. sophronia
C. evelynae
C. haxaireiACX8621
C. haxairei
C. cydippe
C. cydippueblensis
C. cydippeADP2950
C. cydippeAAB2021
C. chimaltenangensis
C. purulhensis
C. sanmarcosiana
C. cydiphondurensis
C. trottierorum
C. rufijalcoensis
C. rufimichoacanensis
C. pararufinans
C. koenigiAAB7176
C. yungaskoenigoides
C. koenigi
C. apollinairei
C. winbrechlini
C. niepelti
C. moinieri
C. curvilineaDHJ01
C. curvilinea
C. rufstralica
C. rufinansAAA3393
C. rufinansAAA2517
C. wolfei
C. arianae
C. yarumala
C. witti
C. yungescens
C. decrescensADH8114
C. troetschi
C. marona
C. decrescensADH8602
C. decrescens
C. metescens
C. decrescensADL5284
C. andescens
C. expandens
C. parexpandens
C. inti
C. andorientalis
C. winbrechliniani
C. brunnaeocaeca
C. amazpandens
C. yungaspandens
C. expandensABY5341
C. satipoexpandens
C. cuscoexpandens
C. andcarchensis
C. cuscoandensis
C. cochandensis
C. andpichinchensis
C. andchinchao
C. pascoandensis
C. marquezae
C. virgensis
C. andensis
C. cabrera
C. gachala
C. machadoi
C. sumacensis
C. silkae
C. yungallita
C. madrediosiana
C. satellitoidea
C. junfenstrata
C. vanschaycki
C. satellitaAAC4710
C. mielkeorum
C. dariensis
C. litensis
C. australoescalantei
C. escalantei
C. satellita
C. antiollita
C. denhezi
C. wernermeisteri
C. cineraceaAAF1305
C. novocineracea
C. cineracea
C. mazaorum
C. conlani
C. rufotinctaACF2616
C. rufotincta
C. kitchingi
C. siriae
C. garciorum
C. mexcentralis
C. multifenestrata
C. knorkei
C. hondfenestrata
C. multifenestrataACE2986
C. schmiti

In striking contrast with these clades that experienced a limited spatial and temporal diversification, we identified two independent dispersal events towards the South American landmass during the Miocene period, after the closure of the Central American Seaway (CAS) *ca.* 10 Ma. The earliest event happened along the branch leading to clade A_2; then the second one along the branch leading to the clade formed by *C. trottierorum* and sub-clade B_2. The analyses performed with RPANDA revealed that both events were associated with positive diversification rate shifts (Figure 4, Table 1).

Table 1 - Results of time-dependent models of diversification fitted on the different shift configurations with RPANDA. For each subtree and for the backbone, only the best model is shown and the following column indicate the percentage of posterior dated trees for which we recovered a similar model of diversification. According to the AICc scores, the best shift configuration is that considering two diversification rate shifts at nodes A_2 and B_2. BCST: constant speciation model; BVAR: time-dependent speciation model; λ: speciation rate at present; α: coefficient of time variation of the speciation rate.

Shifts	Subtree	Best Model	% of replicates supporting the same model	Nb. Param	logL (consensus tree)	AICc	λ	α
No Shift	whole tree	BCST	100 %	1	-277.05	**556.12**	0.408	-
A_2	backbone	BCST	98 %	1	-203.73	409.49	0.412	-
	A_2	BCST	100 %	1	-141.69	141.69	0.391	-
	total			2	-345.42	**551.18**		
B_2	backbone	BCST	100 %	1	-131.43	264.93	0.315	-
	B_2	BVAR	100 %	2	-134.38	272.91	0.346	0.243
	total			3	-265.81	**537.84**		
$A_2 + B_2$	backbone	BCST	100 %	1	-59.42	121.00	0.229	-
	A_2	BCST	100 %	1	-141.69	141.69	0.391	-
	B_2	BVAR	100 %	2	-134.38	272.91	0.346	0.243
	total			4	-335.49	**535.60**		

Among the four shift configurations tested, the scenario considering two shifts is the model with the lowest AICc (ΔAICc>20 in comparison to the null model; Table 1). In this model, we estimated that the phylogeny backbone fits a model with a constant speciation rate λ=0.229 speciation event per lineage per million years. It is also the case of the A_2 sub-clade, but with a significantly higher speciation rate λ = 0.391. We estimated that the B_2 sub-clade best fits a model where the speciation rates are time dependent (BVAR): speciation rates are very high near the root of the B_2 sub-clade and exponentially decrease with time to reach similar rates to those of A_2 (see chart in Figure 4). We obtained similar results for each tree sampled in the posterior distribution of the MCMCtree analysis (Table 1; chart in Figure 4). These results are congruent with those obtained using the RevBayes model that inferred branch-specific diversification rates (Figure 4). The A_2 and B_2 lineages indeed exhibit much higher diversification rates compared to other *Copaxa* clades. This is especially conspicuous in clade B_2 with most lineages exhibiting speciation rates higher than 0.5 speciation event/million years. These rates are

more heterogeneous within the clade A_2: the *canella* and *medea* species group diversified more slowly than the *sapatoza* group, whose rates are comparable to those of clade B_2.

The ancestors of clade A_2 colonized the Andes between 12.7 and 8.6 Ma [13.6-7.3 Ma] (Figure 4). Then its three species groups experienced very contrasted spatial and temporal diversification. On one hand, the *sapatoza* and *medea* groups remained in the Andes, both derived from a shared ancestor inferred as having been inhabiting high elevation environments (Figs. 4 and S7). Indeed, all extant species are found at very high elevation (several species well above 3000m asl). The *sapatoza* group however experienced higher diversification than the *medea* group. No known member of these groups ever dispersed out of the Andes and most speciation events occurred within the same biogeographical region. On the other hand, the lineage currently represented by species of the *canella* group specialized in inhabiting lowland environments (Figs. 4 and S7) and experienced multiple dispersal events toward the different areas of South America. A lineage reached the Atlantic forest *ca.* 4Ma [4.9-3.1 Ma], probably before the formation of the dry Cerrado region acting as a barrier to dispersal. This is the earliest colonization of the Atlantic forest by genus *Copaxa* and this lineage is now the most diverse in this region (5 out of the 7 species). It is noteworthy that members of the *canella* lineage successfully traveled across the Cerrado to colonize again the Amazonian forest and the Central Andes *ca.* 2.5 Ma [4.0-0.7 Ma].

Our results highlight a second colonization of South America in the B clade by the common ancestor of clade B_2 and *C. trottierorum*. This event is younger (*ca.* 7 Ma [8.1-5.6 Ma]) than that described above in clade A, but it still predates the Panama strait closure. Sub-clade B_2 initially diversified at a high pace, rapidly splitting into three lineages now represented by species of the *expandens*, *decrescens* and *multifenestrata* groups. Interestingly the latter two followed a similar pattern of spatial and temporal diversification. Both lineages experienced tremendous biogeographical dynamics through South and Central America and comprise species occurring at various altitudinal ranges. Whereas poorly resolved relationships within these groups prevent us to reliably infer the number and the timing of biogeographical events (Figure 4), our analyses nonetheless imply multiple northward colonization events of Central America (*e.g. C. troetschi*, *C. escalantei* or the common ancestors of *C. conlani* and *C. knorkei*). These are most likely posterior to the closure of the Panama strait. We also evidenced one independent colonization event of the Atlantic forest within each of the *decrescens* and *multifenestrata* lineages. Our analyses suggest that these events likely happened through the Amazonian forest, unlike the older colonization of the Atlantic forest region within the *canella* group, likely achieved through the Central Andes. They also largely postdate the formation of the Cerrado, thus raising to three the number of dispersal event across the Cerrado. Mirroring our previous observations in sub-clade A_2, these many dispersal events between biogeographical regions can be linked to the ability of lineages to inhabit lowland habitats (Figures 4 and S7; note that *C. mielkeorum*, categorized as 'mid-elevation' species, has an altitudinal median of 800m and thus could be also considered a lowland species). Overall, the altitudinal preferences (Figure S7) are rather phylogenetically conserved within all species groups,

except for the *decrescens* and *multifenestrata* ones that appear very labile with respect to this environmental preference. For example, the close relationships of *C. cineraceaAAF1305* (median: 2000m) or *C. satellita* (2080m), with species inhabiting lowland or mid-elevation environments (Figure S7) suggest that altitudinal preferences can evolve quickly. The third species group of the clade B_2, the *expandens* group, differs from the two others in being restricted to mid to high-elevation areas, exclusively in the three Andean regions. This group mirrors our observations for the *sapatoza* and *medea* species groups, also inhabiting high-elevation environments and where closely related species tend to occur in the same biogeographical region, thus bringing further evidence of a link between niche conservatism and spatial dynamics (see Discussion).

Discussion

Our study produced a clear overview of the spatial and temporal diversification dynamics of *Copaxa* moths following the colonization of the Neotropics by their ancestors. These dynamics were characterized through the reconstruction of a phylogeny combining genomic and genetic markers and including all known extant species in the genus. DNA barcodes were available for all described valid species and several newly recognized OTUs, and they were instrumental in defining the evolutionary units considered in our analyses, offering a dramatically improved understanding of species diversity if compared to taxonomic treatment of the genus in the early 2000's (Figure 1; 41 species vs. 130 species + 19 OTUs). RADSeq markers, frequently used in population genetics and phylogeography (e.g. Dupuis et al. 2018), proved useful for resolving inter-specific relationships, as predicted *in silico* (Cariou et al. 2013) and as shown in other organisms (e.g. Cruaud et al. 2014; Hipp et al. 2014). Our analyses showed however that they should be analyzed carefully, especially with respect to the influence on the topologies of analytical parameters used to assemble nucleotide matrices (see also Rubin et al. 2012; Leaché et al. 2015). Here we combined them with UCE loci to gain resolution at deeper nodes and with DNA barcodes to complete taxon sampling. The latter brought additional phylogenetic resolution at shallower nodes (e.g. in the *cydippe* and *canella* groups; Figure 3), albeit limited in groups that diversified at higher pace (e.g. in the *decrescens* group; see Figure 4) or in isolated taxa that lacked sufficient coverage for genomic markers (e.g. the position *C. denda* and *C. dendoidea*; Figure 3). This is the first time, to our knowledge, that such combined approach is employed in resolving phylogenetic relationships in a diverse group, taking advantage of DNA barcode data to improve the selection of samples for the genomic approach and to permit the inference of the position of all species based on an objective analysis of characters.

The resulting phylogenetic hypothesis for the genus is very resolved, except for the position of clade A_1 (Figure 3), and it permitted to track back the history and dynamics of diversification of these moths after their lineage diverged from its sister-group, the genus *Saturnia*. Arnal et al. (in prep.; Chapitre 1) inferred that the common ancestor to both *Saturnia* and *Copaxa* colonized the Nearctic region (and from there

Central America) through Beringia and its mixed deciduous coniferous forests (Norris 1982; Wolfe 1987) about 27Ma (30.8-23.5Ma), probably benefiting – like several other groups of organisms (Jiang et al. 2019) - of climate warming (Zachos et al. 2001, 2008) during the late Oligocene and of the emergence of a dispersal corridor formed by a deciduous, humid forest along the Eastern Palearctic coasts (Sun & Wang 2005). Our knowledge of the food plants used by *Copaxa* caterpillars in the wild remains very fragmentary (Wolfe, 1993), but the known associations of several species with oak (e.g. *C. mannana, C. muellerana, C. copaxoides, C. lavendera*; Wolfe, 1993), alder (*C. sapatoza*; Wolfe et al. 2003b) or *Pinus* trees (e.g. *C. medea* (see Wolfe et al. 2003a), *C. cydippe* (see Wolfe, 1988), *C. haxairei* (unpublished observation)) support this scenario of the origin of genus *Copaxa* and of the ability of its ancestor to thrive in the Nearctic and Central American regions. It is likely that, while Central American lineages were settling and diversifying, the Nearctic *Copaxa* lineages went extinct after the mid-Miocene and the CAS closure (Sepulchre et al. 2014; Montes 2012, 2015; Jaramillo et al. 2017; Zachos et al. 2008) drastically impacted Nearctic climate (Wolfe 1994), with decreasing temperatures.

In Central America the *Copaxa* ancestors diverged into two clades *ca.* 13 Ma (clades A and B, Figure 4) that experienced remarkably similar patterns of spatial and temporal diversification. In both clades, most of the early-diverging lineages remained in Central America and experienced limited diversification; some are only represented by a single extant species, like *C. mannana, C. muellerana, C. copaxoides, C. trottierorum* and may deserve special consideration for their conservation because of their evolutionary "uniqueness" (Vane-Wright et al. 1991; Isaac et al. 2007). One noticeable exception is the *cydippe* group (Figure 4), whose diversification rates were found to be significantly higher (at the exception of *C. evelynae*) than in other representatives of the B_1 sub-clade. This increase in diversification rate may be linked to a food plant shift and fits an 'escape and radiate' diversification mechanism (Erhlich & Raven 1964) after species in this group specialized to feed on Gymnosperms (genus *Pinus*; see Wolfe 1993).

The results of our biogeographical analyses revealed that *Copaxa* dispersed eight times across the Panama strait (Figure 4) but only three times toward South-America. One is recent (Figure 4; less than 3Ma, in the *multifenestrata* group), but the two others date back to *ca.* 8.5-12.7 Ma and 6.3-7.6 Ma along the branches leading to clade A_2 and to *C. trottierorum* + clade B_2, respectively (Figure 4), likely occurring after the CAS closure, 10 Ma (Sepulchre et al. 2014; Montes 2012, 2015; Jaramillo et al. 2017). These results support the hypothesis that the Panama strait acted as an efficient geographical barrier after the CAS closure and until 4-3Ma, preventing further interchanges in both directions despite its narrowness (Coates & Stallard, 2013). As most *Copaxa* moths are associated to montane environments, dispersal across the Panama strait might also have long been impeded by the lower elevation of the Central American Volcanic Arc in Miocene (Coates et al. 2004), thus offering no connection between the Andes and the Cordillera de Talamanca until the upraise of the Panama isthmus.

Interestingly, the two *Copaxa* lineages that managed to disperse early into South-America experienced a significant positive shift of diversification (Figure 4 and Table 1). This is similar to patterns documented by Moore & Donoghue (2007) and termed 'dispersification' to describe positive shifts of diversification associated with a dispersal event into new geographical regions. Dispersification events were found in several groups of organisms and on different continents (*e.g.* Lemuriformes in Madagascar; Springer et al. 2012), but also specifically in the Andean region for several plant groups (e.g. in genus *Bartsia* (Orobanchaceae) (Uribe-Convers & Tank 2015), *Viburnum* (Adoxaceae) or *Valeriana* (Valerianaceae) genera (Moore & Donoghue 2007)). However, our results represent, to our knowledge, the first documented case in insects of dispersification events following colonization of South-America from Central America by a group of Nearctic origin. Interestingly, *Copaxa* moths are largely associated to montane environments (126 species or OTUs (85%) occur at mid- or high-elevation; Figure 4) and our reconstruction of ancestral states unambiguously infer that the *Copaxa* ancestor inhabited mid-elevation habitats in Central America (Figures 4 and S7). Considering the Nearctic origin of the ancestor of *Copaxa*, we hypothesize that this strong link to montane environments is the result of Phylogenetic Niche Conservatism (PNC) and of the inheritance of traits from this ancestor adapted to temperate habitats (Donoghue, 2008). These traits might be related to physiological adaptations, linked to cooler climate and seasonality (e.g. regulation of flight activity, phenology, diapause), as well as to adaptations of their caterpillars and of adult females (when selecting oviposition sites) to food plants, themselves constrained to this mountain climatic niche (e.g. oak or alder trees). PNC here appears as having played a major role in the evolutionary history of *Copaxa*, both in driving their diversification dynamics in Central American mountains where the genus originated, but also after reaching South-America and the Andean subregions where the genus thrived and became most diverse. As described by Donoghue (2008) for plants, PNC promoted the dispersal of pre-adapted *Copaxa* moths to new emerging mountain environments of the South-American continent where speciation occurred mostly without major niche shifts but rather through dispersal or vicariance. Furthermore, such mechanism is exacerbated in tropical regions where seasonal temperature amplitude does not exceed that of the elevational temperature gradient, thus making topographic barriers more important in tropical than in temperate regions (Janzen 1967). For organisms adapted to life in mountain habitats, the Neotropical mountains can be compared to large and isolated archipelago systems characterized by altitudinal gradients and various exposure, geologic and edaphic conditions where the gene flow is limited (Graves 1988; Fjeldså et al. 2012). In *Copaxa*, these groups that diversified at high elevation (*sapatoza* and *expandens* species groups) exhibit high diversification rates (Figure 4) and each of their extant representatives are geographically restricted to a unique Andean sub-region (except *C. medea*); also, closely related species tend to be distributed in the same sub-region.

While PNC played a major role in driving the diversification dynamics of *Copaxa* moths, our results also highlighted that several lineages in the genus independently succeeded in colonizing lowland

habitats (23 species or OTUs (15%), in the *canella*, *multifenestrata* and *decrescens* species groups; Figures. 4 and S7). Such niche shift did not have a significant impact on diversification rates, but interestingly these three groups experienced the strongest spatial dynamics, colonizing new biogeographical regions (Atlantic Forest, Amazonia), and dispersing back into Central America. We note that one of these lowland species, *C. bireni* (*canella* group) stands out as being the sole species distributed over 4 biogeographical regions (Figure 4).

Conclusions

Our detailed reconstruction of the spatial and temporal evolutionary history of genus *Copaxa* revealed a very clear picture of the diversification dynamics of this genus 'into the Neotropics' after it originated from a Holarctic ancestor. It is marked by two key events of dispersification and a predominant role of PNC. These dynamics are similar to those already documented in plant groups (Hughes & Eastwood 2006; Moore & Donoghue 2007; Sklenář et al. 2011), but were rarely reported in animals. A similar 'into the tropics' spatial diversification pattern was found in other groups of Lepidoptera (Mullen et al. 2011; Condamine et al. 2012), but those thrived in lowland Amazonia rather than in mountainous regions of Central and South-America. This instance of a Holarctic insect group successfully colonizing and diversifying in Neotropical mountains may be exceptional. In fact, the Andes may have been mostly colonized by autochthonous Neotropical lineages as documented for Godyridina butterflies, from Amazonia (Chazot et al. 2016, 2019), or *Eois* moths, from lowland Central America (Jahner et al 2017). We note that within Saturniidae Arnal et al (in prep.; Chapitre 1) reported two other genera that colonized South America and originated from an Holarctic ancestor: *Rothschildia* (Saturniinae: Attacini), with dispersification dynamics suspected to be comparable, at least in part, to that of *Copaxa*, and *Antheraea* (subgenus *Telea;* Saturniinae: Saturniini), that in contrast experienced a very limited diversification. A thorough investigation of the spatial and temporal dynamics of these genera and of autochthonous Neotropical saturniid lineages that colonized Andean mountains will bring further insight on the respective contributions of these different sources and routes, as well as PNC and adaptive radiation in generating the world richest region for saturniid moths.

References

Aberer, A. J., Krompass, D., & Stamatakis, A. (2013). Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Systematic Biology, 62(1), 162-166.

Andersen, M. J., McCullough, J. M., Friedman, N. R., Peterson, A. T., Moyle, R. G., Joseph, L., & Nyári, Á. S. (2019). Ultraconserved elements resolve genus-level relationships in a major Australasian bird radiation (Aves: Meliphagidae). Emu - Austral Ornithology, 119(3), 218-232.

Antonelli, A., & Sanmartin, I. (2011). Why are there so many plant species in the Neotropics? Taxon, 60(2), 403-414.

Antonelli, A., Zizka, A., Carvalho, F. A., Scharn, R., Bacon, C. D., Silvestro, D., & Condamine, F. L. (2018). Amazonia is the primary source of Neotropical biodiversity. Proceedings of the National Academy of Sciences USA, 115(23), 6034-6039.

Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P., & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proceedings of the National Academy of Sciences USA, 112(19), 6110-6115.

Barat, F., Mercier de Lépinay, B., Sosson, M., Müller, C., Baumgartner, P. O., & Baumgartner-Mora, C. (2014). Transition from the Farallon Plate subduction to the collision between South and Central America: Geological evolution of the Panama Isthmus. Tectonophysics, 622, 145-167.

Barber, J. R., Leavell, B. C., Keener, A. L., Breinholt, J. W., Chadwell, B. A., McClure, C. J. W., ... & Kawahara, A. Y. (2015). Moth tails divert bat attack: Evolution of acoustic deflection. Proceedings of the National Academy of Sciences USA, 201421926.

Bello, M. A., Chase, M. W., Olmstead, R. G., Rønsted, N., & Albach, D. (2002). The páramo endemic *Aragoa* is the sister genus of *Plantago* (Plantaginaceae; Lamiales): evidence from plastid rbcL and nuclear ribosomal ITS sequence data. Kew Bulletin, 57(3), 585-597.

Blaimer, B. B., Lloyd, M. W., Guillory, W. X., & Brady, S. G. (2016). Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PLoS One, 11(8), e0161531.

Borgardt, S. J., & Pigg, K. B. (1999). Anatomical and developmental study of petrified *Quercus* (Fagaceae) fruits from the Middle Miocene, Yakima Canyon, Washington, USA. American Journal of Botany, 86(3), 307-325.

Bouvier, E.-L. (1929). *Seconde contribution à la connaissance des Saturnioïdes du Hill Museum*.

Bouvier, E.-L. (1936). Etude des Saturnioides normaux. Famille des Saturniidés. Mémoires du Muséum National d'Histoire Naturelle, Nouvelle Série, 3, 1-354.

Brechlin, R. (2018). Two new species of the genus *Copaxa* Walker, 1855 from Colombia and Panama (Lepidoptera: Saturniidae). Entomo-Satsphingia, 11(3), 5-7.

Brechlin, R. (2019a). *Copaxa wittiana* n. sp., a new saturniid from South America (Lepidoptera). Entomo-Satsphingia, 12(2), 5-9.

Brechlin, R. (2019b). Two new species in the genus *Copaxa* Walker, 1855 (Lepidoptera: Saturniidae). Entomo-Satsphingia, 12(3), 57-61.

Brito, D. (2010). Overcoming the Linnean shortfall: data deficiency and biological survey priorities. Basic and Applied Ecology, 11(8), 709-713.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. (2013). Stacks: an analysis tool set for population genomics. Mol Ecol, 22.

Chazot, N., Willmott, K. R., Condamine, F. L., De-Silva, D. L., Freitas, A. V., Lamas, G., ... & Elias, M. (2016). Into the Andes: multiple independent colonizations drive montane diversity in the Neotropical clearwing butterflies Godyridina. Mol Ecol, 25(22), 5765-5784.

Chazot, N., Willmott, K. R., Lamas, G., Freitas, A. V. L., Piron-Prunier, F., Arias, C. F., ... & Elias, M. (2017). Renewed diversification following Miocene landscape turnover in a Neotropical butterfly radiation. BioRxiv.

Coates, A. G., & Stallard, R. F. (2013). How old is the Isthmus of Panama? Bulletin of Marine Science, 89(4), 801-813.

Condamine, F. L., Silva-Brandão, K. L., Kergoat, G. J., & Sperling, F. A. (2012). Biogeographic and diversification patterns of Neotropical Troidini butterflies (Papilionidae) support a museum model of diversity dynamics for Amazonia. BMC Evolutionary Biology, 12(1), 82-82.

Cruaud, A., Gautier, M., Galan, M., Foucaud, J., Sauné, L., Genson, G., ... & Rasplus, J.-Y. (2014). Empirical assessment of RAD sequencing for interspecific phylogeny. Mol Biol Evol, 31(5), 1272-1274.

Cruaud, A., Gautier, M., Rossi, J.-P., Rasplus, J.-Y., & Gouzy, J. (2016). RADIS: Analysis of RAD-seq data for InterSpecific phylogeny. Bioinformatics.

Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., ... & Rasplus, J. Y. (2019). Optimized DNA extraction and library preparation for minute arthropods: Application to target enrichment in chalcid wasps used for biocontrol. Mol Ecol Resour.

Daghlian, C. P., & Crepet, W. L. (1983). Oak catkins, leaves and fruits from the oligocene catahoula formation and their evolutionary significance. American Journal of Botany, 70(5), 639-649.

de Vienne, D. M., Ollier, S., & Aguileta, G. (2012). Phylo-MCOA: A Fast and Efficient Method to Detect Outlier Genes and Species in Phylogenomics Using Multiple Co-inertia Analysis. Mol Biol Evol, 29(6), 1587-1598.

Deuve, T., Cruaud, A., Genson, G., & Rasplus, J.-Y. (2012). Molecular systematics and evolutionary history of the genus *Carabus* (Col. Carabidae). Mol Phylogenet Evol, 65(1), 259-275.

deWaard, J., Ivanova, N. V., Hajibabaei, M., & Hebert, P. D. N. (2008). Assembling DNA barcodes: analytical methods. In C. C. Martin (Ed.), *Methods in Molecular Biology 410: Environmental Genetics* (pp. 275-293). Totowa, USA: Humana Press Inc.

Diniz-Filho, J. A., Loyola, R. D., Raia, P., Mooers, A. O., & Bini, L. M. (2013). Darwinian shortfalls in biodiversity conservation. Trends in Ecology and Evolution, 10.1016/j.tree.2013.09.003.

Diniz-Filho, J. A. F., De Marco Jr, P., & Hawkins, B. A. (2010). Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. Insect Conservation and Diversity, 3, 172-179.

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. Proceedings of the National Academy of Sciences USA, 105 Suppl 1, 11549-11555.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7, 214.

Dupuis, J. R., Peigler, R. S., Geib, S. M., & Rubinoff, D. (2018). Phylogenomics supports incongruence between ecological specialization and taxonomy in a charismatic clade of buck moths. Mol Ecol, 27(22), 4417-4429.

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: a study in coevolution. Evolution, 18(4), 586-608.

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods in Molecular Biology, 772, 157-178.

Fang, Y., Shi, W.-Q., & Zhang, Y. (2017). Molecular phylogeny of *Anopheles hyrcanus* group (Diptera: Culicidae) based on mtDNA COI. Infectious diseases of poverty, 6(1), 61-61.

FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution, 3(6), 1084-1092.

Fjeldså, J., Bowie, R. C. K., & Rahbek, C. (2012). The role of mountain ranges in the diversification of birds. Annual Review of Ecology, Evolution, and Systematics, 43(1), 249-265.

Graves, G. (1988). Linearity of geographic range and its possible effect on the population structure of Andean birds. The Auk, 105, 47-52.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307-321.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences, 270, 313-321.

Hebert, P. D. N., & Gregory, T. R. (2005). The promise of DNA barcoding for taxonomy. Systematic Biology, 54(5), 852-859.

Hershkovitz, P. (1969). The recent mammals of the Neotropical region: a zoogeographic and ecological review. The Quarterly Review of Biology, 44(1), 1-70.

Hipp, A. L., Eaton, D. A., Cavender-Bares, J., Fitzek, E., Nipper, R., & Manos, P. S. (2014). A framework phylogeny of the American oak clade based on sequenced rad data. PLoS One, 9(4), e93975.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol, 35(2), 518-522.

Höhna, S., Freyman, W. A., Nolen, Z., Huelsenbeck, J. P., May, M. R., & Moore, B. R. (2019). A Bayesian approach for estimating branch-specific speciation and extinction rates. BioRxiv, 555805.

Hohna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., ... & Ronquist, F. (2016). RevBayes: bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic Biology, 65(4), 726-736.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. Annual Review of Ecology, Evolution, and Systematics, 46(1), 523-549.

Hugues, C., & Eastwood, R. O. D. (2006). Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. PNAS, 103(27), 10334-10339.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. PLoS One, 2(3), e296.

Iturralde-Vinent, M. A. (2006). Meso-cenozoic caribbean paleogeography: implications for the historical biogeography of the region. International Geology Review, 48(9), 791-827.

Jahner, J. P., Forister, M. L., Parchman, T. L., Smilanich, A. M., Miller, J. S., Wilson, J. S., ... & Dyer, L. A. (2017). Host conservatism, geography, and elevation in the evolution of a Neotropical moth radiation. Evolution, 71(12), 2885-2900.

Janzen, D. H. (1967). Why mountain passes are higher in the tropics. The American Naturalist, 101(919), 233-249.

Janzen, D. H. (1984). Two ways to be a tropical big moth: Santa Rosa saturniids and sphingids. Oxford Surveys in Evolutionary Biology, 1, 85-140.

Janzen, D. H. (1985). A host plant is more than its chemistry. Illinois Natural History Survey Bulletin, 33(3), 141-174.

Janzen, D. H., Hallwachs, W., Harvey, D. J., Darrow, K., Rougerie, R., Hajibabaei, M., ... & Hebert, P. D. N. (2012). What happens to the traditional taxonomy when a well-known tropical saturniid moth fauna is DNA barcoded? Invertebrate Systematics, 26(6), 478-505.

Jaramillo, C., Montes, C., Cardona, A., Silvestro, D., Antonelli, A., & Bacon, C. D. (2017). Comment (1) on "Formation of the Isthmus of Panama" by O'Dea et al. Science Advances, 3(6), e1602321.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. Nature, 491(7424), 444-448.

Jiang, D., Klaus, S., Zhang, Y.-P., Hillis, D. M., & Li, J.-T. (2019). Asymmetric biotic interchange across the Bering land bridge between Eurasia and North America. National Science Review, 6(4), 739-745.

Jokat, W., Boebel, T., König, M., & Meyer, U. (2003). Timing and geometry of early Gondwana breakup. Journal of Geophysical Research: Solid Earth, 108(B9).

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14(6), 587-589.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol, 30(4), 772-780.

Kitching, I., Rougerie, R., Zwick, A., Hamilton, C., St Laurent, R., Naumann, S., ... & Kawahara, A. (2018). A global checklist of the Bombycoidea (Insecta: Lepidoptera). Biodiversity Data Journal, 6, e22236.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol, 34(3), 772-773.

Leache, A. D., Chavez, A. S., Jones, L. N., Grummer, J. A., Gottscho, A. D., & Linkem, C. W. (2015). Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biol Evol, 7(3), 706-719.

Lemaire, C. (1978). *Les Attacidae Américains : Attacinae* (Lemaire C. ed.). Neuilly-sur-Seine.

Matzke, N. J. (2013). BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis in R Scripts. R package, version 0.2, 1, 2013.

McCormack, J. E., Maley, J. M., Hird, S. M., Derryberry, E. P., Graves, G. R., & Brumfield, R. T. (2012). Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Mol Phylogenet Evol, 62(1), 397-406.

Meseguer, A. S., Aldasoro, J. J., & Sanmartín, I. (2013). Bayesian inference of phylogeny, morphology and range evolution reveals a complex evolutionary history in St. John's wort (Hypericum). Mol Phylogenet Evol, 67(2), 379-403.

Meseguer, A. S., Lobo, J. M., Ree, R., Beerling, D. J., & Sanmartín, I. (2014). Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: the case of *Hypericum* (Hypericaceae). Systematic Biology, 64(2), 215-232.

Molnar, P. (2008). Closing of the Central American Seaway and the ice age: a critical review. Paleoceanography, 23(2).

Monteiro, A., & Pierce, N. E. (2001). Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) Inferred from COI, COII, and EF-1α Gene Sequences. Mol Phylogenet Evol, 18(2), 264-281.

Montes, C., Bayona, G., Cardona, A., Buchs, D. M., Silva, C. A., Morón, S., ... & Valencia, V. (2012). Arc-continent collision and orocline formation: closing of the Central American seaway. Journal of Geophysical Research: Solid Earth, 117(B4).

Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J. C., Valencia, V., ... & Niño, H. (2015). Middle Miocene closure of the Central American Seaway. Science, 348(6231), 226.

Moore, B. R., & Donoghue, M. J. (2007). Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. The American Naturalist, 170 Suppl 2, S28-55.

Morlon, H., Lewitus, E., Condamine, F. L., Manceau, M., Clavel, J., & Drury, J. (2015). RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. Methods in Ecology and Evolution, n/a-n/a.

Morlon, H., Parsons, T. L., & Plotkin, J. B. (2011). Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences USA, 108(39), 16327-16332.

Mullen, S. P., Savage, W. K., Wahlberg, N., & Willmott, K. R. (2011). Rapid diversification and not clade age explains high diversity in neotropical Adelpha butterflies. Proceedings of the Royal Society B, 278(1713), 1777-1785.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403, 853-858.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol, 32(1), 268-274.

Norris, G. (1982). Spore–pollen evidence for early Oligocene high-latitude cool climatic episode in northern Canada. Nature, 297(5865), 387-389.

Nürk, N., Scheriau, C., & Madriñán, S. (2013). Explosive radiation in high Andean *Hypericum*—rates of diversification among New World lineages. Frontiers in Genetics, 4(175).

Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., ... & Faircloth, B. C. (2019). Earth history and the passerine superradiation. Proceedings of the National Academy of Sciences USA, 116(16), 7916-7925.

Parada, A., Pardiñas, U. F., Salazar-Bravo, J., D'Elía, G., & Palma, R. E. (2013). Dating an impressive Neotropical radiation: Molecular time estimates for the Sigmodontinae (Rodentia) provide insights into its historical biogeography. Mol Phylogenet Evol, 66(3), 960-968.

Prebus, M. (2017). Insights into the evolution, biogeography and natural history of the acorn ants, genus *Temnothorax* Mayr (hymenoptera: Formicidae). BMC Evolutionary Biology, 17(1), 250.

Pyron, R. A., Costa, G. C., Patten, M. A., & Burbrink, F. T. (2015). Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. Biological Reviews, 90(4), 1248-1262.

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One, 9(2), e89543.

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Natire Communication, 4, 1958.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using Tracer 1.7. Systematic Biology, 67(5), 901-904.

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes, 7(3), 355-364.

Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-based registry for all animal species: the Barcode Index Number (BIN) System. PLoS One, 8(7), e66213.

Regier, J. C., Mitter, C., Peigler, R. S., & Friedlander, T. P. (2002). Monophyly, composition, and relationships within Saturniinae (Lepidoptera: Saturniidae): evidence from two nuclear genes. Insect Systematics & Evolution, 33(1), 9-21.

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., ... & Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. Nature Ecology & Evolution, 1(11), 1677-1682.

Rolland, J., Silvestro, D., Schluter, D., Guisan, A., Broennimann, O., & Salamin, N. (2018). The impact of endothermy on the climatic niche evolution and the distribution of vertebrate diversity. Nature Ecology & Evolution, 2(3), 459-464.

Rubin, B. E., Ree, R. H., & Moreau, C. S. (2012). Inferring phylogenies from RAD sequence data. PLoS One, 7(4), e33394.

Schultz, J. (2005). *The ecozones of the world*. Berlin Heidelberg: Springer-Verlag.

Seo, T.-K., Kishino, H., & Thorne, J. L. (2005). Incorporating gene-specific variation when inferring and evaluating optimal evolutionary tree topologies from multilocus sequence data. Proceedings of the National Academy of Sciences USA, 102(12), 4436.

Sepulchre, P., Arsouze, T., Donnadieu, Y., Dutay, J. C., Jaramillo, C., Le Bras, J., ... & Waite, A. J. (2014). Consequences of shoaling of the Central American Seaway determined from modeling Nd isotopes. Paleoceanography, 29(3), 176-189.

Simon, M. F., Grether, R., de Queiroz, L. P., Skema, C., Pennington, R. T., & Hughes, C. E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. Proceedings of the National Academy of Sciences USA, 106(48), 20359-20364.

Sklenář, P., Dušková, E., & Balslev, H. (2011). Tropical and temperate: evolutionary history of páramo flora. The Botanical Review, 77(2), 71-108.

Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., ... & Murphy, W. J. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PLoS One, 7(11), e49521.

Stehli, F. G., & Webb, S. D. (2013). *The great American biotic interchange* (Vol. 4): Springer Science & Business Media.

Sun, X., & Wang, P. (2005). How old is the Asian monsoon system?—Palaeobotanical records from China. Palaeogeography, Palaeoclimatology, Palaeoecology, 222(3), 181-222.

Tagliacollo, V. A., & Lanfear, R. (2018). Estimating improved partitioning schemes for ultraconserved elements. Mol Biol Evol, 35(7), 1798-1811.

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology, 56(4), 564-577.

Tavares, E. S., Baker, A. J., Pereira, S. L., & Miyaki, C. Y. (2006). Phylogenetic relationships and historical biogeography of neotropical parrots (Psittaciformes: Psittacidae: Arini) inferred from mitochondrial and nuclear DNA sequences. Systematic Biology, 55(3), 454-470.

Thorne, J. L., Kishino, H., & Painter, I. S. (1998). Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol, 15(12), 1647-1657.

Torsvik, T. H., & Cocks, L. R. M. (2004). Earth geography from 400 to 250 Ma: a palaeomagnetic, faunal and facies review. Journal of the Geological Society, 161(4), 555-572.

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biology, 17(12), e3000494.

Uribe-Convers, S., & Tank, D. C. (2015). Shifts in diversification rates linked to biogeographic movement into new areas: an example of a recent radiation in the Andes. American Journal of Botany, 102(11), 1854-1869.

Vane-Wright, R. I., Humphries, C. J., & Williams, P. H. (1991). What to protect? - Systematics and the agony of choice. Biological Conservation, 55, 235-254.

Wiens, J. (2004). The role of morphological data in phylogeny reconstruction. Systematic Biology, 53(4), 653-661.

Wilson, E. O. (1959). Adaptive shift and dispersal in a tropical ant fauna. Evolution, 13(1), 122-144.

Wilson, E. O. (1961). The nature of the taxon cycle in the Melanesian ant fauna. The American Naturalist, 95(882), 169-193.

Wilson, J. J. (2010). Assessing the value of DNA barcodes and other priority gene regions for molecular phylogenetics of Lepidoptera. PLoS One, 5(5), e10525.

Wolfe, J. A. (1987). Late Cretaceous-Cenozoic history of deciduousness and the terminal Cretaceous event. Paleobiology, 13(2), 215-226.

Wolfe, J. A. (1994). An analysis of Neogene climates in Beringia. Palaeogeography, Palaeoclimatology, Palaeoecology, 108(3-4), 207-216.

Wolfe, K. L. (1993). The *Copaxa* of Mexico and their immature stages (Lepidoptera : Saturniidae). Tropical Lepidoptera, 4(suppl.1), 26 pp.

Wolfe, K. L. (2005a). A resurrection of *Copaxa canella flavobrunnea* Bouvier, 1930 and elevation to species status, with illustration of early stages. Nachrichten des Entomologischen Vereins Apollo, 26(1/2), 31-33.

Wolfe, K. L. (2005b). Revision of the *Copaxa semioculata* and *Copaxa medea* groups, with corrigenda of Wolfe et al. (2003a), descriptions of three new species, and notes on their early stages (Lepidoptera: Saturniidae). Nachrichten des Entomologischen Vereins Apollo, 26(3), 121-136.

Wolfe, K. L., Bonilla, D., Ramirez, L. D., & Decaëns, T. (2003b). Rediscovery of *Copaxa sapatoza* and revealing of its immature stages (Lepidoptera: Saturniidae, Saturniinae). Nachrichten des Entomologischen Vereins Apollo, 24(3), 143-146.

Wolfe, K. L., & Conlan, C. A. (2002). A new *Copaxa* from Ecuador and its immature stages (Lepidoptera: Saturniidae, Saturniinae). Nachrichten des Entomologischen Vereins Apollo, 22(4), 235-238.

Wolfe, K. L., Lemaire, C., Amarillo S., A., & Conlan, C. A. (2003a). A contribution to the systematics of the *Copaxa semioculata* species-group (Saturniidae), with notes on the early stages, and a description of *Copaxa lunula*, new species. Journal of the Lepidopterists' Society, 57(1), 54-61.

Yang, Z. (1997). PAML: a program for package for phylogenetic analysis by maximum likelihood. CABIOS, 15, 555-556.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 ma to present. Science, 292, 686-693.

Zachos, J. C., Dickens, G. R., & Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature, 451, 279-283.

Supplementary Material

Figure S1 (Online resource) - Neighbor joining tree built with BOLD tools from the 1455 DNA barcodes of the DS-COPAX1 dataset. We used uncorrected pairwise distances, BOLD aligner and pairwise deletion option.

Figure S2 - Main node supports obtained with the phylogenetic analyses performed of the different RAD matrices. RAD matrices in column, main nodes in row. Supports are indicated as such: SH-aLRT/UFBoot.

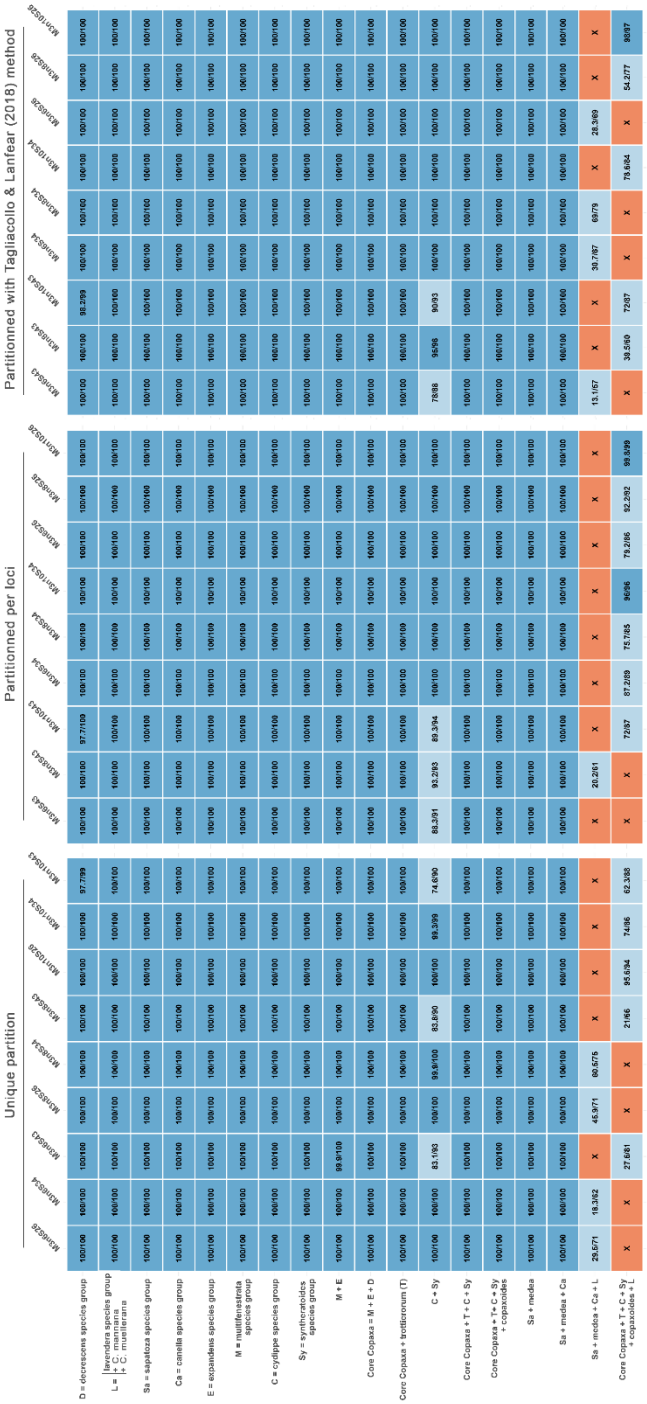
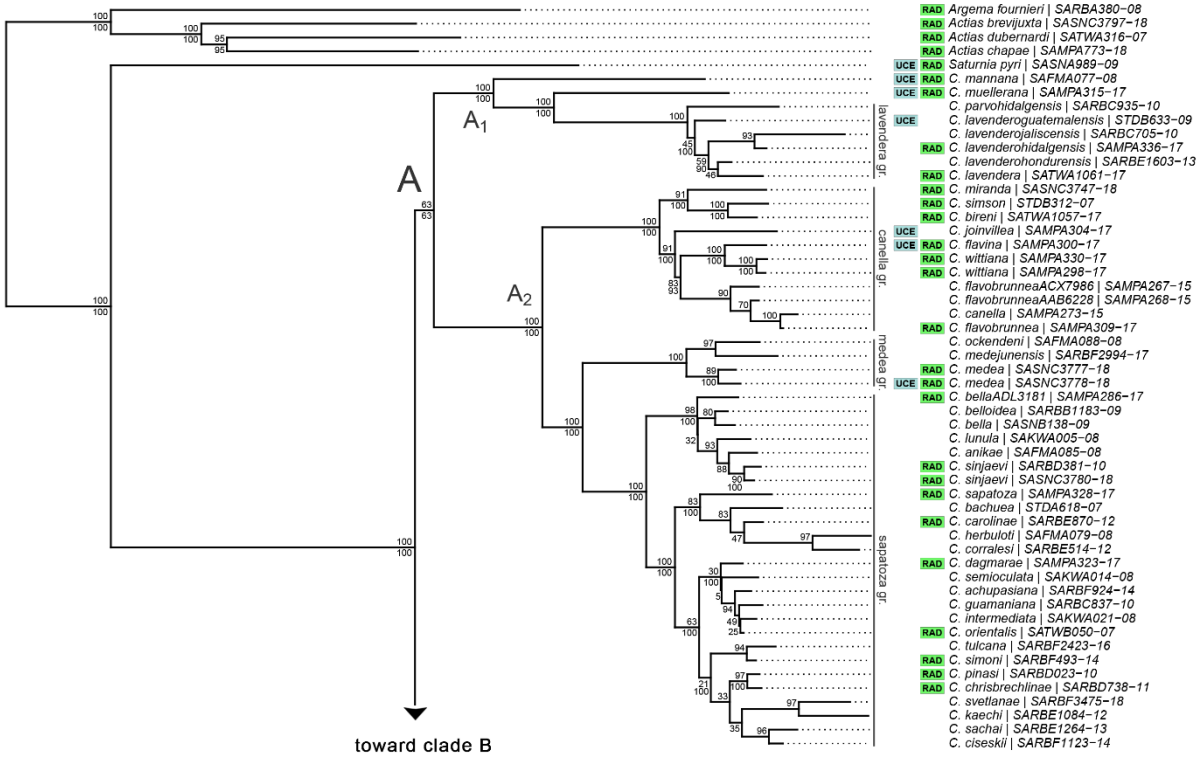
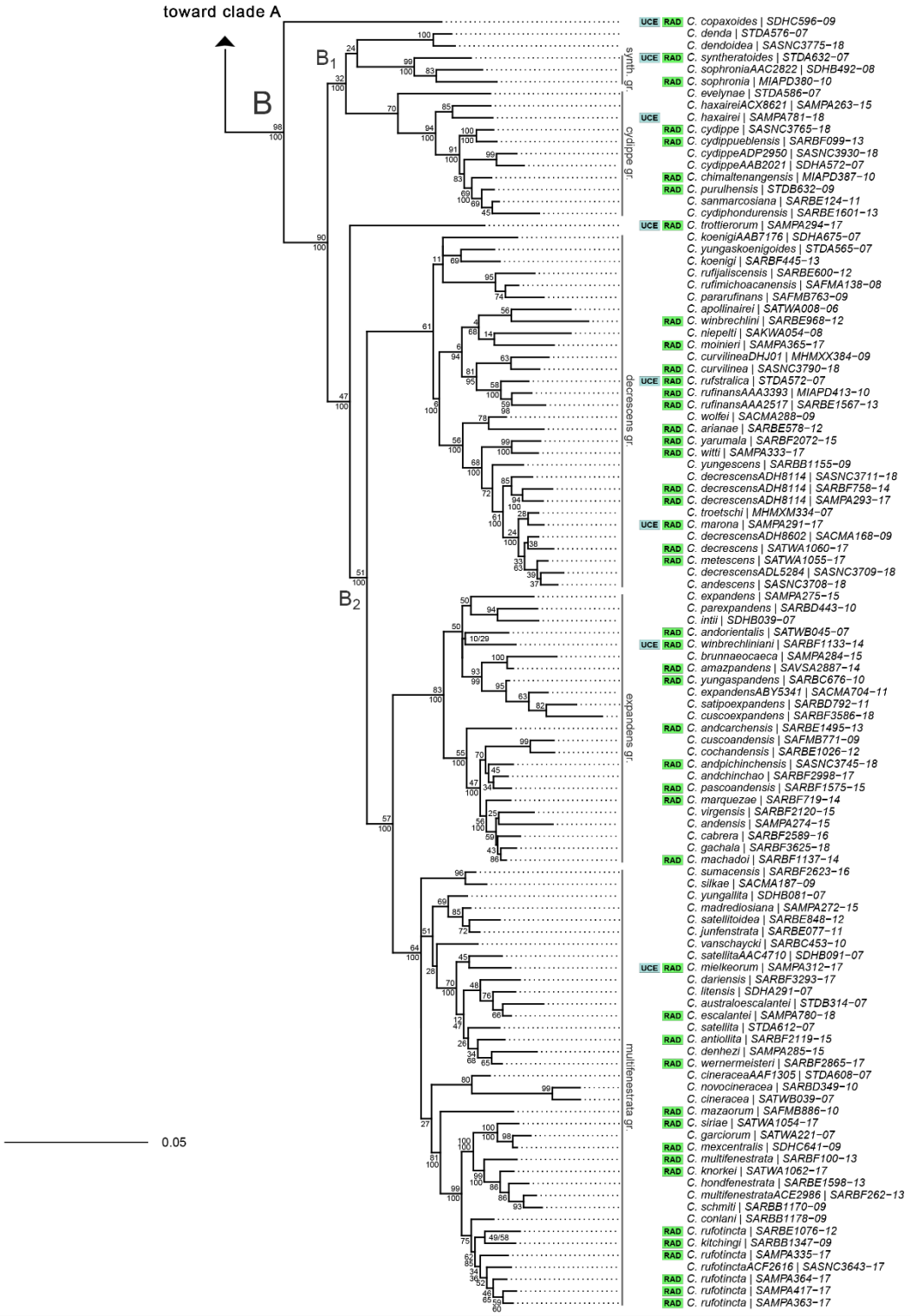


Figure S3 (Online resource) - Phylogenies inferred from 27 RAD matrices with IQ-TREE. Species names, sample codes and Process-ID are depicted at tips. Supports are indicated as such: SH-aLRT/UFBoot.

Figure S4 (two following pages) - Maximum likelihood phylogenetic hypothesis for genus *Copaxa* built with IQTREE for 5 outgroups and 157 samples of *Copaxa* representing all 130 valid species within the genus and 18 OTUs (1 rogue taxa was excluded); taxon names are preceded by symbols representing availability of genomic markers (RADSeq (green) and UCE (blue)); they are followed by the ProcessID code of the DNA barcode record used in the analysis (available for all terminals in BOLD dataset DS-COPAX1). The five main lineages discussed in the text are characterized as clades/sub-clades A, A_1, A_2, B, B_1, B_2 on branches, and species groups are indicated as vertical lines on the right side of the tree. Two-steps bootstrap support (BS) values are depicted above every branch. We also indicated below the former BS value, when relevant, the computed BS supports after discarding the samples for which the DNA barcode was the only marker available.




toward clade A
B
B1
B2
synth. gr.
cydippe gr.
decrescens gr.
expandens gr.
multifenestrata gr.
0.05
UCE RAD C. copaxoides | SDHC596-09
C. denda | STDA576-07
C. dendoidea | SASNC3775-18
UCE RAD C. syntheratoides | STDA632-07
C. sophroniaAAC2822 | SDHB492-08
RAD C. sophronia | MIAPD380-10
C. evelynae | STDA586-07
C. haxaireiACX8621 | SAMPA263-15
UCE C. haxairei | SAMPA781-18
RAD C. cydippe | SASNC3765-18
RAD C. cydippueblensis | SARBF099-13
C. cydippeADP2950 | SASNC3930-18
C. cydippeAAB2021 | SDHA572-07
RAD C. chimaltenangensis | MIAPD387-10
RAD C. purulhensis | STDB632-09
C. sanmarcosiana | SARBE124-11
C. cydiphondurensis | SARBE1601-13
UCE RAD C. trottierorum | SAMPA294-17
C. koenigiAAB7176 | SDHA675-07
C. yungaskoenigoides | STDA565-07
C. koenigi | SARBF445-13
C. rufijaliscensis | SARBE600-12
C. rufimichoacanensis | SAFMA138-08
C. pararufinans | SAFMB763-09
C. apollinairei | SATWA008-06
RAD C. winbrechlini | SARBE968-12
C. niepelti | SAKWA054-08
RAD C. moinieri | SAMPA365-17
C. curvilineaDHJ01 | MHMXX384-09
RAD C. curvilinea | SASNC3790-18
UCE RAD C. rufstralica | STDA572-07
RAD C. rufinansAAA3393 | MIAPD413-10
RAD C. rufinansAAA2517 | SARBE1567-13
C. wolfei | SACMA288-09
RAD C. arianae | SARBE578-12
RAD C. yarumala | SARBF2072-15
RAD C. witti | SAMPA333-17
C. yungescens | SARBB1155-09
C. decrescensADH8114 | SASNC3711-18
RAD C. decrescensADH8114 | SARBF758-14
RAD C. decrescensADH8114 | SAMPA293-17
C. troetschi | MHMXM334-07
UCE RAD C. marona | SAMPA291-17
C. decrescensADH8602 | SACMA168-09
RAD C. decrescens | SATWA1060-17
RAD C. metescens | SATWA1055-17
C. decrescensADL5284 | SASNC3709-18
C. andescens | SASNC3708-18
C. expandens | SAMPA275-15
C. parexpandens | SARBD443-10
C. intii | SDHB039-07
RAD C. andorientalis | SATWB045-07
UCE RAD C. winbrechliniani | SARBF1133-14
C. brunnaeocaeca | SAMPA284-15
RAD C. amazpandens | SAVSA2887-14
RAD C. yungaspandens | SARBC676-10
C. expandensABY5341 | SACMA704-11
C. satipoexpandens | SARBD792-11
C. cuscoexpandens | SARBF3586-18
RAD C. andcarchensis | SARBE1495-13
C. cuscoandensis | SAFMB771-09
C. cochandensis | SARBE1026-12
RAD C. andpichinchensis | SASNC3745-18
C. andchinchao | SARBF2998-17
RAD C. pascoandensis | SARBF1575-15
RAD C. marquezae | SARBF719-14
C. virgensis | SARBF2120-15
C. andensis | SAMPA274-15
C. cabrera | SARBF2589-16
C. gachala | SARBF3625-18
RAD C. machadoi | SARBF1137-14
C. sumacensis | SARBF2623-16
C. silkae | SACMA187-09
C. yungallita | SDHB081-07
C. madrediosiana | SAMPA272-15
C. satellitoidea | SARBE848-12
C. junfenstrata | SARBE077-11
C. vanschaycki | SARBC453-10
C. satellitaAAC4710 | SDHB091-07
UCE RAD C. mielkeorum | SAMPA312-17
C. dariensis | SARBF3293-17
C. litensis | SDHA291-07
C. australoescalantei | STDB314-07
RAD C. escalantei | SAMPA780-18
C. satellita | STDA612-07
RAD C. antiollita | SARBF2119-15
C. denhezi | SAMPA285-15
RAD C. wernermeisteri | SARBF2865-17
C. cineraceaAAF1305 | STDA608-07
C. novocineracea | SARBD349-10
C. cineracea | SATWB039-07
RAD C. mazaorum | SAFMB886-10
RAD C. siriae | SATWA1054-17
C. garciorum | SATWA221-07
RAD C. mexcentralis | SDHC641-09
RAD C. multifenestrata | SARBF100-13
RAD C. knorkei | SATWA1062-17
C. hondfenestrata | SARBE1598-13
C. multifenestrataACE2986 | SARBF262-13
C. schmiti | SARBB1170-09
C. conlani | SARBB1178-09
RAD C. rufotincta | SARBE1076-12
RAD C. kitchingi | SARBB1347-09
RAD C. rufotincta | SAMPA335-17
C. rufotinctaACF2616 | SASNC3643-17
RAD C. rufotincta | SAMPA364-17
RAD C. rufotincta | SAMPA417-17
RAD C. rufotincta | SAMPA363-17

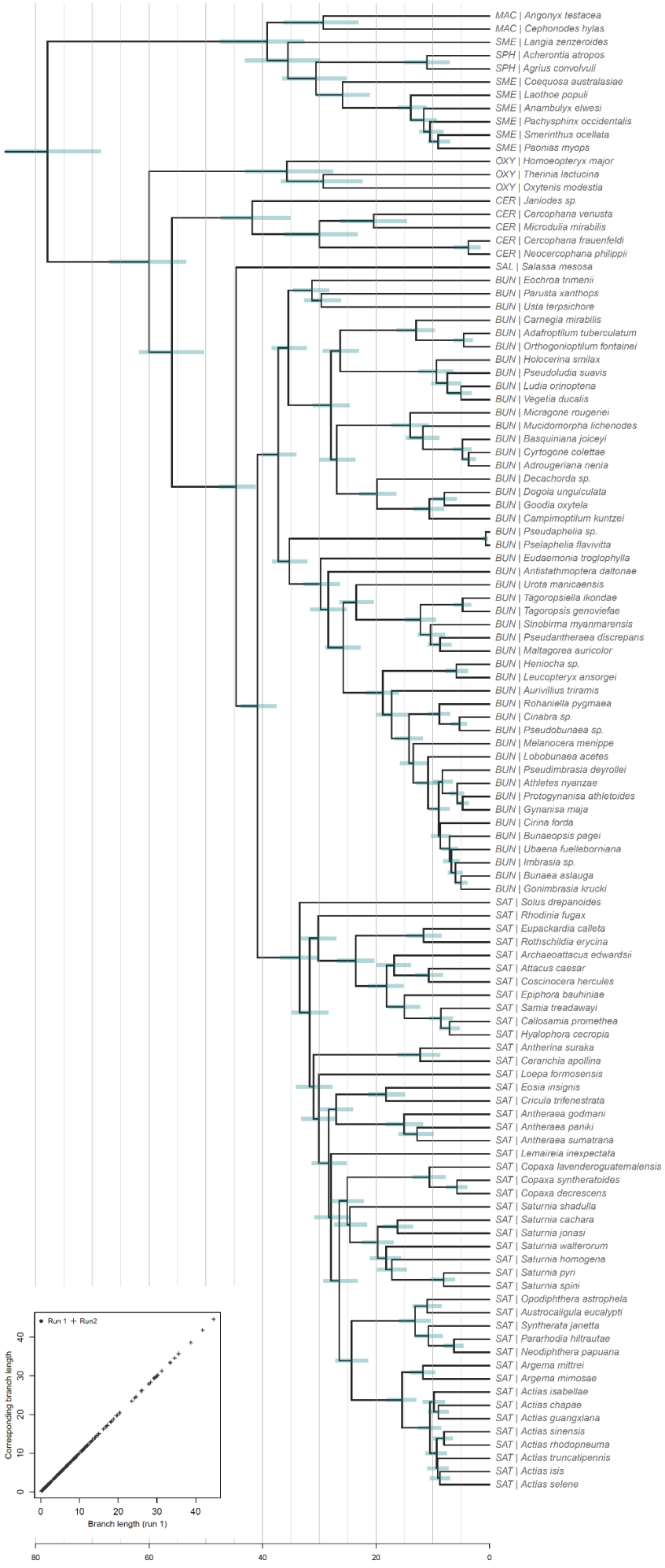


Figure S5 - Divergence time estimation of the Oxyteninae, Cercophaninae, Salassinae, Bunaeinae and Saturniinae subfamilies (Saturniidae) performed with MCMC from Arnal et al. (in prep.) UCEs dataset. Contrary to Arnal et al (in prep.), we included all three *Copaxa* samples they sequenced to obtain calibration points within the *Copaxa* genus. 95% credibility intervals of the estimates are represented with blue horizontal bars. The bottom scale unit is the million of years. The bottom left plot shows that the two independent runs led to similar estimates, thus assessing convergence (also checked with Tracer).

Figure S6 - Convergence of the divergence time analyses performed on the *Copaxa* phylogeny with MCMCtree. The two randomly sampled loci dataset (A and B) led to similar age estimates and the analyses run on the same dataset resulted in identical parameter values, assessing convergence (also checked with Tracer).

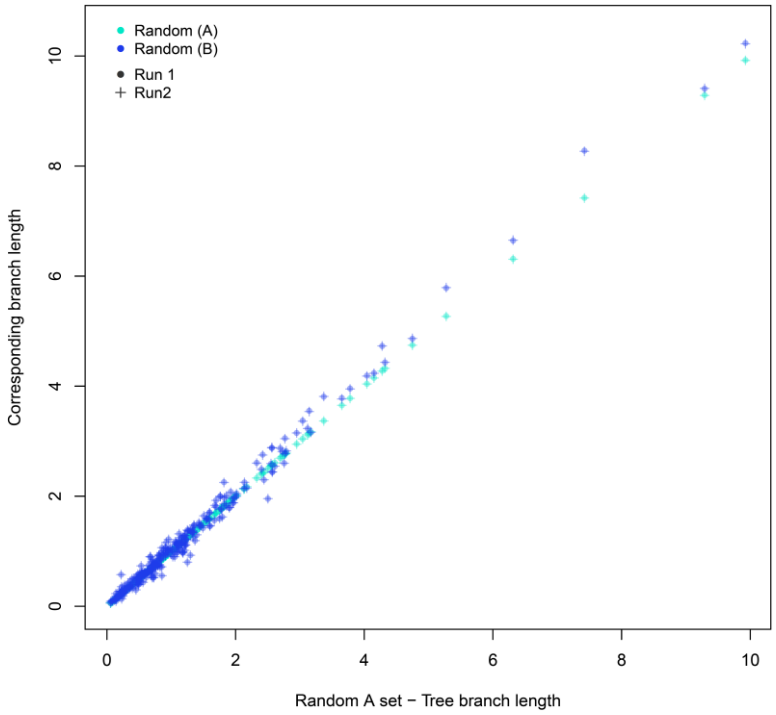


Figure S7 (following page) - Altitudinal preferences of the *Copaxa* lineages. Elevation medians are depicted for extant species along with their names. Ancestral states were estimates with the diversitree R package and node pies represent uncertainties about the estimations.

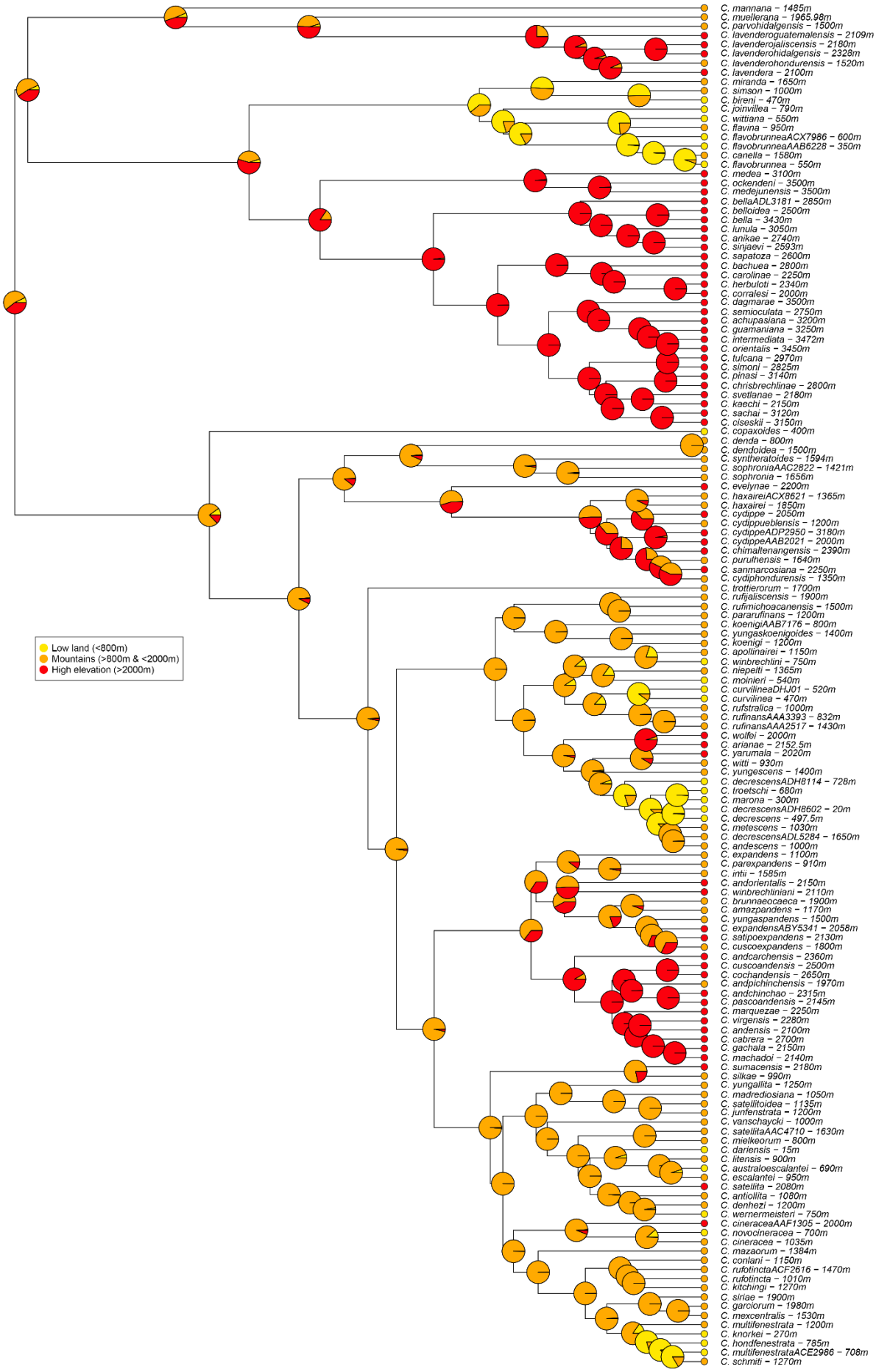

Low land (<800m)
Mountains (>800m & <2000m)
High elevation (>2000m)
C. mannana – 1485m
C. muellerana – 1965.98m
C. parvohidalgensis – 1500m
C. lavenderoguatemalensis – 2109m
C. lavenderojaliscensis – 2180m
C. lavenderohidalgensis – 2328m
C. lavenderohondurensis – 1520m
C. lavendera – 2100m
C. miranda – 1650m
C. simson – 1000m
C. bireni – 470m
C. joinvillea – 790m
C. wittiana – 550m
C. flavina – 950m
C. flavobrunneaACX7986 – 600m
C. flavobrunneaAAB6228 – 350m
C. canella – 1580m
C. flavobrunnea – 550m
C. medea – 3100m
C. ockendeni – 3500m
C. medejunensis – 3500m
C. bellaADL3181 – 2850m
C. belloidea – 2500m
C. bella – 3430m
C. lunula – 3050m
C. anikae – 2740m
C. sinjaevi – 2593m
C. sapatoza – 2600m
C. bachuea – 2800m
C. carolinae – 2250m
C. herbuloti – 2340m
C. corralesi – 2000m
C. dagmarae – 3500m
C. semioculata – 2750m
C. achupasiana – 3200m
C. guamaniana – 3250m
C. intermediata – 3472m
C. orientalis – 3450m
C. tulcana – 2970m
C. simoni – 2825m
C. pinasi – 3140m
C. chrisbrechlinae – 2800m
C. svetlanae – 2180m
C. kaechi – 2150m
C. sachai – 3120m
C. ciseskii – 3150m
C. copaxoides – 400m
C. denda – 800m
C. dendoidea – 1500m
C. syntheratoides – 1594m
C. sophroniaAAC2822 – 1421m
C. sophronia – 1656m
C. evelynae – 2200m
C. haxaireiACX8621 – 1365m
C. haxairei – 1850m
C. cydippe – 2050m
C. cydippueblensis – 1200m
C. cydippeADP2950 – 3180m
C. cydippeAAB2021 – 2000m
C. chimaltenangensis – 2390m
C. purulhensis – 1640m
C. sanmarcosiana – 2250m
C. cydiphondurensis – 1350m
C. trottierorum – 1700m
C. rufijaliscensis – 1900m
C. rufimichoacanensis – 1500m
C. pararufinans – 1200m
C. koenigiAAB7176 – 800m
C. yungaskoenigoides – 1400m
C. koenigi – 1200m
C. apollinairei – 1150m
C. winbrechtini – 750m
C. niepelti – 1365m
C. moirieri – 540m
C. curvilineaDHJ01 – 520m
C. curvilinea – 470m
C. rufstralica – 1000m
C. rufinansAAA3393 – 832m
C. rufinansAAA2517 – 1430m
C. wolfei – 2000m
C. arianae – 2152.5m
C. yarumala – 2020m
C. witti – 930m
C. yungescens – 1400m
C. decrescensADH8114 – 728m
C. troetschi – 680m
C. marona – 300m
C. decrescensADH8602 – 20m
C. decrescens – 497.5m
C. metescens – 1030m
C. decrescensADL5284 – 1650m
C. andescens – 1000m
C. expandens – 1100m
C. parexpandens – 910m
C. intii – 1585m
C. andorientalis – 2150m
C. winbrechliniani – 2110m
C. brunnaeocaeca – 1900m
C. amazpandens – 1170m
C. yungaspandens – 1500m
C. expandensABY5341 – 2058m
C. satipoexpandens – 2130m
C. cuscoexpandens – 1800m
C. andcarchensis – 2360m
C. cuscoandensis – 2500m
C. cochandensis – 2650m
C. andpichinchensis – 1970m
C. andchinchao – 2315m
C. pascoandensis – 2145m
C. marquezae – 2250m
C. virgensis – 2280m
C. andensis – 2100m
C. cabrera – 2700m
C. gachala – 2150m
C. machadoi – 2140m
C. sumacensis – 2180m
C. silkae – 990m
C. yungallita – 1250m
C. madrediosiana – 1050m
C. satellitoidea – 1135m
C. juntenstrata – 1200m
C. vanschaycki – 1000m
C. satellitaAAC4710 – 1630m
C. mielkeorum – 800m
C. dariensis – 15m
C. litensis – 900m
C. australoescalantei – 690m
C. escalantei – 950m
C. satellita – 2080m
C. antiolita – 1080m
C. denhezi – 1200m
C. wernermeisteri – 750m
C. cineraceaAAF1305 – 2000m
C. novocineracea – 700m
C. cineracea – 1035m
C. mazaorum – 1384m
C. conlani – 1150m
C. rufotinctaACF2616 – 1470m
C. rufotincta – 1010m
C. kitchingi – 1270m
C. siriae – 1900m
C. garciorum – 1980m
C. mexcentralis – 1530m
C. multifenestrata – 1200m
C. knorkei – 270m
C. hondfenestrata – 785m
C. multifenestrataACE2986 – 708m
C. schmiti – 1270m

Table S1 (Online resource) - Specific diversity of the *Copaxa* moths.

Table S2 (Online resource) - Voucher information.

Table S3 (following page) - Number of loci for every sample in each of the RAD matrices.

Sample (RAD code)	M3n6S26	M3n8S26	M3n10S26	M3n6S34	M3n8S34	M3n10S34	M3n6S43	M3n8S43	M3n10S43
total	**2954**	**3686**	**4142**	**1359**	**1790**	**2090**	**358**	**496**	**582**
BC-Dec1537_C_purulhensis	1746	2129	2320	881	1151	1310	271	382	437
BC-EvS_3837_C_amazpandens	194	233	246	97	119	136	35	45	56
BC-Her2641_C_mexcentralis	1771	2123	2329	898	1155	1332	284	386	447
BC-Her4039_C_sophronia	2363	2907	3236	1165	1513	1762	326	451	525
BC-Her4046_C_chimaltenangensis	2243	2747	3039	1121	1469	1695	331	457	532
BC-Her4072_C_rufinans	2502	3053	3374	1226	1607	1853	341	470	549
BC-HKT_0184_C_litensis	2130	2595	2835	1049	1363	1563	318	431	500
BC-RBP_3372_C_kitchingi	2488	3035	3350	1191	1558	1799	340	468	546
BC-RBP_4105_C_yungaspandens	2019	2454	2691	971	1274	1480	297	412	479
BC-RBP_4402_C_bachuea	1166	1545	1796	809	1064	1219	280	377	440
BC-RBP_4760_C_sinjaevi	36	47	53	31	38	42	16	16	17
BC-RBP_5117_C_chrisbrechlinae	798	1050	1246	559	739	871	204	272	319
BC-RBP_6482_C_arianae	1119	1350	1469	577	746	859	211	289	330
BC-RBP_6713_C_carolinae	1156	1526	1754	803	1061	1229	276	379	432
BC-RBP_7315_C_nsp1	915	1096	1203	458	598	684	181	233	269
BC-RBP_7387_C_rufinans	850	1009	1101	443	581	651	147	206	236
BC-RBP_7629_C_cydippe	243	276	306	134	165	188	54	72	84
BC-RBP_7630_C_multifenestrata	2045	2486	2747	999	1301	1510	307	420	493
BC-RBP_8023_C_hoehni	1445	1909	2215	966	1280	1481	312	424	498
BC-RBP_8249_C_andensis	1050	1256	1352	528	668	762	186	265	311
BC-RBP_8288_C_ignescens	1533	1841	2003	796	1021	1154	261	359	409
BC-RBP_8646_C_nsp2	650	775	847	341	437	515	128	183	215
BC-RBP_8650_C_nsp3	799	933	1002	399	522	596	154	217	256
BC-RBP_9086_C_pascoandensis	1891	2270	2493	948	1215	1414	292	399	472
BC-RBP_9583_C_witti	2379	2920	3238	1185	1558	1803	343	473	551
BC-RBP_9630_C_sat_sp2	96	116	123	46	62	75	18	28	31
GUELPHA01	131	221	347	105	168	249	59	89	124
JRAS06156_0101	964	1294	1499	598	825	976	211	300	356
RROU00003	145	214	303	123	180	242	73	103	131
RROU00023	454	735	1004	332	522	712	151	227	308
RROU00038	2656	3292	3686	1277	1680	1955	353	487	570
RROU00040	2586	3203	3576	1260	1660	1924	351	487	564
RROU00041	2366	2991	3375	1166	1550	1827	328	454	534
RROU00047	1554	2158	2606	1056	1446	1715	337	471	555
RROU00049	1623	2241	2704	1081	1472	1752	339	472	558
RROU00061	1608	2208	2683	1064	1450	1730	339	469	551
RROU00064	2573	3190	3556	1219	1610	1883	330	459	539
RROU00068	1853	2474	2868	1060	1444	1708	313	449	522
RROU00069	2734	3390	3783	1291	1696	1980	351	486	568
RROU00133	1769	2368	2777	1128	1520	1774	332	461	541
RROU00139	1598	2117	2466	1059	1396	1634	328	449	526
RROU00145	71	94	101	45	65	70	14	22	24
RROU00148	2607	3214	3584	1264	1660	1922	350	481	558
RROU00150	179	236	263	134	170	189	57	73	83
RROU00163_C_lavenderohidalgensi	928	1223	1395	597	792	922	216	304	355
RROU00167	2696	3330	3721	1287	1678	1952	348	474	560
RROU00171	285	437	637	223	337	473	115	173	232
RROU00281_C_arianae	95	109	119	53	62	73	24	29	29
RROU00283_C_andescens	326	397	421	171	226	260	65	89	105
RROU00285_C_kitchingi	172	217	243	91	120	142	27	44	54
RROU00300_C_conlani	312	379	409	168	228	268	58	88	105
RROU00301_C_rufotincta	153	188	208	65	81	104	28	38	45
RROU00302_C_moinieri	457	538	601	233	298	346	81	108	133
RROU00354_C_kitchingi	120	150	166	63	79	96	17	28	36
SNB_RR_0004	2600	3198	3571	1258	1646	1910	345	476	557
SNB_RR_0032	380	472	520	199	251	286	73	101	112
SNB_RR_0038	117	131	144	69	80	86	27	33	40
SNB_RR_0041	2603	3184	3535	1247	1619	1886	337	469	552
SNB_RR_0043	1630	2223	2682	1061	1444	1710	332	462	542
SNB_RR_0047	1499	2081	2506	1003	1371	1632	318	443	518
SNB_RR_0057	2130	2709	3043	1119	1470	1705	320	436	511
SNB_RR_0065	2384	2949	3292	1159	1535	1790	329	462	538
SNB_RR_0072	813	973	1054	398	516	595	136	176	207
SNB_RR_0073	1261	1688	1970	818	1080	1277	247	349	416
SNB_RR_0074	1862	2499	2934	1166	1544	1813	335	465	544
SNB_RR_0076	1527	2064	2417	1001	1333	1548	308	426	498
SNB_RR_0078	231	325	372	158	225	264	48	73	88
SNB_RR_0080	1226	1718	2065	855	1180	1396	292	411	482
SNB_RR_0081	1913	2527	2948	1086	1458	1723	318	447	522
SNB_RR_0084	1286	1549	1701	612	790	920	203	275	328
SNB_RR_0087	2544	3128	3463	1243	1629	1883	346	476	556
SNB_RR_0089	2765	3435	3840	1305	1716	2006	352	487	570
SNB_RR0090	2622	3219	3573	1255	1650	1910	346	479	561
SNB_RR0101	246	359	532	196	284	400	100	144	196

Chapitre 3

Inférence de la mégaphylogénie des Saturniidae

Préambule

Les études de diversification réalisées à partir de phylogénies *higher-level* (Chapitre 1) et à partir de phylogénies de clades restreints (Chapitre 2) sont complémentaires, car elles permettent de soulever des questions à des échelles distinctes, notamment sur le rôle de variations phénotypiques ou de traits d'histoire de vie sur les dynamiques spatiales et temporelles de cette diversification. Cependant, l'estimation de l'influence relative des différents facteurs, ceux qui s'observent à des échelles larges comme ceux étudiés à partir dse phylogénies intra-générique, est difficile à tester objectivement et peut mener à des interprétations arbitraires et possiblement erronées. Une façon de réconcilier ces deux échelles serait la génération de grandes phylogénies datées (*i.e.* des « mégaphylogénies ») intégrant l'ensemble des espèces du groupe d'étude. Ce type d'approche est encore rare et limitée aux grands groupes de Vertébrés ; chez les insectes des déficits majeurs dans la connaissance de leur biodiversité ne permettent pas l'inférence de mégaphylogénies. De fait, les Saturniidae représentent l'une des premières opportunités d'inférer une phylogénie presque complète d'un groupe diversifié d'insectes à distribution mondiale. Au cours de ce Chapitre, je présente la toute première phylogénie à l'espèce des Saturniidae ; il s'agit également de la première mégaphylogénie datée inférée de manière robuste, sur la base de données génomiques et génétiques pour un clade d'insecte.

Les résultats présentés ici sont exploratoires et visent à tester la faisabilité opérationnelle de l'inférence d'une phylogénie complète des Saturniidae. L'échantillonnage génomique utilisé ici sera complété dans les prochains mois dans le cadre de l'ANR SPHINX (PI : Rodolphe Rougerie) afin d'inférer avec plus de précision certains nœuds que nous avons identifiés comme non résolus. A terme, cette phylogénie nous permettra d'étudier précisément l'origine des patrons globaux de la diversité des Saturniidae et fera l'objet d'une soumission dans un journal généraliste.

Résumé

La génération de larges phylogénies à l'échelle de l'espèce (*i.e.* mégaphylogénies) est primordiale pour la compréhension des patrons spatiaux et temporels de la biodiversité. Si de telles phylogénies ont déjà été proposées pour de grands groupes de Vertébrés, aucune n'a jusqu'alors été inférée pour un clade d'insectes alors même qu'ils constituent une part majeure de la biomasse et de la biodiversité terrestre. Ce Chapitre présente un travail exploratoire dont l'objectif était de tester la pertinence de l'association de marqueurs UCE (Éléments Ultra-Conservés du génome - *Ultra Conserved Elements* - UCEs) et du code-barres ADN dans l'inférence de mégaphylogénies, objectif majeur de l'ANR SPHINX. Dans un premier temps, nous avons défini 3947 OTUs de Saturniidae à partir de données biologiques et géographiques ainsi que des codes-barres disponibles sur la plateforme BOLD. Nous avons ensuite séquencé des UCEs pour 1007 d'entre eux, en maximisant les distances phylogénétiques. Finalement, nous avons utilisé une approche *backbone + subtree*, déjà employée par plusieurs auteurs pour inférer des phylogénies de taille comparable, pour inférer la plus grande mégaphylogénie inférée pour un groupe d'insecte. Nos résultats mettent en exergue la pertinence de la combinaison des marqueurs UCE et du code-barres ADN dans l'inférence des nœuds les plus profonds comme les plus superficiels. La répétabilité de ces analyses est assurée par un *pipeline* phylogénomique dont la rapidité (10j de computation) est assurée par l'utilisation des topologies des procédures de bootstrap afin de prendre en compte l'incertitude liée à la topologie et aux temps de divergence. Enfin, nous discutons de la difficulté d'inférer des taux de diversification spécifiques aux branches de l'arbre à partir de telles phylogénies.

Introduction

Les phylogénies sont la représentation des liens d'apparentement entre les lignées du monde vivant. Elles jouent un rôle central dans la recherche en Écologie et Évolution puisqu'elles constituent le support de nombreuses analyses cherchant à lier des processus écologiques et/ou évolutifs aux patrons spatiaux et temporels de la biodiversité. Elles sont depuis longtemps utilisées pour corriger la non-indépendance des phénotypes de différentes lignées évolutives dans des analyses comparatives (Felsenstein 1985). Plus récemment, les phylogénies ont aussi été utilisées dans le domaine de l'Écologie pour identifier les zones (Pollock *et al.* 2017) ou clades (Isaac *et al.* 2007) à protéger en priorité, pour étudier les invasions biologiques (Gallien *et al.* 2016) ou pour comparer les niches écologiques d'espèces apparentées (Joly *et al.* 2014). En Biologie Évolutive elles constituent le socle des analyses qui tentent de comprendre l'évolution des taux de spéciation et d'extinction (Morlon 2014), d'étudier l'évolution de traits (par ex. Rabosky *et al.* 2013) ou d'inférer les dynamiques spatiales de la biodiversité (par ex. Chazot *et al.* 2016). L'inflation du nombre de phylogénies publiées, largement induite par les progrès des technologies et des méthodes de séquençage (Delsuc *et al.* 2005 ; Lemmon & Lemmon 2013), a permis de mieux comprendre l'évolution et la distribution de la biodiversité (Condamine *et al.* 2012 ; Economo *et al.* 2015 ; Moeller *et al.* 2017). Cependant, la majorité des études s'intéressant à ces questions réalisent des analyses basées sur des phylogénies de taille limitée, restreintes à des clades spécifiques et dont la puissance statistique est limitée. Les résultats obtenus peuvent ainsi être biaisés par les spécificités intrinsèques des clades étudiés. Pour mieux comprendre les patrons globaux de biodiversité, la recherche en écologie et évolution bénéficierait largement d'un développement de mégaphylogénies[4] pour des groupes d'organismes variés : des phylogénies datées, à l'échelle de l'espèce, de taille conséquente permettant de bénéficier d'un grand pouvoir statistique (>1000 feuilles) et dont la résolution et la complétion sont suffisamment élevées pour constituer une représentation fidèle de l'histoire évolutive des groupes étudiés.

Inférer une mégaphylogénie constitue un travail long et complexe nécessitant une connaissance approfondie du groupe d'intérêt. De nombreuses années de travail sont tout d'abord nécessaires pour obtenir un échantillonnage suffisamment complet, taxonomiquement et géographiquement. De fait, il ne serait pas envisageable aujourd'hui de générer des mégaphylogénies pour de nombreux groupes taxonomiques dont la majeure partie de la diversité est inconnue (Hamilton *et al.* 2010, Mora *et al.* 2011). Ensuite, il faut générer des marqueurs génétiques suffisamment informatifs pour résoudre à la fois des nœuds profonds, anciens, et des nœuds superficiels, récents. Cette étape peut s'avérer

[4] Le terme *mega-phylogeny* avait déjà été utilisé par Smith *et al.* (2009) pour désigner un *pipeline* (‘*mega-phylogeny approach*’) destiné à générer de larges matrices nucléotidiques à partir de données disponibles en ligne. L'intérêt du *pipeline* est qu'il permet de limiter l'intervention de l'utilisateur lors du nettoyage des données, éliminant de fait les potentielles erreurs humaines, courantes sur des jeux de données de grande dimension. Tel qu'il avait été défini par Smith *et al.* (2009), le terme *mega-phylogeny* ne désigne donc pas une phylogénie. Pour plus de clarté, nous récusons ce terme, peu utilisé au sein de la littérature.

particulièrement délicate lorsqu'il s'agit de travailler sur des espèces rares, dont les spécimens connus sont disséminés dans les collections et muséums du monde entier et conservés, dans des conditions diverses, depuis plusieurs dizaines d'années (Cooper 1994 ; Burrell *et al.* 2015). Enfin, il faut inférer des milliers de relations phylogénétiques et estimer un nombre considérable de temps de divergence, deux étapes complexes qui impliquent des temps de calcul conséquents (par ex. Upham *et al.* 2019). Prises dans leur ensemble, ces étapes demandent des années de travail pour les naturalistes, taxonomistes, techniciens de laboratoire et phylogénéticiens, ce qui explique le nombre très limité de mégaphylogénies publiées à ce jour. A notre connaissance, les seules mégaphylogénies datées ont jusqu'alors été inférées pour des groupes emblématiques de Vertébrés (Tableau 1) : Mammifères (Bininda-Emonds *et al.* 2007 ; Upham *et al.* 2019), Oiseaux (Jetz *et al.* 2012), Squamates (Tonini *et al.* 2016), Amphibiens (Jetz & Pyron 2018) et Actinoptérygiens (Rabosky *et al.* 2018) (voir aussi Särkinen *et al.* 2013 ; Pyron & Wiens 2013 ; Zheng & Wiens 2016 ; Chazot *et al.* 2020 pour de larges phylogénies dont la complétion s'approche des 50% ; Tableau 1). Leur publication a eu un écho considérable dans la communauté scientifique et elles ont servi de support à de nombreuses analyses en Écologie et Évolution (Liker *et al.* 2013 ; Rolland *et al.* 2014 ; Odom *et al.* 2014 ; Healy *et al.* 2014 ; Condamine *et al.* 2019), démontrant l'intérêt fort de la communauté scientifique pour de tels supports d'analyses. La nécessité de disposer de mégaphylogénies pour comprendre les patrons globaux de diversité est telle que certains auteurs prennent le parti de générer de tels arbres à partir de jeux de données pourtant très parcellaires. C'est le cas de l'étude de Economo *et al.* (2018) qui s'est intéressée aux patrons de distribution des espèces de fourmis (Hymenoptera : Formicidae) en utilisant une phylogénie de près de 15 000 feuilles inférée à partir d'une matrice nucléotidique comprenant seulement 673 espèces (Tableau 1).

Auteur(s) et année	Clade	Nombre d'OTUs	% de complétion	(% OTUs avec données moléculaires)	Type de données	Méthode résumée	Mesure de l'incertitude phylogénétique et temporelle
Bininda-Emonds et al. 2007	Mammalia (class)	4 510	**99%**	Données trop diverses	morphologiques, taxonomiques et génétiques	Combinaison d'arbres de 10 publications distinctes à l'aide de la méthode d'inférence de *supertree* MRP. Datation à l'aide d'horloges moléculaires locales et de fossiles (30 au total).	temporelle (partielle : 3 arbres)
Jetz et al. 2012	Aves (class)	9 993	**100%**	67%	taxonomiques et génétiques	*backbone + subtrees* non chevauchants. Définition des lignées du backbone en se basant sur Ericson et al. (2006) et Hackett et al. (2008). Utilisation de la méthode implémentée par la suite dans PASTIS pour placer les espèces sans données génétiques, une première analyse permettant d'établir les contraintes pour leur branchement.	phylogénétique et temporelle (1000 arbres)
Pyron & Wiens 2013	Amphibia (class)	2 871	44%	100%	génétiques	Datation de la phylogénie inférée par Pyron & Wiens (2011). Ces derniers ont inféré une unique phylogénie par maximum de vraisemblance. Pyron & Wiens (2013) ont daté cette phylogénie à l'aide de treePL et ainsi ne prennent pas en compte l'incertitude liée à la topologie et aux temps de divergence.	non
Särkinen et al. 2013	Solanaceae (family)	1 075	45%	100%	génétiques	Inférence de la topologie et des temps de divergence à l'aide de BEAST (10 analyses de 10 millions de générations combinées) à partir d'une matrice concaténée (supermatrice).	phylogénétique et temporelle (distribution postérieure des arbres non disponible)
Zheng & Wiens 2016	Squamata (order)	4 162	43%	100%	génétiques	Inférence de la topologie à partir d'une matrice concaténée (supermatrice). Datation de la phylogénie inférée à l'aide de treePL, de façon similaire à l'approche utilisée par Pyron & Wiens 2013.	non
Tonini et al. 2016	Squamata (order)	9 754	**100%**	55%	taxonomiques et génétiques	*backbone + subtrees* non chevauchants. Définition des lignées du backbone et des subtrees à partir d'un arbre inféré par maximum de vraisemblance à partir d'une matrice concaténée. Cet arbre a également été utilisé afin de définir les contraintes de position des espèces sans données génétiques, utilisées dans PASTIS. Contrairement à Jetz et al. (2012) auquel il se réfèrent, Tonini et al. 2016 ont contraint la topologie de l'arbre à celle inférée par maximum de vraisemblance, ainsi, l'ensemble d'arbres qu'ils ont publiés ne rendent pas compte de l'incertitude liée aux positions des clades pour lesquels des données génétiques sont disponibles.	phylogénétique (très partiellement) et temporelle sur 10 000 arbres
Economo et al. 2018	Formicidae (family)	14 594	**~100%**	5%	taxonomiques et génétiques	Inférence de la topologie d'un arbre *higher-level* à partir d'une matrice concaténée de 673 espèces. Datation de cette topologie à l'aide de BEAST (2 méthodes distinctes utilisées). Greffe de clades générés aléatoirement avec le package R *phytools* et représentant les différents clades terminaux considérés (principalement des genres). Deux positions de greffe ont été considérées: (i) au stem ages ; (ii) au crown ages.	(100 trees)

Tableau 1 (part. 1) – Synthèse bibliographique organisée par ordre chronologique des grandes phylogénies (>1000 feuilles) disponibles dans la littérature. Les mégaphylogénies (en gras), telles que définies dans ce Chapitre, sont celles dont le pourcentage de complétion est ≥50%. La mégaphylogénie des Saturniidae inférée dans ce Chapitre est la seule inférée à partir de données génomiques (*le nombre d'OTUs indiqué ne correspond pas à la mégaphylogénie des Saturniidae représentée dans la Figure 4 du fait du retrait des genre *Dirphiopsis* et *Othorene* (voir la partie Résultats)).

Jetz & Pyron 2018	Amphibia (class)	7 238	**100%**	56%	taxonomiques et génétiques	*backbone* + *subtrees* non chevauchants. Définition des lignées du backbone et des subtrees à l'aide d'un arbre inféré par maximum de vraisemblance à partir d'une matrice concaténée. Cet arbre a également été utilisé afin de définir les contraintes de position des espèces sans données génétiques, utilisées dans PASTIS. Lors de l'analyse effectuée avec MrBayes 3,2 (Ronquist et al. 2012) la topologie de l'arbre a été contrainte à être celle inférée par maximum de vraisemblance. Ainsi, l'ensemble d'arbres qu'ils ont publiés ne rendent pas compte de l'incertitude liée aux positions des clades pour lesquels des données génétiques sont disponibles (similairement à Tonini et al. 2016).	phylogénétique (très partiellement) et temporelle sur 10 000 arbres
Rabosky et al. 2018	Actinopterygii (class)	31 526	**100%**	37%	taxonomiques et génétiques	Inférence de la topologie à partir d'une matrice concaténée (supermatrice). Datation de la phylogénie inférée à l'aide de treePL. Placement des espèces sans données génétiques à l'aide d'un script Python développé par les auteurs, en fixant la topologie à correspondre à celle inférée à partir des données moléculaires. Ainsi, dans l'ensemble de 100 arbres générés, l'incertitude de topologie ne concerne que les espèces placées à partir des informations taxonomiques.	phylogénétique (très partiellement) sur 100 arbres
Upham et al. 2019	Mammalia (class)	5 911	**93%**	69%	taxonomiques et génétiques	*backbone* + *subtrees* non chevauchants. Définition des lignées du *backbone* et des *subtrees* à partir d'un arbre inférée par maximum de vraisemblance dans un premier temps, à partir d'une matrice concaténée. Cet arbre a également été utilisé afin de définir les contraintes de position des espèces sans données génétiques dans PASTIS.	phylogénétique et temporelle sur 10 000 arbres
Chazot et al. 2020	Nymphalidae (family)	2 866	45%	100%	génétiques	*backbone* + *subtrees* non chevauchants. Définition des lignées du *backbone* et des *subtrees* à partir de connaissances taxonomiques.	phylogénétique et temporelle sur 1000 arbres
Arnal et al., Chapitre 3	Saturniidae (family)	3 947*	**de 87,8 à 100%**	**100%**	**génomiques** et génétiques	*backbone* + subtrees non chevauchants, définis sur la base de Arnal et al. (in prep.; Chapitre 1). Contrairement aux autres études qui ont inferré des Mégaphylogenies à l'aide de la méthode *backbone* + *subtree*, nous avons séparé les étapes d'inférence des topologies et des temps de divergence. La prise en compte des incertitudes liées à la topologie et aux temps de divergence a été permise par la datation des topologies des réplicats de bootstrap.	phylogénétique et temporelle (11 arbres)

Tableau 1 (part. 2) – Synthèse bibliographique organisée par ordre chronologique des grandes phylogénies (>1000 feuilles) disponibles dans la littérature. Les mégaphylogénies (en gras), telles que définies dans ce Chapitre, sont celles dont le pourcentage de complétion est ≥50%. La mégaphylogénie des Saturniidae inférée dans ce Chapitre est la seule inférée à partir de données génomiques (*le nombre d'OTUs indiqué ne correspond pas à la mégaphylogénie des Saturniidae représentée dans la Figure 4 du fait du retrait des genre *Dirphiopsis* et *Othorene* (voir la partie Résultats)).

Les insectes constituent la majorité de la biodiversité terrestre décrite (Zhang 2011) mais une grande partie de leur diversité reste aujourd'hui inconnue. Récemment, le nombre d'espèces d'insectes a été estimé à 5,5 millions (de 2,6 à 7,8 millions ; Stork *et al.* 2015) alors qu'aujourd'hui le nombre d'espèces décrites d'insectes s'élève à près de 1 million (Zhang 2011). Moins d'un quart de la diversité des insectes est donc décrite, un chiffre frappant qui témoigne de notre méconnaissance de cette partie du monde vivant. C'est ce qu'on appelle le déficit linnéen (manque de connaissance taxonomique ; Brito 2010). En comparaison, le nombre d'espèces de vertébrés est très stable depuis de nombreuses années et la description d'une nouvelle espèce de grand vertébré terrestre en 2017 a eu un large retentissement scientifique (*i.e. Pongo tapanuliensis* ; Nater *et al.* 2017) et médiatique (Ter Minassian 2017). A ce déficit linnéen abyssal s'ajoute un déficit wallacéen (manque de données de distribution) qui n'est malheureusement pas moins conséquent (Diniz-Filho *et al.* 2010). De fait, tous types d'études confondus, les insectes reçoivent aujourd'hui moins d'attention que les Vertébrés ou les Spermatophytes (voir par exemple Beck *et al.* 2012 concernant les études macroécologiques). Nos connaissances phylogénétiques ont certes progressé lors de la dernière décennie (Misof *et al.* 2014 ; Peters *et al.* 2017 ; Espeland *et al.* 2018) mais aucune mégaphylogénie à la complétion comparable à celles inférées pour les Vertébrés n'a encore été publiée pour un groupe d'insecte, limitant de fait notre compréhension des patrons spatiaux et temporels spécifiques à ce groupe (voir cependant la phylogénie inférée par Chazot *et al.* 2020 dont la complétion spécifique est de 45%). C'est ce qu'on appelle le déficit darwinien (*i.e.* le manque de données phylogénétiques ; Diniz-Filho *et al.* 2013).

Si ces déficits sont tous des obstacles à une meilleure compréhension de la diversité des insectes, le premier frein à l'élaboration de mégaphylogénies est le déficit linnéen : il existe peu de larges clades d'insectes pour lesquels les connaissances taxonomiques sont assez avancées pour envisager d'inférer de tels arbres phylogénétiques. La famille des Saturniidae Boisduval 1837, au sein des Lépidoptères, fait partie de ces rares clades pour lesquels un tel travail est envisageable. La majorité de ces papillons volent de nuit et sont relativement faciles à échantillonner à l'aide d'un piège lumineux ; certains arborent des motifs alaires exceptionnels, des couleurs vives ou des formes étonnantes (par exemple des queues de plus de 10cm, *Eudaemonia argiphontes*) et beaucoup sont relativement faciles à élever. C'est pourquoi, depuis plusieurs siècles, les Saturniidae ont attisé la curiosité des naturalistes et des scientifiques qui les ont collectionnés et étudiés (Packard 1895 ; Bouvier 1927 ; Michener 1952 ; Rougeot 1955 ; Bénéluz 1986 ; Naumann & Peigler 2001). De larges collections, privées et publiques, ont ainsi été établies et constituent aujourd'hui une source d'informations inestimable. La collection du Muséum national d'Histoire naturelle de Paris, qui renferme plus de 60 000 spécimens de Saturniidae, en est un exemple. Celle-ci a servi de support à de nombreuses descriptions ou révisions taxonomiques et contient la majorité des spécimens de la collection Lemaire, auteur des dernières révisions taxonomiques majeures de la famille (Lemaire 1978, 1980, 1988, 2002) pour la faune néotropicale. Ces révisions ont fait considérablement avancer notre connaissance de la famille ; fondées sur des arguments

morphologiques, comportementaux et géographiques, elles font toujours référence aujourd'hui en tant que socle aux études plus récentes intégrant des données génétiques. La nouvelle classification de la famille en 10 sous-familles et 23 tribus par Arnal *et al.* (*in prep.* ; Chapitre 1) reprend ainsi très largement le travail entrepris par Lemaire et ses prédécesseurs. Plus récemment, l'avènement des codes-barres ADN (voir ci-dessous ; également appelés barcodes ADN) a permis une compréhension plus fine de la diversité des Saturniidae. Grâce à leur utilisation, de nombreuses lignées ont été redéfinies en des espèces distinctes, difficilement distinguables morphologiquement, mais que les divergences génétiques ont permis de découvrir et caractériser. Le nombre d'espèces de Saturniidae a ainsi connu une inflation significative : Kitching *et al.* (2018) ont comptabilisé près de 1 500 descriptions d'espèces nouvelles au cours des dix dernières années, élevant le nombre d'espèces à 3 462. Le nombre de descriptions annuelles reste aujourd'hui élevé et, de fait, de nombreux groupes monophylétiques d'individus phylogénétiquement très apparentés (Unités Taxonomiques Opérationnelles, OTUs en anglais) restent à décrire. Mais, comme mentionné plus haut, cela n'implique pas que la diversité des Saturniidae nous reste largement inconnue : la majorité de ces descriptions résultent de la séparation de lignées décrites en plusieurs espèces distinctes et non de la description de spécimens nouvellement échantillonnés. Ainsi, la famille des Saturniidae constitue certainement l'un des groupes d'insectes pour lequel le déficit linnéen est le mieux comblé.

C'est dans le début des années 2000 que fut proposé le principe du *barcoding* ADN proposé en premier lieu comme un outil d'identification des espèces, mais également comme un nouveau moyen de découvrir et caractériser des espèces encore inconnues et non décrites (Hebert *et al.* 2003 ; Stoeckle 2003 ; Hebert *et al.* 2004). Ce principe de l'identification par barcoding est simple : le code-barres ADN d'un échantillon inconnu est séquencé et comparé à l'ensemble des codes-barres d'une base de données associant ces séquences à des identifications spécifiques ; s'il est identique ou très similaire à l'une ou plusieurs de ces séquences, l'échantillon inconnu est identifié comme appartenant à l'espèce qu'elles représentent. Le succès d'une telle approche repose largement sur l'existence, pour le marqueur analysé et dans le groupe taxonomique considéré, d'un « *barcoding gap* » ; c'est-à-dire que la variabilité génétique à l'intérieur des espèces pour ce marqueur n'excède pas les divergences génétiques mesurées entre les espèces. Cette approche n'est cependant pas exempte de biais. Par exemple, la bactérie symbiotique *Wolbachia*, en influençant la transmission des mitochondries d'une génération à l'autre par incompatibilité cytoplasmique, peut induire de fortes divergences génétiques au sein même de populations (Smith *et al.* 2012) et ainsi conduire à une surestimation de la diversité spécifique. Elias *et al.* (2007) ont également démontré les limites de l'utilisation du code-barres ADN dans l'identification de spécimens de clades dont la diversification fut rapide et pour lesquels le phénomène d'*incomplete lineage sorting* est très fort. Malgré ces biais, qui semblent être marginaux (Smith *et al.* 2012), le barcoding a connu un succès retentissant dans la recherche en écologie et en taxonomie. Un fragment d'environ 650 paires de bases du gène Cytochrome Oxydase 1, qui code pour une enzyme impliquée

dans la chaine respiratoire mitochondriale, est le marqueur standard utilisé en barcoding chez les animaux (Hebert *et al.* 2003a, 2004). Son utilisation a été largement facilitée et accélérée par le développement de la plateforme Barcode of Life Data Systems (BOLD, www.boldsystems.org ; Ratnasingham & Hebert 2007) et des campagnes massives de *barcoding* ont permis la génération de codes-barres ADN pour des millions de spécimens de tous types d'organismes. En particulier, la campagne de *barcoding* visant à construire des librairies de référence pour l'ordre des Lépidoptères (Rougerie *et al.* 2007) est l'une des plus avancée et tout particulièrement pour la famille des Saturniidae grâce à la mobilisation de taxonomistes du monde entier. La librairie de codes-barres ADN pour cette famille est aujourd'hui unique par sa taille et son exhaustivité (Rougerie *et al.* 2019) : au 20 juin 2020, la banque de données hébergée par la plateforme en ligne BOLD renfermait les codes-barres ADN de 51729 spécimens de Saturniidae représentant la quasi-totalité de la diversité spécifique de la famille. Cette banque de données constitue une quantité d'information exceptionnelle qui pourrait être valorisée dans des études phylogénétiques. Wilson (2010) a montré que ce fragment du gène COI contenait autant (si ce n'est davantage) de sites informatifs que les gènes WG (*wingless*) et EF1a (codant pour la protéine facteur d'élongation 1-alpha), notamment à l'échelle intra-générique. Monteiro & Pierce (2001) ont également montré son efficacité dans l'inférence des relations au sein du genre *Bicyclus* (Lepidoptera : Numphalidae), malgré les forts niveau d'homoplasie relevés. C'est cependant dans l'inférence des nœuds superficiels, entre des espèces fortement apparentées, que ce marqueur est le plus intéressant, comme en témoigne la phylogénie des abeilles du genre *Lasioglossum* (Hymenoptera: Halictidae) proposée par Danforth (1999) ou, plus récemment, la phylogénie des espèces apparentées au moustique *Anopheles hyrcanus* (Diptera: Culicidae) proposée par Fang *et al.* (2017).

Il demeure cependant évident que le code-barres ADN n'est pas suffisant pour inférer des nœuds plus profonds (Klopfstein *et al.* 2010 ; Wilson 2010 ; Chapitre 2) et qu'il est indispensable de se tourner vers d'autres marqueurs pour la résolution de ces divergences plus anciennes. Ces dernières années, les progrès des techniques de séquençage ont permis la génération de marqueurs moléculaires plus nombreux et mieux adaptés aux différentes échelles temporelles étudiées (Lemmon & Lemmon 2013). Dans le projet ANR SPHINX, nous avons choisi d'utiliser les Éléments Ultra-Conservés du génome (*Ultra Conserved Elements* ; UCEs) et de les associer aux codes-barres ADN. Les UCEs font figures de candidats idéals grâce, notamment, à la diversité de l'information phylogénétique qu'ils contiennent. Un UCE peut en effet être découpé en trois parties qui ont des vitesses d'évolution propres (Tagliacollo & Lanfear 2018) : le cœur de l'UCE est la partie très conservée du marqueur et est informative pour l'inférence des nœuds profonds alors que les deux parties flanquantes ont des positions nucléotidiques bien plus variables et donc informatives pour l'inférence de nœuds superficiels. Proposée initialement par Faircloth *et al.* (2012), leur utilisation s'est avérée pertinente pour l'étude de relations phylogénétiques de profondeurs très variées allant des nœuds interfamiliaux (McCormack *et al.* 2012 ; Faircloth *et al.* 2015) aux nœuds intra-spécifiques (Newman & Austin 2016). Dans le cadre de l'ANR

SPHINX, nous avons donc séquencé massivement des marqueurs UCEs à l'aide de techniques de capture par hybridation (Gnirke *et al.* 2009) et d'un protocole optimisé pour réduire le temps de manipulation et le coût par échantillon (Cruaud *et al.* 2019). Nous avons conçu notre échantillonnage de telle sorte que l'inférence de l'ensemble des nœuds inter-génériques et des nœuds intra-génériques les plus profonds repose très majoritairement sur les marqueurs UCEs. Ainsi, les codes-barres ADN, disponibles pour un nombre beaucoup plus important de taxons, ont vocation à résoudre l'inférence des nœuds les plus superficiels.

Dans ce Chapitre, nous avons inféré la toute première phylogénie à l'espèce de la famille des Saturniidae. Pour cela, nous avons d'abord défini des Unités Taxonomiques Opérationnelles (OTUs) au sein de la Famille des Saturniidae à l'aide de l'ensemble des codes-barres ADN disponibles sur la plateforme BOLD. Puis nous avons séquencé 1381 UCEs pour 1007 OTUs afin d'inférer une mégaphylogénie en combinant UCEs et barcodes ADN. Nous avons adapté l'approche *backbone + subtrees*, consistant à découpler les inférences phylogénétiques entre un « squelette » et des sous-arbres non chevauchants, à notre jeu de données génomique en séparant notamment les étapes d'inférence de la topologie et d'estimation des temps de divergence. Contrairement à de nombreuses mégaphylogénies publiées (Tableau 1), nous avons considéré les incertitudes liées à la topologie et aux temps de divergence en datant les topologies issues des réplicats de bootstrap. L'ensemble des analyses ont été exécutées à l'aide d'un *pipeline* phylogénomique nouvellement établi. Nous discutons enfin de la difficulté d'inférer les taux de diversification spécifique à chaque branche d'une telle phylogénie.

Matériels et Méthodes

L'ensemble du *pipeline* présenté dans ce Chapitre est schématisé dans la Figure 1, représentée avant chaque étape décrite dans la section Matériels et Méthodes.

1. Définition des Unités Taxonomiques Opérationnelles (OTUs)

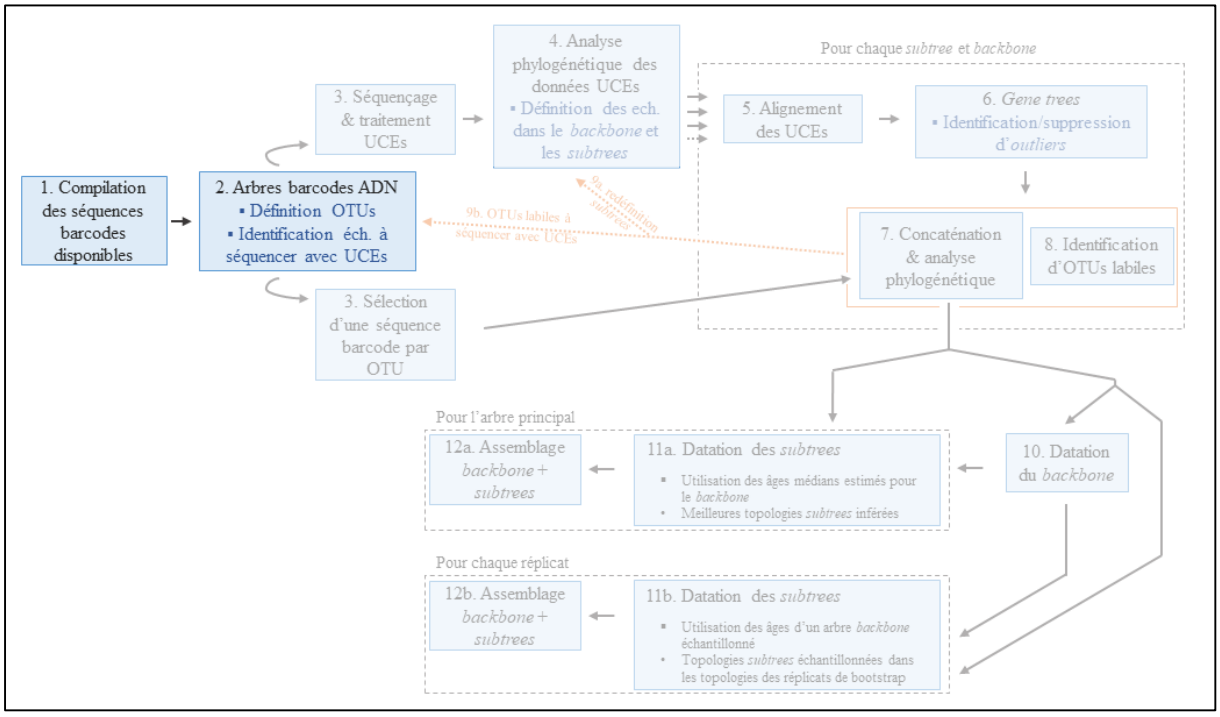


Figure 1A - Schéma du pipeline phylogénomique utilisé dans ce Chapitre pour inférer une mégaphylogénie datée (et des réplicats) de la Famille des Saturniidae. **Etapes 1 et 2** : construction d'arbre à partir de l'ensemble des codes-barres ADN disponible sur BOLD afin de définir des OTUs au sein de la famille des Saturniidae.

Bien que la connaissance taxonomique des Saturniidae soit très avancée, le nombre de nouvelles descriptions d'espèces effectuées ces dernières années témoignent des efforts qu'il nous reste à fournir pour prendre la pleine mesure de la diversité du groupe (Kitching *et al.* 2018). La nomenclature actuelle ne reflète donc pas intégralement la diversité de la famille. En revanche, cette diversité est très bien représentée dans la base de données *Barcode Of Life Data Systems* (BOLD ; www.boldsystems.org ; Ratnasingham & Hebert 2007) où les codes-barres ADN de 51729 spécimens de Saturniidae ont été compilées (au 20/06/2020). Par ailleurs, la majorité des nouvelles descriptions d'espèces, qui résultent de la séparation de lignées décrites en plusieurs espèces, reposent largement sur l'analyse de séquences de cette base de données. Dans le but de considérer la pleine diversité des Saturniidae, nous avons généré des arbres de distances génétiques pour chaque grand groupe (*i.e.* sous-familles ou tribus) à partir de l'ensemble des codes-barres ADN disponibles sur BOLD. Ces arbres ont été inférés avec l'outil '*Taxon ID Tree'* implanté dans la base de données en utilisant l'algorithme d'alignement de BOLD, et la méthode de Neighbour Joining pour analyser et visualiser sous forme d'arbres les distances génétiques corrigées selon le modèle de Kimura à deux paramètres (Kimura 1980).

Pour délimiter les différentes Unités Taxonomiques Opérationnelles (OTUs, pour *Operational Taxonomic Units*) qui seront ensuite considérées pour la construction de la mégaphylogénie nous avons pris en compte les résultats de l'assignation automatique sur BOLD de chaque séquence à un *Barcode Index Number* (BIN ; Ratnasingham & Hebert 2013) et leur correspondance avec les identifications

spécifiques proposées par les taxonomistes ayant contribué aux campagnes de barcoding. L'utilisation de ces BINs, en conjonction avec un ensemble d'informations morphologiques, biologiques (ex. phénologie, comportement, plantes hôtes) et biogéographiques nous a mené parfois à considérer des OTUs représentant des espèces cryptiques et non décrites. Nous avons pu observer empiriquement que la correspondance entre BINs et espèces est bonne (par ex. chez les *Copaxa*, voir Chapitre 2, mais aussi dans d'autres genres dont une révision poussée est en cours, comme les *Lonomia* ou les *Dirphia*). Néanmoins, nous n'avons pas considéré cette correspondance comme une règle et il arrive que l'approche de taxonomie intégrative évoquée plus haut conduise à considérer plusieurs espèces distinctes alors qu'elles partagent pourtant le même BIN ou, inversement, à ne considérer qu'une seule espèce pour des individus assignés par BOLD à des BINs différents.

2. Données nucléotidiques et stratégie de séquençage génomique

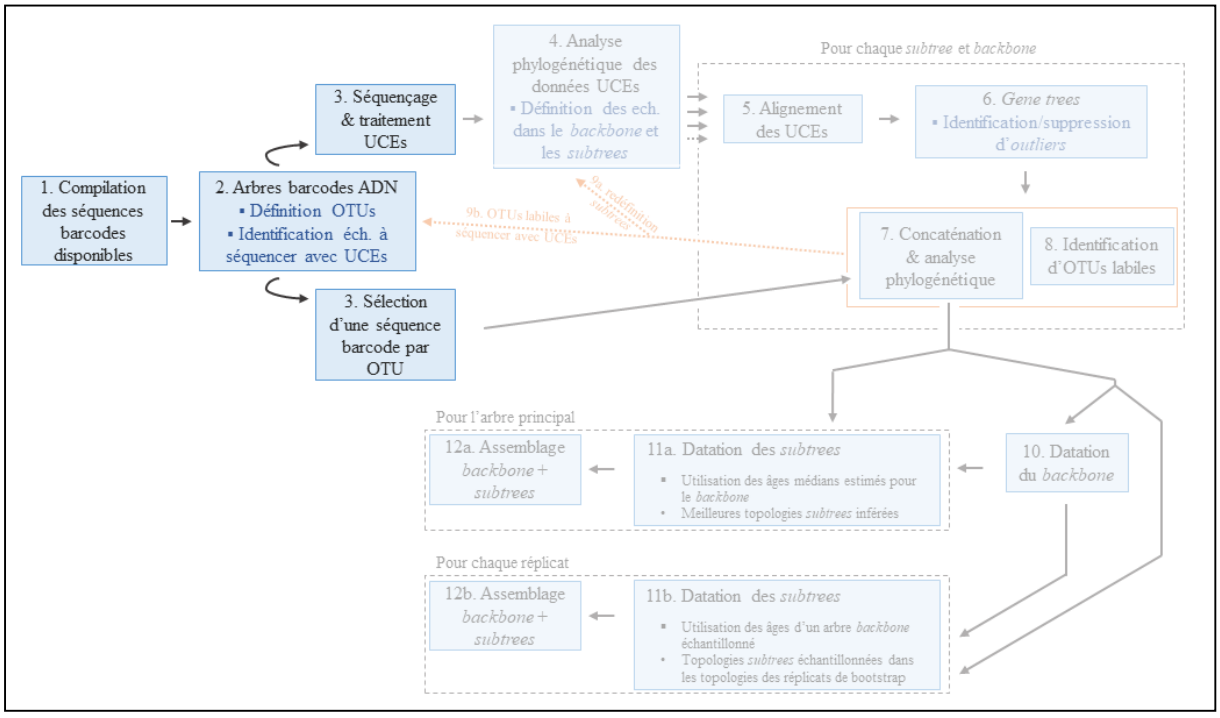


Figure 1B - Schéma du pipeline phylogénomique utilisé dans ce Chapitre pour inférer une mégaphylogénie datée (et des réplicats) de la Famille des Saturniidae. **Etapes 2 et 3** : Définition des OTUs, identification des échantillons à séquencer à l'aide des techniques de capture d'UCEs, séquençage et traitement des séquences brutes d'UCE.

Les marqueurs UCEs sont très performants dans l'inférence à la fois des nœuds profonds (McCormack *et al.* 2012 ; Faircloth *et al.* 2015) et superficiels (Newman & Austin 2016). L'idéal aurait donc été de séquencer ces marqueurs pour chaque OTU définie. Cependant une telle stratégie d'échantillonnage aurait eu un coût considérable (prix des réactifs et du séquençage ≈ 310 000 €; temps de manipulation ≈ 60 semaines). Nous avons donc opté pour une stratégie intermédiaire consistant à combiner les codes-barres ADN de l'ensemble des espèces aux UCE séquencés eux pour une sélection d'OTUs visant à maximiser la diversité phylogénétique.

2.1 Codes-barres ADN

Nous avons compilé une séquence du code-barres ADN pour chacune des OTUs définies dans la partie précédente. De manière générale et autant que possible, nous avons cherché à maximiser la sélection de barcodes issus des mêmes spécimens pour lesquels des UCEs ont été séquencées. Les codes-barres ADN ont été téléchargés depuis BOLD ; ils ont dans leur immense majorité été générés lors de la campagne de barcoding pour les Saturniidae, mais un certain nombre proviennent néanmoins d'autres projets (ex. inventaire des chenilles et papillons de l'Aire de Conservation de Guanacaste (ACG) au Costa-Rica (Janzen *et al.* 2012) ; Lépidoptères de Barro Colorado Island (BCI) au Panama (Basset *et al.* 2017) ; l'inventaire des papillons et chenilles de Papouasie-Nouvelle-Guinée ou du Kenya (Miller *et al.* 2013, 2014, 2016).

2.2 Ultra Conserved Elements (UCEs)

Nous avons cherché à optimiser le choix des échantillons séquencés à l'aide des techniques de capture d'UCEs de sorte qu'ils permettent avant tout l'inférence des nœuds profonds et semi-profonds lors des analyses phylogénétiques. Afin de sélectionner ces OTUs, nous avons utilisé les arbres de distances génétiques inférés à partir des codes-barres ADN pour définir les Unités Taxonomiques Opérationnelles. Ces arbres nous ont permis d'identifier des lignées assez clairement définies dont les espèces sont très proches les unes des autres. Nous avons alors séquencé (ou planifié le séquençage) des UCEs pour au moins un échantillon de chaque clade ainsi défini, maximisant de fait la diversité phylogénétique.

En parallèle, nous avons également séquencé des UCEs pour une grande partie des OTUs de l'ensemble des genres de la tribu des Attacini et pour les genres *Cricula*, *Dirphia*, *Periphoba*, *Eacles* et *Lonomia* afin de tester la performance de ces marqueurs dans l'inférence de nœuds plus superficiels et répondre à des problématiques spécifiques à ces groupes.

Nous ne détaillerons pas ici les étapes du *pipeline* que nous avons utilisé pour traiter les données brutes de séquençage. Celles-ci sont décrites dans le Chapitre 1 ainsi que dans la publication de Cruaud *et al.* (2019).

3. Définition du *backbone* et des *subtrees*

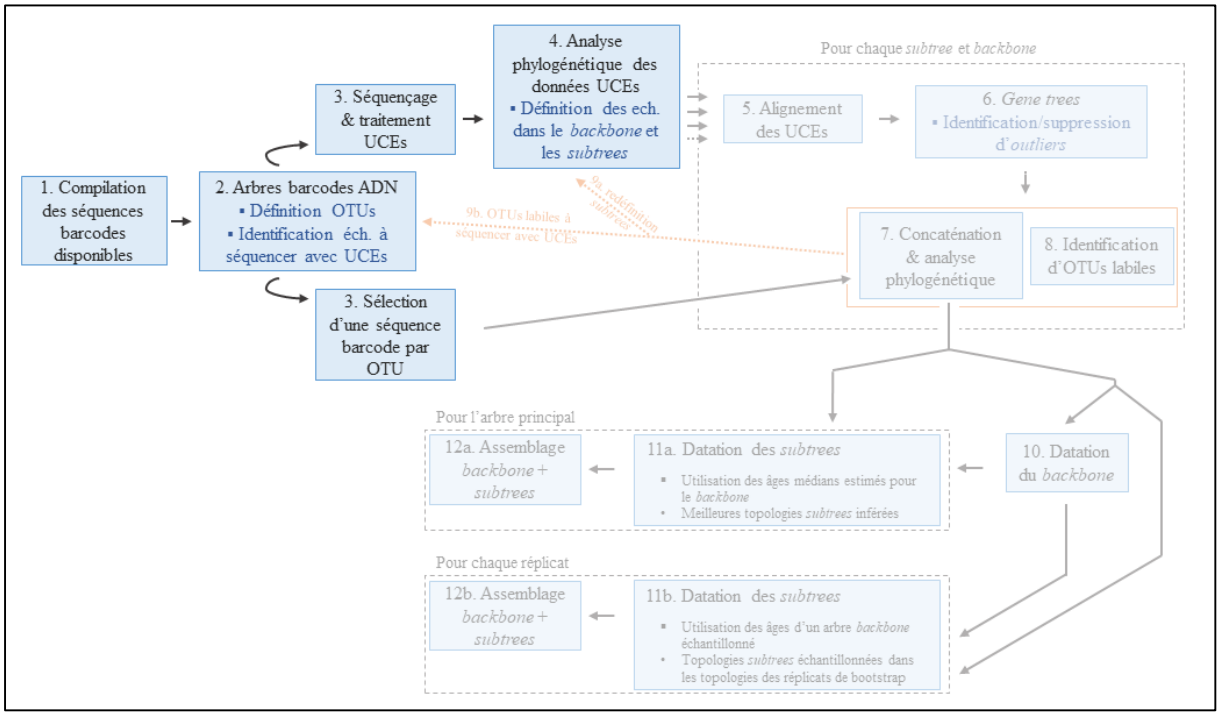


Figure 1C - Schéma du pipeline phylogénomique utilisé dans ce Chapitre pour inférer une mégaphylogénie datée (et des réplicats) de la Famille des Saturniidae. **Etape 4** : les subtrees considérés doivent impérativement être monophylétiques. Le jeu de données backbone doit comporter suffisamment d'échantillons pour que les relations phylogénétiques soient inférées avec précision mais sans que ce nombre soit trop élevé, ce qui impliquerait des temps de calcul élevés.

Afin d'inférer une mégaphylogénie de la famille des Saturniidae à l'aide des données moléculaires évoquées précédemment, nous avons opté pour l'approche *backbone* + *subtree* (littéralement colonne vertébrale + sous-arbres ; également appelée *backbone* + *patchs* par Upham *et al.* 2019). Cette approche repose sur la séparation du jeu de données en de multiples sous-jeux de données à partir desquels sont ensuite inférés indépendamment les *subtrees*. Ces *subtrees* sont non chevauchants et leur assemblage est permis par l'inférence d'un squelette phylogénétique, le *backbone*, représentatif de la diversité du groupe étudié. Nous avons préféré une approche *backbone* + *subtrees* (Jetz *et al.* 2012) aux approches utilisant une matrice unique (supermatrice) malgré que les méthodes par maximum de vraisemblance aient fait des progrès considérables et sont maintenant capables d'inférer des phylogénies à partir d'imposantes matrices (Stamatakis 2014 ; Nguyen *et al.* 2015 ; Zhou *et al.* 2018 ; Minh *et al.* 2020). Une approche par supermatrice a par exemple été utilisée par Rabosky *et al.* (2018) qui ont inféré un arbre unique des Actinoptérygiens (31 526 feuilles ; Tableau 1) et l'ont daté avec treePL (Smith & O'Meara 2012), un logiciel qui utilise le maximum de vraisemblance et qui repose sur les longueurs de branches du phénogramme (et non une matrice nucléotidique) pour estimer les temps de divergence. Or, nous voulions obtenir une mesure de l'incertitude des estimations des temps de divergence, ce qu'aujourd'hui seules les méthodes bayésiennes permettent. Cependant l'estimation des temps de

divergence à l'aide de méthodes bayésiennes serait trop longue pour être envisageable en utilisant une supermatrice, ce qui exclut de fait une telle approche.

L'approche *backbone* + *subtrees* nécessite une bonne connaissance des relations phylogénétiques profondes du groupe d'intérêt afin de définir correctement les différents *subtrees*. Les *subtrees* doivent effectivement être des clades dont la monophylie est établie et il est primordial que le nœud basal de chaque *subtree* soit représenté dans le *backbone*. Lors de tests préliminaires, nous avons effectivement remarqué que si ce dernier critère n'est pas rempli, les estimations des temps de divergence sont faussées et l'assemblage de la mégaphylogénie impossible. Enfin, il est préférable de limiter le nombre d'échantillons par *subtree* ainsi que de définir un nombre raisonnable de *subtrees* pour éviter des temps de calcul élevés dans l'inférences des topologies des différents *subtrees* et lors de la datation du *backbone*, respectivement. Dans ce Chapitre, nous avons défini les différents *subtrees* et les échantillons à intégrer au *backbone* à l'aide de la phylogénie des Saturniidae à l'échelle générique inférée dans le Chapitre 1. Si de telles connaissances sur le *backbone* phylogénétique du groupe manquent, il est toujours possible de les obtenir en effectuant une analyse rapide, sans procédure de bootstrap et sans recherche du meilleur modèle d'évolution nucléotidique, de l'ensemble de la matrice moléculaire (Jetz & Pyron 2018 ont par exemple utilisé cette approche ; Tableau 1). Une telle analyse prendrait moins d'une semaine de calcul à l'aide d'une vingtaine de *threads* sur un jeu de données similaire au notre.

4. Inférences de la topologie du *backbone* et des *subtrees*

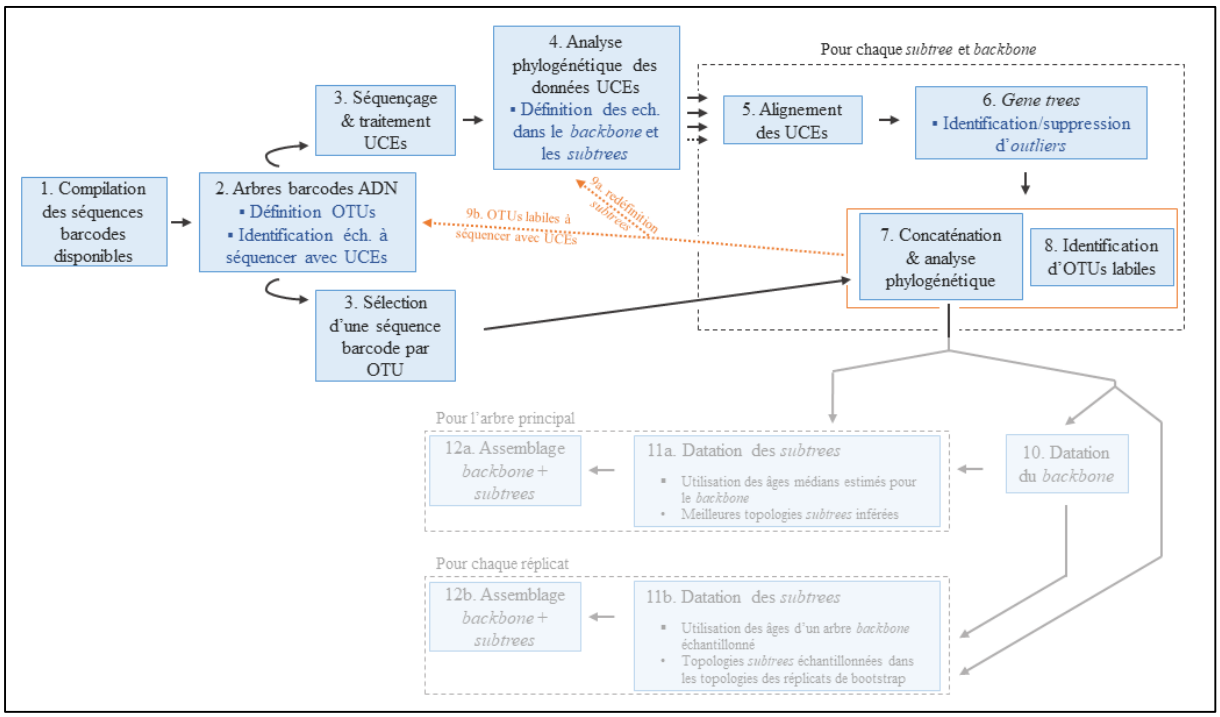


Figure 1D - Schéma du pipeline phylogénomique utilisé dans ce Chapitre pour inférer une mégaphylogénie datée (et des réplicats) de la Famille des Saturniidae. **Etape 5** : Les alignements sont effectués indépendamment pour chaque locus. **Etape 6** : Identification des outliers en utilisant Phylo-PMCOA. **Etape 8** : Identification des OTU labiles (rogue taxa) à l'aide de RogueNaRok. **Etape 9a** : Redéfinition des subtrees en cas de défaut dans la monophylie de ces derniers. **Etape 9b** : Les échantillons dont la position est labile sont priorisés pour les prochains séquençages d'UCEs.

Alignements des séquences

Pour chaque *subtree* ainsi que pour le *backbone*, nous avons considéré les loci pour lesquels plus de 50% des échantillons étaient disponibles. Les alignements ont été exécutés par *locus* à l'aide du logiciel MAFFT v7.313 (Katoh & Standley 2013). Pour les *loci* UCEs, nous avons également utilisé le logiciel Gblocks v0.91b (Castresana 2000) afin de les simplifier en retirant certains gaps ou positions peu partagées.

Identification d'*outliers*

Cette étape, qui ne concerne que les UCEs, repose sur le logiciel Phylo-PMCOA (de Vienne *et al.* 2012) et a pour but d'écarter trois types d'outliers : (i) un gène dont l'information phylogénétique est très différente de l'information des autres loci ; (ii) un échantillon dont la position phylogénétique est très labile d'un gene tree à l'autre ; (iii) la séquence d'un échantillon pour un gène donné lorsque la position de l'échantillon pour le *gene tree* correspondant est très différente des positions dans les autres *gene trees*. Les deux premiers points désignent ce qui est appelé des *complete outliers* alors que le point (iii) défini ce qu'on appelle un *cell outlier*.

Phylo-PMCOA identifie les *outliers* à partir de l'ensemble des *gene trees*. Nous avons donc préalablement inféré ces derniers à l'aide du logiciel IQ-TREE v1.6.7 (Nguyen *et al.* 2015) en sélectionnant le modèle d'évolution qui convient le mieux à chaque *locus* avec ModelFinder (Kalyaanamoorthy 2017), implémenté dans IQ-TREE, et l'option '*TESTNEW*'. Les *outliers* identifiés par Phylo-PMCOA ont ensuite été écartés des jeux de données à l'aide d'un script R.

Inférences de la topologie du *backbone* et des *subtrees*

Une fois les *outliers* écartés, nous avons concaténé les matrices de chaque *locus* UCE ainsi que l'alignement des codes-barres ADN à l'aide d'un script R. Nous avons ensuite inféré des phylogénies pour le *backbone* et l'ensemble des *subtrees* à l'aide de IQ-TREE v1.6.7 (Nguyen *et al.* 2015) en considérant deux partitions : (i) les UCEs concaténés et (ii) le barcode ADN. Les modèles d'évolutions ont été déterminés par ModelFinder (Kalyaanamoorthy 2017) et nous avons mesuré les supports des phylogénies en utilisant 1000 réplicats de *ultrafast bootstrap* (Hoang *et al.* 2018) ainsi que 1000 réplicats du *SH-aLRT branch test* (Guindon *et al.* 2010). Nous avons également enregistré les topologies bootstrap dans un fichier en utilisant le paramètre *-wbtl* de IQ-TREE.

Identification des OTUs labiles (*rogue taxa)*

Afin d'identifier les taxons dont la position est très incertaine (*i.e.* les *rogue taxa* ; Wilkinson 1996), nous avons utilisé RogueNaRok (Aberer *et al.* 2011, 2013) pour chaque *subtree*. RogueNaRok utilise les topologies des réplicats de bootstrap afin d'identifier les échantillons dont la position diffère significativement d'un réplicat à l'autre. Ce sont ces échantillons qui impactent le plus les valeurs de bootstrap. L'influence qu'a le retrait de taxons ou de clades des arbres de bootstrap est mesurée à l'aide

du critère d'optimisation R*elative Bipartition Information Content* (RBIC) sur l'arbre consensus résultant.

Le RBIC est défini ainsi pour un arbre consensus C':

$$RBIC(C') = \frac{\sum_{i=1}^{l} \sup(B_i)}{T-3}$$

avec *sup(Bi)* la fréquence relative du nœud dans les arbres de bootstrap

l le nombre de nœuds dont le support est supérieur au seuil -*c*

T le nombre de taxons initiaux.

Ici, nous avons utilisé –c=50, ce qui signifie que nous avons considéré des arbres de consensus majoritaire. Les nœuds présents dans moins de 50% des réplicats de bootstrap n'influencent donc pas le score *RBIC*. Nous avons aussi utilisé une valeur de 3 pour le paramètre *dropset size* (-s) et ainsi identifié des clades labiles de taille inférieure ou égale à 3 feuilles. Nous avons utilisé l'arbre inféré par IQ-TREE comme '*bestKnownTree*' pour optimiser les supports de bootstrap sur cet arbre probable. Lorsque RogueNaRok écarte un taxon ou un clade, il calcule à nouveau les scores RBIC pour chaque taxon ou clade de la phylogénie et s'arrête lorsque le RBIC ne peut plus être amélioré. Il est attendu que la majorité des *rogue taxa* soient des taxons pour lesquels le code-barres ADN est le seul marqueur génétique disponible. Le séquençage d'UCEs sera prioritaire pour ces échantillons afin d'inférer leur position phylogénétique. Pour de mieux comprendre comment se branchent les *rogue taxa* et dans le cadre exploratoire de ce Chapitre, nous avons gardé ces taxons dans la suite des analyses. A terme, lorsque l'ensemble des librairies UCEs auront été séquencées dans le cadre de l'ANR SPHINX, nous envisagerons d'écarter ces taxons qui peuvent avoir une influence conséquente sur les analyses comparatives.

5. Datation et assemblage de l'arbre

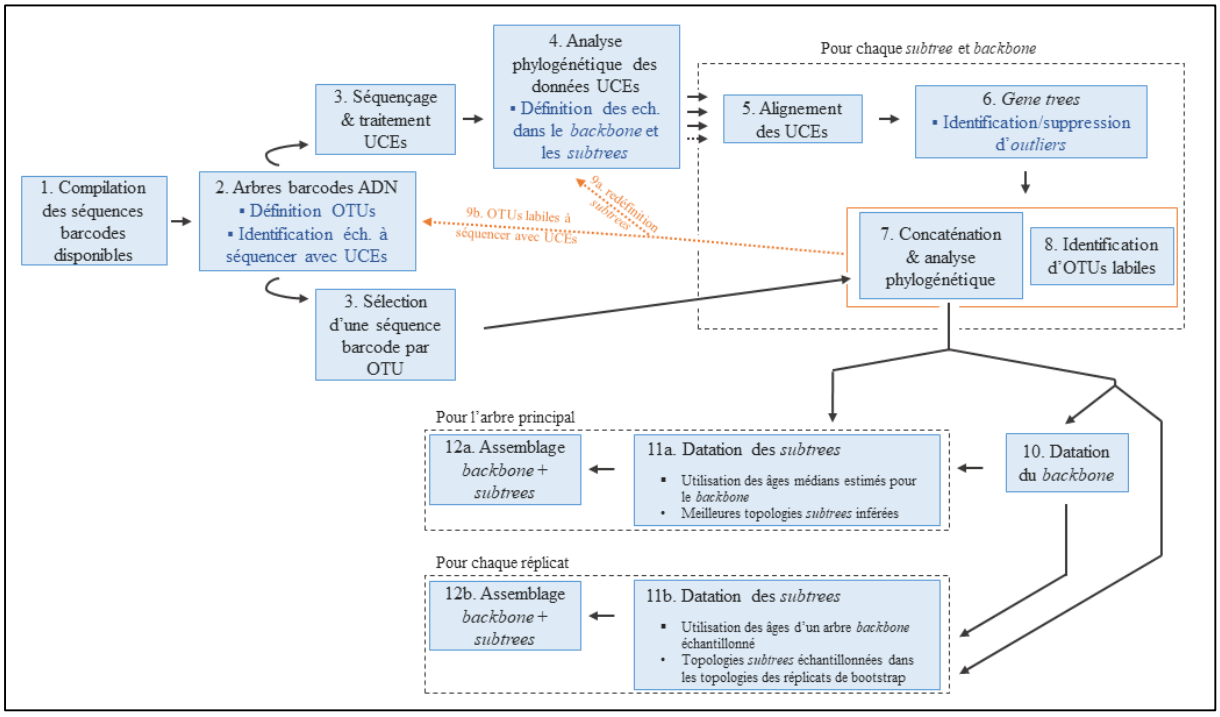


Figure 1E - Schéma du pipeline phylogénomique utilisé dans ce Chapitre pour inférer une mégaphylogénie datée (et des réplicats) de la Famille des Saturniidae. **Etape 10** : Datation du backbone à l'aide du logiciel MCMCtree et de calibrations fossiles et secondaires. **Etape 11** : Les âges médians estimés pour le backbone avec MCMCtree et les topologies subtrees estimées avec IQ-TREE sont utilisées pour dater les subtrees de l'arbre principal. Pour les réplicats, nous considérons les âges d'un arbre échantillonné aléatoirement dans la distribution postérieure produite par MCMCtree et datons des topologies échantillonnées aléatoirement au sein des réplicats de bootstrap des différents subtrees (sauvegardés lors de l'étape 7).

Datation du *backbone* et des *subtrees*

Afin d'inférer les temps de divergence, nous avons utilisé le logiciel MCMCtree appartenant à la suite de logiciels PAML v4.8 (Yang 1997 ; Yang & Rannala 2006). L'utilisation des matrices que nous avons générées (de 1 à 84 millions de nucléotides) impliquerait des temps de calcul considérables (jusqu'à plusieurs semaines). C'est pourquoi, que ce soit pour le *backbone* ou pour les *subtrees*, nous avons considéré des matrices nucléotidiques réduites : 50 UCEs sélectionnés aléatoirement et le code-barres ADN, définissant les deux partitions de l'analyse. Nous avons montré, dans le Chapitre 1, qu'une telle stratégie de réduction du jeu de données résultait en des estimations cohérentes. Lors de l'échantillonnage des 50 UCEs, nous nous sommes assurés que le nombre minimal de loci pour un échantillon était supérieur à 10. Lorsque ce critère n'était pas rempli, nous échantillonnions un autre ensemble d'UCEs jusqu'à validation de la condition. Si ce n'était toujours pas le cas après 100 randomisations, nous considérions l'ensemble de *loci* dont le nombre minimal d'UCEs pour un échantillon était le plus grand.

Afin de date le *backbone*, nous avons considéré deux fossiles et 5 calibrations secondaires (Wahlberg *et al.* 2013 ; voir le Chapitre 1 – Table S2 pour plus de détails). Les calibrations ont été définies à l'aide

d'une loi uniforme couvrant, pour les fossiles, l'incertitude liée à l'âge de fossilisation ou, pour les points de datation secondaires, l'intervalle de confidence à 95%. Afin d'établir l'influence de la sélection des loci sur l'estimation des temps de divergence, nous avons daté indépendamment le *backbone* avec deux ensembles de loci distincts (50_loci_random_A et 50_loci_random_B). Pour s'assurer de la convergence des analyses, nous avons également lancé la datation deux fois sur chaque ensemble de loci (*approx1* et *approx2*). Les calculs de la vraisemblance peuvent être très chronophages pour de larges alignements, nous avons donc, dans un premier temps, approximé les taux d'évolution moléculaire avec la méthode en maximum de vraisemblance introduite par Thorne *et al.* (1998) et qui est implémentée dans PAML v4.5 et les versions plus récentes. Nous avons lancé les analyses bayésiennes sur 2 millions de générations, dont les 300 000 premières furent écartées avant de calculer les moyennes et les intervalles de confidence de chaque paramètre.

Ce n'est que dans un second temps que nous avons daté les *subtrees*. Pour ce faire, nous avons utilisé la médiane des temps de divergence inférés à partir du *backbone* pour les nœuds partagés avec les différents *subtrees*. Contrairement au *backbone*, nous avons utilisé ici des calibrations fixes (*point calibrations*) qui ne prennent donc pas en compte l'incertitude concernant les âges des nœuds datés. Pour tous les *subtrees*, un seul ensemble de loci a été utilisé et l'analyse n'a été lancé qu'une seule fois. Des tests préliminaires ont montré que, d'un ensemble de loci à l'autre, les temps de divergence étaient relativement similaires et que deux analyses d'un même ensemble de *loci* convergeaient vers des estimations strictement identiques. Également, ces tests ont permis de mettre en évidence que les estimations ne variaient pas d'une génération à la suivante. Pour la datation de l'ensemble des *subtrees*, nous avons donc utilisé une seule génération, sans *burnin*.

Afin de prendre en considération les incertitudes liées à la topologie et aux temps de divergence, nous avons reproduit les étapes de datation des *subtrees* et d'assemblage (voir paragraphe suivant) plusieurs fois. Pour ce faire, nous avons échantillonné aléatoirement un arbre parmi la distribution postérieure produite par MCMCtree dont nous avons utilisé les temps de divergence afin de dater un nouvel ensemble de topologies *subtree* échantillonnées parmi les réplicats de bootstrap des analyses IQ-TREE.

Combinaison du *backbone* avec les *subtrees*

Une fois l'ensemble des topologies datées, nous avons combiné le *backbone* avec les *subtrees* à l'aide d'un script R qui repose notamment sur la fonction *bind.tree* du package *ape* (Paradis & Schliep 2019). Les temps de divergences des nœuds communs au *backbone* et aux *subtrees* étant similaires, la juxtaposition de ces éléments est possible. Les petites erreurs d'arrondi qui ont été introduites lors de la phase de datation ont ensuite été corrigées à l'aide de la fonction *force.ultrametric* du package R *phytools* (Revell 2012).

6. Mesure des taux de diversification

Afin de mesurer l'évolution des taux de diversification au sein de la phylogénie des Saturniidae, nous avons utilisé deux méthodes : l'approche développée par Höhna *et al.* (2019) qui utilise l'environnement RevBayes (Höhna *et al.* 2016) et le logiciel Bayesian Analysis of Macroevolutionary Mixtures (BAMM v2.5 ; Rabosky *et al.* 2013, Rabosky 2014a). Les deux méthodes sont des approches bayésiennes qui utilisent la méthode de Monte-Carlo par chaînes de Markov à sauts réversibles pour explorer les différents modèles candidats permettant d'expliquer la diversité observée dans les phylogénies. Elles diffèrent cependant en de nombreux points. L'approche de Höhna *et al.* (2019) a l'avantage de modéliser les shifts de diversification sur les lignées éteintes, ce que ne permet pas BAMM. De plus, cette méthode permet d'inférer des taux de diversification spécifiques à chaque branche contrairement à BAMM. BAMM est quant à lui un logiciel destiné à détecter et quantifier l'hétérogénéité des taux de diversification. BAMM est plus approprié pour détecter de larges changements (*shifts*) dans l'évolution des taux de diversification. Contrairement à RevBayes, BAMM considère un ensemble de régimes de diversification partagés par plusieurs branches successives. Ces régimes sont délimités par des *shifts* dont BAMM détermine dans un second temps la significativité.

Nous avons utilisé le modèle de Höhna *et al.* (2019) en discrétisant le prior sur les taux de diversification (distribution lognormale) en 6 catégories et en utilisant 5000 générations. En parallèle, nous avons lancé une analyse BAMM en considérant les *priors* suggérés par la fonction *setBAMMpriors* du package R *BAMMtools* v2.1.6 (Rabosky *et al.* 2014b). Dans l'analyse BAMM, nous avons utilisé 50 millions de générations. Les résultats obtenus avec BAMM peuvent être fortement influencés par les *priors* utilisés (Moore *et al.* 2016 ; Rabosky *et al.* 2017) mais nous n'avons pas testé ici différentes valeurs du paramètre *expectedNumberOfShifts* dans BAMM. Notre analyse est ici davantage destinée à tester l'applicabilité de ces méthodes et à fournir des mesures préliminaires des taux de diversification qu'à tester d'éventuels *shifts* de diversification.

7. Plateforme de calcul

L'ensemble des analyses ont été effectuées sur la plateforme de calcul Genotoul (Genopole Toulouse). Cette plateforme, qui fait partie du réseau IBiSA (Infrastructures en Biologie Santé et Agronomie), est équipée de 3054 cœurs de calcul (6218 *threads*), de 34 Tera Byte de mémoire et de 3000 Tera Byte d'espace disque. Pour les analyses réalisées dans ce Chapitre nous avons pleinement profité de ces capacités de calcul, utilisant parfois plus de 1500 *threads* simultanément. Cependant, l'ensemble des analyses que nous présentons ici sont réalisables avec un matériel plus modeste.

Résultats

1. Définition des Unités Taxonomiques Opérationnelles (OTUs)

Nous avons identifié 3947 OTUs de Saturniidae. La majorité des OTUs ont été assignées à un nom d'espèce. Cependant, nous n'avons pas été en mesure d'identifier avec suffisamment de confiance 26% des OTUs (Tableau 2). Beaucoup de ces OTUs sont des lignées qui n'ont effectivement pas encore été décrites. Cependant, une part non négligeable des 26% correspond certainement à des lignées qui ont été décrites mais n'ont pas encore été identifiées comme telles dans les librairies de codes-barres ADN. Compte tenu de cette incertitude, la complétion de notre jeu de donnée s'élève de 87,8[5] à 100%. De façon frappante, le pourcentage d'OTUs non nommées varie largement d'un groupe à l'autre. Un patron qui peut en partie s'expliquer par la popularité des différents groupes auprès des collectionneurs amateurs ou des scientifiques. Ainsi, au sein de la tribu des Attacini dont la grande taille et les motifs alaires ont fait la renommée, seulement 7% des OTUs ne sont pas assignées à une espèce décrite. Un tel chiffre contraste avec celui obtenu pour les Cercophaninae qui n'ont pas toujours été considérés comme des Saturniidae et dont les deux tiers des OTUs ne sont pas décrites.

Tableau 2 - Nombre d'OTUs dans la phylogénie des Saturniidae.

Clades (selon la classification proposée dans le Chapitre 1)		Nombre d'espèces selon Kitching et al. 2018	Nombre d'OTUs considérées dans ce chapitre	Nombre d'OTUs non identifiées ou non décrites
Saturniidae		**3462**	**3947**	**1033**
Agliinae		5	5	0
Arsenurinae		100	135	18
	Almeidaiini	2	2	0
	Arsenurini	88	133	18
Bunaeinae		576	633	250
	Antistathmopterini	4	3	0
	Bunaeini	269	324	137
	Eochroini	10	13	5
	Eudaemoniini	3	3	0
	Micragonini	220	231	99
	Pseudapheliini	17	12	2
	Urotini	53	47	7
Ceratocampinae		300	274	84
	Bathyphlebiini	62	88	11
	Ceratocampini	31	51	15
	Dryocampini	207	235	58
Cercophaninae		27	87	58
	Cercophanini	4	5	1
	Janiodini	23	82	57
Hemileucinae		1594	1785	435
	Hemileucini	1433	1610	388
	Lonomiini	154	170	46
	Polythysanini	7	5	1
Hirpidinae		19	30	7
Oxyteninae		70	83	27
Salassinae		32	41	9
Saturniinae		739	774	145
	Attacini	174	164	12
	Saturniini	558	591	130
	Soliini	7	19	3

[5] Si aucune des OTUs non-identifiées n'est une espèce déjà décrite : $\frac{3947}{3462+1033} = 87{,}8\%$.

2. Données nucléotidiques

Nous avons sélectionné un code-barres ADN par OTU préalablement définie, en maximisant la qualité (*i.e.* en minimisant le nombre de positions ambigües) et la longueur des séquences. Au total, nous avons assemblé sur BOLD (http://www.boldsystems.org/) un jeu de données de 3942 séquences de codes-barres ADN qui nous avons ensuite téléchargées pour nos analyses. En revanche, aucun code-barres ADN n'était disponible pour 5 OTUs pour lesquels nous avons séquencé des UCEs (voir ci-dessous). La grande majorité (environ 82%) des séquences de codes-barres sont complètes (658 paires de bases), et la taille minimale est de 405 paires de bases. Seules 3 séquences ont moins de 500 paires de bases, et 12 moins de 600.

Dans le cadre de l'ANR SPHINX, nous avons, à la date du 20 juin 2020, généré des UCEs pour 1007 échantillons représentatifs de l'ensemble des groupes de Saturniidae (Tableau 3). Les disparités dans les efforts d'échantillonnage s'expliquent principalement par la concentration de nos efforts de séquençage pour certains groupes comme la tribu des Attacini ou le genre *Eacles* (Bathyphlebiini). Bien que cet échantillonnage soit déjà très dense, nous avons planifié de nouveaux séquençages pour le renforcer dans des groupes pour lesquels peu de séquences UCEs ont été générées. Comme présenté dans le Chapitre 1, nous avons également généré des séquences UCEs pour 14 *outgroups* appartenant à la super-famille des Bombycoidea : 1 Brahmaeidae, 1 Endromidae, 1 Lemoniidae et 11 Sphingidae. L'ensemble des sous-familles connues de Sphingidae ont été échantillonnées ainsi qu'une espèce de chaque genre de la tribu des Smerinthini, dont un fossile a été décrit. Le nombre d'UCEs par échantillon varie de 92 à 1198. La moyenne du nombre d'UCEs par échantillon est de 802.4 ; ce nombre est homogène d'un groupe à l'autre (Tableau 3).

Subtree	Groupe taxonomique	Nombre d'échantillons	Nombre d'échantillons séquencés avec UCEs	Nombre min. et max. d'UCEs	Moyenne du nombre d'UCEs	Nombre de *cell outliers*	Support moyen SH-ALRT/UFBOOT	Pourcentage de nœuds faiblement soutenus	Nombre de *rogue taxa*
Agliinae	Sous-famille Agliinae	5	4	745-1112	904.4	50	65/86.3	0,75	0
Antheraea	Genre *Antheraea*	77	38	422-1161	847.4	726	82.2/87.7	0,41	1
Antistathmopterini ●	Genre *Antistathmoptera*	3	2	NA	NA	NA	NA	NA	NA
Arsenurinae	Sous-famille Arsenurinae	133	42	118-1172	757.93	371	82.1/88.4	0,5	0
Attacini	Tribu Attacini	163	131	105-1175	738.4	1697	91.5/91.6	0,25	4
Aurivillius	Genre *Aurivillius*	25	4	732-951	845.4	19	76.1/70.7	0,79	0
Automerina_A	Genres *Automeris*, *Hypermerina*, *Leucanella*, *Mexicantha* et *Pseudautomeris*	451	39	360-1137	827.1	458	NA	NA	12
Automerina_B	Genres *Auroraia*, *Austrolippa*, *Automerella*, *Erythromeris* et *Molippa*	74	11	360-1122	869.2	111	77.6/82.9	0,58	1
Automerina_C	Genres *Gamelia*, *Gamelioides*, *Hylesiopsis* et *Prohylesia*	126	9	511-1138	818.2	56	80.8/90.1	0,4	0
Automerina_D	Genres *Eubergia* et *Hylesia*	285	18	510-1138	819.3	68	72.3/87	0,58	2
Automerina_E	Genres *Automerina*, *Automeropsis*, *Hyperchiria* et *Hyperchirioides*	58	7	759-1149	912.7	37	79.1/89.7	0,6	0
Backbone	**Saturniidae**	**180**	**163**	**351-1189**	**865.8**	**2586**	**96.6/97**	**0,15**	**NA**
Bathyphlebiini	Tribu Bathyphlebiini	88	54	131-1131	713.1	469	87.1/89.4	0,37	2
Bunaeini_A	Genres *Athletes*, *Bunaea*, *Bunaeopsis*, *Cirina*, *Gonimbrasia*, *Gynanisa*, *Imbrasia*, *Lobobunaea*, *Melanocera*, *Protogynanisa*, *Pseudimbrasia* et *Ubaena*	213	64	95-1198	853.1	643	78.6/85.3	0,48	10
Bunaeini_B	Genres *Cinabra*, *Pseudobunaea* et *Rohaniella*	69	11	153-1127	777.2	80	75.6/89.9	0,57	0
Bunaeini_C	Genres *Heniocha* et *Leucopteryx*	17	12	786-913	859.1	149	87.7/82.6	0,44	0
Catacantha	Genre *Catacantha*	22	3	353-1122	862.3	0	79/76	0,81	0
Ceratocampini	Tribu Ceratocampini	51	13	154-1157	811.7	135	76.2/89.3	0,54	0
Cercophaninae	Sous-famille Cercophaninae	87	26	133-1124	760.3	219	78.9/87	0,45	2
Dryocampini_A	Genres *Adeloneivaia*, *Adelowalkeria*, *Almeidella*, *Ceratesa*, *Citioica*, *Giacomellia*, *Jaiba*, *Megaceresa*, *Mielkesia*, *Oiticella*, *Psigida*, *Ptiloscola*, *Rachesa* et *Scolesa*	135	34	92-1007	693.1	265	84.5/91.7	0,39	1
Dryocampini_B	Genres *Anisota*, *Cicia*, *Dacunju*, *Dryocampa*, *Othorene*, *Psilopygida*, *Psilopygoides*, *Syssphinx*	101	22	195-998	682.1	101	77.4/88.1	0,5	1
Eochroini	Tribu Eochroini	13	4	473-1080	839.6	43	84/89.7	0,58	0
Eudaemoniini ●	Tribu Eudaemoniini	3	2	NA	NA	NA	NA	NA	NA

Tableau 3 (part. 1) – Liste synthétique des sous-jeux de données génomiques considérés dans la construction de la mégaphylogénie des Saturniidae. Les *subtrees* marqués par un rond noir ont été inclus dans le *backbone* du fait du faible nombre d'UCEs et/ou d'échantillons. La ligne correspondant au *backbone* a été mise en évidence (gras). Les nombres d'UCEs indiqués correspondent aux nombres de loci obtenus avant d'écarter ceux pour lesquels moins de 50% des échantillons étaient disponibles. Les trois dernières colonnes sont des indicateurs de la résolution des différents *subtrees*. Un nœud non soutenu a été considéré comme tel lorsque SH-ALRT < 80 ou UltraFast Bootstrap < 95.

Subtree	Groupe taxonomique	Nombre d'échantillons	Nombre d'échantillons séquencés avec UCEs	Nombre min. et max. d'UCEs	Moyenne du nombre d'UCEs	Nombre de *cell outliers*	Support moyen SH-ALRT/UFBOOT	Pourcentage de nœuds faiblement soutenus	Nombre de *rogue taxa*
Hemileucina_A	Genres *Arias*, *Cerodirphia*, *Coloradia*, *Dirphiella*, *Hemileuca*, *Hidripa*, *Hispaniodirphia*, *Lemaireodirphia*, *Manodirphia*, *Meroleuca*, *Ormiscodes*, *Paradirphia*, *Rhodirphia*, *Winbrechlinia et Xanthodirphia*	289	32	583-1189	908.3	544	81.2/87.4	0,46	4
Hemileucina_B	Genres *Dirphia*, *Dirphiopsis*, *Eudyaria*, *Heliconisa*, *Ithomisa*, *Kentroleuca*, *Periphoba* et *Pseudodirphia*	302	98	110-1189	801.3	1236	84.7/91.9	0,38	2
Hirpidinae	Sous-famille Hirpidinae	30	2	853-908	888.7	0	75.5/84.6	0,55	0
Lemaireia	Genre *Lemaireia*	14	4	858-935	886.0	35	73.6/74.3	0,77	1
Loepa	Genre *Loepa*	68	20	680-1137	870.0	62	86.3/91.4	0,39	0
Lonomiini	Tribu Lonomiini	170	55	204-971	768.7	591	83.8/88.7	0,38	3
Micragonini	Tribu Micragonini	231	51	98-1099	768.6	507	77.1/82.1	0,61	7
Oxyteninae	Sous-famille Oxyteninae	83	13	355-1062	730.1	64	79/89.3	0,48	2
Polythysanini	Tribu Polythysanini	7	6	786-1162	907.0	19	99.8/99	0,17	0
Pseudapheliini	Tribu Pseudapheliini	12	7	398-954	737.3	18	95.3/93.4	0,36	0
Salassinae	Sous-famille Salassinae	41	11	110-1112	821.4	94	83/89.1	0,4	0
Saturniinae_A	Genres *Actias* et *Argema*	46	18	442-1121	875.8	129	85.9/91.9	0,24	0
Saturniinae_B	Genres *Austrocaligula*, *Neodiphthera*, *Opodiphthera*, *Pararhodia* et *Syntherata*	72	13	302-1113	760.5	99	78.1/83.6	0,56	1
Saturniinae_C	Genres *Copaxa* et *Saturnia*	233	43	112-1168	814.9	458	80.9/88.4	0,42	2
Saturniinae_D	Genres *Cricula* et *Eosia*	76	53	174-1085	811.9	536	89.5/92.4	0,33	0
Saturniinae_E ●	Genres *Antherina* et *Ceranchia*	4	2	NA	NA	NA	NA	NA	NA
Soliini	Tribu Soliini	19	3	834-1139	938.3	0	80/79.5	0,78	0
Travassosula ●	Genre *Travassosula*	4	1	NA	NA	NA	NA	NA	NA
Urotini	Tribu Urotini	47	25	83-1024	727.2	315	88.6/82.2	0,43	2

Tableau 3 (part. 1) – Liste synthétique des sous-jeux de données génomiques considérés dans la construction de la mégaphylogénie des Saturniidae. Les *subtrees* marqués par un rond noir ont été inclus dans le *backbone* du fait du faible nombre d'UCEs et/ou d'échantillons. La ligne correspondant au *backbone* a été mise en évidence (gras). Les nombres d'UCEs indiqués correspondent aux nombres de loci obtenus avant d'écarter ceux pour lesquels moins de 50% des échantillons étaient disponibles. Les trois dernières colonnes sont des indicateurs de la résolution des différents *subtrees*. Un nœud non soutenu a été considéré comme tel lorsque SH-ALRT < 80 ou UltraFast Bootstrap < 95.

3. Définition du *backbone* et des *subtrees*

Nous avons défini 41 *subtrees* dont le nombre d'échantillons est très variable (Tableau 3). Nous avons incorporé les *subtrees* Antistathmopterini, Eudaemoniini, Saturniinae_E et Travassosula au *backbone* car le nombre d'échantillons était très faible et/ou parce que nous n'avions séquencé des UCEs que pour un seul échantillon (Tableau 3). Ces *subtrees* pourront être considérés dans des analyses ultérieures, lorsque des UCEs auront été séquencés pour de nouveaux échantillons. Pour l'inférence du *backbone*, nous avons considéré 180 échantillons de Saturniidae en prenant soin de considérer des échantillons représentatifs de chaque sous-jeux de données pour que les racines de chaque *subtree* soit présentes dans le *backbone*. Si, en théorie, la sélection de deux échantillons par *subtree* permet d'inférer l'ensemble des temps de divergence de la mégaphylogénie, nous avons considéré jusqu'à 12 échantillons d'un même *subtree* pour limiter des biais d'inférence liés à un manque d'échantillonnage (Poe 1998 ; Heath *et al.* 2008).

4. Inférences de la topologie du backbone et des *subtrees*

Identification d'*outliers*

Phylo-MCOA n'a identifié aucun *complete outlier* dans les différents sous-jeux de données (gène dont l'information phylogénétique est discordante ; espèce dont la position varie drastiquement d'un *gene tree* à l'autre). En revanche, nous avons identifié des *cell outliers* (*i.e.* séquences dont l'information phylogénétique est discordante avec celles des autres loci de l'espèce) dans l'ensemble des sous-jeux de données sauf pour les *subtrees* Catacantha, Hirpidinae et Soliini dont le nombre d'échantillons séquencés avec des marqueurs UCEs était très limité. Le nombre de *cell outliers* est globalement très limité (environ 2% des UCEs). Logiquement, le nombre de *cell outliers* identifiés apparaît corrélé au nombre d'échantillons pour lesquels nous disposons d'UCEs. Les *cell outliers* ont été écartés des alignements.

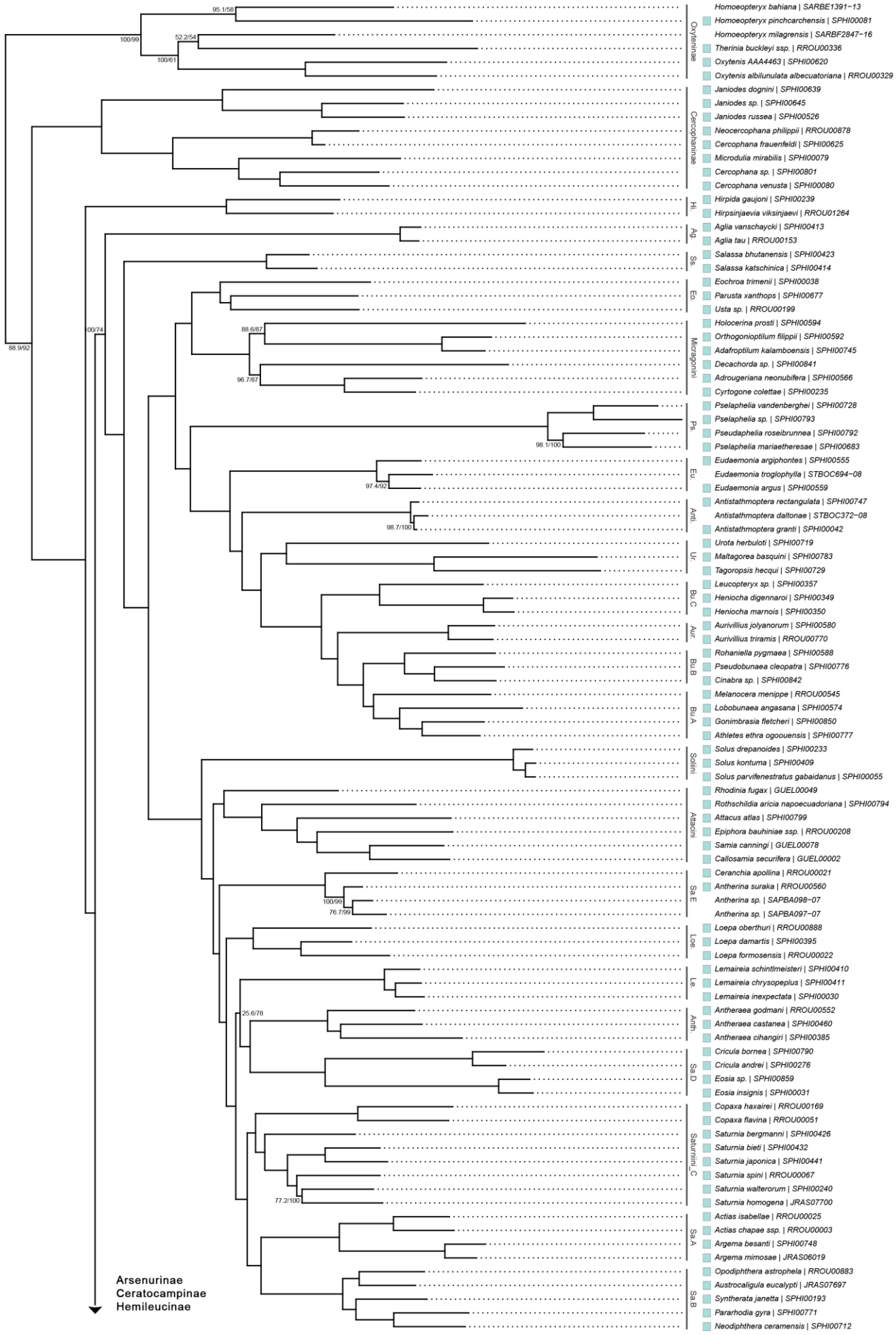
Topologie du *backbone*

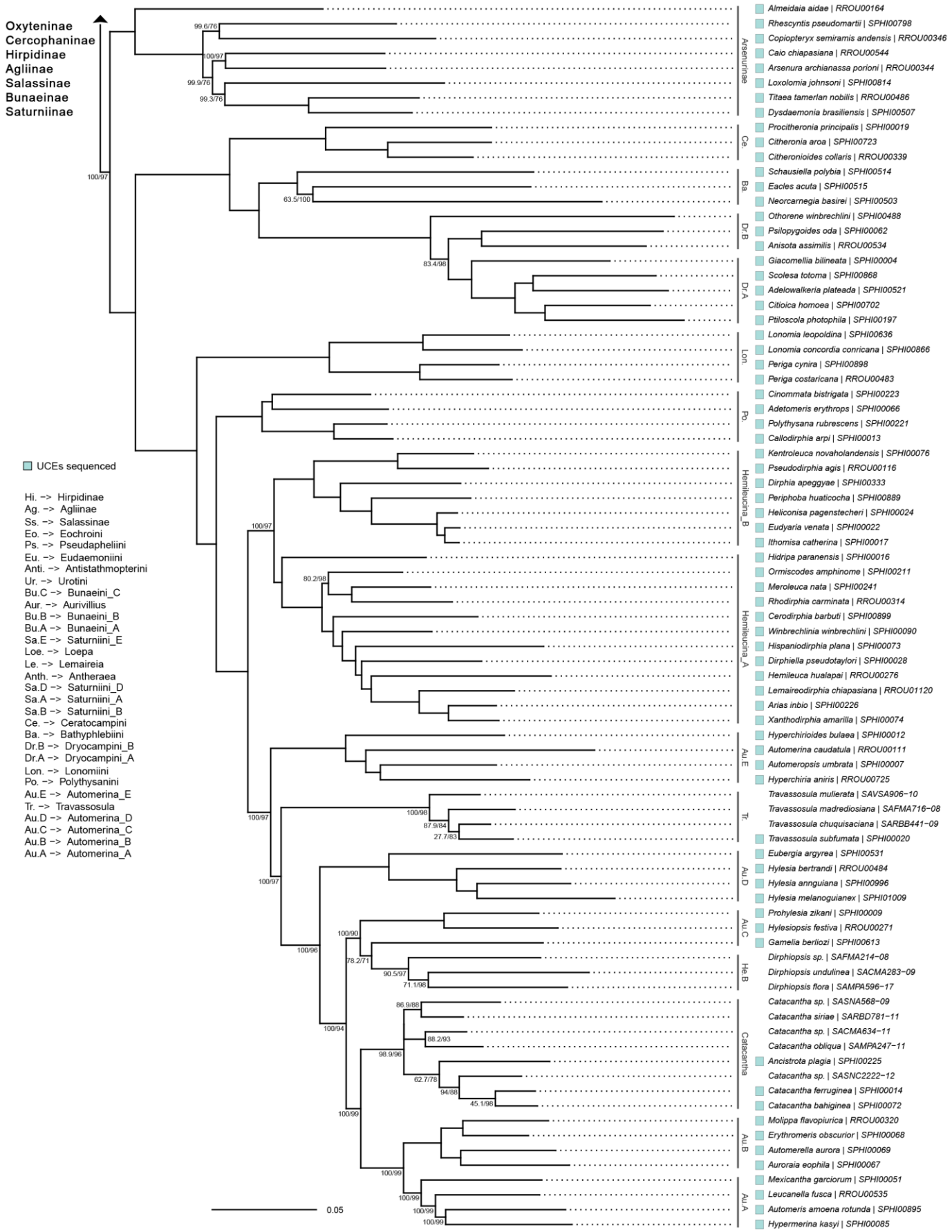
La topologie *backbone* que nous avons inférée dans ce Chapitre est bien soutenue et très semblable à celle inférée dans le Chapitre 1 (Figure 2). Cependant, certains points diffèrent de façon remarquable. Nos résultats suggèrent une position basale de la sous-famille Hirpidinae vis-à-vis du groupe composé par les sous-familles Agliinae, Salassinae, Saturniinae, Bunaeinae, Arsenurinae, Ceratocampinae et Hemileucinae alors que dans le Chapitre 1 nous avions inférée pour les Hirpidinae une position de groupe frère aux grandes sous-familles Néotropicales uniquement. La position de la sous-famille Agliinae s'en retrouve également modifiée : celle-ci est inférée comme apparentée aux autres sous-familles distribuées dans l'Ancien Monde (Salassinae, Bunaeinae et Saturniinae) alors que, dans les analyses par concaténation effectuées dans le Chapitre 1 nous avions estimé que les sous-familles Agliinae et Hirpidinae étaient groupes frères. Cette position avait déjà été inférée à l'aide du logiciel

ASTRAL (Mirarab & Warnow 2015) qui utilise le Modèle de Coalescence Multi Espèces (Multi-Species Coalescent Model - MSCM) dans le Chapitre 1. Nous avons également inféré avec de bons supports que le genre *Antheraea* est apparenté au clade *Eosia*/*Cricula*. La position du genre *Lemaireia,* ici positionné en groupe frère du clade *Antheraea*/*Eosia*/*Cricula,* est quant à elle non résolue.

L'ensemble des *subtrees* définis sont retrouvés comme monophylétiques au sein du *backbone* à l'exception des *subtrees* Dryocampini_B et Hemileucina_B. En se basant sur les résultats obtenus dans le Chapitre 1, nous avions défini le genre *Othorene* comme faisant partie du *subtree* Dryocampini_B. Or, nous avons inféré ici une position alternative, en dehors de ce *subtree,* dans une position de groupe frère du reste des membres de la tribu Dryocampini. Le cas du *subtree* Hemileucina_B est différent et est lié à la position du genre *Dirphiospsis*. Dans le Chapitre 1, nous avions utilisé les séquences d'un spécimen identifié comme *Dirphiopsis sp.* que nous n'avons pas été en mesure d'identifier à l'espèce. Ce spécimen étant le seul représentant du genre *Dirphiopsis* pour lequel nous avions généré des séquences UCEs, nous ne disposions donc que des codes-barres ADN pour inférer la position du genre dans le *backbone*. Or sa position ici est très différente de celle estimée dans le Chapitre 1 et implique la paraphylie du *subtree* Hemileucina_B. Puisque la monophylie des *subtrees* est un prérequis du *pipeline* d'analyse, nous avons, pour la suite des analyses, écarté les genres *Dirphiopsis* et *Othorene*.

Figure 2 (2 pages suivantes) - Phylogénie du backbone inférée à partir d'une matrice combinant les marqueurs UCEs et les codes-barres ADN. Les supports SH-ALRT (gauche) et UltraFast Boostrap (droite) sont indiqués pour chaque nœud sauf s'ils sont égaux à 100/100. A chaque feuille de l'arbre nous avons indiqué le nom de l'espèce et le code de l'échantillon. Des rectangles bleus signalent les échantillons pour lesquels nous disposions d'UCEs. Des barres verticales délimitent les différents subtrees. Si une abréviation a été utilisée, le nom complet du subtree est indiqué à gauche de la figure.


Homoeopteryx bahiana | SARBE1391−13
Homoeopteryx pinchcarchensis | SPHI00081
Homoeopteryx milagrensis | SARBF2847−16
Therinia buckleyi ssp. | RROU00336
Oxytenis AAA4463 | SPHI00620
Oxytenis albilunulata albecuatoriana | RROU00329
Janiodes dognini | SPHI00639
Janiodes sp. | SPHI00645
Janiodes russea | SPHI00526
Neocercophana philippii | RROU00878
Cercophana frauenfeldi | SPHI00625
Microdulia mirabilis | SPHI00079
Cercophana sp. | SPHI00801
Cercophana venusta | SPHI00080
Hirpida gaujoni | SPHI00239
Hirpsinjaevia viksinjaevi | RROU01264
Aglia vanschaycki | SPHI00413
Aglia tau | RROU00153
Salassa bhutanensis | SPHI00423
Salassa katschinica | SPHI00414
Eochroa trimenii | SPHI00038
Parusta xanthops | SPHI00677
Usta sp. | RROU00199
Holocerina prosti | SPHI00594
Orthogonioptilum filippii | SPHI00592
Adafroptilum kalamboensis | SPHI00745
Decachorda sp. | SPHI00841
Adrougeriana neonubifera | SPHI00566
Cyrtogone colettae | SPHI00235
Pselaphelia vandenberghei | SPHI00728
Pselaphelia sp. | SPHI00793
Pseudaphelia roseibrunnea | SPHI00792
Pselaphelia mariaetheresae | SPHI00683
Eudaemonia argiphontes | SPHI00555
Eudaemonia troglophylla | STBOC694−08
Eudaemonia argus | SPHI00559
Antistathmoptera rectangulata | SPHI00747
Antistathmoptera daltonae | STBOC372−08
Antistathmoptera granti | SPHI00042
Urota herbuloti | SPHI00719
Maltagorea basquini | SPHI00783
Tagoropsis hecqui | SPHI00729
Leucopteryx sp. | SPHI00357
Heniocha digennaroi | SPHI00349
Heniocha marnois | SPHI00350
Aurivillius jolyanorum | SPHI00580
Aurivillius triramis | RROU00770
Rohaniella pygmaea | SPHI00588
Pseudobunaea cleopatra | SPHI00776
Cinabra sp. | SPHI00842
Melanocera menippe | RROU00545
Lobobunaea angasana | SPHI00574
Gonimbrasia fletcheri | SPHI00850
Athletes ethra ogoouensis | SPHI00777
Solus drepanoides | SPHI00233
Solus kontuma | SPHI00409
Solus parvifenestratus gabaidanus | SPHI00055
Rhodinia fugax | GUEL00049
Rothschildia aricia napoecuadoriana | SPHI00794
Attacus atlas | SPHI00799
Epiphora bauhiniae ssp. | RROU00208
Samia canningi | GUEL00078
Callosamia securifera | GUEL00002
Ceranchia apollina | RROU00021
Antherina suraka | RROU00560
Antherina sp. | SAPBA098−07
Antherina sp. | SAPBA097−07
Loepa oberthuri | RROU00888
Loepa damartis | SPHI00395
Loepa formosensis | RROU00022
Lemaireia schintlmeisteri | SPHI00410
Lemaireia chrysopeplus | SPHI00411
Lemaireia inexpectata | SPHI00030
Antheraea godmani | RROU00552
Antheraea castanea | SPHI00460
Antheraea cihangiri | SPHI00385
Cricula bornea | SPHI00790
Cricula andrei | SPHI00276
Eosia sp. | SPHI00859
Eosia insignis | SPHI00031
Copaxa haxairei | RROU00169
Copaxa flavina | RROU00051
Saturnia bergmanni | SPHI00426
Saturnia bieti | SPHI00432
Saturnia japonica | SPHI00441
Saturnia spini | RROU00067
Saturnia walterorum | SPHI00240
Saturnia homogena | JRAS07700
Actias isabellae | RROU00025
Actias chapae ssp. | RROU00003
Argema besanti | SPHI00748
Argema mimosae | JRAS06019
Opodiphthera astrophela | RROU00883
Austrocaligula eucalypti | JRAS07697
Syntherata janetta | SPHI00193
Pararhodia gyra | SPHI00771
Neodiphthera ceramensis | SPHI00712
Oxyteninae
Cercophaninae
Hi.
Ag.
Ss.
Eo.
Micragonini
Ps.
Eu.
Anti.
Ur.
Bu. C
Aur.
Bu. B
Bu. A
Solini
Attacini
Sa. E
Loe.
Le.
Anth.
Sa. D
Saturniini_C
Sa. A
Sa. B
95.1/58
100/99
52.2/54
100/61
00/74
88.9/92
88.6/87
96.7/87
98.1/100
97.4/92
98.7/100
100/99
76.7/99
25.6/78
77.2/100
Arsenurinae
Ceratocampinae
Hemileucinae


Oxyteninae
Cercophaninae
Hirpidinae
Agliinae
Salassinae
Bunaeinae
Saturniinae
UCEs sequenced
Hi. -> Hirpidinae
Ag. -> Agliinae
Ss. -> Salassinae
Eo. -> Eochroini
Ps. -> Pseudapheliini
Eu. -> Eudaemoniini
Anti. -> Antistathmopterini
Ur. -> Urotini
Bu.C -> Bunaeini_C
Aur. -> Aurivillius
Bu.B -> Bunaeini_B
Bu.A -> Bunaeini_A
Sa.E -> Saturniini_E
Loe. -> Loepa
Le. -> Lemaireia
Anth. -> Antheraea
Sa.D -> Saturniini_D
Sa.A -> Saturniini_A
Sa.B -> Saturniini_B
Ce. -> Ceratocampini
Ba. -> Bathyphlebiini
Dr.B -> Dryocampini_B
Dr.A -> Dryocampini_A
Lon. -> Lonomiini
Po. -> Polythysanini
Au.E -> Automerina_E
Tr. -> Travassosula
Au.D -> Automerina_D
Au.C -> Automerina_C
Au.B -> Automerina_B
Au.A -> Automerina_A
Arsenurinae
Ce.
Ba.
Dr.B
Dr.A
Lon.
Po.
Hemileucina_B
Hemileucina_A
Au.E
Tr.
Au.D
Au.C
He.B
Catacantha
Au.B
Au.A
Almeidaia aidae | RROU00164
Rhescyntis pseudomartii | SPHI00798
Copiopteryx semiramis andensis | RROU00346
Caio chiapasiana | RROU00544
Arsenura archianassa porioni | RROU00344
Loxolomia johnsoni | SPHI00814
Titaea tamerlan nobilis | RROU00486
Dysdaemonia brasiliensis | SPHI00507
Procitheronia principalis | SPHI00019
Citheronia aroa | SPHI00723
Citheronioides collaris | RROU00339
Schausiella polybia | SPHI00514
Eacles acuta | SPHI00515
Neorcarnegia basirei | SPHI00503
Othorene winbrechlini | SPHI00488
Psilopygoides oda | SPHI00062
Anisota assimilis | RROU00534
Giacomellia bilineata | SPHI00004
Scolesa totoma | SPHI00868
Adelowalkeria plateada | SPHI00521
Citioica homoea | SPHI00702
Ptiloscola photophila | SPHI00197
Lonomia leopoldina | SPHI00636
Lonomia concordia conricana | SPHI00866
Periga cynira | SPHI00898
Periga costaricana | RROU00483
Cinommata bistrigata | SPHI00223
Adetomeris erythrops | SPHI00066
Polythysana rubrescens | SPHI00221
Callodirphia arpi | SPHI00013
Kentroleuca novaholandensis | SPHI00076
Pseudodirphia agis | RROU00116
Dirphia apeggyae | SPHI00333
Periphoba huaticocha | SPHI00889
Heliconisa pagenstecheri | SPHI00024
Eudyaria venata | SPHI00022
Ithomisa catherina | SPHI00017
Hidripa paranensis | SPHI00016
Ormiscodes amphinome | SPHI00211
Meroleuca nata | SPHI00241
Rhodirphia carminata | RROU00314
Cerodirphia barbuti | SPHI00899
Winbrechlinia winbrechlini | SPHI00090
Hispaniodirphia plana | SPHI00073
Dirphiella pseudotaylori | SPHI00028
Hemileuca hualapai | RROU00276
Lemaireodirphia chiapasiana | RROU01120
Arias inbio | SPHI00226
Xanthodirphia amarilla | SPHI00074
Hyperchirioides bulaea | SPHI00012
Automerina caudatula | RROU00111
Automeropsis umbrata | SPHI00007
Hyperchiria aniris | RROU00725
Travassosula mulierata | SAVSA906−10
Travassosula madrediosiana | SAFMA716−08
Travassosula chuquisaciana | SARBB441−09
Travassosula subfumata | SPHI00020
Eubergia argyrea | SPHI00531
Hylesia bertrandi | RROU00484
Hylesia annguiana | SPHI00996
Hylesia melanoguianex | SPHI01009
Prohylesia zikani | SPHI00009
Hylesiopsis festiva | RROU00271
Gamelia berliozi | SPHI00613
Dirphiopsis sp. | SAFMA214−08
Dirphiopsis undulinea | SACMA283−09
Dirphiopsis flora | SAMPA596−17
Catacantha sp. | SASNA568−09
Catacantha siriae | SARBD781−11
Catacantha sp. | SACMA634−11
Catacantha obliqua | SAMPA247−11
Ancistrota plagia | SPHI00225
Catacantha sp. | SASNC2222−12
Catacantha ferruginea | SPHI00014
Catacantha bahiginea | SPHI00072
Molippa flavopiurica | RROU00320
Erythromeris obscurior | SPHI00068
Automerella aurora | SPHI00069
Auroraia eophila | SPHI00067
Mexicantha garciorum | SPHI00051
Leucanella fusca | RROU00535
Automeris amoena rotunda | SPHI00895
Hypermerina kasyi | SPHI00085
0.05

Support bootstrap du *backbone* et des *subtrees* et identification des *rogue taxa*

Nos résultats témoignent, dans leur globalité, de la pertinence de l'utilisation combinée des marqueurs UCEs et des codes-barres ADN pour inférer les nœuds profonds comme superficiels. Nous avons mesuré la qualité des supports de deux manières différentes : la moyenne des supports sur l'ensemble des phylogénies et le pourcentage de nœuds faiblement soutenus (*i.e.* SH-ALRT < 80 ou UltraFast Bootstrap < 95). Si les supports moyens mesurés dans les différents *subtrees* sont globalement élevés, les résultats sont hétérogènes (Tableau 3) : la topologie inférée à partir du jeu de données Attacini est par exemple bien soutenue (SH-ALRT : 91.5 ; UltraFast Boostrap : 91.6) alors que celle estimée pour le *subtree* Aurivillius l'est modérément (76.1/70.7). Par ailleurs, de nombreux nœuds sont faiblement soutenus dans les différents *subtrees*. Ces mauvais supports peuvent être expliqués par (i) l'effet de l'*incomplete lineage sampling*, (ii) par le manque d'informations phylogénétiques des marqueurs UCEs et des codes-barres ADN utilisés, notamment lorsque les taux de diversification s'accélèrent ou, plus probablement, (iii) par un échantillonnage d'UCEs insuffisant (Tableau 3). Les topologies les plus soutenues ont en effet été obtenues à partir de jeux de données pour lesquels nous disposions d'un nombre important d'échantillons avec UCEs : par ex. Polythysanini (6/8 des échantillons avec UCEs), Saturniidae_A (18/47) et Attacini (131/165). Ces deux mesures nous permettent d'identifier quels sous-jeux de données doivent concentrer nos futurs efforts de séquençage d'UCEs.

L'utilisation des codes-barres ADN s'est avérée globalement efficace pour inférer les nœuds superficiels (Annexe 1). Cependant, lorsque les taux de spéciation étaient plus élevés (branches internes plus courtes), comme c'est le cas au sein du genre *Bunaeopsis* (*subtree* Bunaeini_A), l'utilisation de ce marqueur atteint ses limites. Les UCEs en revanche ont été très performants dans l'inférence de nœuds profonds comme superficiels. Les topologies des *subtrees* pour lesquels nous avions séquencé ces marqueurs pour presque l'ensemble des OTUs sont particulièrement bien soutenues (par ex. topologies des genres *Eacles, Lonomia*).

A noter que nous n'avons pas été en mesure de mesurer les supports de la topologie Automerina_A car les UltraFast Bootstrap n'ont pas convergé après plus de 96h de calcul sur 40 processeurs. Le jeu de données Automerina_A est celui rassemblant le plus grand nombre d'échantillons (452) mais nous disposions d'un nombre limité d'échantillons avec UCEs (39). La présence de certains taxons à la position très labile (*i.e. rogue taxa*) pour lesquels nous n'avons pas encore séquencé d'UCEs explique certainement la non-convergence de la mesure des supports.

En utilisant les topologies des différents réplicats de bootstrap, nous avons identifié, à l'aide du logiciel *RogueNaRoK*, un total de 60 *rogue taxa* répartis dans 19 des 37 sous-jeux de données. Le jeu de données Automerina_A est celui pour lequel nous avons identifié le plus grand nombre de *rogue taxa* (Tableau 3). Bien que les Bootstraps n'aient pas convergé pour ce *subtree*, nous estimons que ce résultat est informatif car les *outliers* identifiés sont des échantillons phylogénétiquement isolés et pour lesquels le

code-barres ADN est le seul marqueur disponible. C'est également le cas de la très grande majorité des *rogue taxa* identifiés dans les différents *subtrees*.

Identification de paraphylies

Le but de ce travail n'est pas de discuter exhaustivement de l'ensemble des topologies des *subtrees*. Cependant, il nous paraît pertinent de discuter des paraphylies identifiées :

- [*Subtree* Attacini] Le genre *Samia* est paraphylétique du fait d'un clade composé par l'espèce *Samia watsoni* et de deux OTUs apparentées, non décrites. Cette espèce avait précédemment été placée dans un genre distinct, *Archaeosamia* (Brechlin 2007), qui a ensuite été synonymisé avec le genre *Samia*. Nos résultats soutiennent la validité du genre *Archaeosamia*.
- [*Subtree* Automerina_A] Le genre *Automeris*, qui est le plus riche en espèces au sein des saturnidés, est paraphylétique. Cependant, nous ne pouvons pas discuter des différents clades d'*Automeris* tant que nous n'avons pas généré des UCEs pour l'ensemble des groupes d'espèces.
- [*Subtree* Automerina_A] Le genre *Mexicantha* apparait paraphylétique. Mais seulement un échantillon sur trois a été séquencé avec des UCEs. Ce point nécessite d'être approfondi.
- [*Subtree* Automerina_A] Le genre *Pseudautomeris* apparaît paraphylétique du fait de la position de l'espèce *Automeris masti*. Cette dernière ferait donc partie des *Pseudautomeris*. Ce résultat devra néanmoins être confirmé une fois que nous aurons généré des UCEs pour cette OTU.
- [*Subtree* Automerina_B] Nous inférons trois clades distincts pour le genre *Molippa*. Si le groupe d'espèces *superba* pour lequel nous disposons d'UCEs se branche avec le genre *Automerella*, la position du groupe d'espèces apparentées à *M. bertrandi* doit être confirmée une fois que des UCEs auront été séquencés pour l'une des espèces du groupe.
- [*Subtree* Automerina_C] L'unique espèce du genre *Hylesiopsis* se branche à l'intérieur du genre *Prohylesia*. Ces deux genres pourraient être synonymisés.
- [*Subtree* Bunaeini_A] Le genre très diversifié *Gonimbrasia* est paraphylétique du fait d'un groupe de trois OTUs dont *G. anna* qui vient se brancher avec le genre *Ubaena*. Un autre groupe d'espèce, celui de *G. eblis*, se branche avec le genre *Cirina* mais nous ne pouvons pas tirer de conclusion sur sa position car nous ne disposons pas d'UCEs pour ce groupe.
- [*Subtree* Bunaeini_A] L'espèce *Protogynanisa probsti* se branche à l'intérieur du genre *Gynanisa* et doit être considéré comme un *Gynanisa*. De fait, le genre bispécifique *Protogynanisa* est paraphylétique et doit être redéfini.
- [*Subtree* Bunaeini_C] Le genre *Leucopteryx* se branche à l'intérieur du genre *Heniocha*, le rendant paraphylétique. Ce clade est l'objet d'une révision taxonomique en collaboration avec Stefan Naumann (Berlin).
- [*Subtree* Catacantha] Le genre *Catacantha* a été inféré comme paraphylétique mais cette inférence devra être confirmée après avoir séquencé des UCEs pour les groupes d'espèces concernés.

- [*Subtree* Cercophaninae] Le genre *Cercophana*, comme nous l'avions indiqué dans le Chapitre 1, est paraphylétique. Il nécessite une révision taxonomique.
- [*Subtree* Dryocampini_A] Le genre *Ceratesa* a été inféré à l'intérieur du genre *Scolesa*. Mais ce résultat, soutenu par la seule utilisation des codes-barres ADN, nécessite d'être approfondi à l'aide d'UCEs.
- [*Subtree* Dryocampini_B] Au sein de la tribu des Dryocampini, le genre *Dacunju* se branche à l'intérieur du genre *Cicia*. Des analyses supplémentaires sont cependant nécessaires pour confirmer cette hypothèse.
- [*Subtree* Dryocampini_B] Nous avons inféré une polyphylie du genre *Syssphinx*. Bien que la position de deux des quatre clades de *Syssphinx* n'ait pas été inférée à l'aide d'UCEs, la position de *S. molina* implique la paraphylie du genre qui doit être redéfini.
- [*Subtree* Hemileucina_A] Le genre *Dirphiella* a été inféré paraphylétique du fait de l'espèce *Dirphiella niobe* qui ne se branche pas avec le reste des espèces du genre. Cependant, ce résultat doit être confirmé une fois que nous aurons généré des UCEs pour cette espèce.
- [*Subtree* Hemileucina_A] La position de l'espèce *Ormiscodes schmidtnielseni* implique la paraphylie du genre. Mais ce résultat devra être confirmé une fois que des UCEs seront disponible pour cette OTU.
- [*Subtree* Hemileucina_B] Lors du Chapitre 1, nous avions constaté la paraphylie du genre *Dirphia* mais nous nous refusions à la moindre conclusion du fait du faible échantillonnage taxonomique. Les résultats obtenus lors de ce Chapitre nous permettent d'affirmer que le genre *Dirphia* est polyphylétique. Les genres apparentés, *Heliconisa*, *Ithomisa* et *Eudyaria* se branchent à l'intérieur du genre *Dirphia* et, bien que leur *habitus* soient très différents, doivent être considéré comme faisant partie de ce dernier. Le genre *Periphoba*, bien qu'il soit monophylétique, vient également se brancher au sein du genre *Dirphia*. Tous ces genres pourraient être synonymisés ou, les distances phylogénétiques étant relativement importantes, le clade pourrait être séparé en trois genres distincts. Un travail spécifique à cette lignée est en cours et a été l'objet d'un stage de Master en 2020 (Reboud *et al.* in prep).
- [*Subtree* Hemileucina_B] Nous inférons que le genre *Kentroleuca* est polyphylétique mais ce résultat devra être confirmé par une analyse utilisant des UCEs pour chaque groupe d'espèces. Pour l'instant, nous ne disposons d'UCEs que pour un groupe d'espèces de *Kentroleuca*. Mais celui-ci se branche à l'intérieur du genre *Pseudodirphia* et implique sa paraphylie.
- [*Subtree* Micragonini] Le genre *Vegetia* se branche à l'intérieur du genre *Ludia* ce qui suggèrent que ces deux genres doivent être synonymisés.
- [*Subtree* Micragonini] Le genre *Cyrtogone* est paraphylétique du fait de l'espèce *C. martinae* qui se branche à la base du clade composé par les genres *Basquiniana* et *Cyrtogone*. Les longueurs de branches à l'intérieur de ce clade sont relativement courtes et peuvent justifier le fait de synonymiser ces deux genres dont l'usage est relativement récent (Darge, 2015).

- [*Subtree* Micragonini] Le genre *Adrougeriana* est paraphylétique du fait de la position de l'espèce *A. nenia* qui se branche au sein du clade formé par les genres *Cyrtogone* et *Basquiniana* (voir ci-dessus).
- [*Subtree* Pseudapheliini] Au sein de la petite tribu des Pseudapheliini, la position du genre *Pseudaphelia* implique la paraphylie du genre *Pselaphelia*. La Tribu nécessite une révision taxonomique.
- [*Subtree* Saturniini_B] Les trois genres de saturnidés australiens les plus diversifiés (*Neodiphthera, Opodiphthera* et *Syntherata*) ont besoin d'une révision. Le genre *Syntherata* est monophylétique si on considère *Neodiphthera excavus* comme faisant partie du genre. Le genre *Opodiphthera* est lui aussi monophylétique si *Opodiphthera pristina* est considérée comme un *Neodiphthera*. Quant au genre *Neodiphthera,* il est paraphylétique du fait de la position du genre bispécifique *Pararhodia*.
- [*Subtree* Urotini] Le branchement des espèces *Tagoropsis flavinata* et *T. hanningtoni* avec le genre *Tagoropsiella* implique la paraphylie du genre *Tagoropsis*. Cependant cette relation est peu soutenue et devra être confirmée à l'aide d'UCEs.

5. Datation et concaténation de l'arbre

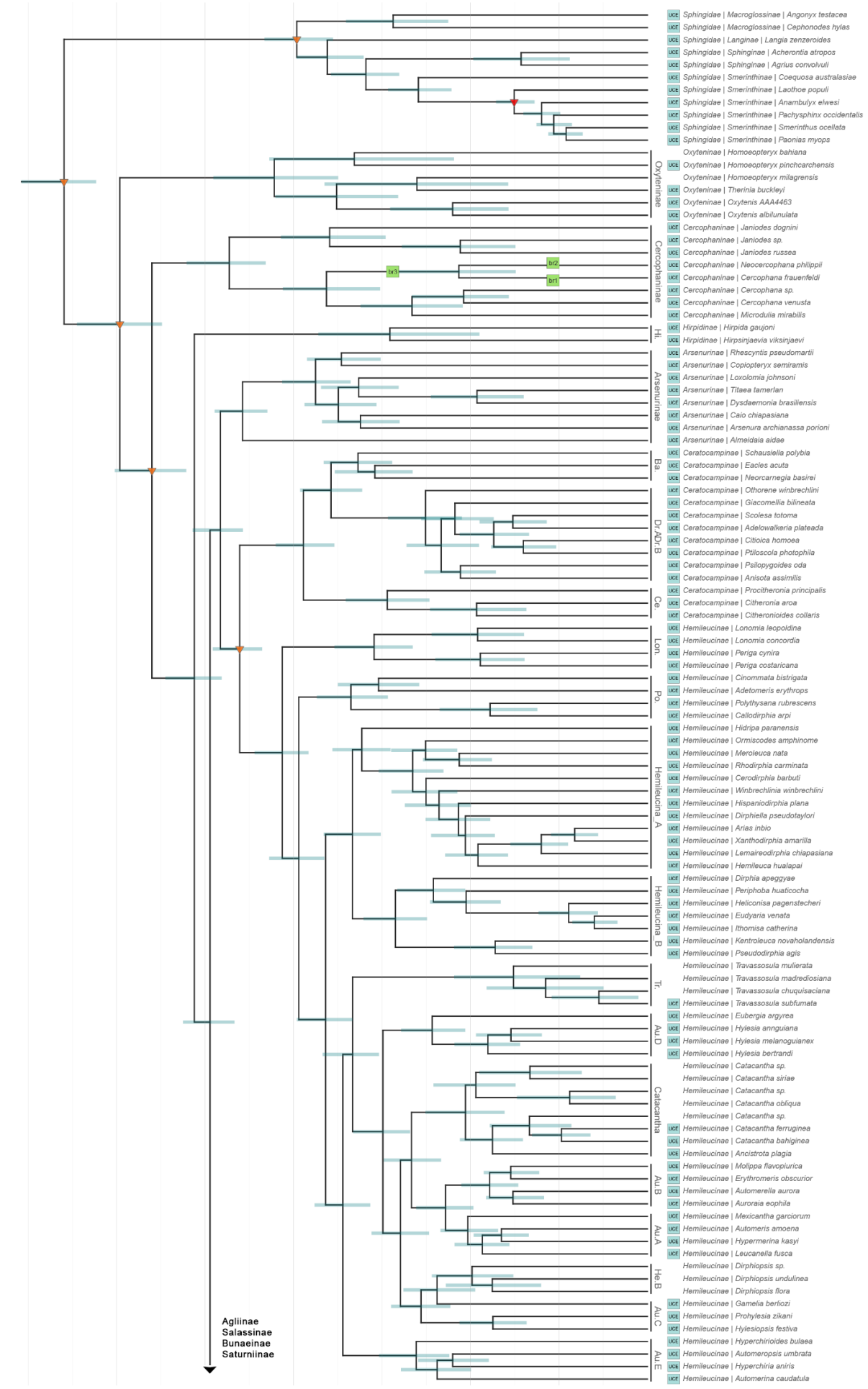
Datation du backbone

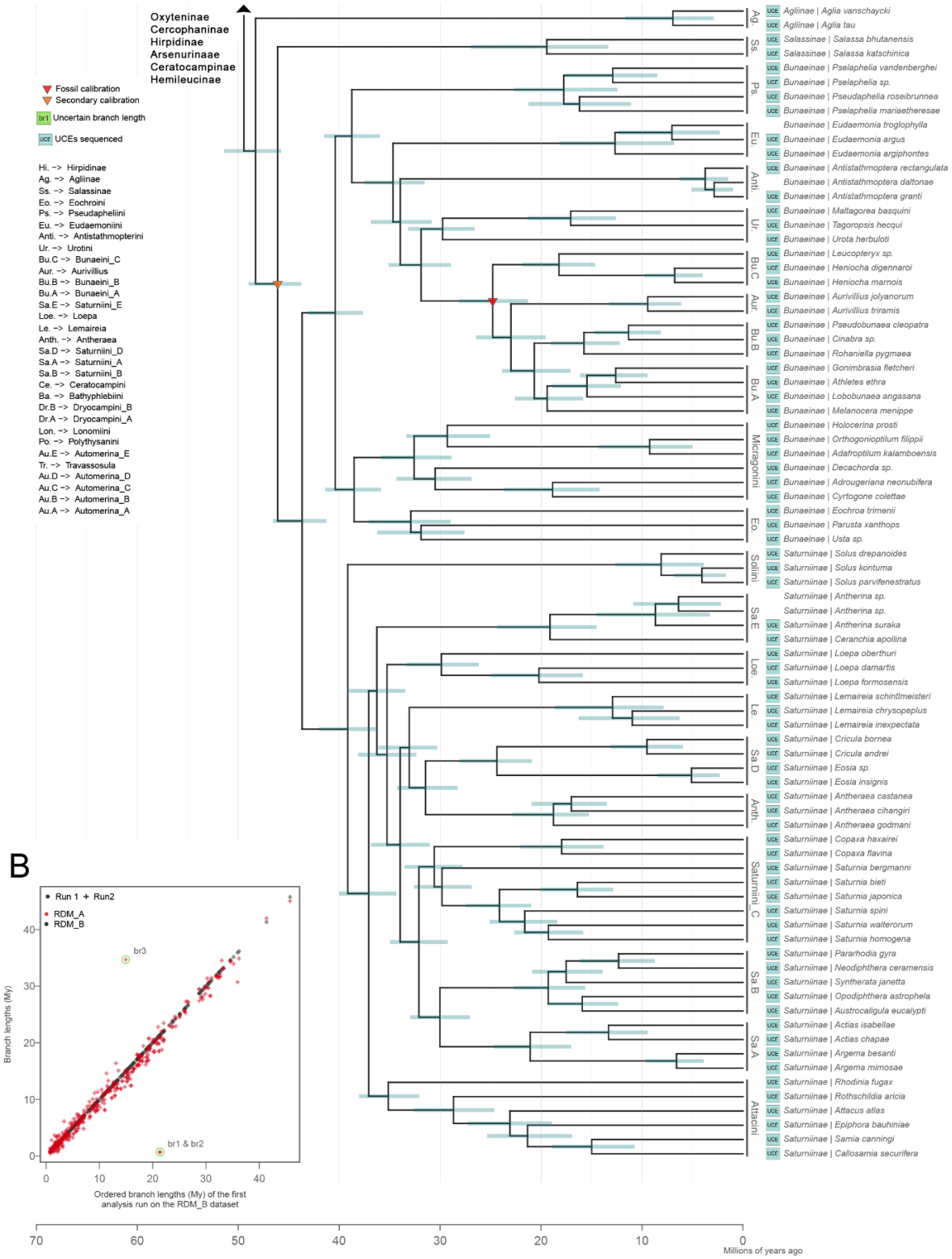
L'ensemble des analyses de datation ont convergées : les valeurs d'ESS (Effective Sample Size) dépassent 200 pour tous les paramètres dans l'ensemble des analyses et les analyses indépendantes lancées à partir des mêmes jeux de données ont convergé vers des estimations similaires (Figure 3 - B). Les datations effectuées à partir des deux jeux de données composés de 50 UCEs sélectionnés indépendamment au hasard ont résulté en des estimations très similaires pour presque l'ensemble des nœuds de la topologie du *backbone* (Figure 3 - B). Cependant, d'un jeu de données à l'autre, nous avons estimé des longueurs de branches très différentes au sein de la sous-famille des Cercophaninae (identifiées par un rectangle vert sur la Figure 3 - A). Ainsi, pour les branches terminales menant aux taxons *Neocercophana philippii* et *Cercophana frauenfeldi*, nous avons estimé une longueur de moins d'un million d'année à partir du jeu de données A quand nous estimions des branches longues de plus de 20 millions d'années à partir du jeu de données B. Ces différences sont étonnantes car nous disposions d'un nombre conséquent d'UCEs pour ces deux taxons. Dans le Chapitre 1, nous avions estimé le temps de divergence entre *Neocercophana philippii* et *Cercophana frauenfeldi* à 7.6 Ma [4.5-11.5 Ma] à partir d'une matrice composée de 50 UCEs sélectionnés différemment. Lors des tests effectués dans le Chapitre 1, l'âge de ce nœud était déjà variable d'un sous-jeu de données à l'autre puisque nous avions aussi estimé des temps de divergence plus anciens (12.7 Ma [8.6-17.6]) à partir d'une autre sélection aléatoire de 50 UCEs. Cependant, ces comparaisons sont limitées car nous n'avions pas utilisé le code-barres ADN dans les matrices du Chapitre 1 et nous avions utilisé un autre échantillon pour l'espèce *Cercophana frauenfeldi*. Mis à part ce nœud, les calibrations effectuées dans ce Chapitre sont très

cohérentes avec celles effectuées dans le Chapitre 1 (Chapitre 1 – Figure 4 ; Figure S5). Afin de limiter les différences dans les estimations des temps de divergence d'un sous-jeu de données à l'autre, nous envisageons d'augmenter le nombre d'UCEs échantillonnés aléatoirement dans les futures versions du *pipeline*. Afin de dater les *subtrees*, nous avons considéré les résultats obtenus à partir du jeu de données B.

Figure 3 (deux pages suivantes) – A) Phylogénie datée du backbone. Les résultats représentés ont été obtenus à partir du sous-jeu de données B (50 UCEs échantillonnés aléatoirement + codes-barres ADN). Les barres bleues horizontale représentent les intervalles de confidence des inférences effectuées pour chaque nœud. A chaque feuille de l'arbre nous avons indiqué le nom de l'espèce et sa sous-famille. Pour les outgroups, la famille est également indiquée. Un rectangle bleu marque les échantillons pour lesquels nous disposions d'UCEs. Les subtrees auxquels ont été assignées les différentes feuilles de l'arbre sont représentés verticalement en bout de branches. Si une abréviation a été utilisée, le nom complet du subtree est indiqué à gauche de la figure. Les calibrations utilisées sont indiquées par un triangle renversé, qu'elles soient fossiles (en rouge) ou secondaires (en orange). Nous avons signalé par un rectangle vert les branches pour lesquelles les estimations diffèrent le plus d'un sous-jeu de données à l'autre (se référer au (B)). B) Vérification de la convergence des analyses de datation effectuées à partir des différents runs et à partir des différents sous-jeux de données. Nous avons indiqué en vert les trois branches pour lesquelles les longueurs estimées diffèrent le plus d'un sous-jeu de données à l'autre.

A


Sphingidae | Macroglossinae | Angonyx testacea
Sphingidae | Macroglossinae | Cephonodes hylas
Sphingidae | Langiinae | Langia zenzeroides
Sphingidae | Sphinginae | Acherontia atropos
Sphingidae | Sphinginae | Agrius convolvuli
Sphingidae | Smerinthinae | Coequosa australasiae
Sphingidae | Smerinthinae | Laothoe populi
Sphingidae | Smerinthinae | Anambulyx elwesi
Sphingidae | Smerinthinae | Pachysphinx occidentalis
Sphingidae | Smerinthinae | Smerinthus ocellata
Sphingidae | Smerinthinae | Paonias myops
Oxyteninae | Homoeopteryx bahiana
Oxyteninae | Homoeopteryx pinchcarchensis
Oxyteninae | Homoeopteryx milagrensis
Oxyteninae | Therinia buckleyi
Oxyteninae | Oxytenis AAA4463
Oxyteninae | Oxytenis albilunulata
Cercophaninae | Janiodes dognini
Cercophaninae | Janiodes sp.
Cercophaninae | Janiodes russea
Cercophaninae | Neocercophana philippii
Cercophaninae | Cercophana frauenfeldi
Cercophaninae | Cercophana sp.
Cercophaninae | Cercophana venusta
Cercophaninae | Microdulia mirabilis
Hirpidinae | Hirpida gaujoni
Hirpidinae | Hirpsinjaevia viksinjaevi
Arsenurinae | Rhescyntis pseudomartii
Arsenurinae | Copiopteryx semiramis
Arsenurinae | Loxolomia johnsoni
Arsenurinae | Titaea tamerlan
Arsenurinae | Dysdaemonia brasiliensis
Arsenurinae | Caio chiapasiana
Arsenurinae | Arsenura archianassa porioni
Arsenurinae | Almeidaia aidae
Ceratocampinae | Schausiella polybia
Ceratocampinae | Eacles acuta
Ceratocampinae | Neorcarnegia basirei
Ceratocampinae | Othorene winbrechlini
Ceratocampinae | Giacomellia bilineata
Ceratocampinae | Scolesa totoma
Ceratocampinae | Adelowalkeria plateada
Ceratocampinae | Citioica homoea
Ceratocampinae | Ptiloscola photophila
Ceratocampinae | Psilopygoides oda
Ceratocampinae | Anisota assimilis
Ceratocampinae | Procitheronia principalis
Ceratocampinae | Citheronia aroa
Ceratocampinae | Citheronioides collaris
Hemileucinae | Lonomia leopoldina
Hemileucinae | Lonomia concordia
Hemileucinae | Periga cynira
Hemileucinae | Periga costaricana
Hemileucinae | Cinommata bistrigata
Hemileucinae | Adetomeris erythrops
Hemileucinae | Polythysana rubrescens
Hemileucinae | Callodirphia arpi
Hemileucinae | Hidripa paranensis
Hemileucinae | Ormiscodes amphinome
Hemileucinae | Meroleuca nata
Hemileucinae | Rhodirphia carminata
Hemileucinae | Cerodirphia barbuti
Hemileucinae | Winbrechlinia winbrechlini
Hemileucinae | Hispaniodirphia plana
Hemileucinae | Dirphiella pseudotaylori
Hemileucinae | Arias inbio
Hemileucinae | Xanthodirphia amarilla
Hemileucinae | Lemaireodirphia chiapasiana
Hemileucinae | Hemileuca hualapai
Hemileucinae | Dirphia apeggyae
Hemileucinae | Periphoba huaticocha
Hemileucinae | Heliconisa pagenstecheri
Hemileucinae | Eudyaria venata
Hemileucinae | Ithomisa catherina
Hemileucinae | Kentroleuca novaholandensis
Hemileucinae | Pseudodirphia agis
Hemileucinae | Travassosula mulierata
Hemileucinae | Travassosula madrediosiana
Hemileucinae | Travassosula chuquisaciana
Hemileucinae | Travassosula subfumata
Hemileucinae | Eubergia argyrea
Hemileucinae | Hylesia annguiana
Hemileucinae | Hylesia melanoguianex
Hemileucinae | Hylesia bertrandi
Hemileucinae | Catacantha sp.
Hemileucinae | Catacantha siriae
Hemileucinae | Catacantha sp.
Hemileucinae | Catacantha obliqua
Hemileucinae | Catacantha sp.
Hemileucinae | Catacantha ferruginea
Hemileucinae | Catacantha bahiginea
Hemileucinae | Ancistrota plagia
Hemileucinae | Molippa flavopiurica
Hemileucinae | Erythromeris obscurior
Hemileucinae | Automerella aurora
Hemileucinae | Auroraia eophila
Hemileucinae | Mexicantha garciorum
Hemileucinae | Automeris amoena
Hemileucinae | Hypermerina kasyi
Hemileucinae | Leucanella fusca
Hemileucinae | Dirphiopsis sp.
Hemileucinae | Dirphiopsis undulinea
Hemileucinae | Dirphiopsis flora
Hemileucinae | Gamelia berliozi
Hemileucinae | Prohylesia zikani
Hemileucinae | Hylesiopsis festiva
Hemileucinae | Hyperchiroides bulaea
Hemileucinae | Automeropsis umbrata
Hemileucinae | Hyperchiria aniris
Hemileucinae | Automerina caudatula
Oxyteninae
Cercophaninae
Hi.
Arsenurinae
Ba.
Dr.
AD
Ce.
Lon.
Po.
Hemileucina_A
Hemileucina_B
Tr.
Au. D
Catacantha
Au. B
Au. A
He. B
Au. C
Au. E
br2
br3
br1
Agliinae
Salassinae
Bunaeinae
Saturniinae


Oxyteninae
Cercophaninae
Hirpidinae
Arsenurinae
Ceratocampinae
Hemileucinae
Fossil calibration
Secondary calibration
br1 Uncertain branch length
UCE UCEs sequenced
Hi. -> Hirpidinae
Ag. -> Agliinae
Ss. -> Salassinae
Eo. -> Eochroini
Ps. -> Pseudapheliini
Eu. -> Eudaemoniini
Anti. -> Antistathmopterini
Ur. -> Urotini
Bu.C -> Bunaeini_C
Aur. -> Aurivillius
Bu.B -> Bunaeini_B
Bu.A -> Bunaeini_A
Sa.E -> Saturniini_E
Loe. -> Loepa
Le. -> Lemaireia
Anth. -> Antheraea
Sa.D -> Saturniini_D
Sa.A -> Saturniini_A
Sa.B -> Saturniini_B
Ce. -> Ceratocampini
Ba. -> Bathyphlebiini
Dr.B -> Dryocampini_B
Dr.A -> Dryocampini_A
Lon. -> Lonomiini
Po. -> Polythysanini
Au.E -> Automerina_E
Tr. -> Travassosula
Au.D -> Automerina_D
Au.C -> Automerina_C
Au.B -> Automerina_B
Au.A -> Automerina_A
Agliinae | Aglia vanschaycki
Agliinae | Aglia tau
Salassinae | Salassa bhutanensis
Salassinae | Salassa katschinica
Bunaeinae | Pselaphelia vandenberghei
Bunaeinae | Pselaphelia sp.
Bunaeinae | Pseudaphelia roseibrunnea
Bunaeinae | Pselaphelia mariaetheresae
Bunaeinae | Eudaemonia troglophylla
Bunaeinae | Eudaemonia argus
Bunaeinae | Eudaemonia argiphontes
Bunaeinae | Antistathmoptera rectangulata
Bunaeinae | Antistathmoptera daltonae
Bunaeinae | Antistathmoptera granti
Bunaeinae | Maltagorea basquini
Bunaeinae | Tagoropsis hecqui
Bunaeinae | Urota herbuloti
Bunaeinae | Leucopteryx sp.
Bunaeinae | Heniocha digennaroi
Bunaeinae | Heniocha marnois
Bunaeinae | Aurivillius jolyanorum
Bunaeinae | Aurivillius triramis
Bunaeinae | Pseudobunaea cleopatra
Bunaeinae | Cinabra sp.
Bunaeinae | Rohaniella pygmaea
Bunaeinae | Gonimbrasia fletcheri
Bunaeinae | Athletes ethra
Bunaeinae | Lobobunaea angasana
Bunaeinae | Melanocera menippe
Bunaeinae | Holocerina prosti
Bunaeinae | Orthogonioptilum filippii
Bunaeinae | Adafroptilum kalamboensis
Bunaeinae | Decachorda sp.
Bunaeinae | Adrougeriana neonubifera
Bunaeinae | Cyrtogone colettae
Bunaeinae | Eochroa trimenii
Bunaeinae | Parusta xanthops
Bunaeinae | Usta sp.
Saturniinae | Solus drepanoides
Saturniinae | Solus kontuma
Saturniinae | Solus parvifenestratus
Saturniinae | Antherina sp.
Saturniinae | Antherina sp.
Saturniinae | Antherina suraka
Saturniinae | Ceranchia apollina
Saturniinae | Loepa oberthuri
Saturniinae | Loepa damartis
Saturniinae | Loepa formosensis
Saturniinae | Lemaireia schintlmeisteri
Saturniinae | Lemaireia chrysopeplus
Saturniinae | Lemaireia inexpectata
Saturniinae | Cricula bornea
Saturniinae | Cricula andrei
Saturniinae | Eosia sp.
Saturniinae | Eosia insignis
Saturniinae | Antheraea castanea
Saturniinae | Antheraea cihangiri
Saturniinae | Antheraea godmani
Saturniinae | Copaxa haxairei
Saturniinae | Copaxa flavina
Saturniinae | Saturnia bergmanni
Saturniinae | Saturnia bieti
Saturniinae | Saturnia japonica
Saturniinae | Saturnia spini
Saturniinae | Saturnia walterorum
Saturniinae | Saturnia homogena
Saturniinae | Pararhodia gyra
Saturniinae | Neodiphthera ceramensis
Saturniinae | Syntherata janetta
Saturniinae | Opodiphthera astrophela
Saturniinae | Austrocaligula eucalypti
Saturniinae | Actias isabellae
Saturniinae | Actias chapae
Saturniinae | Argema besanti
Saturniinae | Argema mimosae
Saturniinae | Rhodinia fugax
Saturniinae | Rothschildia aricia
Saturniinae | Attacus atlas
Saturniinae | Epiphora bauhiniae
Saturniinae | Samia canningi
Saturniinae | Callosamia securifera
Ag.
Ss.
Ps.
Eu.
Anti.
Ur.
Bu.C
Aur.
Bu.B
Bu.A
Micragonini
Eo.
Solini
Sa.E
Loe
Le.
Sa.D
Anth.
Saturniini_C
Sa.B
Sa.A
Attacini
B
Run 1 + Run2
RDM_A
RDM_B
br3
br1 & br2
Branch lengths (My)
Ordered branch lengths (My) of the first analysis run on the RDM_B dataset
Millions of years ago

Assemblage de la mégaphylogénie des Saturniidae

Nous avons identifié quelques erreurs marginales de datation d'un *subtree* à l'autre qui sont liées aux arrondis utilisés lorsque de la datation des *subtrees*. Ces erreurs sont très limitées (<0.001%) et ont été lissées à l'aide de la fonction *force.ultrametric* du package *phytools* (Revell 2012) sur R. Comme nous le mentionnions plus haut, les mégaphylogénies assemblées ne contiennent aucune espèce des genres *Dirphiopsis* et *Othorene*. Dans les prochaines versions de la mégaphylogénie, nous intégrerons ces genres, après avoir séquencé des UCEs d'échantillons du genre *Dirphiopsis* et considéré ces deux genres dans des *subtrees* qui leur seront propres. La mégaphylogénie assemblée dans ce Chapitre contient 3892 OTUs (Figure 4). Nous avons généré 10 réplicats de la mégaphylogénie en échantillonnant et datant différentes topologies de bootstrap, permettant la prise en compte des incertitudes liées à la topologie et à l'estimation des temps de divergence dans de futures analyses macroévolutives. Un nombre plus élevé de réplicats pourra être généré dans les futures versions.

6. Mesure des taux de diversification

Nous n'avons pas réussi à inférer des taux de diversification cohérents à l'aide de RevBayes. La méthode que nous avons utilisée infère un taux de diversification propre à chaque branche de l'arbre. Sur une telle mégaphylogénie, cela implique un nombre de paramètres considérable. A titre, d'exemple, nous avons obtenu des taux de diversification plus élevés pour certaines branches longues que dans certains clades dans lesquels le nombre de spéciations récentes est très élevé. De plus, les difficultés rencontrées par les fonctions du package R *RevGadgets* (https://github.com/revbayes/RevGadgets) pour résumer les fichiers résultats de RevBayes sont révélatrices du fait que RevBayes n'est pas conçu pour traiter de tels jeux de données.

L'analyse BAMM n'a pas convergé après les 50 millions de générations. Les résultats discutés ici et représentés sur la Figure 4 sont donc indicatifs et ont été extraits après avoir écarté 90% des générations de la distribution postérieure. A la racine, BAMM a estimé un taux de diversification net de 0.18 (spéciation nette par branche par million d'année) et inféré que la majorité des branches de la phylogénie ont un taux qui varie entre 0.13 et 0.18. Seuls certains clades présenteraient des taux de diversification bien plus élevés que le reste de la phylogénie, atteignant parfois des taux supérieurs à 2 spéciations nette par branche par million d'année. C'est par exemple le cas d'une partie des *Hylesia*, des *Automeris* et des *Gonimbrasia*. La sous-famille des Hemileucinae concentre la majorité des clades qui présentent des forts taux de diversification : les genres *Automeris* et *Hylesia*, comme mentionné plus haut, mais aussi *Periga*, *Paradirphia*, *Pseudodirphia* et *Gamelia*. Les hémileucines à part, les plus hauts taux de diversification inférés se trouvent dans la tribu des Bunaeini, au sein des genres *Gonimbrasia* et *Bunaeopsis*. Au sein de la tribu Micragonini, les genres frères *Orthogonioptilum* et *Adafroptilum* ont également diversifié bien plus rapidement que la majorité des Saturniidae. Comme leur diversité limitée le laissaient penser, nous n'avons pas retrouvé de forts taux de diversification au sein des sous-familles

basales Oxyteninae, Cercophaninae, Hirpidinae, Arsenurinae, Agliinae et Salassinae. Enfin, seulement un clade présente des taux élevés au sein des Ceratocampinae (au sein du genre *Syssphinx*) et deux au sein des Saturniinae (au sein des genres *Copaxa* et *Cricula*).

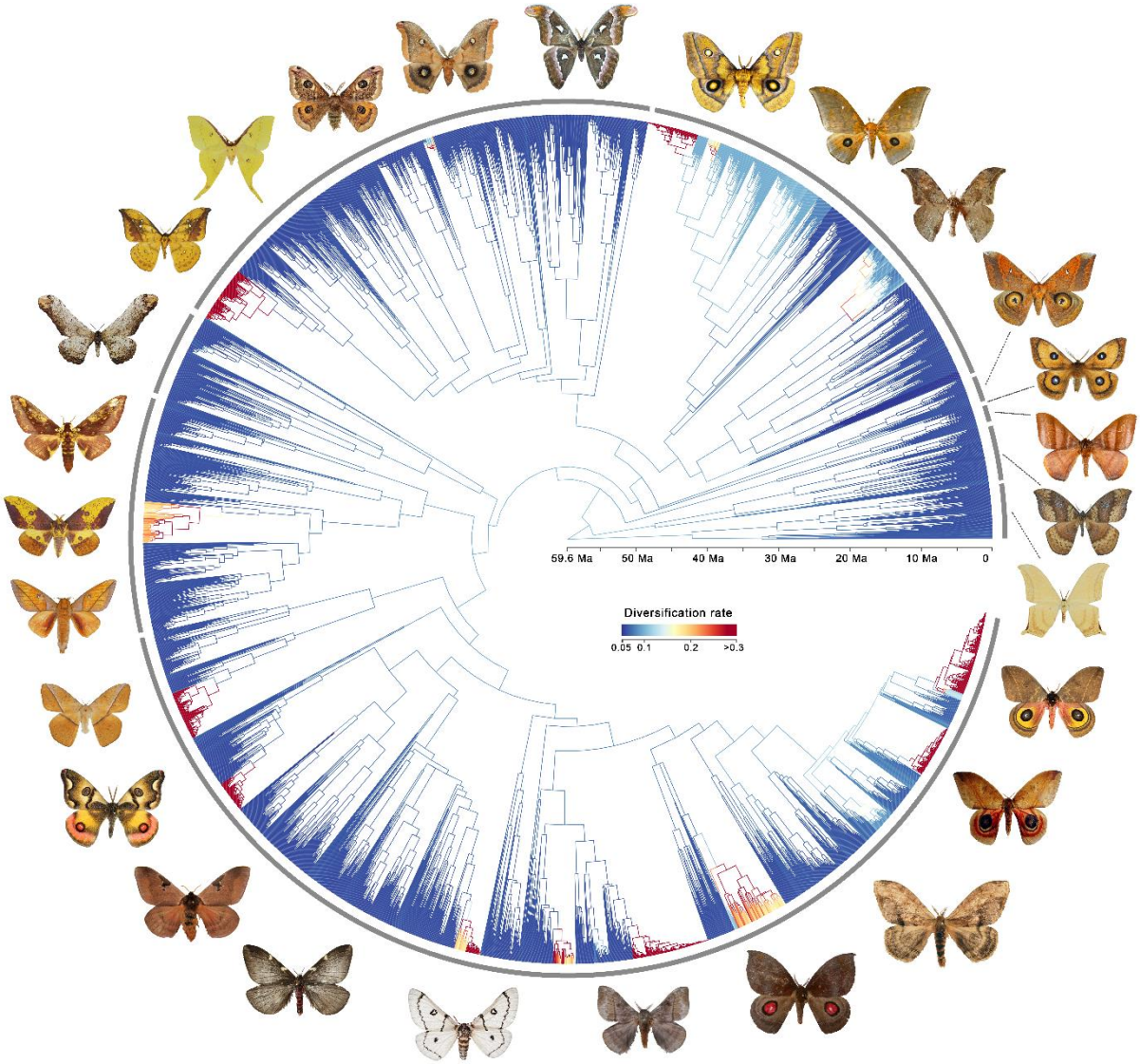


Figure 4 – Mégaphylogénie des Saturniidae inférée à l'aide de notre pipeline, pour 3892 OTUs. Le cercle gris délimite les différentes sous-familles. Les branches de l'arbre sont colorées en fonction des taux de diversification inférés à l'aide du logiciel BAMM (résultats préliminaires, voir résultats). Les photos représentées ne sont pas à la même échelle.

Discussion

Ce Chapitre présente les résultats préliminaires de la construction de la première phylogénie à l'échelle spécifique de la famille des Saturniidae, l'un des objectifs principaux de l'ANR SPHINX (ANR-16-CE02-0011-01). Une fois ce travail finalisé, la mégaphylogénie des Saturniidae sera, à notre connaissance, la plus grande phylogénie inférée pour un groupe d'insecte. Le *pipeline* phylogénomique que nous avons développé est détaillé dans l'Annexe 2. Celui-ci permettra de générer de nouvelles versions de la mégaphylogénie des Saturniidae une fois que de nouveaux UCEs auront été séquencés dans le cadre de l'ANR SPHINX. Une fois la mégaphylogénie des Saturniidae finalisée, nous disposerons d'un support analytique considérable pour aborder un large panel de questions macroécologiques et macroévolutives.

Définition des Unités Taxonomiques Opérationnelles

La première étape de ce travail a été de définir des entités taxonomiques au sein des Saturniidae. Pour cela, nous avons utilisé la très grande quantité de séquences de codes-barres ADN produites dans le cadre de la campagne de DNA barcoding ciblant la faune mondiale des Saturniidae. Ces séquences, compilées sur la plateforme BOLD (http://www.boldsystems.org/), nous ont permis de disposer d'arbres de distances pour chaque grand groupe de saturnidés. La génération de ces codes-barres ADN ayant été particulièrement intensive pour les Saturniidae, nous estimons que le jeu de données utilisé rassemble, à peu de chose près, l'ensemble des lignées de Saturniidae échantillonnés par les générations successives de naturalistes et scientifiques travaillant sur la diversité de cette famille. A partir de ces arbres, en utilisant le *Barcode Index Number* (BIN, Ratnasingham & Hebert 2013) et les informations géographiques et biologiques, nous avons identifié 3947 OTUs. Nous avons donc estimé une diversité spécifique légèrement supérieure au nombre d'espèces reconnues dans la *checklist* établie par Kitching *et al.* (2018) qui reconnaissaient 3462 espèces et 355 sous-espèces valides. Certaines délimitations taxonomiques effectuées ici méritent d'être revisitées, mais ce nombre est cohérent avec la tendance à l'augmentation du nombre d'espèces de Saturniidae de ces dernières années (Kitching *et al.* 2018). La majorité des OTUs considérées correspondent à des espèces listées dans la *checklist* mais 1033 restent non-identifiées. Ce nombre doit néanmoins être relativisé car il inclut certainement de nombreuses espèces et sous-espèces listées dans la *checklist* mais que nous n'avons pas été en mesure d'assigner avec certitude à une OTU.

Combinaison des marqueurs UCE et du code-barres ADN

L'utilisation combinée des marqueurs UCE et du code-barres ADN s'est avérée être une stratégie particulièrement efficace dans l'inférence de la mégaphylogénie de la famille des Saturniidae. Nous avons effectivement inféré avec succès les nœuds les plus profonds grâce à l'utilisation des UCEs (voir discussion sur la topologie du *backbone*) et les codes-barres ADN se sont avérés être suffisamment informatifs pour résoudre la majorité des nœuds superficiels de la phylogénie. A terme, dans le projet

de construction de la mégaphylogénie des Saturniidae, nous n'aurons séquencé que près d'un tiers des échantillons à l'aide des techniques de capture d'UCE, la position des deux tiers des OTUs restantes sera, elle, inférée à l'aide du seul code-barres ADN. Cette stratégie combinant des marqueurs génomiques avec le code-barres ADN peut être appliquée avec d'autres marqueurs génomiques, à condition qu'ils soient suffisamment informatifs pour résoudre les nœuds profonds. La quantité de séquences de codes-barres ADN disponibles en ligne constitue une opportunité considérable de réduire le nombre de spécimens à séquencer avec des marqueurs génomiques, et ainsi de diminuer les coûts et le temps de manipulation d'un projet de construction de mégaphylogénie. Grâce au développement de programmes internationaux (Adamowicz, 2015) et d'applications diverses de cet outil d'identification (Padial & De la Riva 2007 ; Wilson 2010 ; Andújar *et al.* 2018) les librairies de codes-barres ADN ne cessent de croître : par exemple pour les seuls Lépidoptères, en juin 2020, 97 281 espèces étaient représentées par 1 459 760 séquences dans BOLD.

Le gène codant pour la protéine cytochrome c oxydase I (COI) a été l'un des premiers marqueurs génétiques utilisés par les biologistes lors de l'émergence des phylogénies moléculaires (par ex. Brower 1994 ; Howland & Hewitt 1995). Bien qu'il ait été rapidement suggéré que les données moléculaires pouvaient être utilisées dans l'identification des espèces (Wilson 1995), ce n'est que dans le début des années 2000 qu'il a été proposé d'utiliser à grande échelle des données génétiques, en l'occurrence les séquences d'une partie du gène COI, dans un but taxonomique (Hebert *et al.* 2003). Ce marqueur standard a notamment été choisi pour la quantité de signal phylogénétique qu'il renferme, couvrant l'échelle intra-spécifique (Cox & Hebert 2001) comme inter-générique (Wilson 2010). Le code-barres ADN a depuis largement démontré son efficacité dans l'identification des espèces dans des groupes très divers (Hebert 2004 ; Clare *et al.* 2007) et est aujourd'hui très largement utilisé en taxonomie intégrative (Seifert *et al.* 2017) ou en génomique environnementale (Andújar *et al.* 2018). La quantité considérable de séquence de codes-barres ADN disponibles dans les banques de données en ligne (BOLD mais également GenBank, www.ncbi.nlm.nih.gov/genbank ; Benson *et al.* 2012) constitue une source d'information phylogénétique unique. Ici, nous avons pleinement utilisé cette quantité d'information en utilisant les codes-barres ADN pour résoudre la grande majorité des nœuds superficiels de la phylogénie des Saturniidae. Nos résultats confirment que le code-barres ADN est un marqueur suffisamment informatif pour résoudre de tels nœuds (Annexe 1 ; par ex. au sein des genres *Adelowalkeria, Citioica, Prohylesia, Ptiloscola*) mais démontrent également que, lorsque les taux de spéciation s'accélèrent, son utilisation atteint ses limites (par ex. *Bunaeopsis, Orthogonioptilum, Hylesia*).

En revanche, il est évident que l'information phylogénétique qu'il comporte sature lorsqu'on s'intéresse à des nœuds plus profonds (par ex. Miranda *et al.* 1997 ; Carlini & Graves 1999). De fait, des nœuds anciens de plusieurs dizaines de millions d'années ne peuvent être résolus qu'à l'aide de jeux de données plus grands. Dans le cadre de l'ANR SPHINX, nous avons choisi d'utiliser des marqueurs ultra-conservés du génome (UCEs) pour inférer les nœuds profonds et semi-profonds de la mégaphylogénie

des Saturniidae (ainsi que des Sphingidae pour lesquels un travail similaire a été initié). L'avantage des UCEs réside notamment dans la diversité de l'information phylogénétique qu'ils contiennent, pertinente à la fois pour des nœuds très anciens et récents. Proposée initialement par Faircloth *et al.* (2012), leur utilisation s'est effectivement avérée adapté à l'étude de relations phylogénétiques inter-familiales (McCormack *et al.* 2012 ; Faircloth *et al.* 2015) comme intra-spécifiques (Newman & Austin 2016). Afin de se concentrer sur l'inférence de nœuds profonds, nous avons maximisé la diversité phylogénétique lors de la sélection des taxons pour la capture et le séquençage des UCEs, une stratégie qui s'est avérée efficace pour résoudre les nœuds les plus profonds de la phylogénie, comme en témoignent les bons supports de la topologie du *backbone,* et pour inférer avec précision les longueurs de branches qui séparent les différents groupes d'espèces (Annexe 1).

Pour autant, le travail présenté ici devra d'être poursuivi en séquençant les UCEs d'échantillons supplémentaires. Avant la génération de cette mégaphylogénie, nous avions planifié la génération d'UCEs pour 150 autres échantillons afin de mieux couvrir la diversité de la famille avec des données génomiques. A ces échantillons, nous ajouterons les échantillons identifiés ici comme *rogue taxa* afin d'augmenter la précision de nos inférences.

Topologie du *backbone*

La topologie que nous avons inférée pour le *backbone* est très congruente avec celle présentée dans le Chapitre 1. Bien que nous n'ayons pas utilisé les même data (meilleure couverture générique dans le Chapitre 1) ni effectué d'analyses aussi approfondies que lors du Chapitre 1 les deux topologies présentent cependant certaines divergences qui méritent d'être discutées. La position de la sous-famille Hirpidinae est notamment distincte de ce que nous avions pu inférer à partir des différents jeux de données du Chapitre 1. Alors que les Hirpidinae se branchaient systématiquement avec les sous-familles néotropicales Arsenurinae, Ceratocampinae et Hemileucinae, nous inférons ici une position basale à ce qui peut être défini comme le cœur de la diversité des Saturniidae, *i.e.* le clade regroupant Arsenurinae, Ceratocampinae, Hemileucinae, Agliinae, Bunaeinae, Salassinae et Saturniinae. Cette inférence a certainement été influencée par l'ajout du taxon *Hirpsinjaevia viksinjaevi*, récemment décrit (Brechlin 2019), qui redéfinit la racine de la sous-famille. Cette espèce est la seule représentante connue de ce genre nouvellement défini et n'était pas présente dans le jeu de données du Chapitre 1. L'ajout de ce taxon semble aussi avoir influencé la position des Agliinae : dans le Chapitre 1, les inférences effectuées à partir des matrices concaténées résultaient en le groupement des sous-familles Agliinae et Hirpidinae qui se branchaient avec les grandes sous-familles néotropicales. Ici, nous inférons que la sous-famille des Agliinae se branche avec le reste des sous-familles de l'Ancien Monde, une position déjà inférée à l'aide des méthodes de coalescence multi-espèces (Chapitre 1) et qui est plus parcimonieuse d'un point de vue biogéographique. Ces résultats témoignent de l'importance fondamentale de la complétion de l'échantillonnage dans l'inférence de phylogénies.

Nous avons également établi avec des supports élevés que le genre *Antheraea* est le groupe frère du clade formé par les genres *Eosia* et *Cricula*, un clade auquel se brancherait le genre *Lemaireia*. Mais si le clade *Antheraea*/*Eosia*/*Cricula* est bien soutenu, ce n'est pas le cas de la position du genre *Lemaireia* qui devra être investiguée de manière plus approfondie. Au sein des Ceratocampinae, la position du genre *Othorene* diffère de celle présentée dans le Chapitre 1, obtenue à l'aide d'une méthode utilisant la Coalescence Multi Espèces ; une position similaire à celle présentée ici avait cependant été inféré avec IQ-TREE. Cependant, du fait du nombre réduit de genres échantillonnés au sein de la tribu des Dryocampini dans ce Chapitre, la position de *Othorene* telle qu'elle a été inférée dans le Chapitre 1 doit être préférée (mais nécessite d'être plus largement analysée). Notre stratégie de réduction du nombre d'échantillons dans le jeu de données *backbone* (pour limiter les temps de computation) expose ici ses limites. Nous adapterons l'échantillonnage du *backbone* dans les futures versions de la mégaphylogénie des Saturniidae.

Lors de l'étape d'inférence de la topologie du *backbone*, nous avons également été confrontés au problème de la position du genre *Dirphiopsis*. Nous avions défini ce genre comme faisant partie du *subtree Hemileucina_B* en se basant sur les résultats du Chapitre 1 or, nous avons inféré ici une position radicalement différente. Mais, à la différence des autres genres mentionnés plus haut, cette différence est due au fait que nous ne disposions pas d'UCEs pour les OTUs du genre *Dirphiopsis*. Pour la suite des analyses, nous avons donc écarté le genre *Dirphiopsis* et avons priorisé le séquençage d'UCEs pour des OTUs de ce genre pour le futur. La position de *Dirphiopsis* en groupe frère des genres *Kentroleuca* et *Pseudodirphia*, doit être préférée puisque celle-ci a été inférée à l'aide de données génomiques.

Puisque dans notre approche *backbone + subtrees* nous ne datons que la meilleure topologie *backbone* inférée par IQ-TREE, cette topologie doit être particulièrement bien explorée et soutenue. Les résultats exploratoires présentés doivent ainsi être approfondis en améliorant notre échantillonnage et en utilisant, par exemple, des stratégies de partitionnement plus élaborées (par ex. Tagliacollo & Lanfear 2018) ou des méthodes de coalescence multi-espèces (Mirarab & Warnow 2015).

Utilisation des UCEs dans l'inférence des nœuds intragénériques

L'utilisation des marqueurs UCEs a également été pertinente dans l'inférence de relations phylogénétiques plus récentes, comme en témoignent les topologies obtenues pour les différents *subtrees*. Les UCEs nous ont effectivement permis d'inférer, avec de forts supports, les relations entre les différents groupes d'espèces et au sein de ceux-ci pour les genres *Eacles*, *Dirphia*, *Lonomia* et pour la tribu des Attacini dans lesquels l'échantillonnage des UCEs est très dense (Annexe 1, *subtrees* Bathyphlebiini, Hemileucina_B, Lonomiini et Attacini). Dans de très rares cas, nous n'avons pas été en mesure d'obtenir de bons supports pour certains nœuds lorsque les taux de diversification s'accéléraient et ce, alors même que nous disposions d'UCEs pour l'ensemble des clades. C'est notamment le cas pour les nœuds inter-génériques de la sous-famille des Arsenurinae (Annexe 1). La topologie des Arsenurinae

est très différente de celle inférée avec un échantillonnage générique dans le Chapitre 1 et devra être investiguée de façon plus approfondie. Il est notamment envisagé un travail en collaboration avec Chris Hamilton (Université de l'Idaho) et Akito Kawahara (Université de Floride) qui ont reconstruit la phylogénie de cette sous-famille à partir de capture d'exons (AHE – Hamilton *et al.* 2020 - in review) et rencontrent également d'importants problèmes de résolution des divergences initiales de la sous-famille. Notre échantillonnage UCE a été conçu pour être combiné au jeu de données de cette équipe.

Grâce à la complétion inédite de notre jeu de données, nous avons pu identifier une multitude de points taxonomiques nécessitant une révision. Par exemple, nos résultats ont confirmé la nécessité de révisions taxonomiques pour le clade défini par les genres *Dirphia*, *Periphoba*, *Eudyaria*, *Ithomisa* et *Heliconisa* (Annexe 1, Subtree Hemileucina_B). Dans la majeure partie des cas, il ne s'agit que d'une poignée d'espèces assignées à un mauvais genre (par ex. *Neodiphthera excavus* qui se branche à l'intérieur du genre *Syntherata* ; Annexe 1, *subree* Saturniini_B) mais dans certains cas plus complexes, les genres concernés nécessitent d'être redéfinis plus précisément (par ex. le genre *Automeris* ; Annexe 1, subtree Automerina_A). Ces révisions devront cependant faire l'objet de travaux taxonomiques spécifiques.

De la difficulté de mesurer les taux de diversification

Les mégaphylogénies constituent un support dont l'important pouvoir statistique permet de tester un spectre d'hypothèses très diverses. Pour autant, les nombreux outils d'analyses macroévolutives développés cette dernière décennie sont assez peu adaptés à de tels jeux de données. Dans ce Chapitre, nous souhaitions estimer les taux de diversification de chaque branche de la mégaphylogénie inférée. De nombreux logiciels ont été développés pour cela mais tous utilisent des approches bayésiennes. Or, les approches bayésiennes et les algorithmes de Monte Carlo, dans l'état actuel de leur développement, rencontrent de grandes difficultés avec les grands jeux de données pour lesquels les temps de calcul deviennent inconcevables (Lartillot 2020). Nous souhaitions par exemple utiliser le logiciel RevBayes et l'approche bayésienne développée par Höhna *et al.* (2019) qui modélise notamment les shifts de diversification sur les lignées éteintes. Mais la taille de la phylogénie implique un nombre considérable de paramètres à estimer. RevBayes, dans sa version actuelle, ne semble pas adapté à de telles phylogénies : les taux estimés étaient incohérents et le package R *RevGadgets* n'a pas été en mesure de traiter les fichiers *output* produits par le logiciel (nous avons développé notre propre script pour cela). Nous avons ensuite utilisé le logiciel BAMM qui apparaissait comme une alternative intéressante malgré le fait qu'il ne mesure pas des taux spécifiques à chaque branche. Bien que BAMM ne considère pas les shifts sur les lignées éteintes, ce qu'il lui a été reproché (Moore *et al.* 2016), il semble plus adapté aux mégaphylogénies. Rabosky *et al.* (2013) ont par exemple mis en évidence une corrélation des taux de spéciation et des taux d'évolution morphologiques en appliquant le modèle BAMM sur une mégaphylogénie des Actinoptérygiens de 6760 feuilles. Cependant, après 50 millions de générations et 96h de calcul sur 40 cœurs, nous n'avons pas atteint de convergence des résultats (les taux ont tout de même été représentés sur la Figure 4).

Le développement récent de ClaDS (Maliet *et al.* 2019), qui infère les taux de diversification en considérant un ensemble de petits shifts apparait comme prometteur. Le logiciel a par exemple été testé par ses auteurs sur les phylogénies des différentes familles d'Oiseaux les plus diversifiées. Mais lors des tests préliminaires que nous avons effectués, le logiciel, qui repose également sur des méthodes bayésiennes, nécessita beaucoup de temps pour inférer les taux de diversification de phylogénies de taille moyenne (200 feuilles). Nous serons peut-être contraints de découper la phylogénie en plusieurs clades distincts avant d'inférer ces taux comme l'ont fait Maliet *et al.* (2019) pour les familles d'Oiseaux. Malheureusement, une telle approche ne permettrait pas l'inférence des taux de diversification des branches les plus profondes de la mégaphylogénie. La recherche en macroévolution bénéficierait largement du développement de méthodes suffisamment puissantes pour mesurer les taux de diversification à partir de phylogénies de plusieurs milliers de feuilles.

Avantages et désavantages du *pipeline backbone + subtrees* présenté ici

L'approche *backbone + subtree* est apparue comme l'approche la plus adéquate pour inférer des mégaphylogénies. Elle permet une diminution drastique des temps de calcul en comparaison aux approches par supermatrice et par *supertree* et évite les problèmes d'incompatibilité des nœuds que l'on peut rencontrer avec une approche par *supertree.*

Les approches par supermatrice utilisent une matrice unique rassemblant les données nucléotidiques et/ou morphologiques de l'ensemble des taxons disponibles. Une telle approche a été largement appliquée dans les années 2000 afin d'inférer de larges phylogénies (par ex. Driskell *et al.* 2004 ; McMahon & Sanderson 2006). Cependant, les auteurs de ces études n'ont pas daté les topologies, une étape particulièrement chronophage (voir plus bas) mais essentielle pour effectuer des analyses macroévolutives et/ou macroécologiques. Par exemple, Smith *et al.* (2009) ont utilisé une l'approche par supermatrice afin d'inférer une phylogénie de 4954 espèces d'Asterales à partir de 6 gènes mais il est fort probable que l'inférence des temps de divergence à partir d'une telle matrice aurait posé des problèmes de convergence et des temps de calcul considérables. L'approche par supermatrice n'est en réalité pas compatible avec les méthodes de datation bayésiennes qui sont pourtant les plus appropriées (voir plus bas).

Alternativement, l'approche *supertree* a aussi été utilisée pour inférer des mégaphylogénies. C'est notamment le cas de Bininda-Emonds *et al.* (2007) qui ont inféré une phylogénie des Mammifères qui a largement été utilisée par des études macroévolutives ou macroécologiques malgré les nombreuses polytomies qu'elle contient. L'approche par *supertree* est intéressante lorsque les marqueurs utilisés sont très différents d'un clade à l'autre car cela limite la quantité globale de données manquantes, ou pour combiner des phylogénies de différentes publications. Cependant, les approches *supertree* amalgament les nœuds contradictoires d'un arbre à l'autre, produisant des polytomies et, par conséquent,

biaisant les longueurs de branches. L'approche par supertree ne facilite pas non plus l'inférence des temps de divergence et, en pratique, peu se sont essayés à l'inférence de temps de divergence absolus.

Lors de ce Chapitre, nous avons donc privilégié l'approche *backbone + subtrees*. Cette approche est dérivée de l'approche par *supertree* et permet un gain de temps de calcul considérable lors de l'inférence des temps de divergence. Contrairement à l'approche *supertree*, les sous-jeux de données considérés ne se chevauchent pas, ce qui permet d'éviter des conflits entre les topologies, et l'utilisation d'un *backbone* résolu permet un assemblage simple et rapide de la mégaphylogénie. De plus, les nœuds communs entre le *backbone* et les *subtrees* permettent une datation indépendante de l'ensemble des sous-jeux de données avant assemblage. Cette approche, théorisée par Mishler (1994), a été appliquée pour inférer des phylogénies des Oiseaux (Jetz *et al.* 2012), des Squamates (Tonini *et al.* 2016), des Amphibiens (Jetz & Pyron 2018) et des Mammifères (Upham *et al.* 2019). Il faut cependant souligner que les approches présentées par Jetz *et al.* (2012) et Upham *et al.* (2019) sont plus complètes que les inférences effectuées par Tonini *et al.* (2016) et Jetz & Pyron (2018) qui ont d'abord inféré une topologie de leur groupe avant de fixer celle-ci lors de l'inférence des temps de divergence. L'ensemble d'arbres qu'ils ont inféré ne couvre donc pas l'incertitude liée à la topologie mais seulement celle liée à la datation. Ces quatre études ont par ailleurs utilisé un nombre réduit de marqueurs génétiques (moins de 20) et une partie conséquente des OTUs a été positionnée sans aucun marqueur génétique (mais à l'aide du package R PASTIS (Thomas *et al.* 2013) et de MrBayes (Ronquist *et al.* 2012)). Les temps de calcul nécessaires pour inférer ces mégaphylogénies furent néanmoins considérables. Il nous fallait donc modifier les approches existantes pour éviter une inflation des temps de calcul liée à notre jeu de données génomiques.

L'originalité de notre approche réside dans la séparation de l'inférence des topologies et de l'estimation des temps de divergence. A notre connaissance, les seuls logiciels permettant l'inférence simultanée de la topologie et des temps de divergences utilisent les approches bayésiennes (par ex. BEAST v2.5 ; Bouckaert *et al.* 2019). Or, ces approches deviennent extrêmement chronophages lorsqu'appliquées sur des jeux de données génomiques (Lartillot 2020). De plus, l'inférence simultanée de la topologie et des temps de divergence par une approche bayésienne est problématique (Rannala 2016) : bien qu'ils ne moyennent que les estimations effectuées pour des nœuds regroupant les mêmes taxons, ces logiciels bayésiens ne considèrent pas que le reste de la topologie peut différer et donc influencer la datation des nœuds d'intérêts. Pour éviter ces biais, nous avons fait le choix d'inférer dans un premier temps les topologies à l'aide du logiciel IQ-TREE puis de les dater à l'aide du logiciel MCMCtree qui considère une topologie fixée. Cette approche implique cependant de considérer une topologie unique du *backbone* dans les différents réplicats de la mégaphylogénie (comme Jetz *et al.* 2012 mais contrairement à Upham *et al.* 2019). Une alternative, qui impliquerait une augmentation conséquente des temps de calcul, serait la datation de 100 réplicats de bootstrap du *backbone* et d'échantillonner l'un des arbres *backbone* datés avant chaque datation des *subtrees*. Considérée dans son ensemble, notre approche, bien que s'appuyant

sur des données génomiques, est largement plus rapide que celle de Jetz *et al.* (2012) et de Upham *et al.* (2019). Ces derniers ont effectivement eu besoin de plus de 3 mois d'inférence pour inférer un ensemble d'arbres des Mammifères alors que nous avons été en capacité de générer une mégaphylogénie de la famille des Saturniidae en près de 10 jours.

Conclusion

Dans ce Chapitre, nous avons présenté un *pipeline* phylogénomique destiné à l'élaboration de mégaphylogénies à partir d'une combinaison de marqueurs génomiques et génétiques. Les analyses exploratoires que nous avons effectuées démontrent l'intérêt d'utiliser la grande quantité de codes-barres ADN disponibles en ligne dans le but (i) d'identifier les OTUs et (ii) d'inférer les relations phylogénétiques superficielles. Notre travail confirme également la pertinence des UCEs dans l'inférence de divergences anciennes de plusieurs dizaines de millions d'années comme dans l'inférence de relations phylogénétiques récentes. Bien qu'il soit appliqué sur un jeu de données génomiques, la rapidité du *pipeline* présenté ici surpasse celle des *pipelines* équivalents publiés jusqu'alors, pourtant appliqués à des données génétiques. Nous prévoyons le séquençage d'UCEs pour davantage d'échantillons (dans le cadre de l'ANR SPHINX) afin de corriger les défauts identifiés ici et ainsi obtenir une mégaphylogénie des Saturniidae résolue dont la complétion sera inédite, tous type d'organismes considérés.

Références

Aberer, A. J., & Stamatakis, A. (2011). A simple and accurate method for rogue taxon identification. In 2011 IEEE International Conference on Bioinformatics and Biomedicine (pp. 118-122).

Aberer, A. J., Krompass, D., & Stamatakis, A. (2013). Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Systematic biology, 62(1), 162-166.

Adamowicz, S. J. (2015). International Barcode of Life: Evolution of a global research community. Genome, 58(5), 151-162.

Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular Ecology, 27(20), 3968-3975.

Basset, Y., Lamarre, G. P., Ratz, T., Segar, S. T., Decaëns, T., Rougerie, R., ... & Ramirez, J. A. (2017). The Saturniidae of Barro Colorado Island, Panama: A model taxon for studying the long-term effects of climate change? Ecology and Evolution, 7(23), 9991-10004.

Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., ... & Schneider, A. K. (2012). What's on the horizon for macroecology? Ecography, 35(8), 673-683.

Bénéluz, F. (1986). Description d'un *Copaxa* inédit du Costa Rica (Lepidoptera, Saturniidae). Revue française d'entomologie (1979), 8(2), 88-90.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. Nucleic acids research, 41(D1), D36-D42.

Bininda-Emonds, O. R., Cardillo, M., Jones, K. E., MacPhee, R. D., Beck, R. M., Grenyer, R., ... & Purvis, A. (2007). The delayed rise of present-day mammals. Nature, 446(7135), 507-512.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., ... & Matschiner, M. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS computational biology, 15(4), e1006650.

Bouvier, E. L. (1927). Les Saturniens du genre *Aurivillius*. Bulletin du Muséum National d'Histoire Naturelle, 33, 71-75.

Brechlin, R. (2007). Some notes on the genus *Samia* Hübner 1819 ("1816") with description of a new species (Lepidoptera: Saturniidae). Entomofauna 1, 56-63.

Brechlin, R. (2019). *Hirpsinjaevia* gen. nov. *viksinjaevi* sp. nov., eine neue Saturniide aus Peru (Lepidoptera). Entomo-Satsphingia, 12(1), 65-68.

Brito, D. (2010). Overcoming the Linnean shortfall: data deficiency and biological survey priorities. Basic and Applied Ecology, 11(8), 709-713.

Brower, A. V. Z. (1994). Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). Molecular Phylogenetics and Evolution, 3(2), 159-174.

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. Journal of human evolution, 79, 35-44.

Carlini, D. B., & Graves, J. E. (1999). Phylogenetic analysis of cytochrome c oxidase I sequences to determine higher-level relationships within the coleoid cephalopods. Bulletin of Marine Science, 64(1), 57-76.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution, 17(4), 540-552.

Chazot, N., Willmott, K. R., Condamine, F. L., De-Silva, D. L., Freitas, A. V., Lamas, G., ... & Mallet, J. (2016). Into the Andes: multiple independent colonizations drive montane diversity in the Neotropical clearwing butterflies Godyridina. Molecular Ecology, 25(22), 5765-5784.

Chazot, N., Condamine, F., Dudas, G., Peña, C., Matos-Maraví, P., Freitas, A. V., ... & Lohman, D. J. (2020). The latitudinal diversity gradient in brush-footed butterflies (Nymphalidae): conserved ancestral tropical niche but different continental histories. bioRxiv.

Clare, E. L., Lim, B. K., Engstrom, M. D., Eger, J. L., & Hebert, P. D. (2007). DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Molecular Ecology Notes, 7(2), 184-190.

Condamine, F. L., Silva-Brandão, K. L., Kergoat, G. J., & Sperling, F. A. (2012). Biogeographic and diversification patterns of Neotropical Troidini butterflies (Papilionidae) support a museum model of diversity dynamics for Amazonia. BMC evolutionary biology, 12(1), 82.

Condamine, F. L., Rolland, J., & Morlon, H. (2019). Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. Ecology letters, 22(11), 1900-1912.

Cooper, A. (1994). DNA from museum specimens. Ancient DNA, 149-165. Springer, New York, NY.

Cox, A. J., & Hebert, P. D. (2001). Colonization, extinction, and phylogeographic patterning in a freshwater crustacean. Molecular ecology, 10(2), 371-386.

Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., ... & Rasplus, J. Y. (2019). Optimized DNA extraction and library preparation for minute arthropods: application to target enrichment in chalcid wasps used for biocontrol. Molecular ecology resources, 19(3), 702-710.

Danforth, B. N. (1999). Phylogeny of the bee genus *Lasioglossum* (Hymenoptera: Halictidae) based on mitochondrial COI sequence data. Systematic Entomology, 24(4), 377-393.

Darge, P. (2015). Observations sur les genres *Micragone* Walker, 1855, et *Cyrtogone* Walker, 1855, avec description de trois genres nouveaux : *Basquiniana*, *Adrougeriana* et *Mucidomorpha* (Lepidoptera, Saturniidae, Saturniinae, Micragonini). Saturnafrica, 22, 21-27.

de Vienne, D. M., Ollier, S., & Aguileta, G. (2012). Phylo-MCOA: a fast and efficient method to detect outlier genes and species in phylogenomics using multiple co-inertia analysis. Molecular Biology and Evolution, 29(6), 1587-1598.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics, 6(5), 361-375.

Diniz-Filho, J. A. F., De Marco, P., & Hawkins, B. A. (2010). Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. Insect Conservation and Diversity, 3, 172–179.

Diniz-Filho, J. A. F., Loyola, R. D., Raia, P., Mooers, A. O., & Bini, L. M. (2013). Darwinian shortfalls in biodiversity conservation. Trends in Ecology & Evolution, 28(12), 689-695.

Driskell, A. C., Ané, C., Burleigh, J. G., McMahon, M. M., O'Meara, B. C., & Sanderson, M. J. (2004). Prospects for building the tree of life from large sequence databases. Science, 306(5699), 1172-1174.

Economo, E. P., Klimov, P., Sarnat, E. M., Guénard, B., Weiser, M. D., Lecroq, B., & Knowles, L. L. (2015). Global phylogenetic structure of the hyperdiverse ant genus *Pheidole* reveals the repeated evolution of macroecological patterns. Proceedings of the Royal Society B: Biological Sciences, 282(1798), 20141416.

Economo, E. P., Narula, N., Friedman, N. R., Weiser, M. D., & Guénard, B. (2018). Macroecology and macroevolution of the latitudinal diversity gradient in ants. Nature communications, 9(1), 1-8.

Elias, M., Hill, R. I., Willmott, K. R., Dasmahapatra, K. K., Brower, A. V., Mallet, J., & Jiggins, C. D. (2007). Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proceedings of the Royal Society B: Biological Sciences, 274(1627), 2881-2889.

Espeland, M., Breinholt, J., Willmott, K. R., Warren, A. D., Vila, R., Toussaint, E. F., ... & Jarzyna, M. A. (2018). A comprehensive and dated phylogenomic analysis of butterflies. Current Biology, 28(5), 770-778.

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic biology, 61(5), 717-726.

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Molecular ecology resources, 15(3), 489-501.

Fang, Y., Shi, W. Q., & Zhang, Y. (2017). Molecular phylogeny of *Anopheles hyrcanus* group (Diptera: Culicidae) based on mtDNA COI. Infectious diseases of poverty, 6(1), 61.

Felsenstein, J. (1985). Phylogenies and the comparative method. The American Naturalist, 125(1), 1-15.

Gallien, L., Saladin, B., Boucher, F. C., Richardson, D. M., & Zimmermann, N. E. (2016). Does the legacy of historical biogeography shape current invasiveness in pines? New Phytologist, 209(3), 1096-1105.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., ... & Gabriel, S. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature biotechnology, 27(2), 182-189.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology, 59(3), 307-321.

Hamilton, A. J., Basset, Y., Benke, K. K., Grimbacher, P. S., Miller, S. E., Novotný, V., ... & Yen, J. D. (2010). Quantifying uncertainty in estimation of tropical arthropod species richness. The American Naturalist, 176(1), 90-95.

Hamilton, C. A., Winiger, N., Rubin, J. J., Breinholt, J., Rougerie, R., Kitching, I. J., ... & Kawahara, A. Y. (2020). Evolution of body size and wing shape trade-offs in arsenurine silkmoths. bioRxiv.

Healy, K., Guillerme, T., Finlay, S., Kane, A., Kelly, S. B., McClean, D., ... & Cooper, N. (2014). Ecology and mode-of-life explain lifespan variation in birds and mammals. Proceedings of the Royal Society B: Biological Sciences, 281(1784), 20140298.

Heath, T. A., Hedtke, S. M., & Hillis, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. Journal of Systematics and Evolution, 46(3), 239-257.

Hebert, P. D., Cywinska, A., Ball, S. L., & Dewaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1512), 313-321.

Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. Plos biol, 2(10), e312.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Molecular biology and evolution, 35(2), 518-522.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., ... & Ronquist, F. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic biology, 65(4), 726-736.

Höhna, S., Freyman, W. A., Nolen, Z., Huelsenbeck, J. P., May, M. R., & Moore, B. R. (2019). A Bayesian approach for estimating branch-specific speciation and extinction rates. bioRxiv, 555805.

Howland, D. E., & Hewitt, G. M. (1995). Phylogeny of the Coleoptera based on mitochondrial cytochrome oxidase I sequence data. Insect Molecular Biology, 4(3), 203-215.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. PloS one, 2(3), e296.

Janzen, D. H., Hallwachs, W., Harvey, D. J., Darrow, K., Rougerie, R., Hajibabaei, M., ... & Sullivan, J. B. (2012). What happens to the traditional taxonomy when a well-known tropical saturniid moth fauna is DNA barcoded?. Invertebrate Systematics, 26(6), 478-505.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. Nature, 491(7424), 444-448.

Jetz, W., & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nature ecology & evolution, 2(5), 850-858.

Joly, S., Heenan, P. B., & Lockhart, P. J. (2014). Species radiation by niche shifts in New Zealand's rockcresses (Pachycladon, Brassicaceae). Systematic Biology, 63(2), 192-202.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14(6), 587-589.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of molecular evolution, 16(2), 111-120.

Kitching, I. J., Rougerie, R., Zwick, A., Hamilton, C. A., St Laurent, R. A., Naumann, S., ... & Kawahara, A. Y. (2018). A global checklist of the Bombycoidea (Insecta: Lepidoptera). Biodiversity Data Journal, (6).

Klopfstein, S., Kropf, C., & Quicke, D. L. (2010). An evaluation of phylogenetic informativeness profiles and the molecular phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae). Systematic Biology, 59(2), 226-241.

Lartillot, N. (2020). The bayesian approach to molecular phylogeny. In Scornavacca, C., Delsuc, F., Galtier, N. Phylogenetics in the Genomic Era. No commercial publisher. Authors open access book, pp.1.4:1–1.4:17, 2020. hal-02535330.

Lemaire, C. (1978). Les Attacidae Américains : Attacinae. C. Lemaire (eds.), Neuilly-sur-Seine, 238pp.

Lemaire, C. (1980). Les Attacidae américains : Arsenurinae. C. Lemaire (eds.), Neuilly-sur-Seine, 199pp.

Lemaire, C. (1988). Les Saturniidae Américains : Ceratocampinae, Museo Nacional de Costa Rica (eds.), San José, 480pp.

Lemaire, C. (2002). The Saturniidae of America. Les Saturniidae américains (= Attacidae). Hemileucinae. Goecke & Evers, Keltern, Germany, 1388 pp., 140 pls.

Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. Annual Review of Ecology, Evolution, and Systematics, 44, 99-121.

Liker, A., Freckleton, R. P., & Székely, T. (2013). The evolution of sex roles in birds is related to adult sex ratio. Nature communications, 4(1), 1-6.

Maliet, O., Hartig, F., & Morlon, H. (2019). A model with many small shifts for estimating species-specific diversification rates. Nature ecology & evolution, 3(7), 1086-1092.

McCormack, J. E., Faircloth, B. C., Crawford, N. G., Gowaty, P. A., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome research, 22(4), 746-754.

McMahon, M. M., & Sanderson, M. J. (2006). Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. Systematic biology, 55(5), 818-836.

Michener, C. D. (1952). The Saturniidae (Lepidoptera) of the Western Hemisphere: morphology, phylogeny, and classification. Bulletin of the AMNH, v. 98, article 5.

Miller, S. E., Hrcek, J., Novotny, V., Weiblen, G. D., & Hebert, P. D. (2013). DNA barcodes of caterpillars (Lepidoptera) from Papua New Guinea. Proceedings of the Entomological Society of Washington, 115(1), 107-109.

Miller, S. E., Martins, D. J., Rosati, M., & Hebert, P. D. (2014). DNA barcodes of moths (Lepidoptera) from Lake Turkana, Kenya. Proceedings of the Entomological Society of Washington, 116(1), 133-136.

Miller, S. E., Hausmann, A., Hallwachs, W., & Janzen, D. H. (2016). Advancing taxonomy and bioinventories with DNA barcodes. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1702), 20150339.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution, 37(5), 1530-1534.

Miranda Jr, H. C., Kennedy, R. S., & Mindell, D. P. (1997). Phylogenetic placement of Mimizuku gurneyi (Aves: Strigidae) inferred from mitochondrial DNA. The Auk, 114(3), 315-323.

Mirarab, S., & Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics, 31(12), i44-i52.

Mishler, B. D. (1994). Cladistic analysis of molecular and morphological data. American Journal of Physical Anthropology, 94(1), 143-156.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., ... & Niehuis, O. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science, 346(6210), 763-767.

Moeller, D. A., Briscoe Runquist, R. D., Moe, A. M., Geber, M. A., Goodwillie, C., Cheptou, P. O., ... & Ree, R. H. (2017). Global biogeography of mating system variation in seed plants. Ecology letters, 20(3), 375-384.

Monteiro, A., & Pierce, N. E. (2001). Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from COI, COII, and EF-1α gene sequences. Molecular phylogenetics and evolution, 18(2), 264-281.

Moore, B. R., Höhna, S., May, M. R., Rannala, B., & Huelsenbeck, J. P. (2016). Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. Proceedings of the National Academy of Sciences, 113(34), 9569-9574.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. (2011). How many species are there on Earth and in the ocean? PLoS Biol, 9(8), e1001127.

Morlon, H. (2014). Phylogenetic approaches for studying diversification. Ecology letters, 17(4), 508-525.

Nater, A., Mattle-Greminger, M. P., Nurcahyo, A., Nowak, M. G., De Manuel, M., Desai, T., ... & Lameira, A. R. (2017). Morphometric, behavioral, and genomic evidence for a new orangutan species. Current Biology, 27(22), 3487-3498.

Naumann, S., & Peigler, R. S. (2001). Four new species of the silkmoth genus *Samia* (Lepidoptera: Saturniidae). Nachrichten des Entomologischen Vereins Apollo, 22, 75-83.

Newman, C. E., & Austin, C. C. (2016). Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. Molecular ecology, 25(24), 6162-6174.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution, 32(1), 268-274.

Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E., & Langmore, N. E. (2014). Female song is widespread and ancestral in songbirds. Nature Communications, 5(1), 1-6.

Packard, A. S. (1895). Monograph of the Bombycine Moths of America North of Mexico: including their transformations and origin of the larval markings and armature, vol. 7,. US Government Printing Office.

Padial, J. M., & De la Riva, I. (2007). Integrative taxonomists should use and produce DNA barcodes. Zootaxa, 1586, 67–68.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526-528.

Peters, R. S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., ... & Diez, P. A. (2017). Evolutionary history of the Hymenoptera. Current Biology, 27(7), 1013-1018.

Poe, S. (1998). Sensitivity of phylogeny estimation to taxonomic sampling. Systematic Biology, 47(1), 18-31.

Pollock, L. J., Thuiller, W., & Jetz, W. (2017). Large conservation gains possible for global biodiversity facets. Nature, 546(7656), 141-144.

Pyron, R. A., & Wiens, J. J. (2013). Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. Proceedings of the Royal Society B: Biological Sciences, 280(1770), 20131622.

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature communications, 4(1), 1-8.

Rabosky, D. L. (2014a). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PloS one, 9(2), e89543.

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., ... & Larson, J. G. (2014b). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods in Ecology and Evolution, 5(7), 701-707.

Rabosky, D. L., Mitchell, J. S., & Chang, J. (2017). Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. Systematic biology, 66(4), 477-498.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... & Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. Nature, 559(7714), 392-395.

Rannala, B. (2016). Conceptual issues in Bayesian divergence time estimation. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1699), 20150134.

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Molecular ecology notes, 7(3), 355-364.

Ratnasingham, S., & Hebert, P. D. (2013). A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PloS one, 8(7), e66213.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in ecology and evolution, 3(2), 217-223.

Rolland, J., Condamine, F. L., Jiguet, F., & Morlon, H. (2014). Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. PLoS Biol, 12(1), e1001775.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology, 61(3), 539-542.

Rougeot, P. (1955). Les Attacides (Saturnidae) de l'Equateur africain français. Encyclopédie Entomologique, 34, 1-116, xii pls.

Rougerie R., Haxaire J., Kitching I. J., Vaglia T., Hebert P. D. N. (2007). Sphingids and Barcodes - The New Taxonomy. 2nd International Barcoding Conference, Taipei (Taiwan), 17-21.IX.2007.

Rougerie R., Ballesteros-Mejia L., Arnal, P. (2019). ACTIAS & SPHINX consortia. From DNA barcode libraries to global macroecology and macroevolutionary studies in insects. 8th international Barcode of Life Conference, Trondheim, Norway, 17-20.VI.2019.

Särkinen, T., Bohs, L., Olmstead, R. G., & Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC evolutionary biology, 13(1), 214.

Seifert, B., d'Eustacchio, D., Kaufmann, B., Centorame, M., Lorite, P., & Modica, M. (2017). Four species within the supercolonial ants of the *Tapinoma nigerrimum* complex revealed by integrative taxonomy (Hymenoptera: Formicidae). Myrmecological News, 24, 123-144.

Smith, M. A., Bertrand, C., Crosby, K., Eveleigh, E. S., Fernandez-Triana, J., Fisher, B. L., ... & Hrcek, J. (2012). *Wolbachia* and DNA barcoding insects: patterns, potential, and problems. PloS one, 7(5), e36514.

Smith, S. A., Beaulieu, J. M., & Donoghue, M. J. (2009). Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC evolutionary biology, 9(1), 37.

Smith, S. A., & O'Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics, 28(20), 2689-2690.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313.

Stoeckle, M. (2003). Taxonomy, DNA, and the bar code of life. BioScience, 53(9), 796-797.

Stork, N. E., McBroom, J., Gely, C., & Hamilton, A. J. (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. Proceedings of the National Academy of Sciences, 112(24), 7519-7523.

Tagliacollo, V. A., & Lanfear, R. (2018). Estimating improved partitioning schemes for ultraconserved elements. Molecular Biology and Evolution, 35(7), 1798-1811.

Ter Minassian, V. (2017). Découverte d'une nouvelle espèce de grand singe en Indonésie. Journal LeMonde, ed. du 5 novembre 2017.

Thomas, G. H., Hartmann, K., Jetz, W., Joy, J. B., Mimoto, A., & Mooers, A. O. (2013). PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences. Methods in Ecology and Evolution, 4(11), 1011-1017.

Thorne, J. L., Kishino, H., & Painter, I. S. (1998). Estimating the rate of evolution of the rate of molecular evolution. Molecular biology and evolution, 15(12), 1647-1657.

Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., & Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biological Conservation, 204, 23-31.

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS biology, 17(12), e3000494.

Wilkinson, M. (1996). Majority-rule reduced consensus trees and their use in bootstrapping. Molecular Biology and evolution, 13(3), 437-444.

Wilson, K. H. (1995). Molecular biology as a tool for taxonomy. Clinical infectious diseases, 20(Supplement_2), S117-S121.

Wilson, J. J. (2010). Assessing the value of DNA barcodes and other priority gene regions for molecular phylogenetics of Lepidoptera. PLoS One, 5(5), e10525.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences, 13(5), 555-556.

Yang, Z., & Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular biology and evolution, 23(1), 212-226.

Zhang, Z. Q. (2011). Animal biodiversity: an introduction to higher-level classification and taxonomic richness. Zootaxa, 3148(1), 7-12.

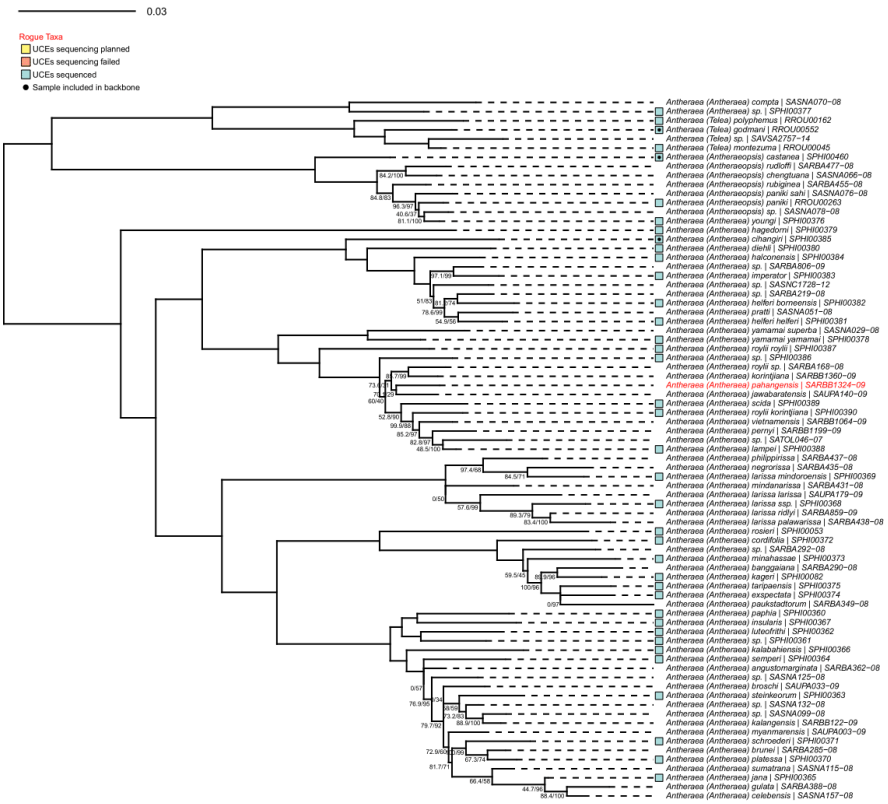
Zheng, Y., & Wiens, J. J. (2016). Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. Molecular phylogenetics and evolution, 94, 537-547.

Zhou, X., Shen, X. X., Hittinger, C. T., & Rokas, A. (2018). Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. Molecular biology and evolution, 35(2), 486-503.

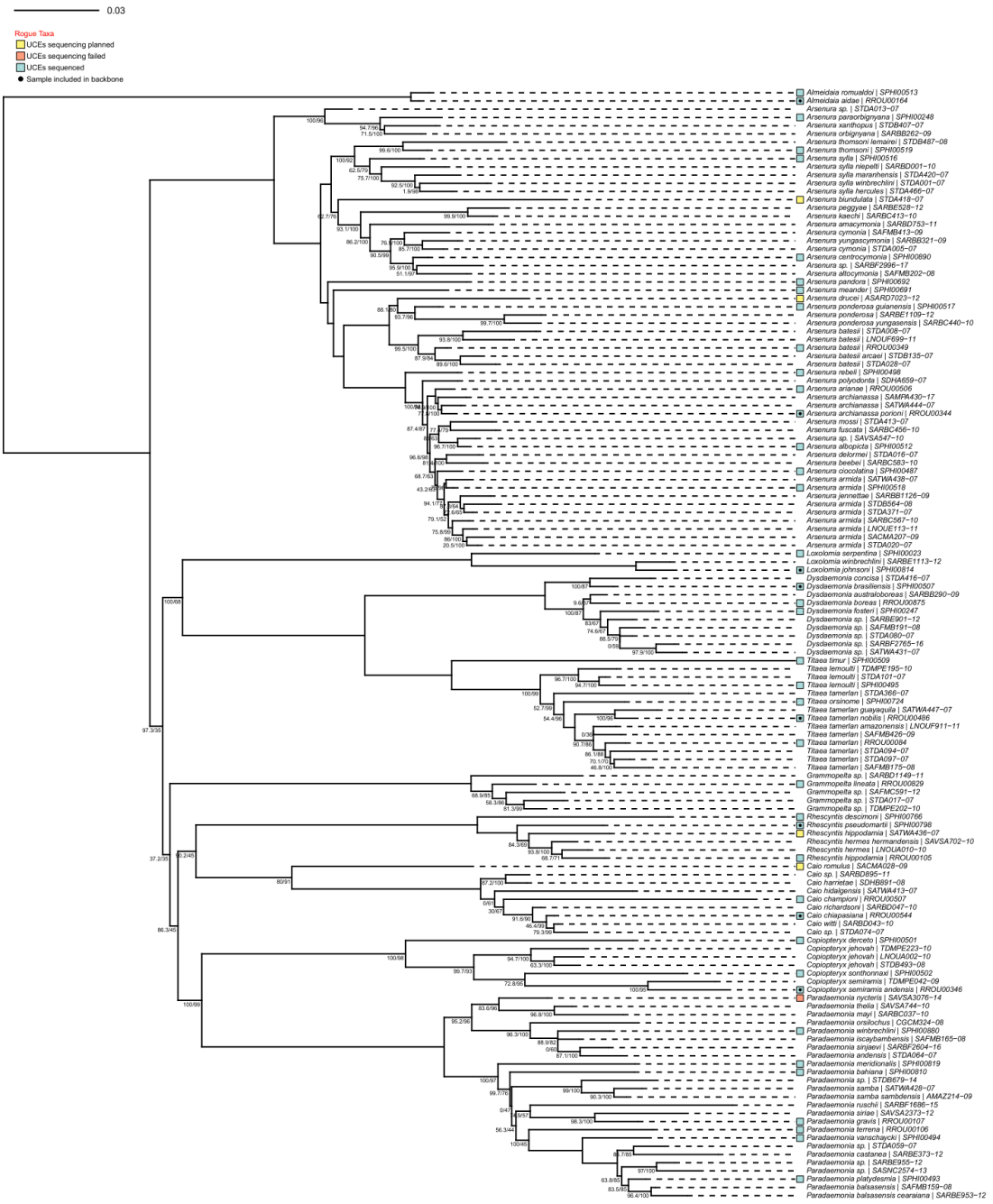
Annexe 1 – Phylogénies des *subtrees*

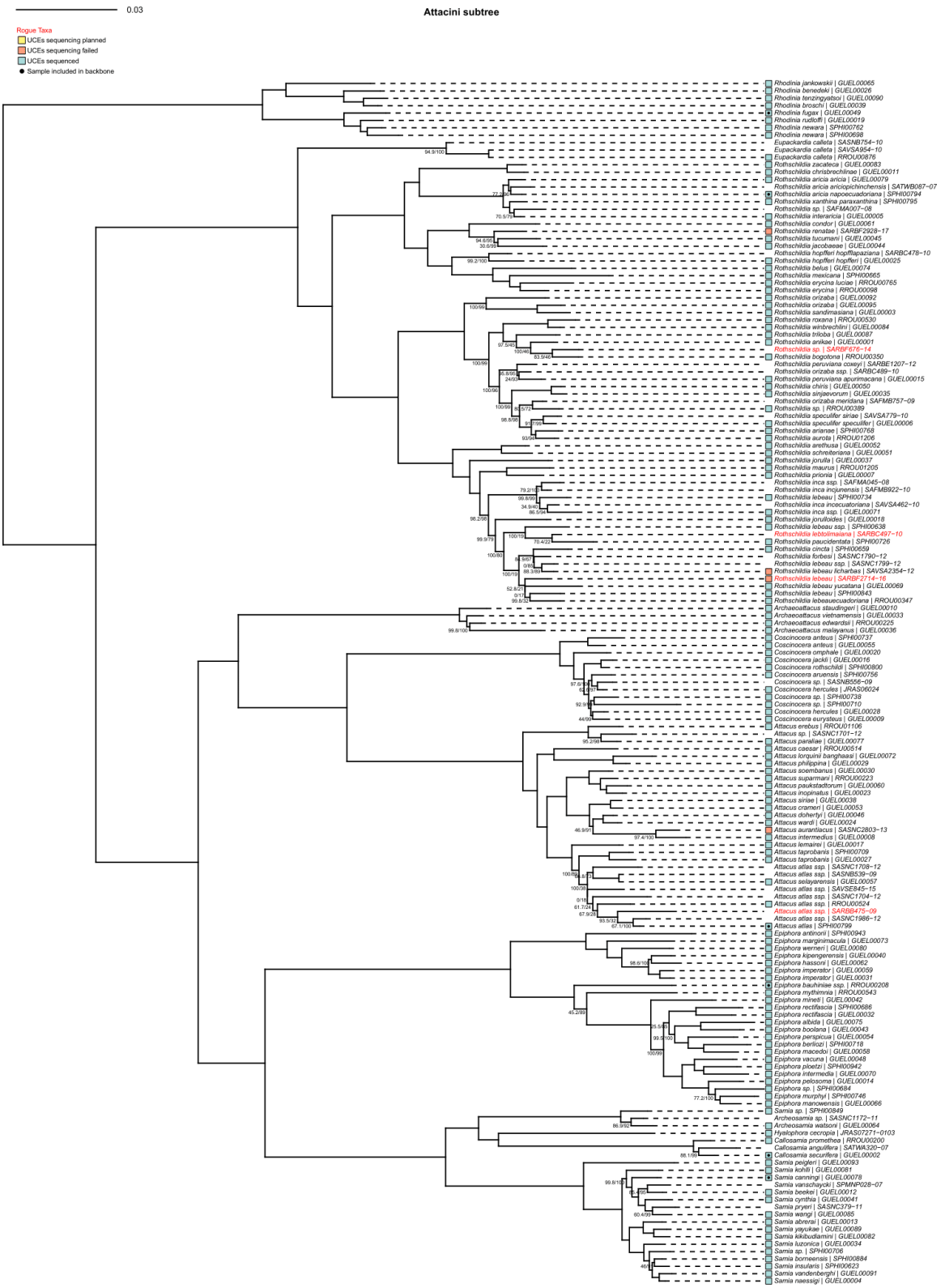


Antheraea subtree

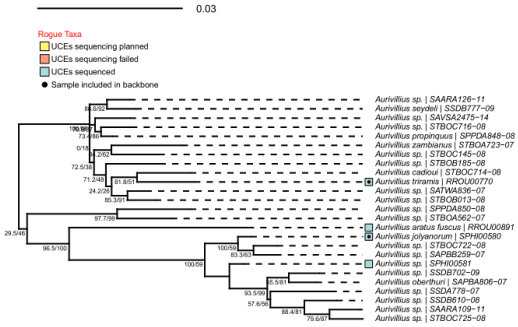


Arsenurinae subtree

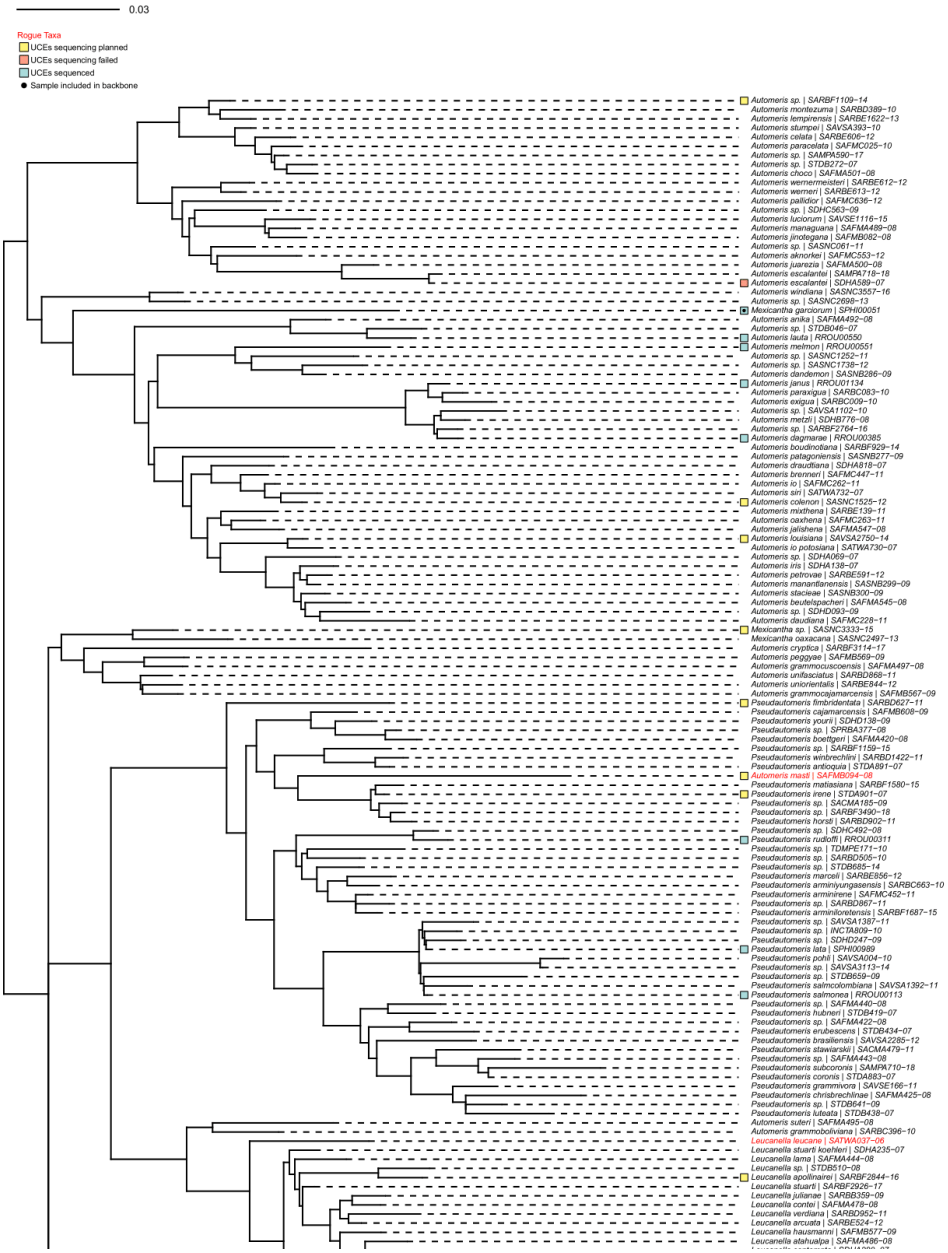



0.03
Attacini subtree
Rogue Taxa
UCEs sequencing planned
UCEs sequencing failed
UCEs sequenced
Sample included in backbone
Rhodinia jankowskii | GUEL00065
Rhodinia benedeki | GUEL00026
Rhodinia tenzingyatsoi | GUEL00090
Rhodinia broschi | GUEL00039
Rhodinia fugax | GUEL00049
Rhodinia rudloffi | GUEL00019
Rhodinia newara | SPHI00762
Rhodinia newara | SPHI00698
Eupackardia calleta | SASNB754−10
Eupackardia calleta | SAVSA954−10
Eupackardia calleta | RROU00876
Rothschildia zacateca | GUEL00083
Rothschildia christbrechlinae | GUEL00011
Rothschildia aricia aricia | GUEL00079
Rothschildia aricia ariciopichinchensis | SATWB087−07
Rothschildia aricia napoecuadoriana | SPHI00794
Rothschildia xanthina paraxanthina | SPHI00795
Rothschildia sp. | SAFMA007−08
Rothschildia interarcia | GUEL00005
Rothschildia constor | GUEL00061
Rothschildia renatae | SARBF2928−17
Rothschildia tucumanni | GUEL00045
Rothschildia jacobaeae | GUEL00044
Rothschildia hopfferi hopfflapaziana | SARBC478−10
Rothschildia hopfferi hopfferi | GUEL00025
Rothschildia belus | GUEL00074
Rothschildia mexicana | SPHI00665
Rothschildia erycina lucae | RROU00765
Rothschildia erycina | RROU00098
Rothschildia orizaba | GUEL00092
Rothschildia orizaba | GUEL00095
Rothschildia sandimasiana | GUEL00003
Rothschildia roxana | RROU00530
Rothschildia winbrechlini | GUEL00084
Rothschildia triloba | GUEL00087
Rothschildia anikae | GUEL00001
Rothschildia sp. | SARBF676−14
Rothschildia bogotona | RROU00350
Rothschildia peruviana coxeyi | SARBE1207−12
Rothschildia orizaba ssp. | SARBC489−10
Rothschildia peruviana apurimacana | GUEL00015
Rothschildia chiris | GUEL00050
Rothschildia sinjaevorum | GUEL00035
Rothschildia orizaba meridana | SAFMB757−09
Rothschildia sp. | RROU00389
Rothschildia speculifer siriae | SAVSA779−10
Rothschildia speculifer speculifer | GUEL00006
Rothschildia arianae | SPHI00768
Rothschildia aurota | RROU01206
Rothschildia arethusa | GUEL00052
Rothschildia schreiteriana | GUEL00051
Rothschildia jorulla | GUEL00037
Rothschildia maurus | RROU01205
Rothschildia prionia | GUEL00007
Rothschildia inca ssp. | SAFMA045−08
Rothschildia inca inquinensis | SAFMB922−10
Rothschildia lebeau | SPHI00734
Rothschildia inca incecuatoriana | SAVSA462−10
Rothschildia inca ssp. | GUEL00071
Rothschildia jorullioides | GUEL00018
Rothschildia lebeau ssp. | SPHI00638
Rothschildia lebtolmaiana | SARBC497−10
Rothschildia paucidentata | SPHI00726
Rothschildia cincta | SPHI00659
Rothschildia forbesi | SASNC1790−12
Rothschildia lebeau ssp. | SASNC1799−12
Rothschildia lebeau licharbas | SAVSA2354−12
Rothschildia lebeau | SARBF2714−16
Rothschildia lebeau yucatana | GUEL00069
Rothschildia lebeau | SPHI00843
Rothschildia lebeauecuadoriana | RROU00347
Archaeoattacus staudingeri | GUEL00010
Archaeoattacus vietnamensis | GUEL00033
Archaeoattacus edwardsii | RROU00225
Archaeoattacus malayanus | GUEL00036
Coscinocera anteus | SPHI00737
Coscinocera anteus | GUEL00055
Coscinocera omphale | GUEL00020
Coscinocera jackli | GUEL00016
Coscinocera rothschildi | SPHI00800
Coscinocera aruensis | SPHI00756
Coscinocera sp. | SASNB556−09
Coscinocera hercules | JRAS036024
Coscinocera sp. | SPHI00738
Coscinocera sp. | SPHI00710
Coscinocera hercules | GUEL00028
Coscinocera eurysteus | GUEL00009
Attacus erebus | RROU01106
Attacus sp. | SASNC1701−12
Attacus paraliae | GUEL00077
Attacus caesar | RROU00514
Attacus lorquinii banghaasi | GUEL00072
Attacus philippina | GUEL00029
Attacus soembanus | GUEL00030
Attacus suparmani | RROU00223
Attacus paukstadtorum | GUEL00060
Attacus inopinatus | GUEL00023
Attacus sinae | GUEL00038
Attacus crameri | GUEL00053
Attacus dohertyi | GUEL00046
Attacus wardi | GUEL00024
Attacus aurantiacus | SASNC2803−13
Attacus intermedius | GUEL00008
Attacus lemairei | GUEL00017
Attacus taprobanis | SPHI00709
Attacus taprobanis | GUEL00027
Attacus atlas ssp. | SASNC1708−12
Attacus atlas ssp. | SASNB539−09
Attacus selayarensis | GUEL00057
Attacus atlas ssp. | SAVSE845−15
Attacus atlas ssp. | SASNC1704−12
Attacus atlas ssp. | RROU00524
Attacus atlas ssp. | SARBB475−09
Attacus atlas ssp. | SASNC1986−12
Attacus atlas | SPHI00799
Epiphora antinorii | SPHI00943
Epiphora marginimacula | GUEL00073
Epiphora werneri | GUEL00080
Epiphora kipengerensis | GUEL00040
Epiphora hassoni | GUEL00062
Epiphora imperator | GUEL00059
Epiphora imperator | GUEL00031
Epiphora bauhiniae ssp. | RROU00208
Epiphora mythimnia | RROU00543
Epiphora mineti | GUEL00042
Epiphora rectifascia | SPHI00686
Epiphora rectifascia | GUEL00032
Epiphora albida | GUEL00075
Epiphora boolana | GUEL00043
Epiphora perspicua | GUEL00054
Epiphora berliozi | SPHI00718
Epiphora macedoi | GUEL00058
Epiphora vacuna | GUEL00048
Epiphora pisetzi | SPHI00942
Epiphora intermedia | GUEL00070
Epiphora pelosoma | GUEL00014
Epiphora sp. | SPHI00684
Epiphora murphyi | SPHI00746
Epiphora manowensis | GUEL00066
Samia sp. | SPHI00849
Archeosamia sp. | SASNC1172−11
Archeosamia watsoni | GUEL00064
Hyalophora cecropia | JRAS07227−0103
Callosamia promethea | RROU00200
Callosamia angulifera | SATWA320−07
Callosamia securifera | GUEL00002
Samia peigleri | GUEL00093
Samia kohlii | GUEL00081
Samia canningi | GUEL00078
Samia vanschayoki | SPMNP028−07
Samia beekei | GUEL00012
Samia cynthia | GUEL00041
Samia pryeri | SASNC379−11
Samia wangi | GUEL00085
Samia abrerai | GUEL00013
Samia yayukae | GUEL00089
Samia kikitsudaimini | GUEL00082
Samia luzonica | GUEL00034
Samia sp. | SPHI00706
Samia borneensis | SPHI00884
Samia insularis | SPHI00623
Samia vandenberghi | GUEL00091
Samia naessigi | GUEL00004

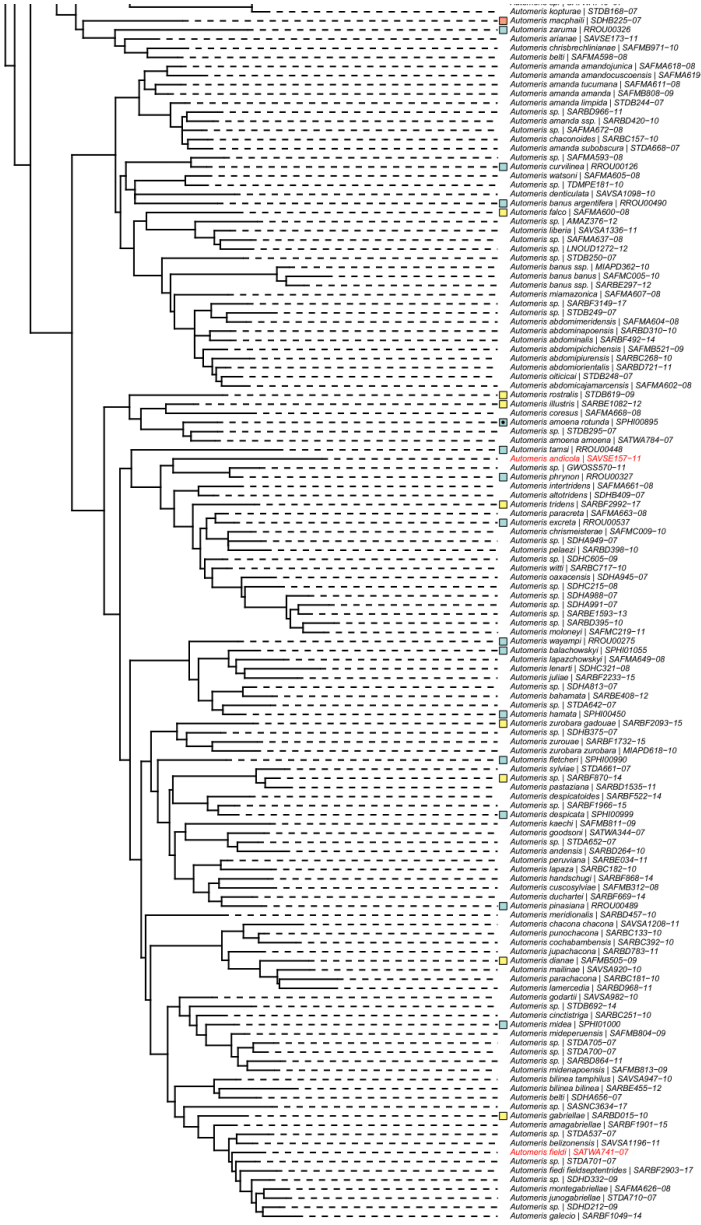
Aurivillius subtree



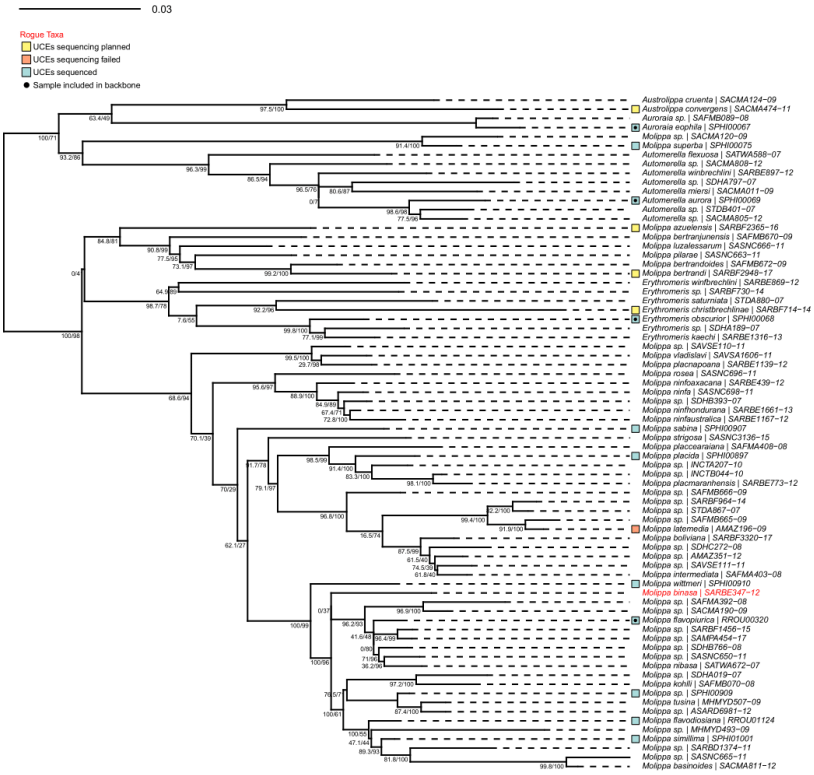
Automerina_A subtree



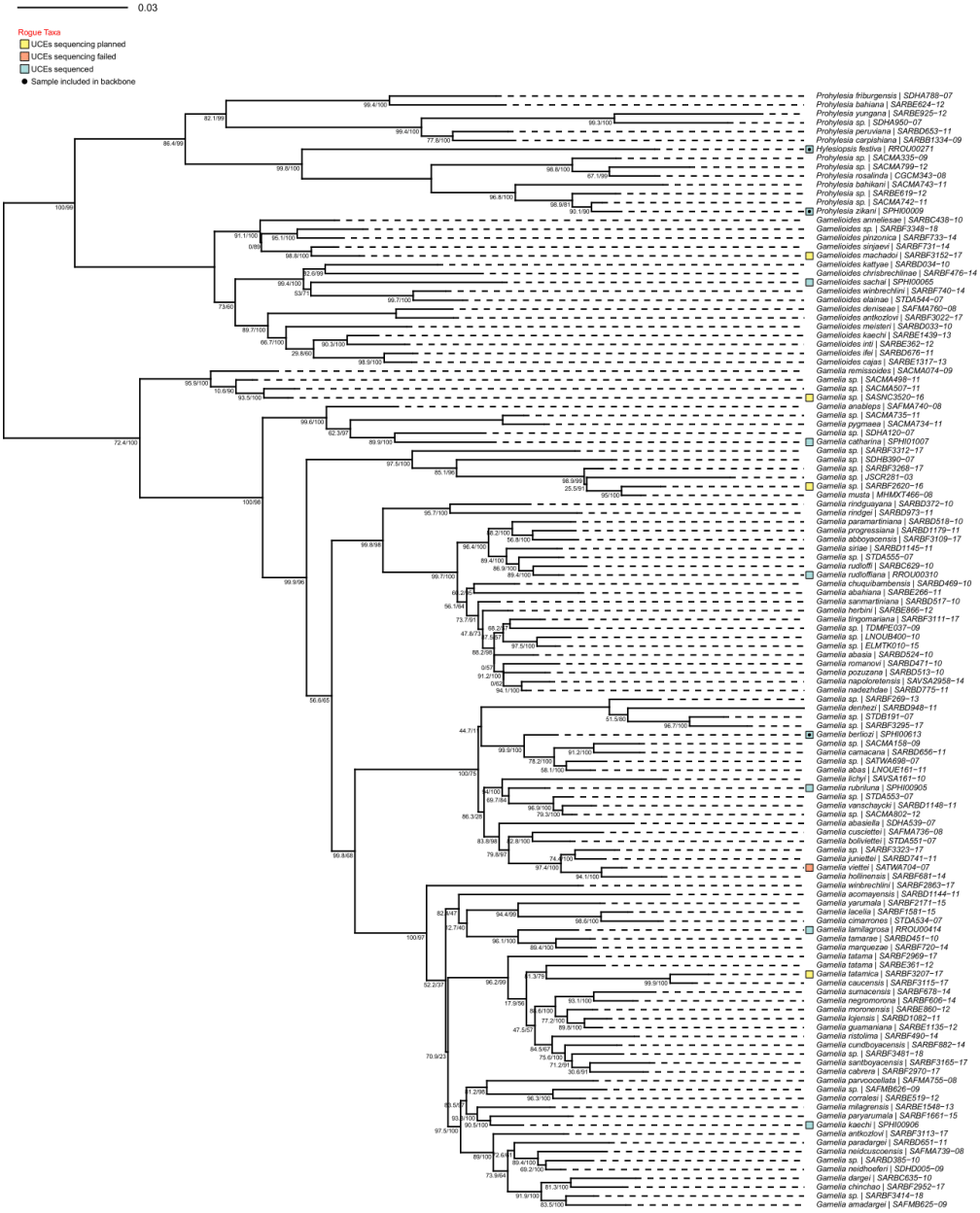

Leucanella arcuata | SARBE524–12
Leucanella hausmanni | SAFMB577–09
Leucanella atahualpa | SAFMA486–08
Leucanella contempta | SDHA280–07
Leucanella neglecta | SARBD603–11
Leucanella contempta flavissima | SDHA233–07
Leucanella gibbosa | SACMA017–09
Leucanella janaira | SPHI01052
Leucanella heisleri | STDB715–14
Leucanella aspera | SARBD109–10
Leucanella sp. | SPHI00912
Leucanella anikae | SARBD104–10
Leucanella viridescens viridior | SAFMB583–09
Leucanella memusae gardineri | SAVSA1413–11
Leucanella manzanoi | SARBE788–12
Leucanella maandensis | SARBB159–09
Leucanella maassarni | SPHI00994
Leucanella rufobolliviana | SARBC640–10
Leucanella viettei | SAVSE130–11
Leucanella winfriechlini | SARBF2963–17
Leucanella evanschayoki | SARBD111–10
Leucanella sp. | SARBF2095–15
Leucanella sp. | RROU00317
Leucanella sp. | SDHA183–07
Leucanella contempta inti | SARBD907–11
Leucanella sp. | SDHA248–07
Leucanella leucane | SDHA255–07
Leucanella contempta windi | SDHA245–07
Leucanella fusca | RROU00535
Leucanella nyctimene | SARBD1068–11
Leucanella hosmera | SDHC224–08
Leucanella flammans | SARBE954–12
Leucanella sp. | SARBE1475–13
Leucanella memusoides | SDHB944–08
Leucanella sp. | SDHC608–09
Leucanella sp. | ASARD6959–12
Leucanella acutissima | SDHB278–07
Automeris heppneri | SARBD580–11
Hypermerina rubrojunensis | SARBD1154–11
Hypermerina kasyi | SPHI00085
Pseudautomeris sp. | SARBD1613–11
Pseudautomeris sp. | SARBD1005–11
Pseudautomeris sp. | SARBF1604–15
Pseudautomeris sp. | SAVSA1579–11
Pseudautomeris sp. | SARBC883–10
Pseudautomeris sp. | SARBE008–11
Automeris sp. | SDHD083–09
Automeris winbrechlinari | SAFMA531–08
Automeris claryi | SAFMA568–08
Automeris basalis | SDHA642–07
Automeris dognini | SPHI00451
Automeris elcearaiana | SARBE701–12
Automeris arminandensis | SARBF1057–14
Automeris arminia | SPHI00995
Automeris musculi | SAFMC065–10
Automeris nubila | SACMA487–11
Automeris umbrosa umbrosa | STDB071–07
Automeris naranja | SAVSA1297–11
Automeris ovalina | SACMA037–11
Automeris parainornata | SAFMA556–08
Automeris inornata | SARBE702–12
Automeris sp. | SASNC2174–12
Automeris castrensis | STDB283–07
Automeris sp. | STDB388–07
Automeris haxaire | SDHA007–07
Automeris paravolanda | SARBE160–11
Automeris sp. | SARBA064–08
Automeris sp. | SAVSA1323–11
Automeris sp. | TDMPE308–10
Automeris sp. | STDB115–07
Automeris inmoxia | STDB113–07
Automeris cosangana | SARBD1518–11
Automeris pomiferoides | SAFMA549–08
Automeris pomifera | SARBD1075–11
Automeris schwartzi | SAVSA3118–14
Automeris parafera | SARBD1593–11
Automeris kitching | SAFMA529–08
Automeris moenchorum | SAFMA581–08
Automeris lachaumei | STDB597–09
Automeris yungasmargaritae | SARBA063–08
Automeris margaritae | SDHA042–07
Automeris sp. | SDHA045–07
Automeris labriquei | SARBC146–10
Automeris sp. | SAMFA589–17
Automeris sp. | SARBF1610–15
Automeris sp. | RROU00431
Automeris sp. | SDHA058–07
Automeris incarnata | STDB085–07
Automeris praemargaritae | SATWA801–07
Automeris maximae | SARBF2483–16
Automeris adusta | SARBF2759–16
Automeris sp. | SARBC722–10
Automeris maeonia | SDHC618–09
Automeris zephyria | RROU01116
Automeris peigleri | SDHA083–07
Automeris cecrops pamina | SATWA796–07
Automeris cecrops cecrops | SDHA055–07
Automeris vomona | SAFMA564–08
Automeris descimoni | SARBD871–11
Automeris viksinjaevi | SARBF3085–17
Automeris sp. | SARBE1133–12
Automeris ecuata | SARBC237–10
Automeris parecuata | SARBD1524–11
Automeris dunschei | SAVSA1238–11
Automeris chanchamayensis | SARBF1917–15
Automeris paramelanops | SACMA642–11
Automeris sp. | TDMPE297–10
Automeris annulata | SDHA147–07
Automeris amasanata | SARBC254–10
Automeris sp. | SACMA046–09
Automeris styx | SAFMA561–08
Automeris cuscoata | SARBC252–10
Automeris janrudloffi | SARBF2594–16
Automeris pinas | SARBD002–10
Automeris zamorana | SARBD679–11
Automeris pacchana | SARBE1285–13
Automeris isabellae | SARBF1656–15
Automeris pichichensis | SARBD007–10
Automeris barragani | SARBF956–14
Automeris latenigra | SDHB613–08
Automeris harricajamarcensis | SAVSA1320–11
Automeris diavolanda | SATWA780–07
Automeris arri | SARBF1540–15
Automeris harriyungasiana | SARBC239–10
Automeris harrisorum | SARBE162–11
Automeris harriamazonica | SAFMA520–08
Automeris sp. | SARBD924–11
Automeris sp. | SARBF545–14
Automeris zozinicareguana | SARBA065–08
Automeris sarapiqui | SARBF2694–16
Automeris zozine | SDHA591–07
Automeris tuxlana | SARBE096–11
Automeris boucardi | SAFMA578–08
Automeris sp. | SAFMC231–11
Automeris sp. | SAVSA1507–11
Automeris parapichinchensis | RROU00315
Automeris manzanoi | SARBD678–11
Automeris lojana | SATWA806–07
Automeris zozimanaguana | RROU01125
Automeris magdaleniana | SARBC228–10
Automeris zugana | STDB354–07
Automeris sp. | SDHD028–09
Automeris tolimaensis | SARBC226–10
Automeris cundinamarcensis | STDB077–07
Automeris sp. | SDHA010–07
Automeris sp. | SAFMA517–08
Automeris lecoulti | SDHA014–07
Automeris granulosa | SACMA042–09
Automeris oberthuri | SARBE159–11
Automeris granulosa | STDB263–07
Automeris beckeri | SPHI01057
Automeris submacula | SARBF2318–16
Automeris bertha | SPHI00002
Automeris silkae | SARBF574–14
Automeris sp. | STDB204–07
Automeris postalbida | SAVSA1079–10
Automeris vanschayoki | SAVSA402–10
Automeris anikmeisterae | SAFMC578–12
Automeris sp. | SARBD959–11
Automeris niepelti | SARBC419–10
Automeris chrisbrechlinae | SARBE1609–13
Automeris larra | SPHI00992
Automeris sp. | TDMPE205–10
Automeris paralarra | SAFMA688–08
Automeris parageus | SARBB1351–09
Automeris sp. | STDB105–07
Automeris egeus | RROU00085
Automeris sp. | SARBD961–11
Automeris boops | SARBF3244–17
Automeris tatiae | SARBC072–10
Automeris mixtus | SARBF3249–17
Automeris sp. | SASNC2454–12
Automeris winbrechlini | SARBD725–11
Automeris huascari | SARBD017–10
Automeris risguindensis | SARBF1812–15
Automeris iwanowitschi | SARBE355–12
Automeris napoensis | SARBB424–09
Automeris conceptiona | SARBF741–14
Automeris doelfi | SARBD723–11
Automeris gunneri | SARBF2353–16
Automeris peggyanae peggyanae | SARBF2356–16
Automeris hausmanni | SARBF2165–15
Automeris africarchensis | SARBD306–10
Automeris rectilinea | SACMA912–12
Automeris tristis | STDB217–07
Automeris sp. | SARBF598–14
Automeris occidentorestes | SAFMB812–09
Automeris sp. | SARBF551–14
Automeris sp. | SAVSA1043–10
Automeris orestes | TIPSY518–12
Automeris themis | SATYE508–09
Automeris morescia | SATWA249–07
Automeris fabiani | LNOUE477–11
Automeris jucunda | SPHI00920
Automeris sp. | SATWA746–07
Automeris kopturae | STDB168–07
Automeris macphaili | SDHB225–07
Automeris zaruma | RROU00326


Automeris kopturae | STDB168–07
Automeris macphaili | SDHB225–07
Automeris zaruma | RROU00326
Automeris arianae | SAVSE173–11
Automeris chrisbrechlinianae | SAFMB971–10
Automeris belti | SAFMA598–08
Automeris amanda amandajunica | SAFMA618–08
Automeris amanda amandocuscoensis | SAFMA619–08
Automeris amanda tucumana | SAFMA611–08
Automeris amanda amanda | SAFMB808–09
Automeris amanda limpida | STDB244–07
Automeris sp. | SARBD966–11
Automeris amanda ssp. | SARBD420–10
Automeris sp. | SAFMA672–08
Automeris chaconoides | SARBC157–10
Automeris amanda subobscura | STDA668–07
Automeris sp. | SAFMA593–08
Automeris curvilinea | RROU00126
Automeris watsoni | SAFMA605–08
Automeris sp. | TDMPE181–10
Automeris denticulata | SAVSA1098–10
Automeris banus argentifera | RROU00490
Automeris falco | SAFMA600–08
Automeris sp. | AMAZ376–12
Automeris liberia | SAVSA1336–11
Automeris sp. | SAFMA637–08
Automeris sp. | LMOUD1272–12
Automeris sp. | STDB250–07
Automeris banus ssp. | MIAPD362–10
Automeris banus banus | SAFMC005–10
Automeris banus ssp. | SARBE297–12
Automeris miamazonica | SAFMA607–08
Automeris sp. | SARBF3149–17
Automeris sp. | STDB249–07
Automeris abdomimeridensis | SAFMA604–08
Automeris abdominapoensis | SARBD310–10
Automeris abdominalis | SARBF492–14
Automeris abdomipichichensis | SAFMB521–09
Automeris abdomipurensis | SARBC269–10
Automeris abdomiorientalis | SARBD721–11
Automeris oticicai | STDB248–07
Automeris abdomicajamarcensis | SAFMA602–08
Automeris rostralis | STDB619–09
Automeris illustris | SARBE1082–12
Automeris coresus | SAFMA668–08
Automeris amoena rotunda | SPHI00895
Automeris sp. | STDB295–07
Automeris amoena amoena | SATWA784–07
Automeris tamsi | RROU00448
Automeris andicola | SAVSE157–11
Automeris sp. | GWOSS570–11
Automeris phrynon | RROU00327
Automeris intertridens | SAFMA661–08
Automeris altotridens | SDHB409–07
Automeris tridens | SARBF2992–17
Automeris paracreta | SAFMA663–08
Automeris excreta | RROU00537
Automeris chrismeisterae | SAFMC009–10
Automeris sp. | SDHA949–07
Automeris pelaezi | SARBD398–10
Automeris sp. | SDHC605–09
Automeris witti | SARBC717–10
Automeris oaxacensis | SDHA945–07
Automeris sp. | SDHC215–08
Automeris sp. | SDHA908–07
Automeris sp. | SDHA991–07
Automeris sp. | SARBE1593–13
Automeris sp. | SARBD395–10
Automeris moloneyi | SAFMC219–11
Automeris wayampi | RROU00275
Automeris balachowskyi | SPHI01055
Automeris lapazchowskyi | SAFMA649–08
Automeris lenarti | SDHC321–08
Automeris juliae | SARBF2233–15
Automeris sp. | SDHA813–07
Automeris bahamata | SARBE408–12
Automeris sp. | STDA642–07
Automeris hamata | SPHI00450
Automeris zurobara gadouae | SARBF2093–15
Automeris sp. | SDHB375–07
Automeris zurouae | SARBF1732–15
Automeris zurobara zurobara | MIAPD618–10
Automeris fletcheri | SPHI00990
Automeris sylviae | STDA661–07
Automeris sp. | SARBF870–14
Automeris pastaziana | SARBD1535–11
Automeris despicatoides | SARBF522–14
Automeris sp. | SARBF1966–15
Automeris despicata | SPHI00999
Automeris kaschi | SAFMB811–09
Automeris goodsoni | SATWA344–07
Automeris sp. | STDA652–07
Automeris andensis | SARBD264–10
Automeris peruviana | SARBE034–11
Automeris lapaza | SARBC182–10
Automeris handschugi | SARBF868–14
Automeris cuscosylviae | SAFMB312–08
Automeris duchartei | SARBF669–14
Automeris pinasiana | RROU00489
Automeris meridionalis | SARBD457–10
Automeris chacona chacona | SAVSA1208–11
Automeris punochacona | SARBC133–10
Automeris cochabambensis | SARBC392–10
Automeris jupachacona | SARBD783–11
Automeris dianae | SAFMB505–09
Automeris malinae | SAVSA920–10
Automeris parachacona | SARBC181–10
Automeris lamercedia | SARBD968–11
Automeris godarti | SAVSA982–10
Automeris sp. | STDB692–14
Automeris cinctistriga | SARBC251–10
Automeris midea | SPHI01000
Automeris midesperansis | SAFMB804–09
Automeris sp. | STDA705–07
Automeris sp. | STDA700–07
Automeris sp. | SARBD864–11
Automeris midenapoensis | SAFMB813–09
Automeris bilinea tamphilus | SAVSA947–10
Automeris bilinea bilinea | SARBE455–12
Automeris belti | SDHA656–07
Automeris sp. | SASNC3634–17
Automeris gabrielae | SARBD015–10
Automeris amagabriellae | SARBF1901–15
Automeris sp. | STDA537–07
Automeris belizonensis | SAVSA1196–11
Automeris fieldi | SATWA741–07
Automeris sp. | STDA701–07
Automeris fieldi fieldseptentrides | SARBF2903–17
Automeris sp. | SDHD332–09
Automeris montegabriellae | SAFMA626–08
Automeris junogabriellae | STDA710–07
Automeris sp. | SDHD212–09
Automeris galecio | SARBF1049–14

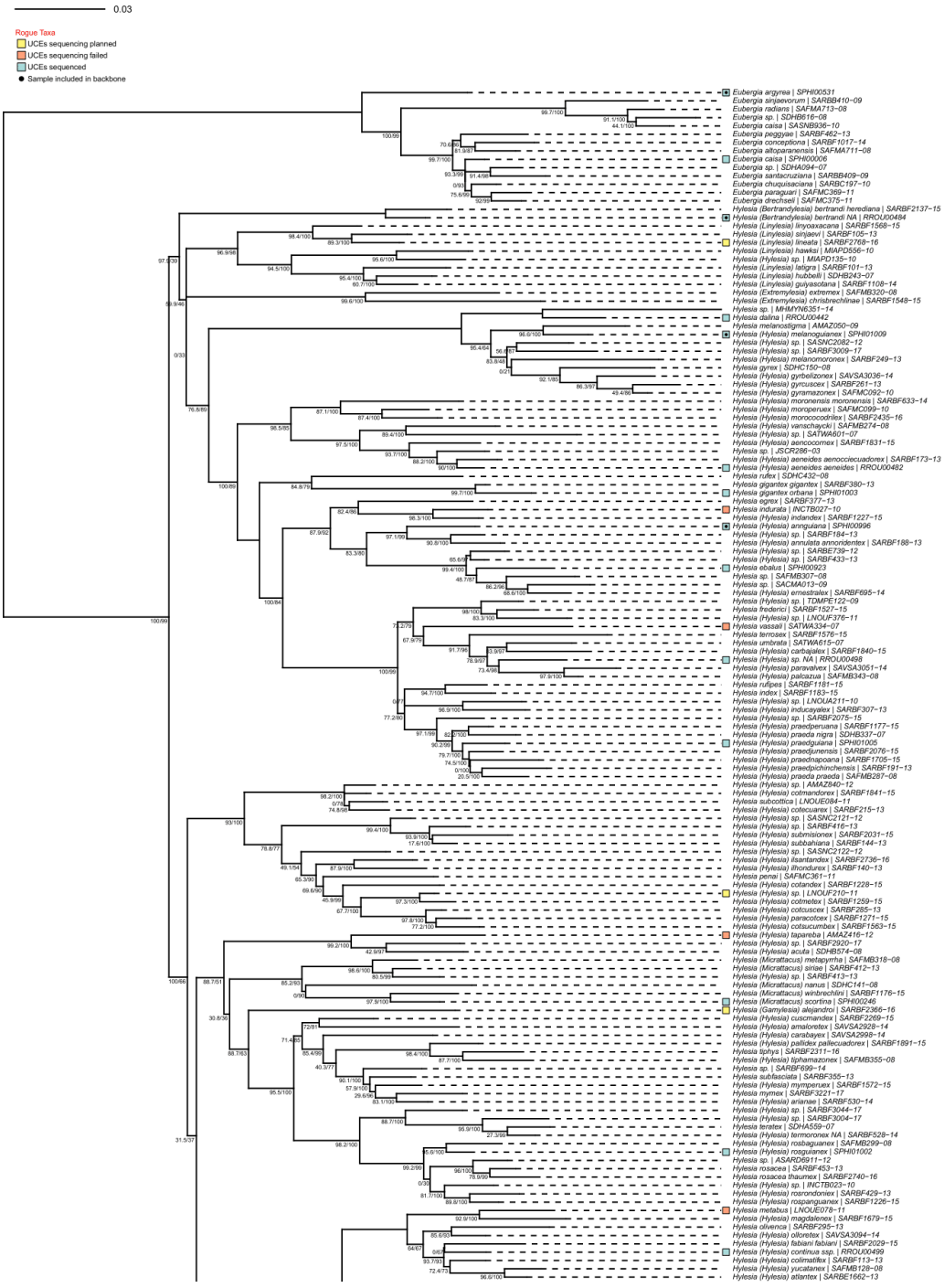
Automerina_B subtree

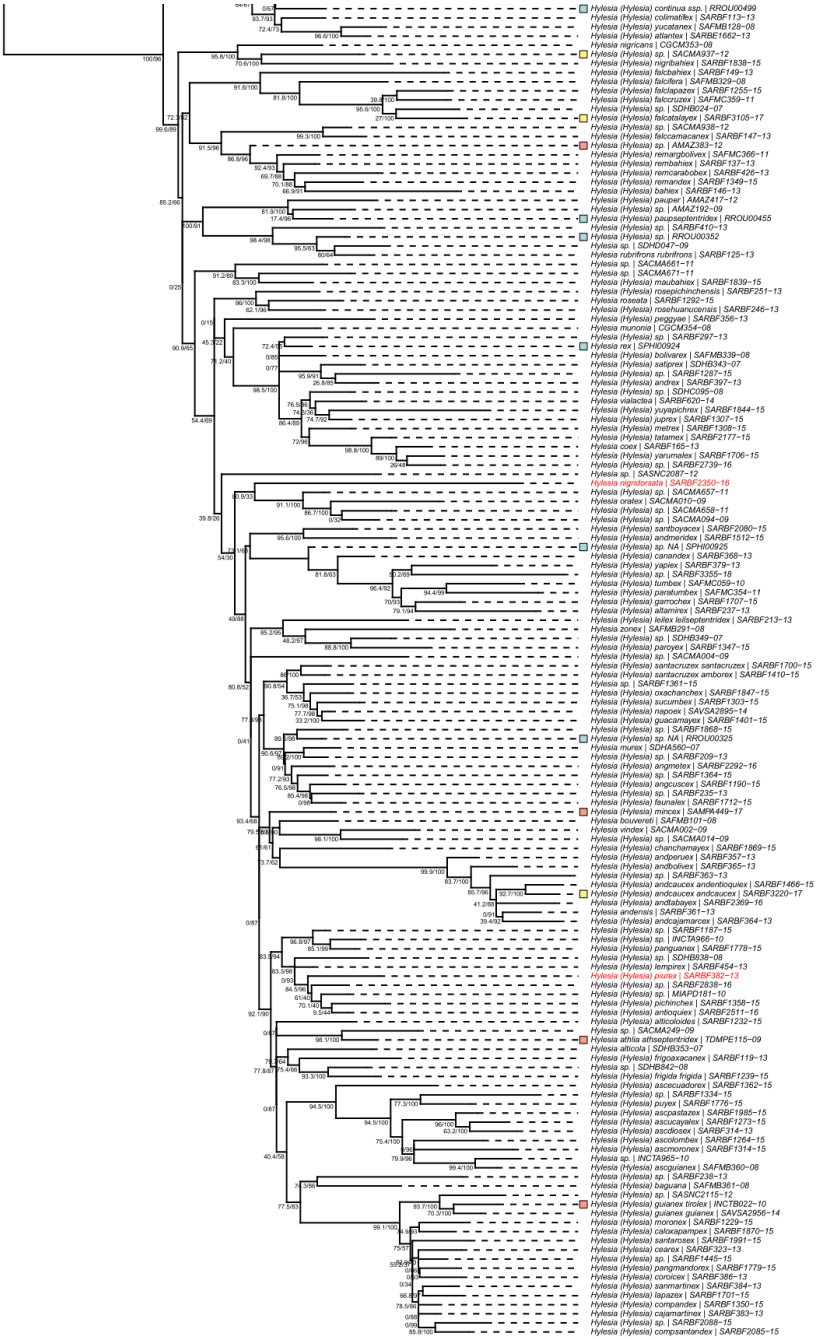


Automerina_C subtree

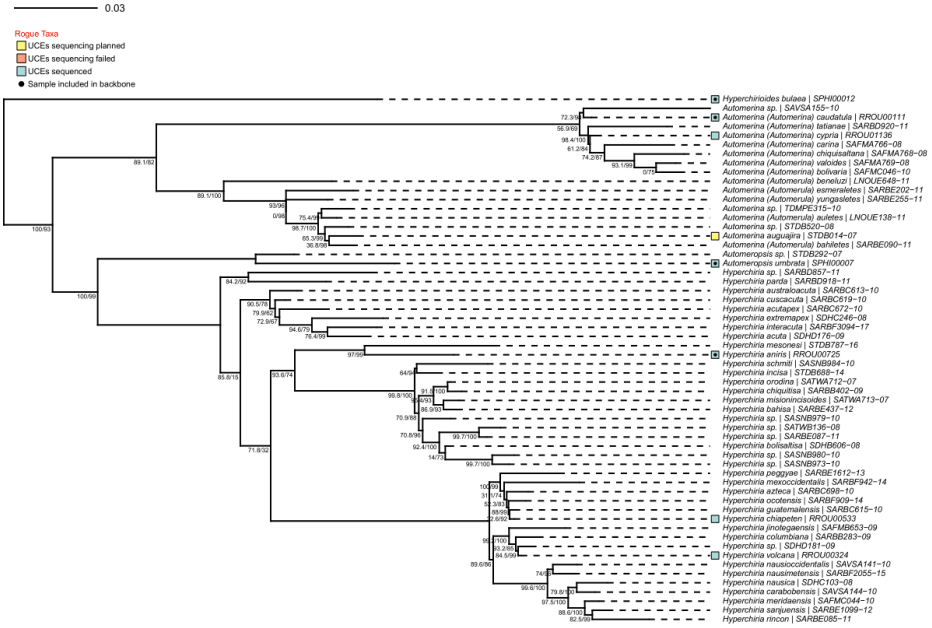


Automerina_D subtree

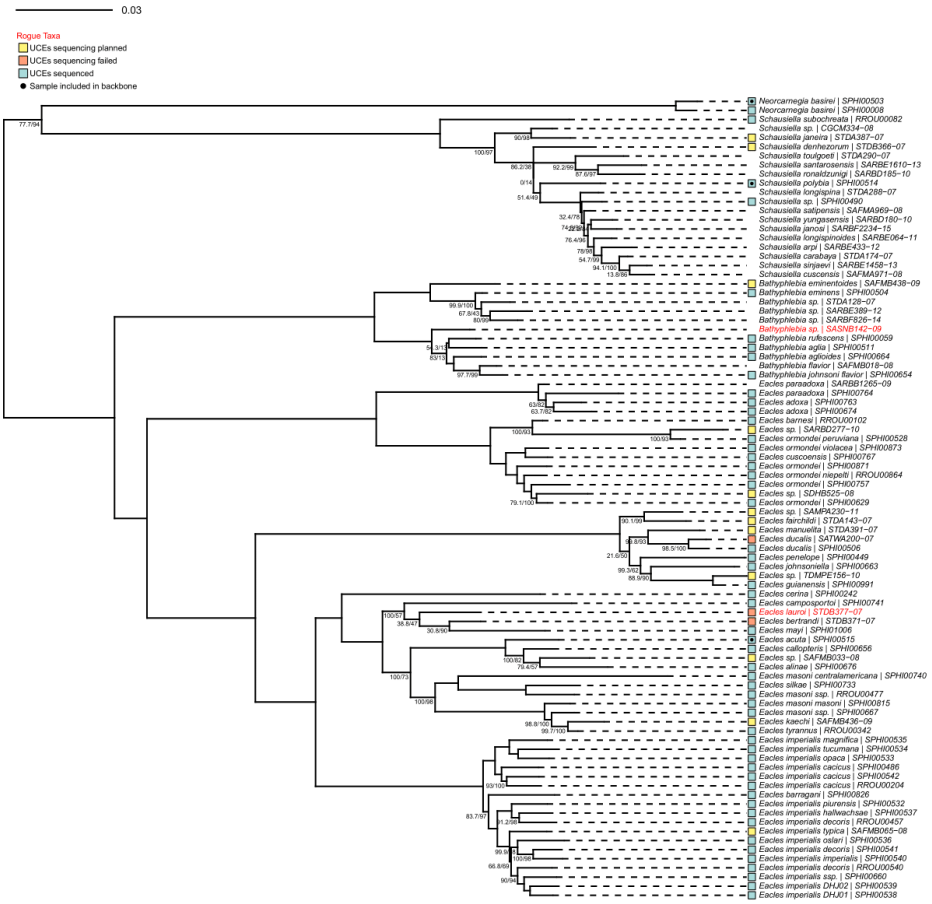



Hylesia (Hylesia) continua ssp. | RROU00499
Hylesia (Hylesia) colimatilex | SARBF113–13
Hylesia (Hylesia) yucatanex | SAFMB128–08
Hylesia (Hylesia) atlantex | SARBE1662–13
Hylesia nigricans | CGCM353–08
Hylesia (Hylesia) sp. | SACMA937–12
Hylesia (Hylesia) nigribahiex | SARBF1838–15
Hylesia (Hylesia) falcbahiex | SARBF149–13
Hylesia (Hylesia) falcifera | SAFMB329–08
Hylesia (Hylesia) falclapazex | SARBF1255–15
Hylesia (Hylesia) falccruzex | SAFMC359–11
Hylesia (Hylesia) sp. | SDHB024–07
Hylesia (Hylesia) falcatalayex | SARBF3105–17
Hylesia (Hylesia) sp. | SACMA938–12
Hylesia (Hylesia) falccamacarex | SARBF147–13
Hylesia (Hylesia) sp. | AMAZ383–12
Hylesia (Hylesia) remargbolivex | SAFMC366–11
Hylesia (Hylesia) rembahiex | SARBF137–13
Hylesia (Hylesia) remcarabobex | SARBF426–13
Hylesia (Hylesia) remandex | SARBF1349–15
Hylesia (Hylesia) bahiex | SARBF146–13
Hylesia (Hylesia) pauper | AMAZ417–12
Hylesia (Hylesia) sp. | AMAZ192–09
Hylesia (Hylesia) paupseptentridex | RROU00455
Hylesia (Hylesia) sp. | SARBF410–13
Hylesia (Hylesia) sp. | RROU00352
Hylesia sp. | SDHD047–09
Hylesia rubrifrons rubrifrons | SARBF125–13
Hylesia sp. | SACMA661–11
Hylesia sp. | SACMA671–11
Hylesia (Hylesia) maubahiex | SARBF1839–15
Hylesia (Hylesia) roseipichinchensis | SARBF251–13
Hylesia roseata | SARBF1292–15
Hylesia (Hylesia) rosehuarucoensis | SARBF246–13
Hylesia (Hylesia) peggyae | SARBF356–13
Hylesia munonia | CGCM354–08
Hylesia (Hylesia) sp. | SARBF297–13
Hylesia rex | SPHI00924
Hylesia (Hylesia) bolivarex | SAFMB339–08
Hylesia (Hylesia) satiprex | SDHB343–07
Hylesia (Hylesia) sp. | SARBF1287–15
Hylesia (Hylesia) andrex | SARBF397–13
Hylesia (Hylesia) sp. | SDHC095–08
Hylesia vialactea | SARBF620–14
Hylesia (Hylesia) yuyapichrex | SARBF1844–15
Hylesia (Hylesia) juprex | SARBF1307–15
Hylesia (Hylesia) metrex | SARBF1308–15
Hylesia (Hylesia) tatamex | SARBF2177–15
Hylesia coex | SARBF165–13
Hylesia (Hylesia) yarumalex | SARBF1706–15
Hylesia (Hylesia) sp. | SARBF2739–16
Hylesia sp. | SASNC2087–12
Hylesia ngridorsata | SARBF2350–16
Hylesia (Hylesia) sp. | SACMA657–11
Hylesia oratex | SACMA010–09
Hylesia (Hylesia) sp. | SACMA658–11
Hylesia (Hylesia) sp. | SACMA094–09
Hylesia (Hylesia) santboyacex | SARBF2080–15
Hylesia (Hylesia) andmeridex | SARBF1512–15
Hylesia (Hylesia) sp. NA | SPHI00925
Hylesia (Hylesia) canandex | SARBF368–13
Hylesia (Hylesia) yapiex | SARBF379–13
Hylesia (Hylesia) sp. | SARBF3355–16
Hylesia (Hylesia) tumbex | SAFMC059–10
Hylesia (Hylesia) paratumbex | SAFMC354–11
Hylesia (Hylesia) garrochex | SARBF1707–15
Hylesia (Hylesia) altaminex | SARBF237–13
Hylesia (Hylesia) lellex leilseptentridex | SARBF213–13
Hylesia zonex | SAFMB291–08
Hylesia (Hylesia) sp. | SDHB349–07
Hylesia (Hylesia) paroyex | SARBF1347–15
Hylesia (Hylesia) sp. | SACMA004–09
Hylesia (Hylesia) santacruzex santacruzex | SARBF1700–15
Hylesia (Hylesia) santacruzex amborex | SARBF1410–15
Hylesia sp. | SARBF1361–15
Hylesia (Hylesia) oxachanchex | SARBF1847–15
Hylesia (Hylesia) sucumbex | SARBF1303–15
Hylesia (Hylesia) napoex | SAVSA2895–14
Hylesia (Hylesia) guacamayex | SARBF1401–15
Hylesia (Hylesia) sp. | SARBF1868–15
Hylesia (Hylesia) sp. NA | RROU00325
Hylesia murex | SDHA560–07
Hylesia (Hylesia) sp. | SARBF209–13
Hylesia (Hylesia) angmotex | SARBF2292–16
Hylesia (Hylesia) sp. | SARBF1364–15
Hylesia (Hylesia) angcuscex | SARBF1190–15
Hylesia (Hylesia) sp. | SARBF235–13
Hylesia (Hylesia) faunalex | SARBF1712–15
Hylesia (Hylesia) mincex | SAMPA449–17
Hylesia bouvertii | SAFMB107–08
Hylesia vindex | SACMA002–09
Hylesia (Hylesia) sp. | SACMA014–09
Hylesia (Hylesia) chanchamayex | SARBF1869–15
Hylesia (Hylesia) andperuex | SARBF357–13
Hylesia (Hylesia) andbolivex | SARBF365–13
Hylesia (Hylesia) sp. | SARBF363–13
Hylesia (Hylesia) andcaucex andentioquiex | SARBF1466–15
Hylesia (Hylesia) andcaucex andcaucex | SARBF3220–17
Hylesia (Hylesia) andtabayex | SARBF2369–16
Hylesia andensis | SARBF361–13
Hylesia (Hylesia) andcajamarcex | SARBF364–13
Hylesia (Hylesia) sp. | SARBF1187–15
Hylesia (Hylesia) sp. | INCTA966–10
Hylesia (Hylesia) panguanex | SARBF1778–15
Hylesia (Hylesia) sp. | SDHB838–08
Hylesia (Hylesia) lempirex | SARBF454–13
Hylesia (Hylesia) piurex | SARBF382–13
Hylesia (Hylesia) sp. | SARBF2838–16
Hylesia (Hylesia) sp. | MIAPD181–10
Hylesia (Hylesia) pichinchex | SARBF1358–15
Hylesia (Hylesia) antioquiex | SARBF2511–16
Hylesia (Hylesia) alticoloides | SARBF1232–15
Hylesia sp. | SACMA249–09
Hylesia athlia athseptentridex | TDMPE115–09
Hylesia alticola | SDHB353–07
Hylesia (Hylesia) frigoaxacanex | SARBF119–13
Hylesia sp. | SDHB842–08
Hylesia (Hylesia) frigida frigida | SARBF1239–15
Hylesia (Hylesia) ascocuadorex | SARBF1362–15
Hylesia (Hylesia) sp. | SARBF1334–15
Hylesia (Hylesia) puyex | SARBF1776–15
Hylesia (Hylesia) ascpastaex | SARBF1985–15
Hylesia (Hylesia) ascucayalex | SARBF1273–15
Hylesia (Hylesia) ascclosex | SARBF314–13
Hylesia (Hylesia) ascolombex | SARBF1264–15
Hylesia (Hylesia) ascmoronex | SARBF1314–15
Hylesia sp. | INCTA965–10
Hylesia (Hylesia) asrguianex | SAFMB360–08
Hylesia (Hylesia) sp. | SARBF238–13
Hylesia (Hylesia) baguana | SAFMB361–08
Hylesia (Hylesia) sp. | SASNC2115–12
Hylesia (Hylesia) guianex tirolex | INCTB022–10
Hylesia (Hylesia) guianex guianex | SAVSA2856–14
Hylesia (Hylesia) moronex | SARBF1229–15
Hylesia (Hylesia) caloxapampex | SARBF1870–15
Hylesia (Hylesia) santarosex | SARBF1991–15
Hylesia (Hylesia) cearex | SARBF323–13
Hylesia (Hylesia) sp. | SARBF1445–15
Hylesia (Hylesia) pangmandonex | SARBF1779–15
Hylesia (Hylesia) coroicex | SARBF386–13
Hylesia (Hylesia) sanmartinex | SARBF384–13
Hylesia (Hylesia) lapazex | SARBF1701–15
Hylesia (Hylesia) compandex | SARBF1350–15
Hylesia (Hylesia) cajamartinex | SARBF383–13
Hylesia (Hylesia) sp. | SARBF2088–15
Hylesia (Hylesia) compsantandex | SARBF2085–15

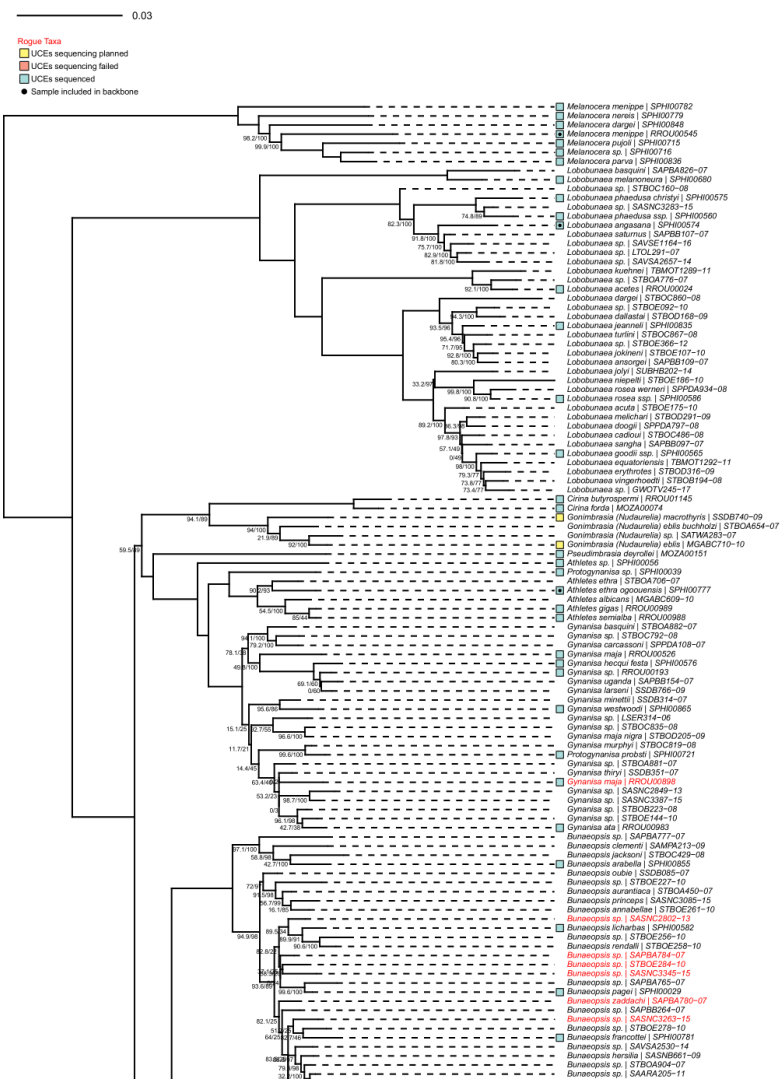
Automerina_E subtree

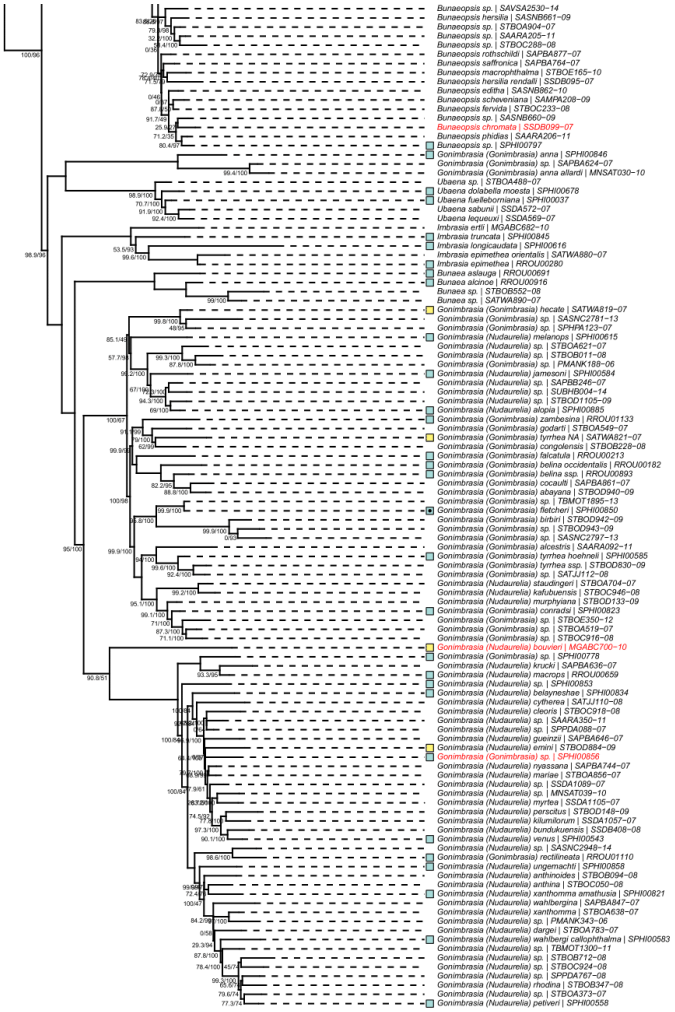


Bathyphlebiini subtree

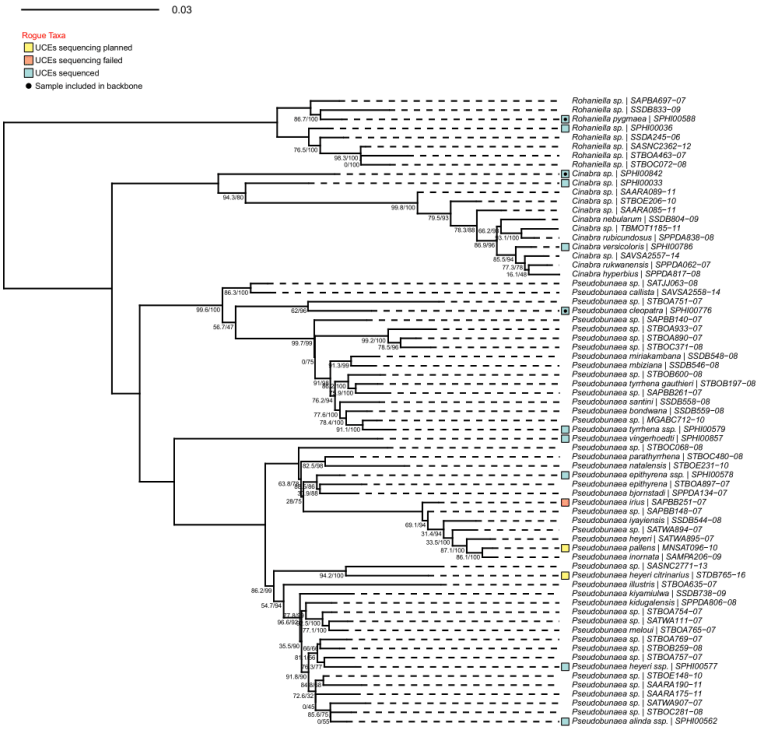


Bunaeini_A subtree

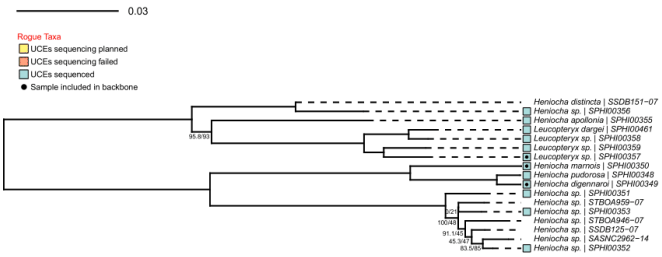



Bunaeopsis sp. | SAVSA2530~14
Bunaeopsis hersilia | SASNB661~09
Bunaeopsis sp. | STBOA904~07
Bunaeopsis sp. | SAARA205~11
Bunaeopsis sp. | STBOC288~08
Bunaeopsis rothschildi | SAPBA877~07
Bunaeopsis saffronica | SAFBA764~07
Bunaeopsis macrophthalma | STBOE165~10
Bunaeopsis hersilia rendalli | SSDB095~07
Bunaeopsis editha | SASNB962~10
Bunaeopsis scheveniana | SAMPA208~09
Bunaeopsis fervida | STBOC233~08
Bunaeopsis sp. | SASNB660~09
Bunaeopsis chromata | SSDB099~07
Bunaeopsis phidias | SAARA206~11
Bunaeopsis sp. | SPHI00797
Gonimbrasia (Gonimbrasia) anna | SPHI00646
Gonimbrasia (Gonimbrasia) sp. | SAPBA624~07
Gonimbrasia (Gonimbrasia) anna allardi | MNSAT030~10
Ubaena sp. | STBOA488~07
Ubaena dolabella moesta | SPHI00678
Ubaena fuelleborniana | SPHI00037
Ubaena sabunii | SSDA572~07
Ubaena lequeuxi | SSDA569~07
Imbrasia ertli | MGABC682~10
Imbrasia truncata | SPHI00845
Imbrasia longicaudata | SPHI00616
Imbrasia epimethea orientalis | SATWA880~07
Imbrasia epimethea | RROU00280
Bunaea alsiaugi | RROU00691
Bunaea alcinoe | RROU00916
Bunaea sp. | STBOB552~08
Bunaea sp. | SATWA890~07
Gonimbrasia (Gonimbrasia) hecate | SATWA819~07
Gonimbrasia (Gonimbrasia) sp. | SASNC2781~13
Gonimbrasia (Gonimbrasia) sp. | SPHPA123~07
Gonimbrasia (Nudaurelia) melanops | SPHI00615
Gonimbrasia (Nudaurelia) sp. | STBOA621~07
Gonimbrasia (Nudaurelia) sp. | STBOB011~08
Gonimbrasia (Gonimbrasia) sp. | PMANK188~06
Gonimbrasia (Nudaurelia) jamesoni | SPHI00584
Gonimbrasia (Nudaurelia) sp. | SAPBB246~07
Gonimbrasia (Nudaurelia) sp. | SUBHB004~14
Gonimbrasia (Nudaurelia) sp. | STBOD1105~09
Gonimbrasia (Nudaurelia) alopia | SPHI00885
Gonimbrasia (Gonimbrasia) zambesina | RROU01133
Gonimbrasia (Gonimbrasia) godarti | STBOA549~07
Gonimbrasia (Gonimbrasia) tyrrhea NA | SATWA821~07
Gonimbrasia (Gonimbrasia) congolensis | STBOB228~08
Gonimbrasia (Gonimbrasia) falcatula | RROU00213
Gonimbrasia (Gonimbrasia) belina occidentalis | RROU00182
Gonimbrasia (Gonimbrasia) belina sagi | RROU00893
Gonimbrasia (Gonimbrasia) cocaulti | SAPBA861~07
Gonimbrasia (Gonimbrasia) abayana | STBOD940~09
Gonimbrasia (Gonimbrasia) sp. | TBMOT1895~13
Gonimbrasia (Gonimbrasia) fletcheri | SPHI00850
Gonimbrasia (Gonimbrasia) biribiri | STBOD942~09
Gonimbrasia (Gonimbrasia) sp. | STBOD943~09
Gonimbrasia (Gonimbrasia) sp. | SASNC2797~13
Gonimbrasia (Gonimbrasia) alcestris | SAARA092~11
Gonimbrasia (Gonimbrasia) tyrrhea hoehnelii | SPHI00585
Gonimbrasia (Gonimbrasia) tyrrhea ssp. | STBOD830~09
Gonimbrasia (Gonimbrasia) sp. | SATJJ112~06
Gonimbrasia (Nudaurelia) staudingeri | STBOA704~07
Gonimbrasia (Nudaurelia) kafubuensis | STBOC946~08
Gonimbrasia (Nudaurelia) murphyiana | STBOD133~09
Gonimbrasia (Gonimbrasia) conradsi | SPHI00823
Gonimbrasia (Gonimbrasia) sp. | STBOE350~12
Gonimbrasia (Gonimbrasia) sp. | STBOA519~07
Gonimbrasia (Gonimbrasia) sp. | STBOC916~08
Gonimbrasia (Nudaurelia) bouvieri | MGABC700~10
Gonimbrasia (Gonimbrasia) sp. | SPHI00778
Gonimbrasia (Nudaurelia) krucki | SAPBA636~07
Gonimbrasia (Nudaurelia) macrops | RROU00659
Gonimbrasia (Nudaurelia) sp. | SPHI00853
Gonimbrasia (Nudaurelia) belayneshae | SPHI00834
Gonimbrasia (Nudaurelia) cytherea | SATJJ110~08
Gonimbrasia (Nudaurelia) cleoris | STBOC918~08
Gonimbrasia (Nudaurelia) sp. | SAARA350~11
Gonimbrasia (Nudaurelia) sp. | SPPDA098~07
Gonimbrasia (Nudaurelia) guenzii | SAPBA646~07
Gonimbrasia (Nudaurelia) ernini | STBOD884~09
Gonimbrasia (Gonimbrasia) sp. | SPHI00856
Gonimbrasia (Nudaurelia) nyassana | SAPBA744~07
Gonimbrasia (Nudaurelia) mariae | STBOA856~07
Gonimbrasia (Nudaurelia) sp. | SSDA1089~07
Gonimbrasia (Nudaurelia) sp. | MNSAT039~10
Gonimbrasia (Nudaurelia) myrtea | SSDA1105~07
Gonimbrasia (Nudaurelia) persoltus | STBOD148~09
Gonimbrasia (Nudaurelia) kilumiorum | SSDA1057~07
Gonimbrasia (Nudaurelia) bundukuensis | SSDB408~08
Gonimbrasia (Nudaurelia) venus | SPHI00543
Gonimbrasia (Nudaurelia) sp. | SASNC2949~14
Gonimbrasia (Gonimbrasia) rectilineata | RROU01110
Gonimbrasia (Nudaurelia) ungemachti | SPHI00858
Gonimbrasia (Nudaurelia) anthinoides | STBOB094~08
Gonimbrasia (Nudaurelia) anthina | STBOC050~08
Gonimbrasia (Nudaurelia) xanthomma amathusia | SPHI00821
Gonimbrasia (Nudaurelia) wahlbergina | SAPBA847~07
Gonimbrasia (Nudaurelia) xanthomma | STBOA638~07
Gonimbrasia (Nudaurelia) sp. | PMANK343~06
Gonimbrasia (Nudaurelia) dargei | STBOA783~07
Gonimbrasia (Nudaurelia) wahlbergi callophthalma | SPHI00583
Gonimbrasia (Nudaurelia) sp. | TBMOT1300~11
Gonimbrasia (Nudaurelia) sp. | STBOB712~08
Gonimbrasia (Nudaurelia) sp. | STBOC924~08
Gonimbrasia (Nudaurelia) sp. | SPPDA767~08
Gonimbrasia (Nudaurelia) rhodina | STBOB347~08
Gonimbrasia (Nudaurelia) sp. | STBOA373~07
Gonimbrasia (Nudaurelia) petiveri | SPHI00558

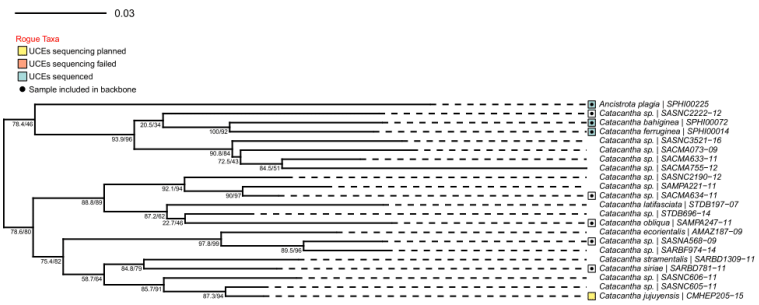
Bunaeini_B subtree



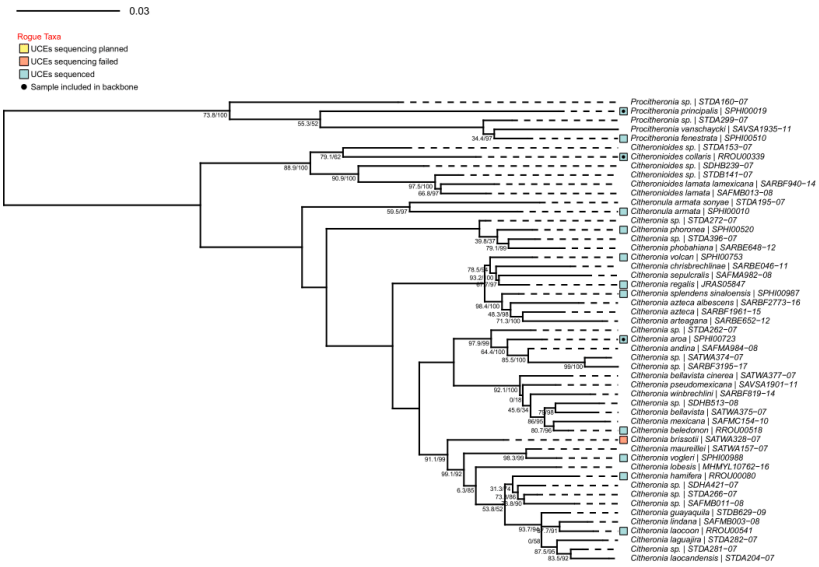
Bunaeini_C subtree



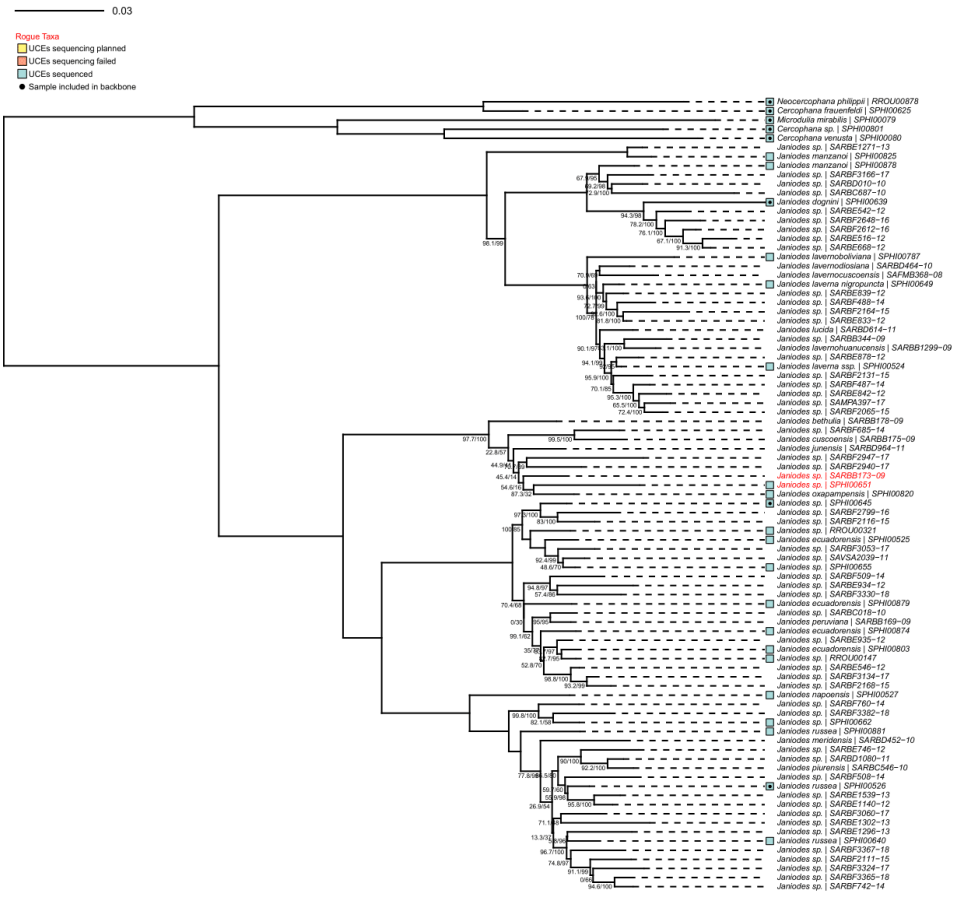
Catacantha subtree



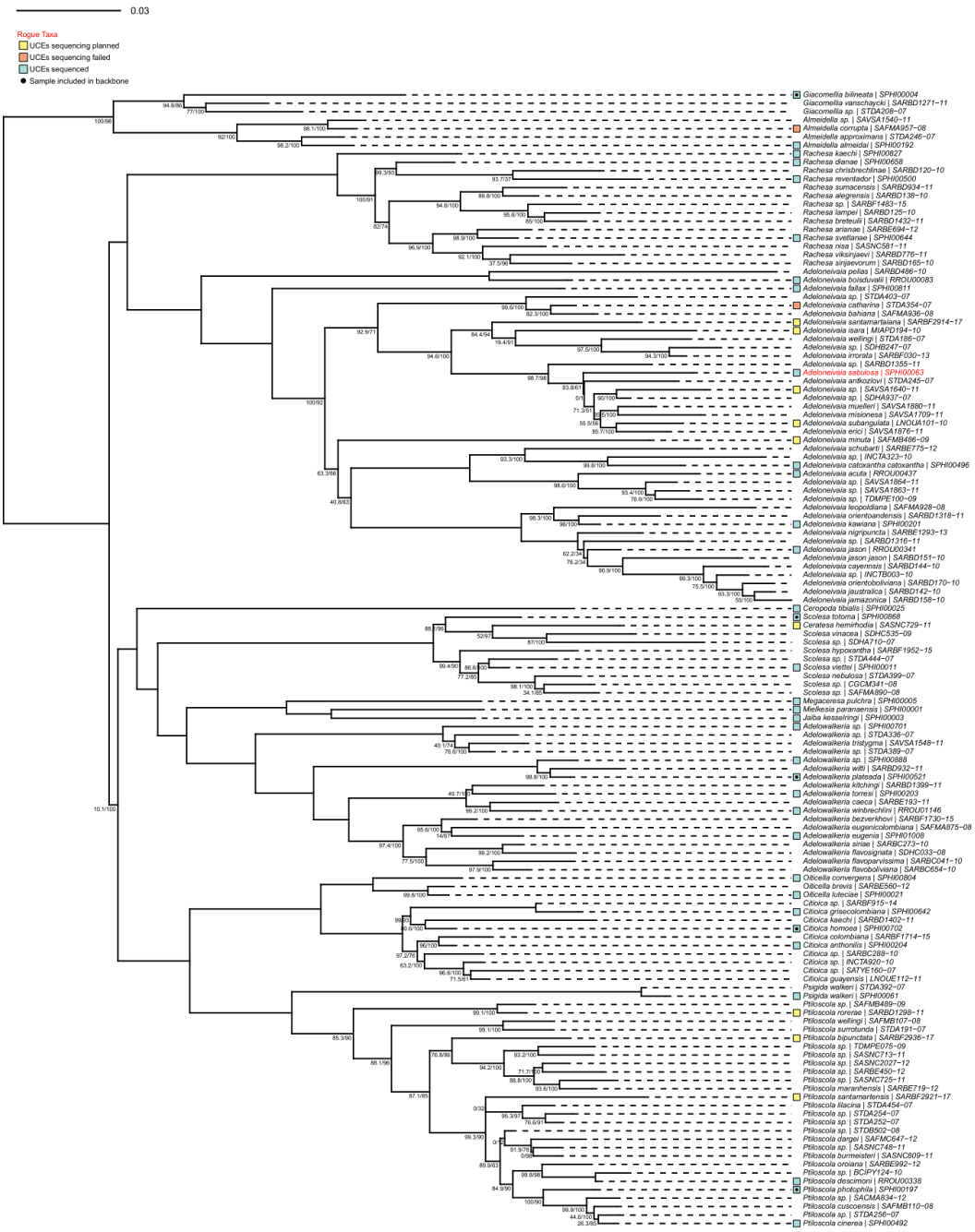
Ceratocampini subtree



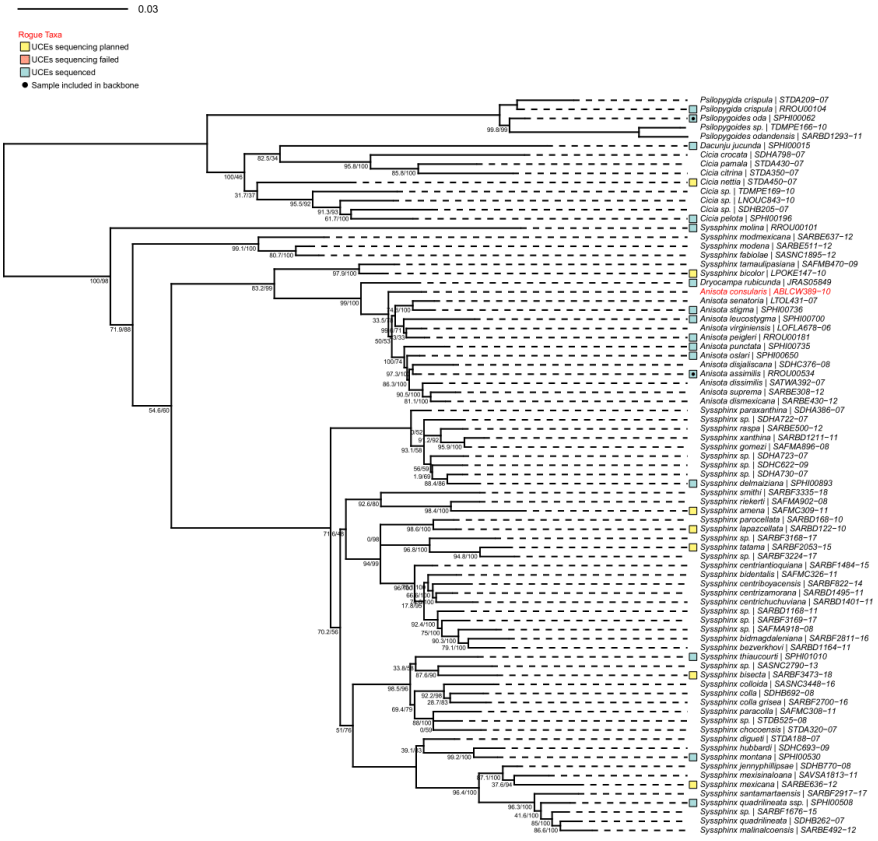
Cercophaninae subtree



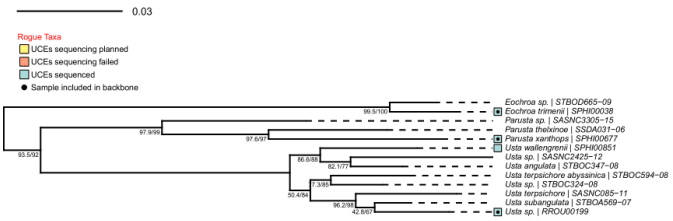
Dryocampini_A subtree



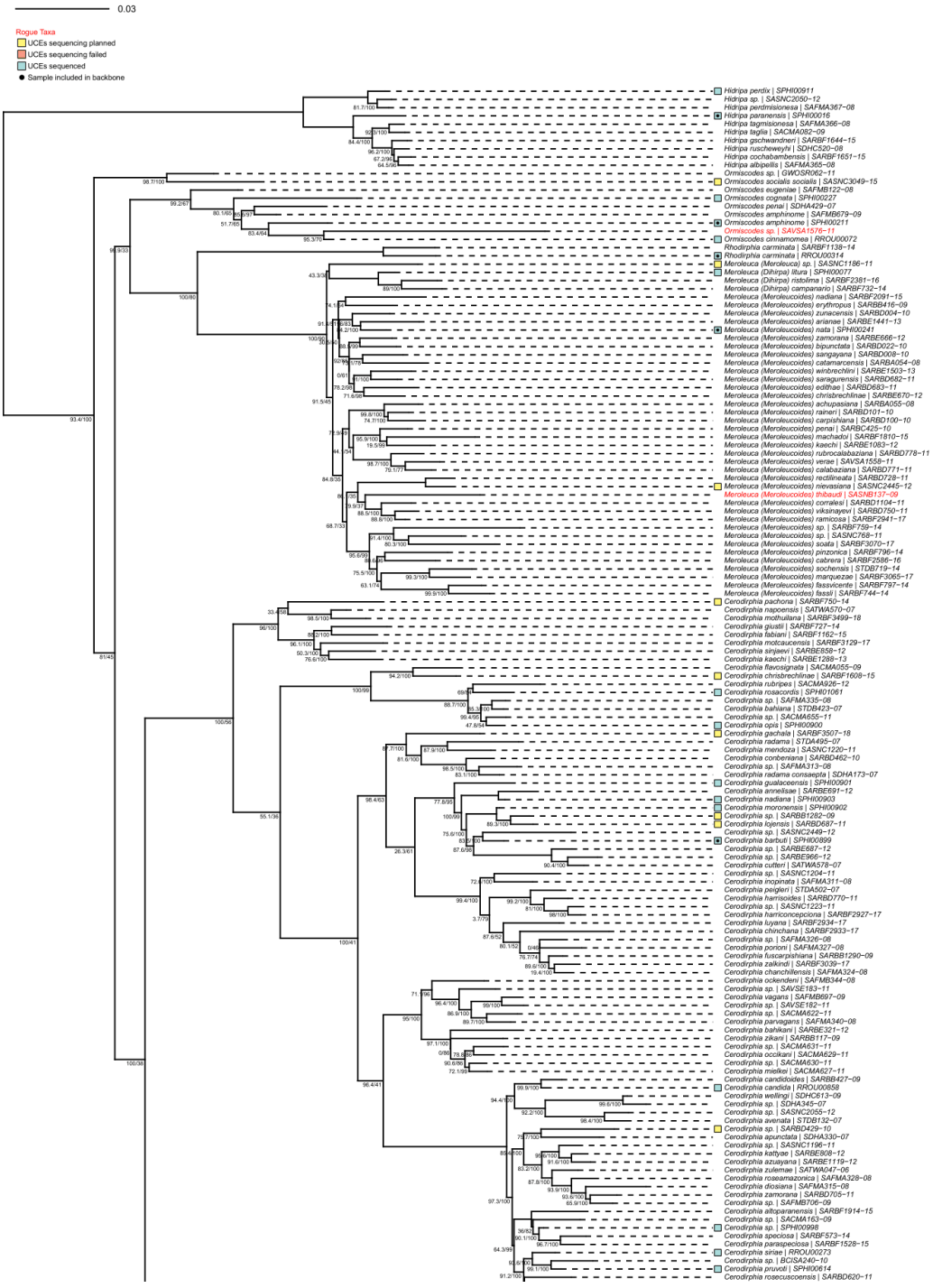
Dryocampini_B subtree

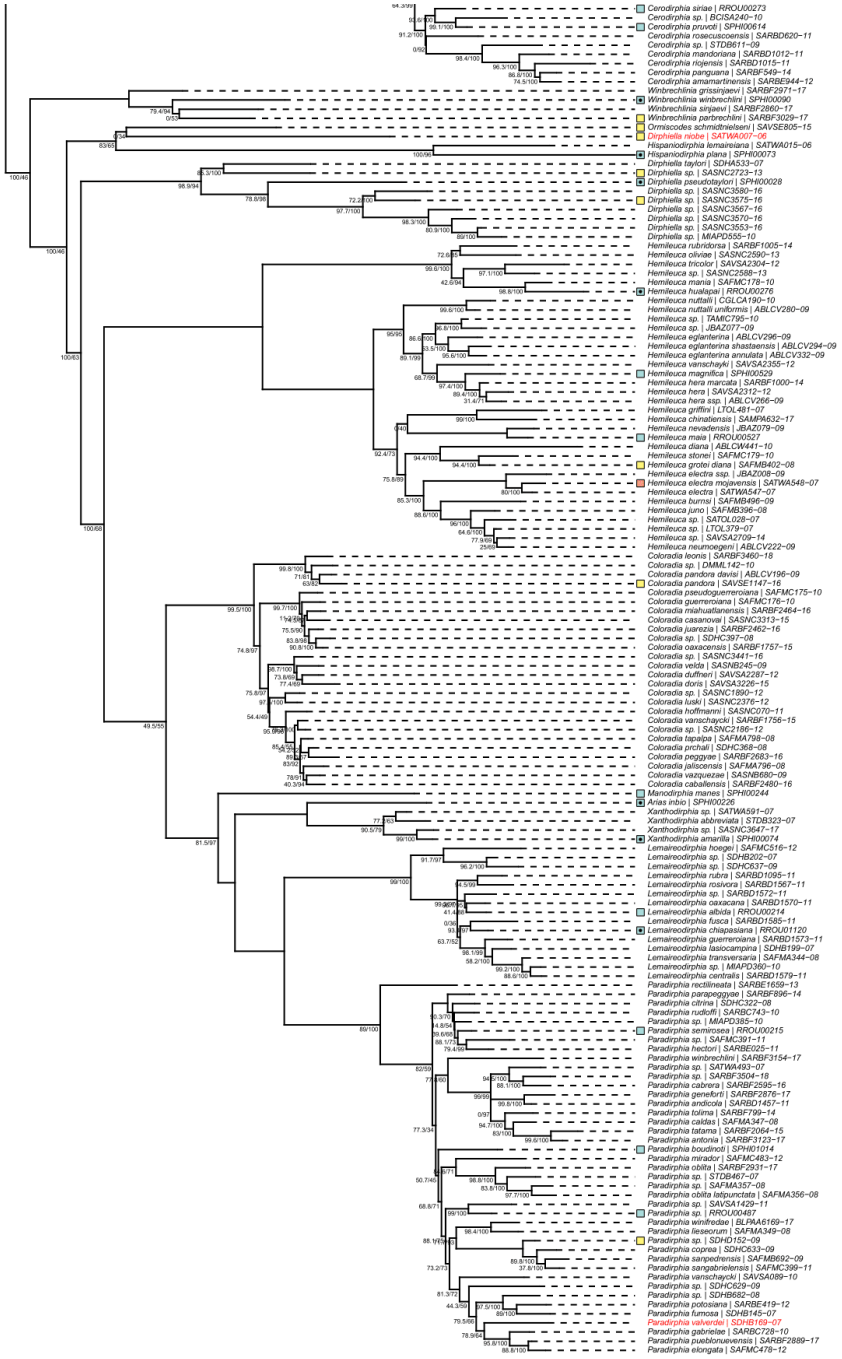


Eochroini subtree

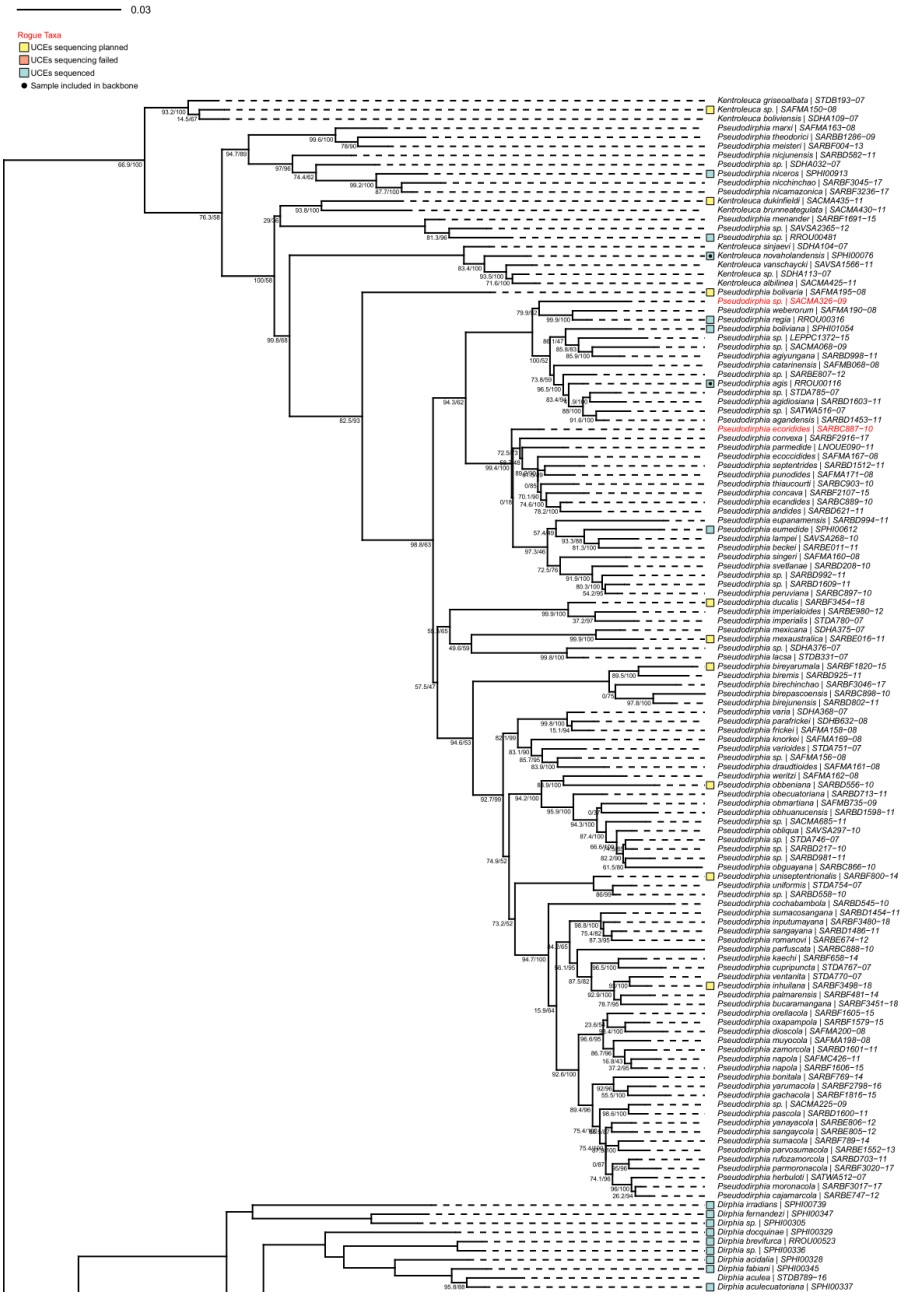


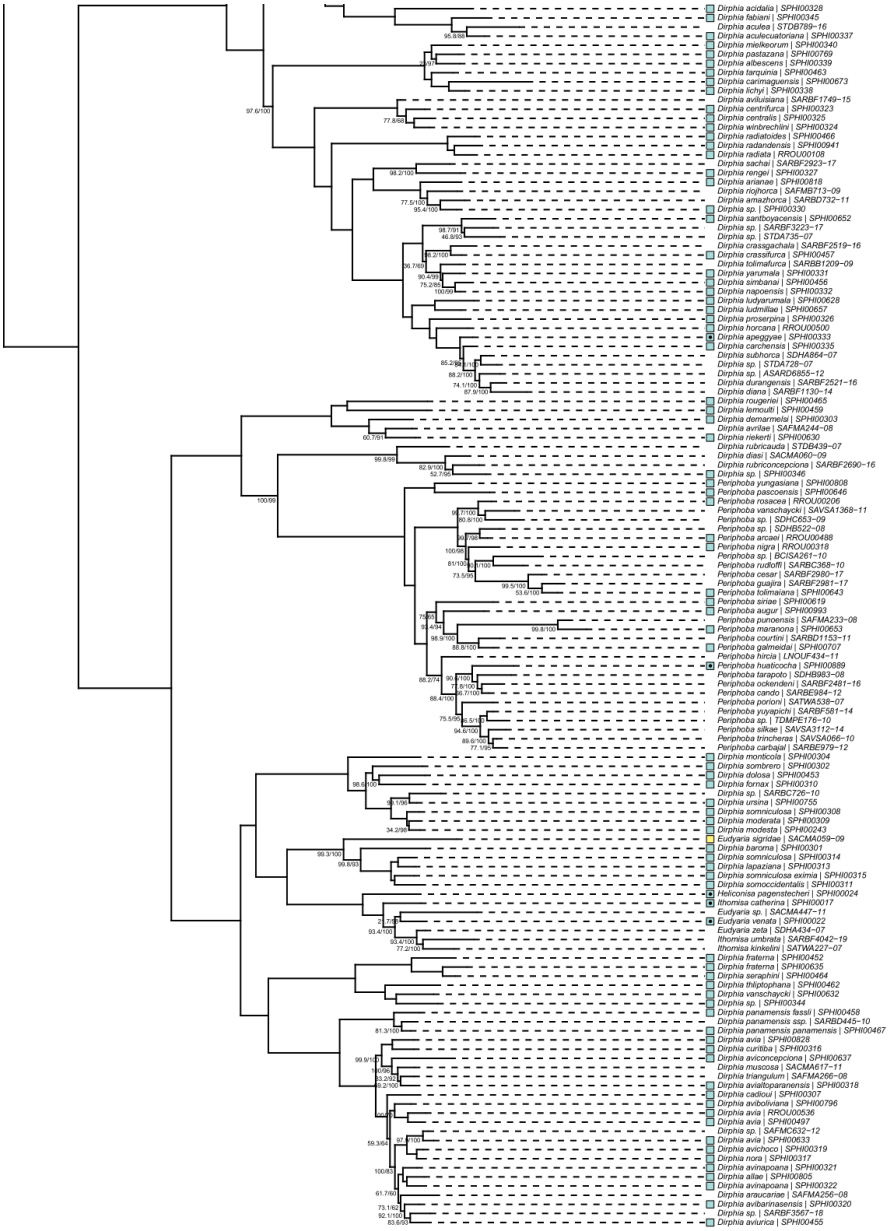
Hemileucina_A subtree



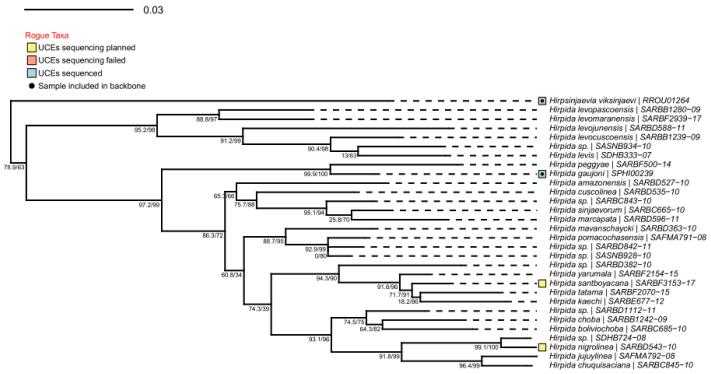

Cerodirphia sirilae | RROU00273
Cerodirphia sp. | BCISA240−10
Cerodirphia pruvoti | SPHI00614
Cerodirphia rosecuscoensis | SARBD620−11
Cerodirphia sp. | STDB611−09
Cerodirphia mandoriana | SARBD1012−11
Cerodirphia rogersis | SARBD1015−11
Cerodirphia panguana | SARBF549−14
Cerodirphia amamartinensis | SARBE944−12
Winbrechlinia grissinjaevi | SARBF2971−17
Winbrechlinia winbrechlini | SPHI00090
Winbrechlinia sinjaevi | SARBF2860−17
Winbrechlinia parbrechlini | SARBF3029−17
Ormiscodes schmidtrielseni | SAVSE805−15
Dirphiella nobe | SATWA007−06
Hispaniodirphia lemairiana | SATWA015−06
Hispaniodirphia plana | SPHI00073
Dirphiella taylori | SDHA533−07
Dirphiella sp. | SASNC2723−13
Dirphiella pseudotaylori | SPHI00028
Dirphiella sp. | SASNC3580−16
Dirphiella sp. | SASNC3575−16
Dirphiella sp. | SASNC3567−16
Dirphiella sp. | SASNC3570−16
Dirphiella sp. | SASNC3553−16
Dirphiella sp. | MIAPD355−10
Hemileuca rubridorsa | SARBF1005−14
Hemileuca oliviae | SASNC2590−13
Hemileuca tricolor | SAVSA2304−12
Hemileuca sp. | SASNC2586−13
Hemileuca mania | SAFMC178−10
Hemileuca hualapai | RROU00276
Hemileuca nuttalli | CGLCA190−10
Hemileuca nuttalli uniformis | ABLCV280−09
Hemileuca sp. | TAMIC795−10
Hemileuca sp. | JBAZ077−09
Hemileuca eglanterina | ABLCV296−09
Hemileuca eglanterina shastaensis | ABLCV294−09
Hemileuca eglanterina annulata | ABLCV332−09
Hemileuca varnchayki | SAVSA2355−12
Hemileuca magnifica | SPHI00529
Hemileuca hera marcata | SARBF1000−14
Hemileuca hera | SAVSA2312−12
Hemileuca hera ssp. | ABLCV266−09
Hemileuca griffini | LTOL481−07
Hemileuca chinatiensis | SAMPA632−17
Hemileuca nevadensis | JBAZ079−09
Hemileuca maia | RROU00527
Hemileuca diana | ABLCW441−10
Hemileuca storei | SAFMC179−10
Hemileuca grotei diana | SAFMB402−08
Hemileuca electra ssp. | JBAZ008−09
Hemileuca electra mojavensis | SATWA548−07
Hemileuca electra | SATWA547−07
Hemileuca burnsi | SAFMB496−09
Hemileuca juno | SAFMB396−08
Hemileuca sp. | SATOL028−07
Hemileuca sp. | LTOL379−07
Hemileuca sp. | SAVSA2709−14
Hemileuca neumoegeni | ABLCV222−09
Coloradia leonis | SARBF3460−18
Coloradia sp. | DMML142−10
Coloradia pandora davisi | ABLCV196−09
Coloradia pandora | SAVSE1147−16
Coloradia pseudoguerreroiana | SAFMC175−10
Coloradia guerreroiana | SAFMC176−10
Coloradia mlahuatlanensis | SARBF2464−16
Coloradia casanovai | SASNC3313−15
Coloradia juarezia | SARBF2462−16
Coloradia sp. | SDHC397−08
Coloradia oaxacensis | SARBF1757−15
Coloradia sp. | SASNC3441−16
Coloradia velda | SMSNB245−09
Coloradia duffneri | SAVSA2287−12
Coloradia doris | SAVSA3226−15
Coloradia sp. | SASNC1890−12
Coloradia luski | SASNC2376−12
Coloradia hoffmanni | SASNC070−11
Coloradia vanschaycki | SARBF1756−15
Coloradia sp. | SASNC2186−12
Coloradia tapalpa | SAFMA798−08
Coloradia prchali | SDHC368−08
Coloradia peggyae | SARBF2683−16
Coloradia jaliscensis | SAFMA796−08
Coloradia vazquezae | SASNB680−09
Coloradia caballensis | SARBF2480−16
Manodirphia manes | SPHI00244
Arias inbio | SPHI00226
Xanthodirphia sp. | SATWA591−07
Xanthodirphia abbreviata | STDB323−07
Xanthodirphia sp. | SASNC3647−17
Xanthodirphia amarilla | SPHI00074
Lemairеodirphia hoegei | SAFMC316−12
Lemairеodirphia sp. | SDHB202−07
Lemairеodirphia sp. | SDHC637−09
Lemairеodirphia rubra | SARBD1095−11
Lemairеodirphia rosivora | SARBD1567−11
Lemairеodirphia sp. | SARBD1572−11
Lemairеodirphia oaxacana | SARBD1570−11
Lemairеodirphia albida | RROU00214
Lemairеodirphia fusca | SARBD1585−11
Lemairеodirphia chiapasiana | RROU01120
Lemairеodirphia guerreroiana | SARBD1573−11
Lemairеodirphia lasiocampina | SDHB199−07
Lemairеodirphia transversaria | SAFMA344−08
Lemairеodirphia sp. | MIAPD360−10
Lemairеodirphia centralis | SARBD1579−11
Paradirphia rectilineata | SARBE1659−13
Paradirphia parapeggyae | SARBF896−14
Paradirphia citrina | SDHC332−08
Paradirphia rudloffi | SARBC743−10
Paradirphia sp. | MIAPD385−10
Paradirphia semirosea | RROU00215
Paradirphia sp. | SAFMC391−11
Paradirphia hectori | SARBE025−11
Paradirphia winbrechlini | SARBF3154−17
Paradirphia sp. | SATWA493−07
Paradirphia sp. | SARBF3504−18
Paradirphia cabrera | SARBF2595−16
Paradirphia geneforti | SARBF2876−17
Paradirphia andicola | SARBD1457−11
Paradirphia tolima | SARBF799−14
Paradirphia caldas | SAFMA347−08
Paradirphia tatama | SARBF2064−15
Paradirphia antonia | SARBF3123−17
Paradirphia boudinoti | SPHI01014
Paradirphia mirador | SAFMC483−12
Paradirphia oblita | SARBF2931−17
Paradirphia sp. | STDB467−07
Paradirphia sp. | SAFMA357−08
Paradirphia oblita latipunctata | SAFMA356−08
Paradirphia sp. | SAVSA1429−11
Paradirphia sp. | RROU00497
Paradirphia winifredae | BLPAA6169−17
Paradirphia lieseorum | SAFMA349−08
Paradirphia sp. | SDHD152−09
Paradirphia coprea | SDHC633−09
Paradirphia sanpedrensis | SAFMB692−09
Paradirphia sangabrielensis | SAFMC399−11
Paradirphia vanschaycki | SAVSA089−10
Paradirphia sp. | SDHC629−09
Paradirphia sp. | SDHB682−08
Paradirphia potosiana | SARBE419−12
Paradirphia fumosa | SDHB145−07
Paradirphia valverdei | SDHB169−07
Paradirphia gabrielae | SARBC728−10
Paradirphia puebloruevensis | SARBF2889−17
Paradirphia elongata | SAFMC478−12

Hemileucina_B subtree

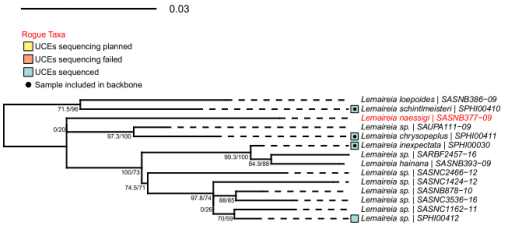



Dirphia acidalia | SPHI00328
Dirphia fabiani | SPHI00345
Dirphia aculea | STDB789–16
Dirphia aculicuatoriana | SPHI00337
Dirphia mielkeorum | SPHI00340
Dirphia pastazana | SPHI00769
Dirphia albescens | SPHI00339
Dirphia tarquinia | SPHI00463
Dirphia carimaguensis | SPHI00673
Dirphia lichyi | SPHI00338
Dirphia aviluisiana | SARBF1749–15
Dirphia centrifurca | SPHI00323
Dirphia centralis | SPHI00325
Dirphia winbrechlini | SPHI00324
Dirphia radiatoides | SPHI00466
Dirphia radandensis | SPHI00941
Dirphia radiata | RROU00108
Dirphia sachai | SARBF2923–17
Dirphia rengei | SPHI00327
Dirphia arianae | SPHI00818
Dirphia riojhorca | SAFMB713–09
Dirphia amazhorca | SARBD732–11
Dirphia sp. | SPHI00330
Dirphia saritboyacensis | SPHI00652
Dirphia sp. | SARBF3223–17
Dirphia sp. | STDA735–07
Dirphia crassgachala | SARBF2519–16
Dirphia crassifurca | SPHI00457
Dirphia tolimafurca | SARBB1209–09
Dirphia yarumala | SPHI00331
Dirphia simbanai | SPHI00456
Dirphia napoensis | SPHI00332
Dirphia ludyarumala | SPHI00628
Dirphia ludmillae | SPHI00657
Dirphia proserpina | SPHI00326
Dirphia horcana | RROU00500
Dirphia apeggyae | SPHI00333
Dirphia carchensis | SPHI00335
Dirphia subhorca | SDHA864–07
Dirphia sp. | STDA728–07
Dirphia sp. | ASARD6855–12
Dirphia durangensis | SARBF2521–16
Dirphia diana | SARBF1130–14
Dirphia rougeriei | SPHI00465
Dirphia lemoulti | SPHI00459
Dirphia demarmela | SPHI00303
Dirphia avrilae | SAFMA244–08
Dirphia riekerti | SPHI00630
Dirphia rubricauda | STDB439–07
Dirphia diasi | SACMA060–09
Dirphia rubriconcepcionis | SARBF2690–16
Dirphia sp. | SPHI00346
Periphoba yungasiana | SPHI00808
Periphoba pascoensis | SPHI00646
Periphoba rosacea | RROU00206
Periphoba vanschaycki | SAVSA1368–11
Periphoba sp. | SDHC653–09
Periphoba sp. | SDHB522–08
Periphoba arcaei | RROU00488
Periphoba nigra | RROU00318
Periphoba sp. | BCISA261–10
Periphoba rudloffi | SARBC368–10
Periphoba cesar | SARBF2980–17
Periphoba guajira | SARBF2981–17
Periphoba tolimaiana | SPHI00643
Periphoba siriae | SPHI00619
Periphoba augur | SPHI00993
Periphoba punoensis | SAFMA233–08
Periphoba maranona | SPHI00653
Periphoba courtini | SARBD1153–11
Periphoba galmeidai | SPHI00707
Periphoba hircia | LNOUF434–11
Periphoba huatiicocha | SPHI00889
Periphoba tarapoto | SDHB983–08
Periphoba ockenderi | SARBF2481–16
Periphoba cando | SARBE984–12
Periphoba porioni | SATWA538–07
Periphoba yuyapichi | SARBF581–14
Periphoba sp. | TDMPE176–10
Periphoba silkae | SAVSA3112–14
Periphoba trincheras | SAVSA066–10
Periphoba carbajai | SARBE979–12
Dirphia monticola | SPHI00304
Dirphia sombrero | SPHI00302
Dirphia dolosa | SPHI00453
Dirphia fornax | SPHI00310
Dirphia sp. | SARBC726–10
Dirphia ursina | SPHI00755
Dirphia somniculosa | SPHI00308
Dirphia moderata | SPHI00309
Dirphia modesta | SPHI00243
Eudyaria sigridae | SACMA059–09
Dirphia baroma | SPHI00301
Dirphia somniculosa | SPHI00314
Dirphia lapaziana | SPHI00313
Dirphia somniculosa eximia | SPHI00315
Dirphia somoccidentalis | SPHI00311
Heliconisa pagenstecheri | SPHI00024
Ithomisa catherina | SPHI00017
Eudyaria sp. | SACMA447–11
Eudyaria venata | SPHI00322
Eudyaria zeta | SDHA434–07
Ithomisa umbrata | SARBF4042–19
Ithomisa kinkelini | SATWA227–07
Dirphia fraterna | SPHI00452
Dirphia fraterna | SPHI00635
Dirphia seraphini | SPHI00464
Dirphia thilptophana | SPHI00462
Dirphia vanschaycki | SPHI00632
Dirphia sp. | SPHI03044
Dirphia panamensis fassli | SPHI00458
Dirphia panamensis ssp. | SARBD445–10
Dirphia panamensis panamensis | SPHI00467
Dirphia avia | SPHI00828
Dirphia curitiba | SPHI00316
Dirphia aviconcepcionai | SPHI00637
Dirphia muscosa | SACMA617–11
Dirphia triangulum | SAFMA256–08
Dirphia avialtoparanensis | SPHI00318
Dirphia cadioui | SPHI00307
Dirphia avibolivianai | SPHI00796
Dirphia avia | RROU00536
Dirphia avia | SPHI00497
Dirphia sp. | SAFMC632–12
Dirphia avia | SPHI00633
Dirphia avichoco | SPHI00319
Dirphia nora | SPHI00317
Dirphia avinapoana | SPHI00321
Dirphia allae | SPHI00805
Dirphia avinapoana | SPHI00322
Dirphia araucariae | SAFMA256–08
Dirphia avibarinasensis | SPHI00320
Dirphia sp. | SARBF3567–16
Dirphia aviurica | SPHI00455

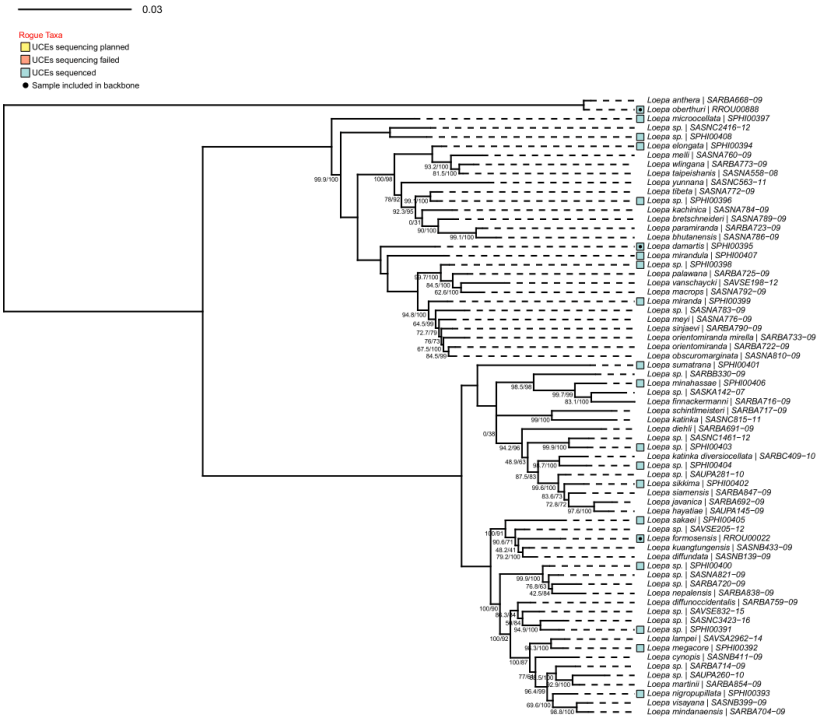
Hirpidinae subtree

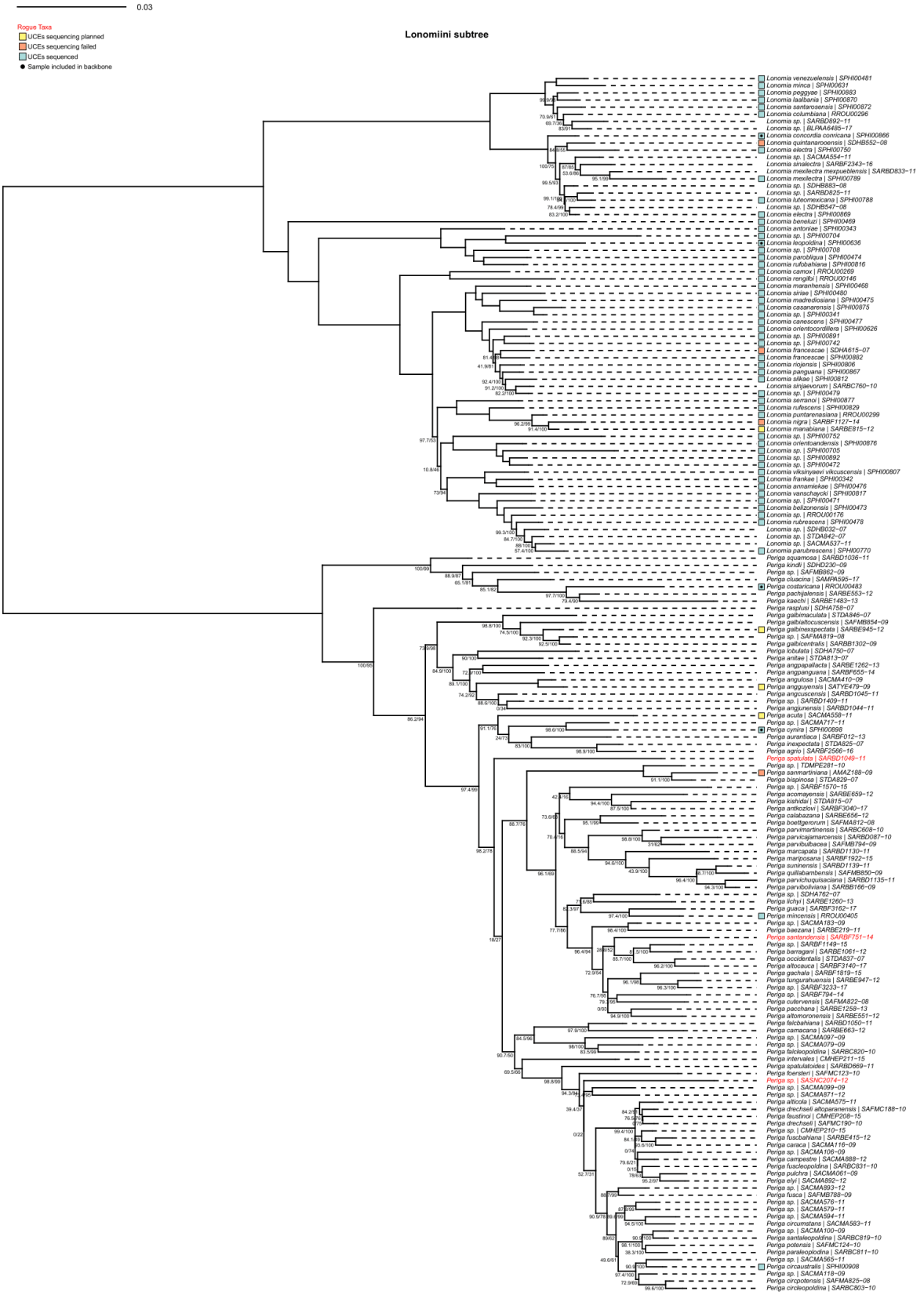


Lemaireia subtree

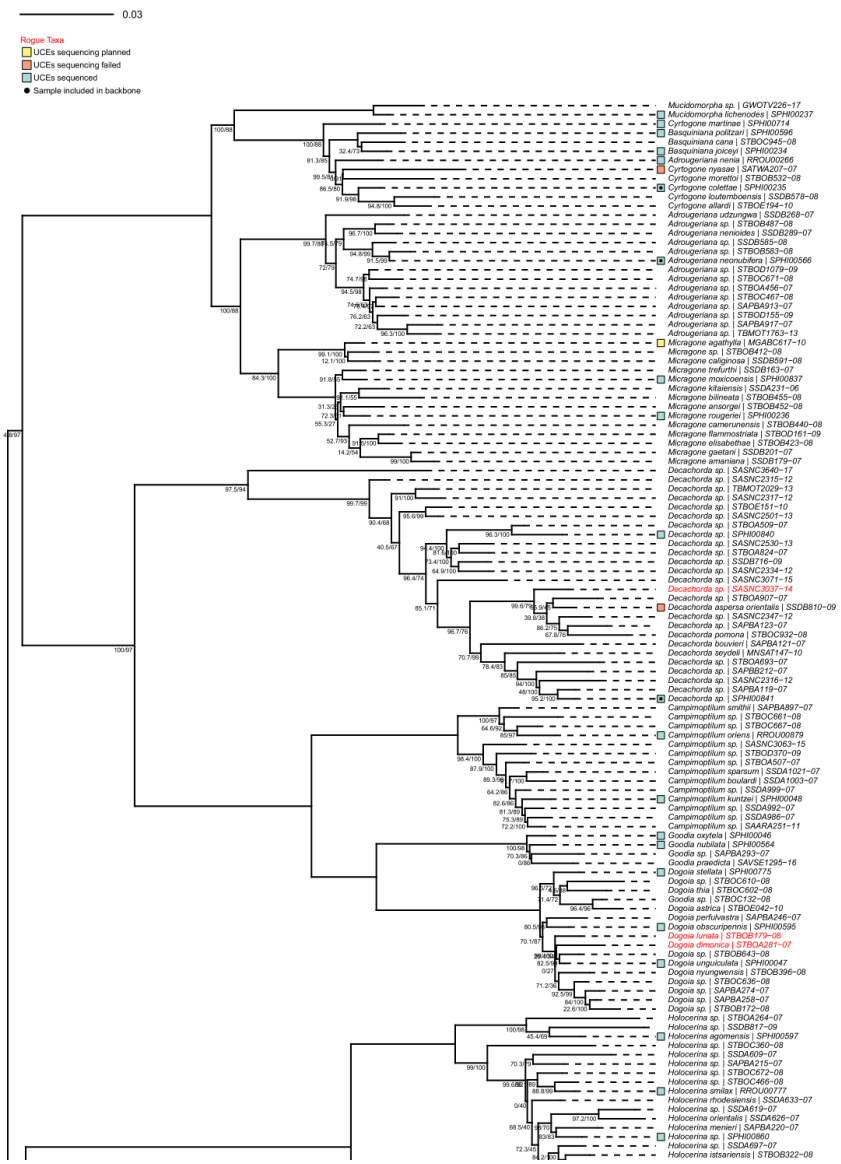


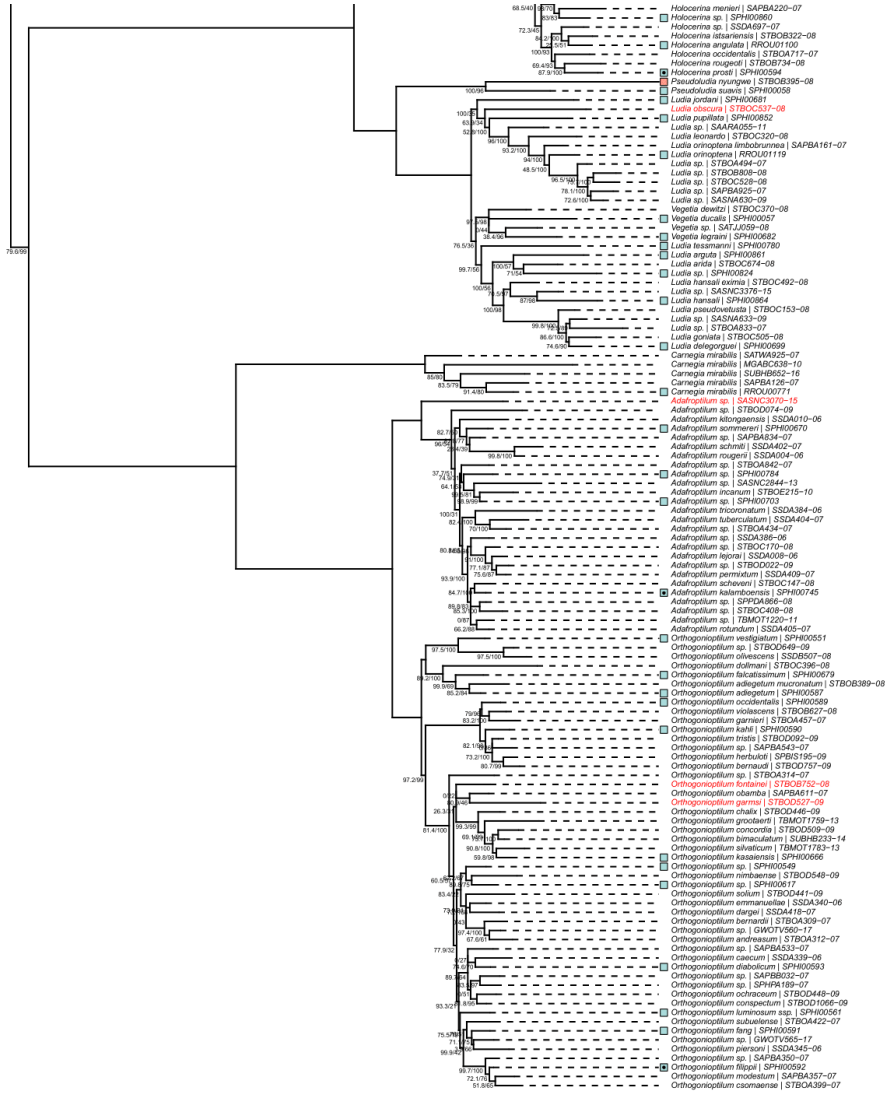
Loepa subtree



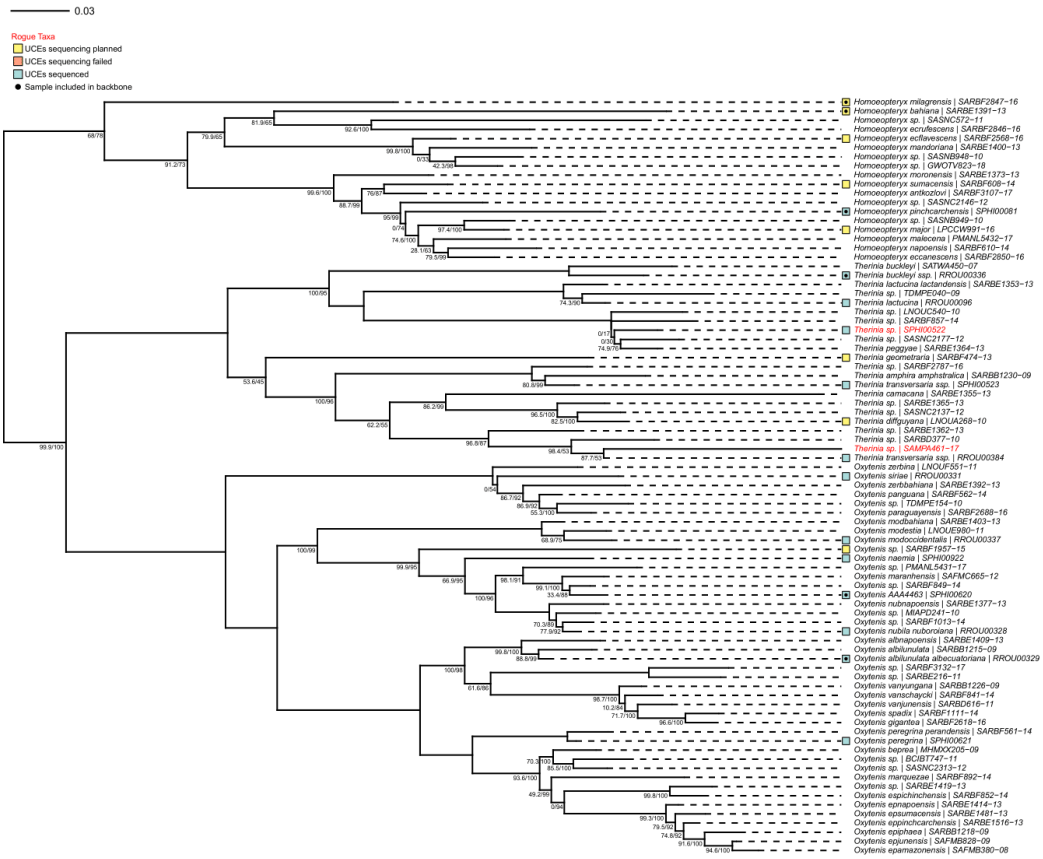

0.03
Rogue Taxa
UCEs sequencing planned
UCEs sequencing failed
UCEs sequenced
Sample included in backbone
Lonomiini subtree
Lonomia venezuelensis | SPHI00481
Lonomia minca | SPHI00631
Lonomia peggyae | SPHI00883
Lonomia laalbaria | SPHI00870
Lonomia santarosensis | SPHI00872
Lonomia columbiana | RROU00296
Lonomia sp. | SARBD892–11
Lonomia sp. | BLPAA6485–17
Lonomia concordia conricana | SPHI00866
Lonomia quintanaroensis | SDHB552–08
Lonomia electra | SPHI00750
Lonomia sp. | SACMA554–11
Lonomia sinalectra | SARBF2343–16
Lonomia mexilectra mexpueblensis | SARBD833–11
Lonomia mexilectra | SPHI00789
Lonomia sp. | SDHB883–08
Lonomia sp. | SARBD825–11
Lonomia luteomexicana | SPHI00788
Lonomia sp. | SDHB547–08
Lonomia electra | SPHI00869
Lonomia beneluzi | SPHI00469
Lonomia antoniae | SPHI00343
Lonomia sp. | SPHI00704
Lonomia leopoldina | SPHI00636
Lonomia sp. | SPHI00708
Lonomia parobliqua | SPHI00474
Lonomia rufobahiana | SPHI00816
Lonomia camox | RROU00269
Lonomia rengifoi | RROU00146
Lonomia maranhensis | SPHI00468
Lonomia siriae | SPHI00480
Lonomia madrediosiana | SPHI00475
Lonomia casanarensis | SPHI00575
Lonomia sp. | SPHI00341
Lonomia canescens | SPHI00477
Lonomia orientocordillera | SPHI00626
Lonomia sp. | SPHI00891
Lonomia sp. | SPHI00742
Lonomia francescae | SDHA615–07
Lonomia francescae | SPHI00882
Lonomia riojensis | SPHI00806
Lonomia panguana | SPHI00867
Lonomia silkae | SPHI00812
Lonomia sinjaevorum | SARBC760–10
Lonomia sp. | SPHI00479
Lonomia serranoi | SPHI00877
Lonomia rufescens | SPHI00829
Lonomia purtarenasiana | RROU00299
Lonomia nigra | SARBF1127–14
Lonomia manabiana | SARBE815–12
Lonomia sp. | SPHI00752
Lonomia orientoandensis | SPHI00876
Lonomia sp. | SPHI00705
Lonomia sp. | SPHI00892
Lonomia sp. | SPHI00472
Lonomia viksinyaevi viknucusensis | SPHI00807
Lonomia frankae | SPHI00342
Lonomia annamiekae | SPHI00476
Lonomia vanschaycki | SPHI00817
Lonomia sp. | SPHI00471
Lonomia belizonensis | SPHI00473
Lonomia sp. | RROU00176
Lonomia rubrescens | SPHI00478
Lonomia sp. | SDHB032–07
Lonomia sp. | STDAB42–07
Lonomia sp. | SACMA537–11
Lonomia parubrescens | SPHI00770
Periga squarmosa | SARBD1036–11
Periga kindli | SDHD230–09
Periga sp. | SAFMB862–09
Periga cluacina | SAMPA595–17
Periga costaricana | RROU00483
Periga pachijalensis | SARBE553–12
Periga kaechi | SARBE1483–13
Periga rasplusi | SDHA758–07
Periga galbimaculata | STDAB46–07
Periga galbiattocuscensis | SAFMB854–09
Periga galbinexspectata | SARBE945–12
Periga sp. | SAFMA819–08
Periga galbicentralis | SARBB1302–09
Periga lobulata | SDHA750–07
Periga anflaei | STDAB13–07
Periga angpapallacta | SARBE1262–13
Periga angpanguana | SARBF555–14
Periga angulosa | SACMA410–09
Periga angguyensis | SATYE479–09
Periga angcuscensis | SARBD1045–11
Periga sp. | SARBD1409–11
Periga angjunensis | SARBD1044–11
Periga acuta | SACMA558–11
Periga sp. | SACMA717–11
Periga cynira | SPHI00898
Periga aurantiaca | SARBF912–13
Periga inexpectata | STDAB25–07
Periga agrio | SARBF2566–16
Periga spatulata | SARBD1049–11
Periga sp. | TDMPE281–10
Periga sanmartiniana | AMAZ188–09
Periga bispinosa | STDAB29–07
Periga sp. | SARBF1570–15
Periga acomayensis | SARBE659–12
Periga kishidai | STDAB15–07
Periga antlozlovi | SARBF3040–17
Periga calabazana | SARBE656–12
Periga boettgeronum | SAFMA812–08
Periga parvimartinensis | SARBC608–10
Periga parvicajamarcensis | SARBD087–10
Periga parvibulbacea | SAFMB794–09
Periga marcapata | SARBD1130–11
Periga mariposana | SARBF1922–15
Periga suninensis | SARBD1139–11
Periga quillabambensis | SAFMB850–09
Periga parvichuquisaciana | SARBD1135–11
Periga parvioliviana | SARBB166–09
Periga sp. | SDHA762–07
Periga lichyi | SARBE1260–13
Periga guaca | SARBF3162–17
Periga mincensis | RROU00405
Periga sp. | SACMA183–09
Periga baezana | SARBE219–11
Periga santandeana | SARBF751–14
Periga sp. | SARBF1149–15
Periga barragani | SARBE1061–12
Periga occidentalis | STDA837–07
Periga altocauca | SARBF3140–17
Periga gachala | SARBF1819–15
Periga tungurahuensis | SARBE947–12
Periga sp. | SARBF3233–17
Periga sp. | SARBF794–14
Periga cutervensis | SAFMA822–08
Periga pacchana | SARBE1258–13
Periga altomoronensis | SARBE551–12
Periga falcbahiana | SARBD1050–11
Periga camacana | SARBE663–12
Periga sp. | SACMA097–09
Periga sp. | SACMA079–09
Periga falcleopoldina | SARBC820–10
Periga intervales | CMHEP211–15
Periga spatulatoides | SARBD869–11
Periga foersteri | SAFMC123–10
Periga sp. | SASNC2074–12
Periga sp. | SACMA099–09
Periga sp. | SACMA871–12
Periga alticola | SACMA575–11
Periga drechaeli altoparanensis | SAFMC188–10
Periga faustinoi | CMHEP208–15
Periga drechaeli | SAFMC190–10
Periga sp. | CMHEP210–15
Periga fuscbahiana | SARBE415–12
Periga caraca | SACMA116–09
Periga sp. | SACMA106–09
Periga campestre | SACMA888–12
Periga fuscleopoldina | SARBC831–10
Periga pulchra | SACMA061–09
Periga elyi | SACMA892–12
Periga sp. | SACMA893–12
Periga fusca | SAFMB788–09
Periga sp. | SACMA576–11
Periga sp. | SACMA579–11
Periga sp. | SACMA594–11
Periga circumstans | SACMA583–11
Periga sp. | SACMA100–09
Periga santaleopoldina | SARBC819–10
Periga potensis | SAFMC124–10
Periga paraleopoldina | SARBC811–10
Periga sp. | SACMA565–11
Periga circaustralis | SPHI00908
Periga sp. | SACMA118–09
Periga circpotensis | SAFMA825–08
Periga circleopoldina | SARBC803–10

Micragonini subtree

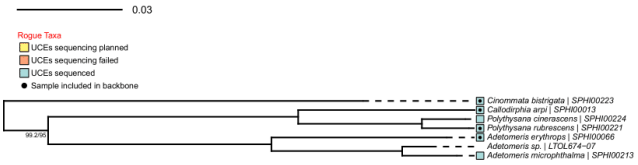



Holocerina menieri | SAPBA220–07
Holocerina sp. | SPHI00860
Holocerina sp. | SSDA697–07
Holocerina latsaniensis | STBOB322–08
Holocerina angulata | RROU01100
Holocerina occidentalis | STBOA717–07
Holocerina rougeoti | STBOB734–08
Holocerina prosti | SPHI00594
Pseudoludia nyungwe | STBOB395–08
Pseudoludia suavis | SPHI00058
Ludia jordani | SPHI00681
Ludia obscura | STBOC537–08
Ludia pupillata | SPHI00852
Ludia sp. | SAARA055–11
Ludia leonardo | STBOC320–08
Ludia orinoptena limbobrunnea | SAPBA161–07
Ludia orinoptera | RROU01119
Ludia sp. | STBOA494–07
Ludia sp. | STBOB808–08
Ludia sp. | STBOC528–08
Ludia sp. | SAPBA925–07
Ludia sp. | SASNA630–09
Vegetia dewitzi | STBOC370–08
Vegetia ducalis | SPHI00057
Vegetia sp. | SATJU059–08
Vegetia tegraini | SPHI00682
Ludia tessmanni | SPHI00780
Ludia arguta | SPHI00861
Ludia arida | STBOC674–08
Ludia sp. | SPHI00824
Ludia hansali eximia | STBOC492–08
Ludia sp. | SASNC3376–15
Ludia hansali | SPHI00864
Ludia pseudovetusta | STBOC153–08
Ludia sp. | SASNA633–09
Ludia sp. | STBOA833–07
Ludia goniata | STBOC505–08
Ludia delegorguei | SPHI00699
Carnegia mirabilis | SATWA925–07
Carnegia mirabilis | MGABC638–10
Carnegia mirabilis | SUBHB652–16
Carnegia mirabilis | SAPBA126–07
Carnegia mirabilis | RROU00771
Adafroptilum sp. | SASNC3079–15
Adafroptilum sp. | STBOD074–09
Adafroptilum kitongaensis | SSDA010–06
Adafroptilum sommereri | SPHI00670
Adafroptilum sp. | SAPBA834–07
Adafroptilum schmiti | SSDA402–07
Adafroptilum rougeoti | SSDA004–06
Adafroptilum sp. | STBOA842–07
Adafroptilum sp. | SPHI00784
Adafroptilum sp. | SASNC2844–13
Adafroptilum incanum | STBOE215–10
Adafroptilum sp. | SPHI00703
Adafroptilum hicoronatum | SSDA384–06
Adafroptilum tuberculatum | SSDA404–07
Adafroptilum sp. | STBOA434–07
Adafroptilum sp. | SSDA386–06
Adafroptilum sp. | STBOC170–08
Adafroptilum legorai | SSDA008–06
Adafroptilum sp. | STBOD022–09
Adafroptilum permixtum | SSDA409–07
Adafroptilum schaverii | STBOC147–08
Adafroptilum kalamboensis | SPHI00745
Adafroptilum sp. | SPPDA866–08
Adafroptilum sp. | STBOC408–08
Adafroptilum sp. | TBMOT1220–11
Adafroptilum rotundum | SSDA405–07
Orthogonioptilum vestigiatum | SPHI00551
Orthogonioptilum sp. | STBOD649–09
Orthogonioptilum olivescens | SSDB507–08
Orthogonioptilum dolfmani | STBOC396–08
Orthogonioptilum falcatissimum | SPHI00679
Orthogonioptilum adlegetum muonsatum | STBOB389–08
Orthogonioptilum adlegetum | SPHI00587
Orthogonioptilum occidentalis | SPHI00589
Orthogonioptilum violascens | STBOB627–08
Orthogonioptilum garnieri | STBOA457–07
Orthogonioptilum kahli | SPHI00590
Orthogonioptilum tristis | STBOD092–09
Orthogonioptilum sp. | SAPBA543–07
Orthogonioptilum herbuloti | SPBIS195–09
Orthogonioptilum bernaudi | STBOD757–09
Orthogonioptilum sp. | STBOA314–07
Orthogonioptilum fontainei | STBOB752–08
Orthogonioptilum obamba | SAPBA611–07
Orthogonioptilum garmsi | STBOD527–09
Orthogonioptilum chalix | STBOD446–09
Orthogonioptilum grootaerti | TBMOT1759–13
Orthogonioptilum concordia | STBOD509–09
Orthogonioptilum bimaculatum | SUBHB233–14
Orthogonioptilum silvaticum | TBMOT1783–13
Orthogonioptilum kasaiensis | SPHI00666
Orthogonioptilum sp. | SPHI00549
Orthogonioptilum nimbaense | STBOD548–09
Orthogonioptilum sp. | SPHI00617
Orthogonioptilum solium | STBOD441–09
Orthogonioptilum emmanuellae | SSDA340–06
Orthogonioptilum dargei | SSDA418–07
Orthogonioptilum bernardi | STBOA309–07
Orthogonioptilum sp. | GWOTV560–17
Orthogonioptilum andreasum | STBOA312–07
Orthogonioptilum sp. | SAPBA533–07
Orthogonioptilum caecum | SSDA339–06
Orthogonioptilum diabolicum | SPHI00593
Orthogonioptilum sp. | SAPBB032–07
Orthogonioptilum sp. | SPHPA189–07
Orthogonioptilum ochraceum | STBOD448–09
Orthogonioptilum conspectum | STBOD1066–09
Orthogonioptilum luminosum ssp. | SPHI00561
Orthogonioptilum subuelense | STBOA422–07
Orthogonioptilum fang | SPHI00591
Orthogonioptilum sp. | GWOTV565–17
Orthogonioptilum piersoni | SSDA345–06
Orthogonioptilum sp. | SAPBA350–07
Orthogonioptilum filippii | SPHI00592
Orthogonioptilum modestum | SAPBA357–07
Orthogonioptilum csomaense | STBOA399–07

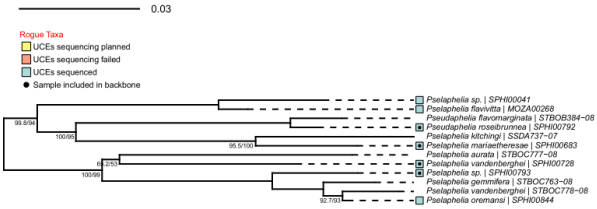
Oxyteninae subtree



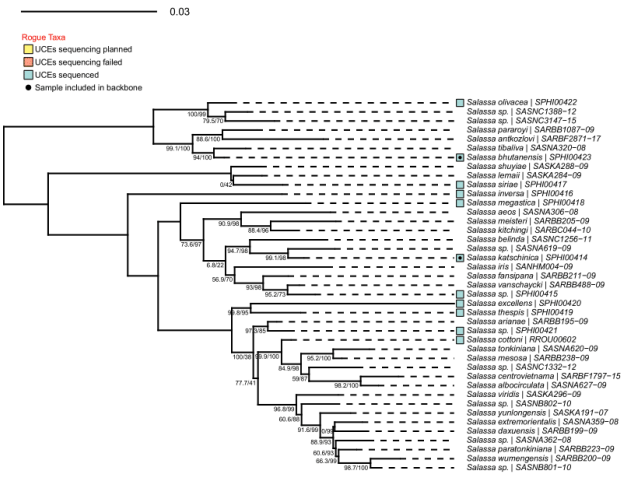
Polythysanini subtree



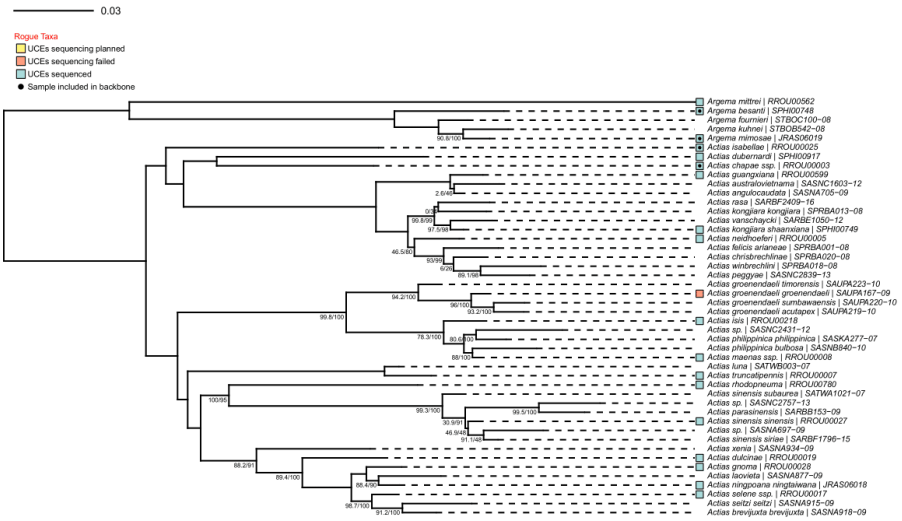
Pseudapheliini subtree



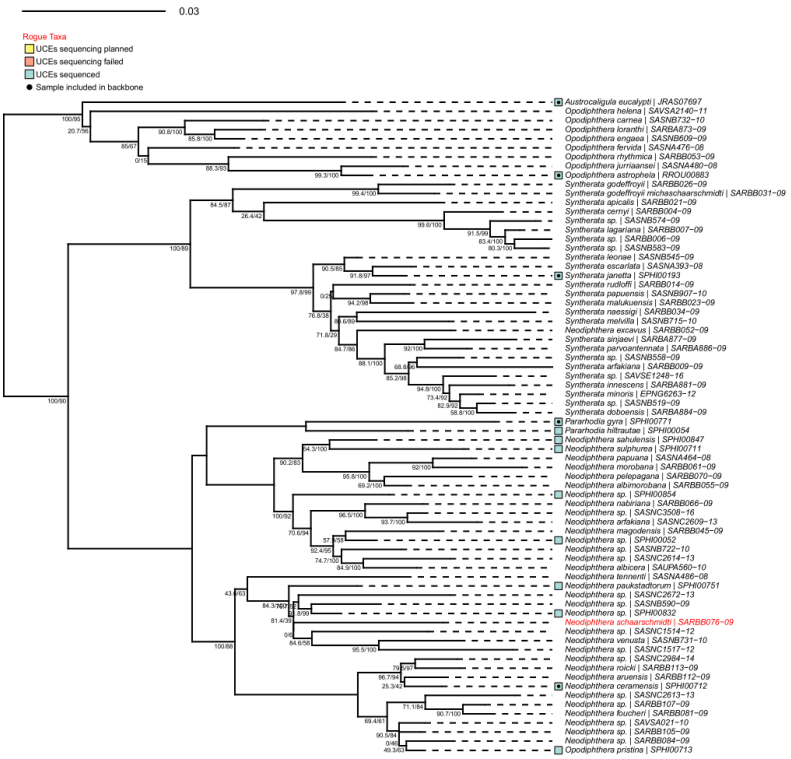
Salassinae subtree



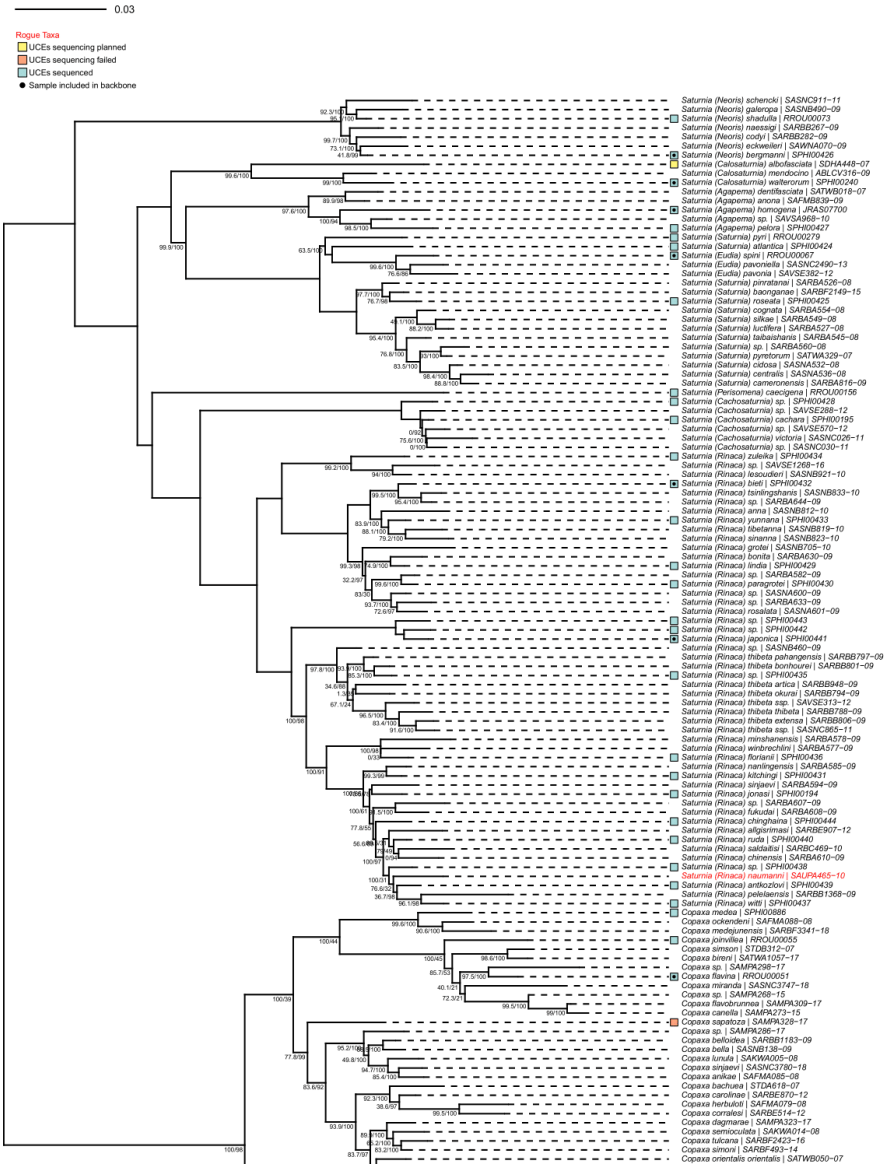
Saturniini_A subtree

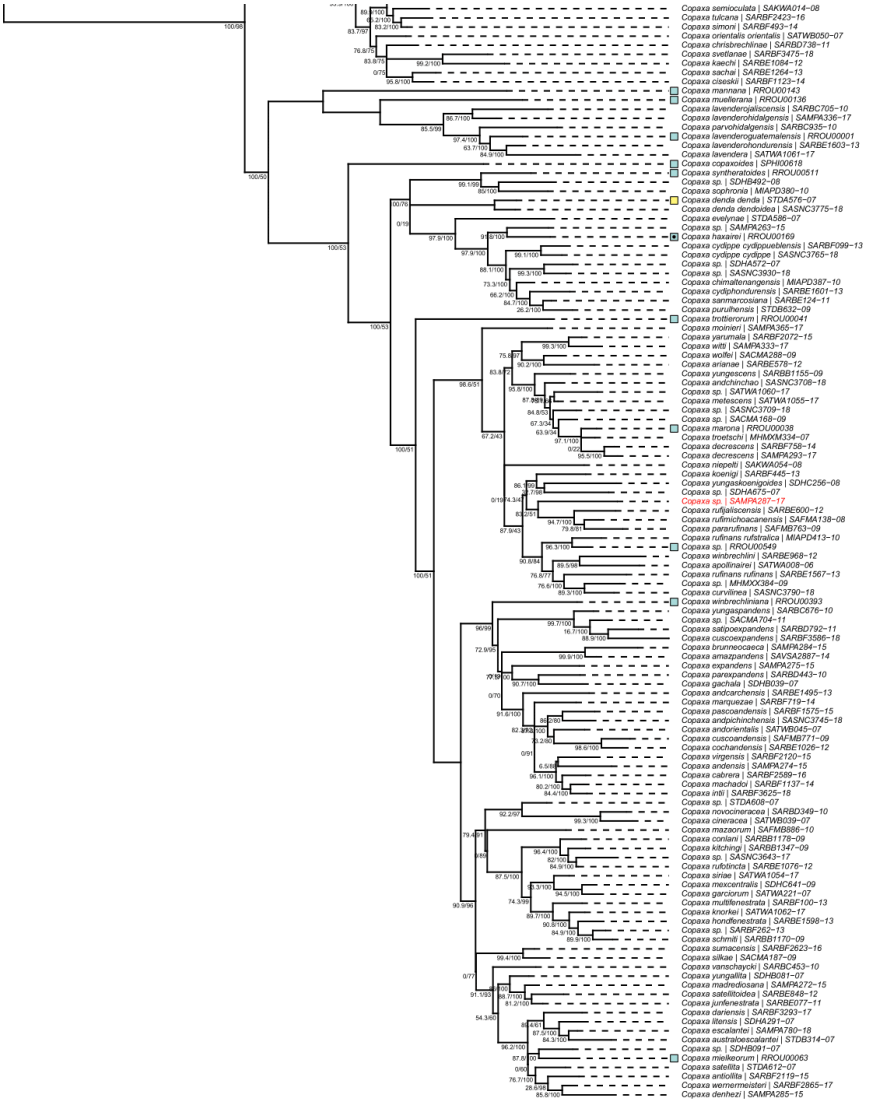


Saturniini_B subtree

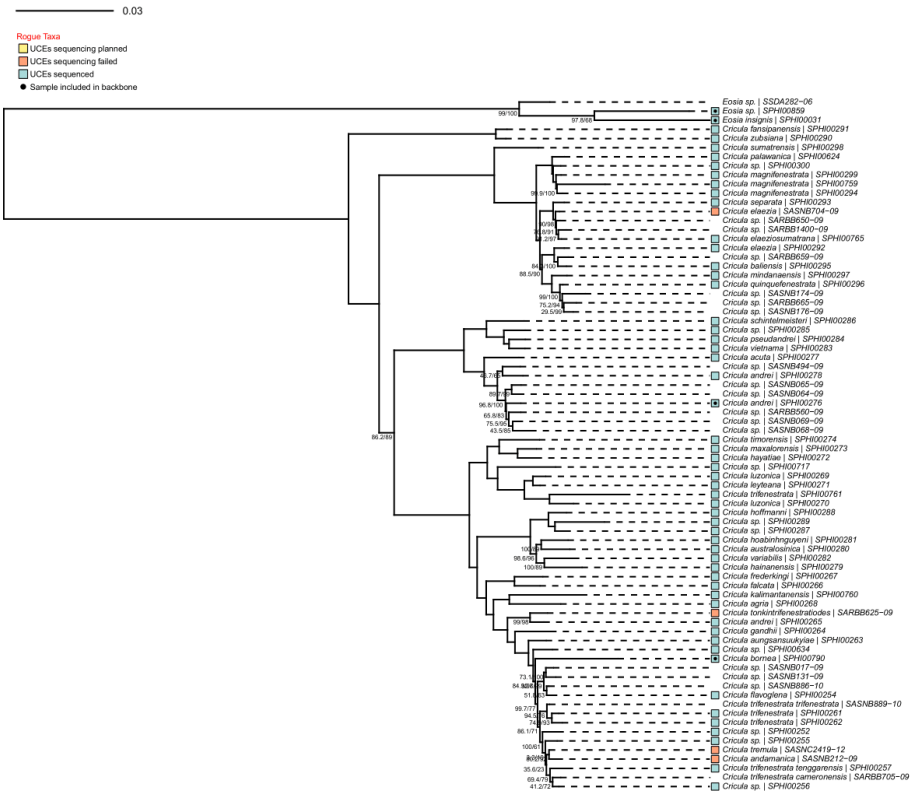


Saturniini_C subtree

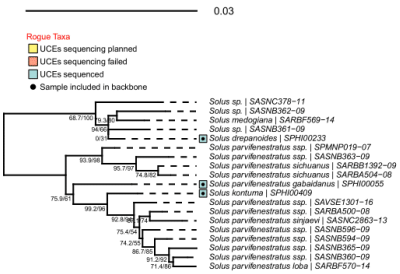



Copaxa semioculata | SAKWA014–08
Copaxa tulcana | SARBF2423–16
Copaxa simoni | SARBF493–14
Copaxa orientalis orientalis | SATWB050–07
Copaxa chrisbrechlinae | SARBD738–11
Copaxa svetlanae | SARBF3475–18
Copaxa kaechi | SARBE1084–12
Copaxa sachai | SARBE1264–13
Copaxa ciasaki | SARBF1123–14
Copaxa mannana | RROU00143
Copaxa muelleriana | RROU00136
Copaxa lavenderojaliscensis | SARBC705–10
Copaxa lavenderohidalgensis | SAMPA336–17
Copaxa parvohidalgensis | SARBC935–10
Copaxa lavenderoguatemalensis | RROU00001
Copaxa lavenderohondurensis | SARBE1603–13
Copaxa lavendera | SATWA1061–17
Copaxa copacoides | SPHI00618
Copaxa syntheratoides | RROU00511
Copaxa sp. | SDHB492–08
Copaxa sophronia | MIAPD380–10
Copaxa denda denda | STDA576–07
Copaxa denda dendoidea | SASNC3775–18
Copaxa evelynae | STDA586–07
Copaxa sp. | SAMPA263–15
Copaxa haxairei | RROU00169
Copaxa cydippe cydippueblensis | SARBF099–13
Copaxa cydippe cydippe | SASNC3765–18
Copaxa sp. | SDHA572–07
Copaxa sp. | SASNC3630–18
Copaxa chimaltenangensis | MIAPD387–10
Copaxa cydiphondurensis | SARBE1601–13
Copaxa sanmarcosiana | SARBE124–11
Copaxa puruihensis | STDB632–09
Copaxa trottierorum | RROU00041
Copaxa moinieri | SAMPA365–17
Copaxa yarumala | SARBF2072–15
Copaxa wittl | SAMPA333–17
Copaxa wolfei | SACMA288–09
Copaxa arianae | SARBE578–12
Copaxa yungascens | SARBB1155–09
Copaxa andchinchao | SASNC3708–18
Copaxa sp. | SATWA1060–17
Copaxa metescens | SATWA1055–17
Copaxa sp. | SASNC3709–18
Copaxa sp. | SACMA168–09
Copaxa marona | RROU00038
Copaxa troetschi | MHMXM334–07
Copaxa decrescens | SARBF758–14
Copaxa decrescens | SAMPA293–17
Copaxa niepelti | SAKWA054–08
Copaxa koenigi | SARBF445–13
Copaxa yungasikoenigoides | SDHC256–08
Copaxa sp. | SDHA675–07
Copaxa sp. | SAMPA287–17
Copaxa rufijaliscensis | SARBE600–12
Copaxa rufimichoacanensis | SAFMA138–08
Copaxa paranufinans | SAFMB763–09
Copaxa rufinans rufstralica | MIAPD413–10
Copaxa sp. | RROU00549
Copaxa winbrechlini | SARBE968–12
Copaxa apollinairei | SATWA008–06
Copaxa rufinans rufinans | SARBE1567–13
Copaxa sp. | MHM00C384–09
Copaxa curvilinea | SASNC3790–18
Copaxa winbrechliniana | RROU00393
Copaxa yungaspandens | SARBC676–10
Copaxa sp. | SACMA704–11
Copaxa satipoexpandens | SARBD792–11
Copaxa cuscoexpandens | SARBF3586–18
Copaxa brunneocaeca | SAMPA284–15
Copaxa amazpandens | SAVSA2887–14
Copaxa expandens | SAMPA275–15
Copaxa parexpandens | SARBD443–10
Copaxa gachala | SDHB039–07
Copaxa andcarchensis | SARBE1495–13
Copaxa marquezae | SARBF719–14
Copaxa pascoandensis | SARBF1575–15
Copaxa andpichinchensis | SASNC3745–18
Copaxa andorientalis | SATWB045–07
Copaxa cuscoandensis | SAFMB771–09
Copaxa cochandensis | SARBE1026–12
Copaxa virgensis | SARBF2120–15
Copaxa andensis | SAMPA274–15
Copaxa cabrera | SARBF2589–16
Copaxa machadoi | SARBF1137–14
Copaxa intti | SARBF3625–18
Copaxa sp. | STDA608–07
Copaxa novocineracea | SARBD349–10
Copaxa cineracea | SATWB039–07
Copaxa mazaorum | SAFMB886–10
Copaxa conlani | SARBB1178–09
Copaxa kitchingi | SARBB1347–09
Copaxa sp. | SASNC3643–17
Copaxa rufotincta | SARBE1076–12
Copaxa siriae | SATWA1054–17
Copaxa mexcentralis | SDHC641–09
Copaxa garciorum | SATWA221–07
Copaxa mutiflenestrata | SARBF100–13
Copaxa knorkei | SATWA1062–17
Copaxa hondflenestrata | SARBE1598–13
Copaxa sp. | SARBF262–13
Copaxa schmitti | SARBB1170–09
Copaxa sumacensis | SARBF2623–16
Copaxa silkae | SACMA187–09
Copaxa vanschaycki | SARBC453–10
Copaxa yungallita | SDHB081–07
Copaxa madrediosana | SAMPA272–15
Copaxa satellitoidea | SARBE848–12
Copaxa jumfenestrata | SARBE077–11
Copaxa darienis | SARBF3293–17
Copaxa litensis | SDHA291–07
Copaxa escalantei | SAMPA780–18
Copaxa australoescalantei | STDB314–07
Copaxa sp. | SDHB091–07
Copaxa mielkeorum | RROU00063
Copaxa satellita | STDA612–07
Copaxa antiolita | SARBF2119–15
Copaxa wernermeisteri | SARBF2865–17
Copaxa denhezi | SAMPA285–15

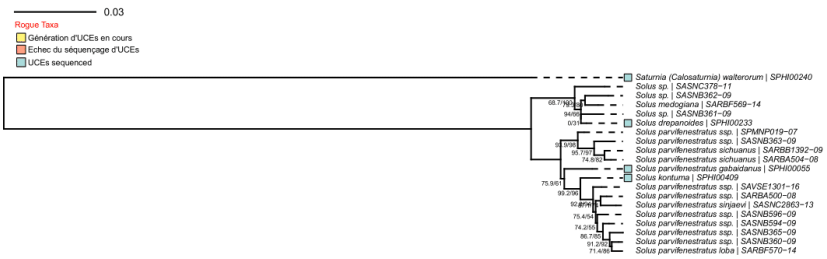
Saturniini_D subtree



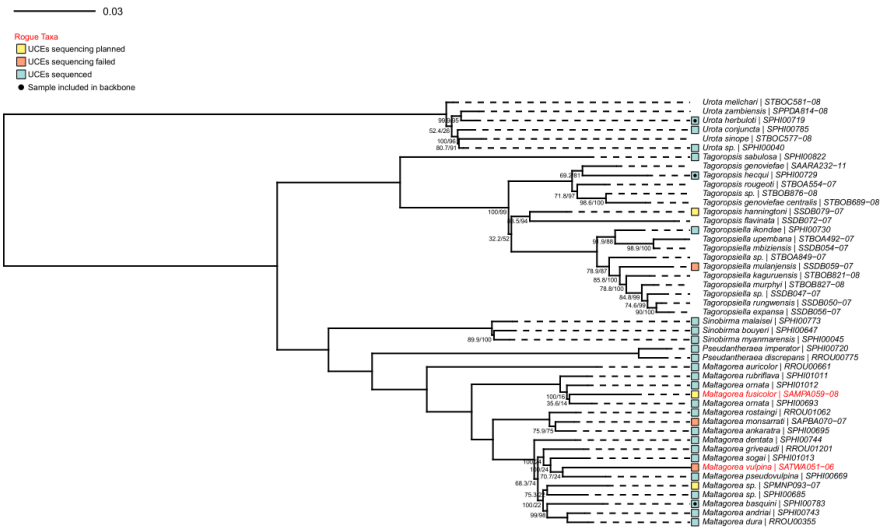
Soliini subtree



Solini subtree



Urotini subtree



Annexe 2 – Pipeline phylogénomique

~

Pipeline conçu pour générer des Mégaphylogénies dans le cadre de l'ANR SPHINX

Pierre Arnal

Juillet 2020

~

Préambule

Le *pipeline* présenté ici a pour but de générer des mégaphylogénies dans le cadre de l'ANR SPHINX. Ces phylogénies sont inférées en combinant deux types de marqueurs : UCE + codes-barres ADN COI (désignés par 'COI' dans le reste du document par commodité). La série de codes est adaptée à ces données mais le principe de l'inférence *backbone + subtrees* peut être appliqué sur des jeux de données distincts. L'exemple présenté ici a permis l'inférence d'une mégaphylogénie de la famille des Saturniidae (Chapitre 3).

NB : Nous avons utilisé la plateforme de bioinformatique genotoul Toulouse Occitanie qui fonctionne avec un environnement SLURM. L'ensemble des scripts sont donc adaptés à cet environnement.

0. Préparation des fichiers d'entrée

En se basant sur les connaissances que nous avons pu accumuler concernant la phylogénie des Saturniidae (Chapitre 1), nous avons défini 41 *subtree*s (clades) dont la monophylie avait été établie préalablement (Chapitre 1 et divers travaux préliminaires). Quatre de ces *subtrees* ont été intégrés directement au *backbone* car le nombre d'échantillons pour lesquels nous avions généré des UCEs était de 1 ou car le nombre d'échantillons dans le *subtree* était très limité (*e.g.* le *subtree* Eudaemoniini ; Tableau 2).

L'ensemble des OTUs considérées ont été identifiées en considérant des arguments biologiques, géographiques et génétiques. Nous avons notamment utilisé des arbres de distance inférés à partir de l'ensemble des séquences COI disponibles pour les Saturniidae dans la base de données BOLD (www.boldsystems.org).

A l'aide d'un script R, nous avons créé trois fichiers d'entrée pour chaque *subtree* ou *backbone* :

1. Une liste des échantillons pour lesquels nous disposions de marqueurs UCEs. Cette liste comporte également le code de l'échantillon utilisé comme *outgroup*. Nous avons défini des *outgroups* différents pour chaque *subtree*. Quand cela était possible, il s'agissait d'un échantillon appartenant à un clade apparenté au *subtree*.
2. Un fichier contenant le code d'un unique échantillon *outgroup*.
3. Un fichier *.fasta* compilant les séquences COI. Si nous disposions de séquences UCEs et COI, nous gardions le code UCE. Mais lorsque le COI était l'unique marqueur disponible, nous avons considéré le code process ID assigné par BOLD.

Un exemple est présenté ici pour le *subtree* correspondant à la Sous-famille Agliinae :

```
$cat Agliinae/Agliinae_UCEs_sample.txt
RROU00154_0101
SPHI00754_0101
RROU00153_0101
SPHI00413_0101
SPHI00414_0101
```

```
$cat Agliinae/Agliinae_outgroup.txt
SPHI00414_0101
```

```
$cat Agliinae/selected_loci_S3/Agliinae_COI_alignment.fas
>RROU00154 0101
aactttatat ttcatctttg […]
>RROU00153_0101
aactttatat ttcatct […]
>SARBB257 09
aactttatat tttatttttg […]
```

Toutes les commandes du *pipeline* sont à reproduire à partir du même dossier de travail. Dans notre exemple, le dossier est '*/work/parnal/SAT_splvl_Phylogeny*'. Si vous souhaitez reproduire le *pipeline*, certaines commandes ou scripts nécessitent d'être modifiés.

1. Génération des alignements *locus* par *locus*

1.1 Copie des fichiers .zip contenant les séquences UCEs.

A la suite du traitement des fichiers issus des séquençages successifs, nous avons regroupé les UCEs par échantillon dans des fichiers compressés. Les fichiers compressés ont été rassemblés dans un dossier unique ('*/work/parnal/SPH_capture/SPHXX/all_locus_all_samples/*').

Dans un premier temps, nous avons créé la liste des fichiers *.zip* à importer (à partir de la liste d'échantillons UCEs). Ici les *jobs* sont très rapides, il n'est donc pas nécessaire de les lancer via la commande *sbatch*.

```
$./create_zip_list.sh
#!/bin/bash

for i in $(ls -d */ | sed 's#/##'); do

 cd $i
 echo $i
 file_name=$(ls *UCEs_sample.txt)
 sed 's/_0101/.zip/g' $file_name > $i"_listzip.txt"
 sed -i 's/_0103/.zip/g'  $i"_listzip.txt"

 cd /work/parnal/SAT_splvl_Phylogeny

done
```

Puis, nous avons importé les fichiers *.zip*.

```
$./import_zip.sh
#!/bin/bash
#SBATCH -t 00:10:00

for i in $(ls -d */ | sed 's#/##'); do

 cd $i
 dos2unix *_listzip.txt
 for j in $(cat *_listzip.txt); do
  cp "/work/parnal/SPH_capture/SPHXX/all_locus_all_samples/"$j .
 done
 cd /work/parnal/SAT_splvl_Phylogeny

done
```

1.2 Compilation des séquences de chaque *locus*.

Nous avons décidé ici de ne sélectionner que les *loci* pour lesquels plus de 50% des échantillons étaient disponibles. Ce seuil peut bien sûr être modifié. Ici nous avons considéré qu'il établit un bon équilibre entre informations phylogénétiques, données manquantes, et taille de la matrice.

A noter que nous avons utilisé un script lanceur ('*launch*') qui appelle un autre script '*run*' qui est le script qui effectue la tâche qui nous intéresse. Ça nous permet de lancer un *job* par *subtree*/*backbone* et donc de gagner beaucoup de temps de calcul. Le *launch* repose ici sur le script *process_zip_merge_alignUCE.sh*. Ce procédé a été régulièrement utilisé dans le *pipeline*. Les séquences *.fas* ont ensuite été rassemblées dans un dossier *selected_loci_SX*. *X* étant le nombre minimum d'échantillons pour conserver un *loci*, il varie d'un sous jeu de données à l'autre puisque le nombre d'échantillons diffère.

```
$./launch_process_zip_merge_alignUCE.sh
```

```
#!/bin/bash
#SBATCH -c 1
#SBATCH -t 00:03:00

for i in $(ls -d */ | sed 's#/##') ; do
 cd $i
 sbatch ../process_zip_merge_alignUCE.sh
 cd /work/parnal/SAT_splvl_Phylogeny
done
```

```
$cat process_zip_merge_alignUCE.sh
```

```
#!/bin/bash
#SBATCH -t 02:00:00

#list all the loci inside the .zip
for i in *.zip ; do unzip -l $i | awk '/-----/ {p = ++p % 2; next} p {print $NF}' ;done
>>listoffiles.txt

#Get the number of sample per loci
cut -d "_" -f 1 listoffiles.txt | sort | uniq -c | sed -e 's/^ *//;s/ /,/' >temp

rm listoffiles.txt

#If the number of sample is under 6, we consider 6 as the threshold
#If a loci is shared by less than 3 samples, it is not informative
sample_nb=$(ls *.zip|wc -l)
cutoff=$(echo $sample_nb | awk '{printf("%d\n",($1/2))}')
my_min=3
if [[ $cutoff -le $my_min ]]
 then
 cutoff=3
fi

for i in $(cat< "temp");do
 number_samples=$(echo $i | cut -d ',' -f 1)
 my_loci=$(echo $i | cut -d ',' -f 2)
 if [[ $number_samples -ge $cutoff ]] ; then
  echo $my_loci >> "selected_loci_S"$cutoff".txt"
 fi;
done

rm temp

for j in $(cat selected*.txt); do
 for i in  *.zip ; do
  sample=$(echo $i | cut -d "." -f 1)
  unzip -o -j $i $j"_"$sample"*.fasta.fa" >> output_zip.o 2>error_zip.e
 done

 for i in $j*".fasta.fa" ; do
  (cat "${i}"; echo) >> $j".fas"
 done

 rm $j*".fasta.fa"
```

```
 echo $j" - Done!"
done

#Create directory and move .fas file
mkdir "selected loci S"$cutoff

mv *.fas "selected_loci_S"$cutoff
```

Quand les *jobs* sont finis, nous avons supprimé les fichiers *.zip*. L'ensemble du *pipeline* produit un nombre important de fichiers. Il est donc important de supprimer les fichiers dont nous n'avons plus besoin.

```
$rm ./*/*.zip
```

1.3 Alignement locus par locus

1.3.1 Alignement

Nous avons utilisé le logiciel MAFFT v7.313 (Katoh & Standley 2013) pour aligner les différents loci. Nous avons lancé un job par *subtree*. Au sein de chaque *subtree*, les alignements sont donc effectués un à un. Nous avons procédé ainsi car si un job par alignement avait été utilisé, le nombre de *jobs* aurait dépassé largement le plafond autorisé sur le cluster.

```
$sbatch launch_all_MAFFT.sh
```

```
#!/bin/bash
#SBATCH -c 1
#SBATCH -t 02:00:00

for i in $(ls -d */selec*/) ; do
 cd $i
 for j in *.fas ; do
  sed -i s/"-"/""/g $j
  sed -i s/Velvet_//g $j
 done
 sbatch /work/parnal/SAT_splvl_Phylogeny/run_MAFFT.sh
 cd /work/parnal/SAT_splvl_Phylogeny
done
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/run_MAFFT.sh
```

```
#!/bin/sh
#SBATCH -o MAFFT.o
#SBATCH -e MAFFT.e
#SBATCH -c 4
#SBATCH -p workq
#SBATCH -t 04:00:00

module load  bioinfo/mafft-7.313
for i in *.fas; do
 mafft $i >$i".al"
done
```

1.3.2 Nettoyage des alignements

Ensuite, nous avons utilisé le logiciel Gblocks 0.91b (Castresana 2000) pour corriger les alignements, supprimer des gaps, etc…

```
$sbatch launch_all_GBLOCK.sh
```

```
#!/bin/bash
#SBATCH -c 1
#SBATCH -t 02:00:00

for i in $(ls -d */selec*/) ; do
 cd $i
 sbatch /work/parnal/SAT splvl Phylogeny/run GBLOCK.sh
 cd /work/parnal/SAT_splvl_Phylogeny
done
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/run_GBLOCK.sh
#!/bin/sh
#SBATCH -o GBLOCK.o
#SBATCH -e GBLOCK.e
#SBATCH --mem=8G
#SBATCH -c 4
#SBATCH -p workq
#SBATCH -t 01:00:00

module load bioinfo/Gblocks_0.91b

for i in locus*.al; do
 Gblocks $i -t=d -b2=b1 -b3=10 -b4=2 -b5=h
Done
```

1.3.3 Conversion en format phylip

Enfin, nous avons converti les alignements vers le format *phylip*.

```
$sbatch launch_all_fasta2phylip.sh
#!/bin/bash
#SBATCH -c 1
#SBATCH -t 02:00:00

for i in $(ls -d */selec*/) ; do

 cd $i

 rename '.fas.al-gb' '_gbrelaxed.fas' *.fas.al-gb

 for j in *_gbrelaxed.fas; do
  sed -i "s/ //ig;" $j
 done

 sbatch /work/parnal/SAT splvl Phylogeny/run fasta2phylip.sh

 cd /work/parnal/SAT_splvl_Phylogeny

 echo $i "- Done !"
done
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/run_fasta2phylip.sh
#!/bin/sh
#SBATCH -o fasta2phylip.o
#SBATCH -e fasta2phylip.e
#SBATCH --mem=2G
#SBATCH -c 1
#SBATCH -p workq
#SBATCH -t 02:00:00

for i in *_gbrelaxed.fas ;do
 /work/parnal/software/Fasta2Phylip.pl $i $i".phy"
done
```

1.3.4 Statistiques

Afin de vérifier que l'alignement s'est bien déroulé pour l'ensemble des échantillons ainsi que pour mieux comprendre notre jeu de données, nous avons compté le nombre d'UCEs par échantillon. Les résultats ont été stocké dans les fichiers '*./-subtree-/selected_loci_SX/nb_loci_per_sample_-subtree-.txt*'.

Le script a également copié les alignements *.phy* dans un dossier '*./-subtree-/selected_loci_SX/per_loci*' dans lequel les *gene trees* ont été inférés (partie 2.1).

```
$sbatch statistics_in_alignment_beforePMCOA.sh
#!/bin/bash
#SBATCH -c 1
#SBATCH -t 03:00:00

for i in $(ls -d */selec*/) ; do
```

```
 cd $i

 #Rename the .phy files
 rename '.fas.phy' '.phy' *_gbrelaxed.fas.phy

 for j in *_gbrelaxed.phy; do cat $j >> my_selected_loci; done

 cat ../*_UCEs_sample.txt |tr "\n" " " |awk 'BEGIN{print "for i in "} {print $0} END{print
"\; do echo $i && grep -c \"$i\" my_selected_loci ; done"}' |tr "\n" " " |sed "s/  / /g"
|bash >> temp

 # Get the subtree name to name the number of loci per sample file
 subtree=$(ls ../*_UCEs_sample.txt | xargs -n 1 basename | sed -r 's/_UCEs_sample.txt//')

 awk 'NR % 2 == 1 { o=$0 ; next } { print o "\t" $0 }'  temp >
"nb loci per sample "$subtree".txt"

 rm temp
 rm my_selected_loci

 # Create a directory for the gene tree analysis
 mkdir per_loci
 cp *.phy per_loci

 cd /work/parnal/SAT_splvl_Phylogeny

done
```

NB : Les étapes de la partie 1.3 pourraient facilement se succéder à l'aide d'une commande *sarray*. C'est l'un des points du *pipeline* à améliorer.

2. Identification d'*outliers*.

Cette étape repose sur le logiciel PMCOA (de Vienne et al. 2012) et a pour but d'écarter des loci, échantillons ou séquences dont le signal phylogénétique est incohérent.

2.1 Inférence des *gene trees*

Dans un premier temps, nous avons inféré les *gene trees* à l'aide de IQTREE v1.6.7 (Nguyen et al. 2015). Pour chaque *gene tree*, nous avons déterminé le meilleur modèle d'évolution à l'aide du programme ModelFinder (Kalyaanamoorthy 2017), implémenté dans IQTREE. A noter qu'avec ce script nous supprimons aussi tous les fichiers issus de la phase 1.3 mis à part les alignements au format *phylip*.

```
$sbatch launch_all_IQTREE_beforePMCOA.sh
```

```
#!/bin/bash
#SBATCH -c 1
#SBATCH -t 05:00:00

for i in $(ls -d */selec*/per loci) ; do

 echo $i "- Start"
 cd $i

 for j in *.phy; do
  sbatch /work/parnal/SAT_splvl_Phylogeny/run_IQTREE_beforePMCOA.sh $j
 done

 echo $i "- Done !"
 nb sample=$(wc -l ../../*UCEs*|cut -d ' ' -f 1)

if [[ $nb_sample -le 10 ]];then sleep 3m ; fi
if [[ $nb_sample -gt 10 && $nb_sample -le 30 ]];then sleep 5m;fi
if [[ $nb_sample -gt 30 ]];then sleep 12m;fi

 cd /work/parnal/SAT splvl Phylogeny

done
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/run_IQTREE_beforePMCOA.sh
#!/bin/bash
#SBATCH -e IQTREE.e
#SBATCH -o IQTREE.o
#SBATCH -p workq
#SBATCH --cpus-per-task=1
#SBATCH -t 01:00:00

module load bioinfo/iqtree-1.6.7

iqtree -nt 1 -s $1 -pre "IQTREE_"$1 -m TESTNEW
```

NB : l'inférence des *gene trees* est l'un des points à améliorer en priorité dans le *pipeline*. L'utilisation d'une commande *sarray* avec un nombre limite de *jobs* simultanés serait une solution au problème du nombre plafond de *jobs* sur le cluster. La solution que nous avons utilisée jusque-là est de lancer l'ensemble des *gene trees* d'un même *subtree* mais en espaçant le lancement des jobs d'un *subtree* au suivant par laps de temps donné (de 3 à 10 min en fonction du nombre d'échantillons UCEs du subtree).

Nous n'avons gardé que les topologies ainsi créées (fichiers *.treefile*) et nous avons supprimé les autres fichiers des dossiers *per_loci/* (250k fichiers).

```
$find . -not -name 'IQTREE*.treefile' | grep "per_loci" | xargs rm -f
```

2.2 Identification des *outliers*

Les topologies inférées sont utilisées par PMCOA pour identifier les *outliers*. PMCOA est un logiciel codé dans R. Ici le *launcher* appelle un script qui lance le script R.

```
$sbatch launch_all_PMCOA.sh
#!/bin/sh
#SBATCH -t 00:10:00

for i in $(ls -d ./*/*/); do
 cd $i
 mkdir PMCOA
 cd PMCOA
 sbatch /work/parnal/SAT_splvl_Phylogeny/run_PMCOA.sh
 cd /work/parnal/SAT_splvl_Phylogeny
done
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/run_PMCOA.sh
#!/bin/bash
#SBATCH -o out.o
#SBATCH -e out.e
#SBATCH -t 08:00:00
#SBATCH --mem=8G

module load system/R-3.4.3
Rscript /work/parnal/SAT_splvl_Phylogeny/pmcoa.R
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/pmcoa.R
library(ape)
library(phylotools)
library(openxlsx)

#software import
source("/work/parnal/software/PMCOA/pmcoa.R")

#import tress
trees_path = list.files("../per_loci",pattern = ".treefile$",full.names = T)
trees=list()
for(i in 1:length(trees_path)) trees[[i]] = read.tree(trees_path[i])
class(trees) = "multiPhylo"

#Loci names
trees_path_names = list.files("../per_loci",pattern = ".treefile$",full.names = F)
loci_names = c()
```

```
for(i in 1:length(trees_path_names)) loci_names =
c(loci_names,gsub("IQTREE_","",strsplit(trees_path_names[i],"_")[[1]][2]))

# Run the PMCOA analysis

step1<-pMCOA(trees,distance="nodal",scannf = T,bvalue = 0) ##performs the first analysis

out1<-detect.complete.outliers(step1$mat2WR, k=1.5, thres=0.5) ##detect complete outliers

newtrees<-rm.gene.and.species(step1$trees, out1$outsp, out1$outgn) ##remove complete outliers

step2<-pMCOA(newtrees) ##second Phylo-MCOA analysis

out2<-detect.cell.outliers(step2$mat2WR,k=3,quiet = T) ##detect cell by cell outliers

save.image("PMCOA results.RData")

results=as.data.frame(out2$outcell,row.names = F)

#create a table with the gene number.
#replace gene number by its id

print(nrow(results))

if(nrow(results)!=0){
 results$Species=as.character(results$Species) ; results$Genes=as.character(results$Genes)
 for(i in 1:length(results$Genes)) {
   temp=as.numeric(strsplit(results$Genes[i],"N")[[1]][2])
   results$Genes[i]=loci_names[temp]
 }

 results=results[,c("Genes","Species")]
 names(results)
 write.table(results,file="to prune PMCOA.txt",sep="\t",row.names = FALSE,col.names =
FALSE,quote = F)

}
if(nrow(results)==0){
 write("",file="to prune PMCOA.txt")

 #Copy the alignment if none to remove
 #Alignements names
 alignments path names = list.files("../per loci",pattern = "^locus.*\\.phy$",full.names = T)

 #Create output dir
 if(!dir.exists("../PMCOA_clean")) dir.create("../PMCOA_clean")

 #Copy files
 for(i in 1:length(alignments_path_names))
file.copy(alignments path names[i],"../PMCOA clean")
}
```

2.3 Délétion des outliers

Lors de cette étape, nous avons écarté les *outliers* identifiés et nous avons créé de nouveaux alignements au format *phylip* dans le dossier '*./-subtree-/selected_loci_SX/PMCOA_clean*'.

```
$cat launch_all_remove_pmcoa.sh
```

```
#!/bin/sh
#SBATCH -t 00:10:00

for i in $(ls -d ./*/*/PMCOA); do
cd $i
sbatch /work/parnal/SAT splvl Phylogeny/run remove PMCOA.sh
cd /work/parnal/SAT_splvl_Phylogeny
done
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/run_remove_PMCOA.sh
```

```
#!/bin/bash
#SBATCH -o remove_pmcoa.o
#SBATCH -e remove_pmcoa.e
#SBATCH -t 02:00:00
#SBATCH --mem=2G
```

```
module load system/R-3.4.3
Rscript /work/parnal/SAT_splvl_Phylogeny/remove_pmcoa_ID.R
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/remove_pmcoa_ID.R
library(ape)
library(phylotools)
library(openxlsx)

#import alignment
alignment_path = list.files("../per_loci",pattern = "^locus.*\\.phy$",full.names = T)
head(alignment_path)
my alignments=list()
for(i in 1:length(alignment_path)) my_alignments[[i]] = read.dna(alignment_path[i],
as.character=T, as.matrix=T)

#Alignements names
alignments path names = list.files("../per loci",pattern = "^locus.*\\.phy$",full.names = F)

#import the PMCOA results
to_prune = read.table("to_prune_PMCOA.txt",header=F,sep="\t")

#Create output dir
if(!dir.exists("../PMCOA_clean")) dir.create("../PMCOA_clean")

for(i in 1:length(alignments_path_names)){
 focal_locus = strsplit(alignments_path_names[i],"_")[[1]][1]

 focal to prune = to prune[to prune$V1==focal locus,]

 if(nrow(focal_to_prune)==0){
  file.copy(alignment_path[i], to="../PMCOA_clean",overwrite=T) #write if no sample to
discard in the given locus
  print(paste(alignments_path_names[i],"- No samples ID by PMCOA"))
 }

 if(nrow(focal_to_prune)!=0){ #if some sample to discard
  #get only the existing samples. PMCOA considers centroids for missing sample so can be in
the resulting table
  sample to prune = paste(focal to prune$V2,"\t",sep="")
  sample to prune = intersect(row.names(my alignments[[i]]),sample to prune)

  if(length(sample_to_prune)==0) {
   file.copy(alignment_path[i], to="../PMCOA_clean",overwrite=T)
   print(paste(alignments_path_names[i],"- No existing samples ID by PMCOA"))
  }

  if(length(sample_to_prune)!=0){
   new_alignment = my_alignments[[i]]
   new_alignment = new_alignment[-
grep(paste(sample to prune,collapse="|"),row.names(new alignment)),]

   to write = list()
   to_write[[1]] = paste(nrow(new_alignment),ncol(new_alignment),sep="\t")
   for(j in 1:nrow(new_alignment)) to_write[[1+j]] =
paste(row.names(new_alignment)[j],paste(new_alignment[j,],collapse = ""),sep="")
   write(unlist(to write),file = paste("../PMCOA clean/",alignments path names[i],sep=""))

   print(paste(alignments_path_names[i],"- Pruned from samples ID by PMCOA"))
  }

 }
}
```

3. Inférence des topologies

3.1 Création des matrices concaténées

Lors de cette étape, nous avons concaténé les alignements par loci afin de créer les matrices d'entrée pour l'analyse IQTREE. Les scripts présentés ici génèrent également un fichier *charset* dans lequel sont enregistrées les positions des deux partitions que nous avons choisi de considérer : une première qui

rassemble tous les UCEs et une deuxième qui ne contient que les séquences COI. Ce fichier *charset* est ensuite utilisé pour lancer des analyses partitionnées dans IQTREE.

A noter que dans ce script, les fichiers issus des étapes d'alignement sont supprimés. Ne reste dans les dossiers '*./-subtree-/selected_loci_SX/*' que les alignements COI et les fichiers *phylip* non nettoyés des *outliers*.

Un script R est utilisé ici pour générer les alignements.

```
$sbatch launch_all_merge_alignment.sh
```

```
#!/bin/sh
#SBATCH -t 00:10:00

for i in $(ls -d ./*/sel*/); do
cd $i
rm locus*.al
rm locus*.fas
rm locus*.htm
sbatch /work/parnal/SAT splvl Phylogeny/run merge alignment.sh
cd /work/parnal/SAT_splvl_Phylogeny
done
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/run_merge_alignment.sh
```

```
#!/bin/bash
#SBATCH -o merge_alignment.o
#SBATCH -e merge_alignment.e
#SBATCH -t 04:00:00
#SBATCH --mem=8G

module load system/R-3.4.3
Rscript /work/parnal/SAT_splvl_Phylogeny/merge_alignment.R
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/merge_alignment.R
```

```
library(ape)
library(phylotools)
library(openxlsx)

# Get the samples list
my_samples_path = list.files("../",pattern="_UCEs_sample.txt$",full.names=T)
my_samples = readLines(my_samples_path)
my samples path = list.files("../",pattern=" UCEs sample.txt$",full.names=F)
my samples = paste(my samples,"\t",sep="")

#import UCE alignments
alignment_path = list.files("./PMCOA_clean",pattern = "^locus.*\\.phy$",full.names = T)
my alignments=list()
for(i in 1:length(alignment_path)) my_alignments[[i]] = read.dna(alignment_path[i],
as.character=T, as.matrix=T)

#import the COI alignment
COI_alignment_path = list.files(pattern="fas.al$")
print(COI_alignment_path)
COI_alignment = read.dna(COI_alignment_path, as.character=T, as.matrix=T, format="fasta")

### First step: merge the UCEs alignments
 to_be_merge_alignments = list() # list to shote the loci with the missing data added

 for(i in 1:length(my_alignments)){
  temp alignment =  my alignments[[i]]
  temp species = row.names(temp alignment)
  temp_ncl = dim(temp_alignment)[2]
  not_in_alignment_sp=setdiff(my_samples,temp_species)

  if(length(not_in_alignment_sp)!=0){
   not in loci alignment = matrix("-",length(not in alignment sp),temp ncl)
   row.names(not_in_loci_alignment) = not_in_alignment_sp
   to_be_merge_alignments[[i]] = rbind(temp_alignment,not_in_loci_alignment)
  }
  if(length(not_in_alignment_sp)==0) to_be_merge_alignments[[i]] = temp_alignment
 }
```

```
 my_merged_alignment = to_be_merge_alignments[[1]] #initalize the merging

 for(i in 2:length(to_be_merge_alignments)){ #merge
  to be merge alignments[[i]] =
to be merge alignments[[i]][row.names(to be merge alignments[[1]]),]
  my_merged_alignment = cbind(my_merged_alignment,to_be_merge_alignments[[i]])

  print(paste(i,"/",length(to_be_merge_alignments),sep="")) #progress
  if(i==length(to_be_merge_alignments)) cat("Done!\n")
 }

### End of first step

### Second step : add the COI alignment
 row.names(my merged alignment) = gsub("\t","",row.names(my merged alignment))

 UCE not COI = setdiff(row.names(my merged alignment),row.names(COI alignment))
 COI_not_UCE = setdiff(row.names(COI_alignment),row.names(my_merged_alignment))

 if(length(UCE_not_COI)!=0){
  nb nucl COI = ncol(COI alignment)
  matrix_toappend_COI = matrix("-",length(UCE_not_COI),nb_nucl_COI)
  row.names(matrix_toappend_COI) = UCE_not_COI
  COI_alignment = rbind(COI_alignment, matrix_toappend_COI)
 }

 if(length(COI_not_UCE)!=0){
  nb_nucl_UCE = ncol(my_merged_alignment)
  matrix_toappend_UCE = matrix("-",length(COI_not_UCE),nb_nucl_UCE)
  row.names(matrix toappend UCE) = COI not UCE
  my merged alignment = rbind(my merged alignment, matrix toappend UCE)
 }


 print(row.names(COI_alignment))
 print(row.names(my_merged_alignment))

 print(sort(table(row.names(COI_alignment))))

 print(setdiff(row.names(COI_alignment),row.names(my_merged_alignment)))
 print(setdiff(row.names(my_merged_alignment),row.names(COI_alignment)))


 if(length(row.names(COI_alignment))!=length(row.names(my_merged_alignment))) stop("Error!
Number of samples COI and UCE different")


 #Get charset
 my charset = list()
 my_charset[[1]] = "#nexus"
 my_charset[[2]] = "begin sets;"
 my_charset[[3]] = paste("charset UCE =",1,"-",ncol(my_merged_alignment),";",sep="")
 my charset[[4]] = paste("charset COI =",ncol(my merged alignment)+1,"-
",ncol(my merged alignment)+ncol(COI alignment),";",sep="")
 my charset[[5]] = "end;"

 #Put in order and cbind COI and UCEs
 COI_alignment = COI_alignment[row.names(my_merged_alignment),]
 my merged alignment = cbind(my merged alignment, COI alignment)

### End of second step

#Write the matrix
if(!dir.exists("IQTREE")) dir.create("IQTREE")
my subtree = gsub(" UCEs sample.txt","",my samples path)

to_write = list()
to_write[[1]] = paste(nrow(my_merged_alignment),ncol(my_merged_alignment),sep="\t")
for(j in 1:nrow(my_merged_alignment)) to_write[[1+j]] =
paste(row.names(my merged alignment)[j],"\t",paste(my merged alignment[j,],collapse =
""),sep="")
write(unlist(to write),file = paste("./IQTREE/", my subtree, " alignment.phy",sep=""))

write(unlist(my_charset),file = paste("./IQTREE/", my_subtree, "_charset.txt",sep=""))
```

Les deux fichiers générés, l'alignement et le fichier *charset* sont stockés dans le dossier '*./-subtree-/selected_loci_SX/IQTREE*'.

NB : une stratégie de partitionnement plus élaborée pourrait être envisagée ici, notamment la méthode de Tagliacollo & Lanfear (2018).

3.2 Inférence des arbres phylogénétiques

Nous avons utilisé IQTREE v1.6.7 (Nguyen et al. 2015) en considérant les deux partitions créées précédemment (3.1). A noter que l'échantillonnage des positions nucléotidiques pour les réplicats de bootstrap se fait à l'intérieur des partitions définies par défaut. Toutes les topologies des réplicats de bootstrap sont écrites (option *–wbtl* de IQTREE).

```
$sbatch launch_all_IQTREE_merged.sh
```

```
#!/bin/bash
#SBATCH -c 1
#SBATCH -t 00:05:00

for i in $(ls -d */selec*/IQTREE) ; do

 cd $i

 my_alignment=$(ls *alignment.phy)
 my_charset=$(ls *charset.txt)

 sbatch /work/parnal/SAT_splvl_Phylogeny/run_IQTREE_merged.sh $my_alignment $my_charset

 cd /work/parnal/SAT_splvl_Phylogeny

done
```

```
$cat run_IQTREE_merged.sh
```

```
#!/bin/bash
#SBATCH -e IQTREE.e
#SBATCH -o IQTREE.o
#SBATCH -p workq
#SBATCH --cpus-per-task=40
#SBATCH -t 48:00:00

module load bioinfo/iqtree-1.6.7

iqtree -nt 40 -s $1 -pre "IQTREE_"$1 -m TESTNEW  -bb 1000 -alrt 1000 -spp $2 -wbtl
```

4. Datation

Nous avons utilisé le logiciel *MCMCtree* appartenant à la suite de logiciels PAML (Yang 1997). Afin de réduire les temps de calcul, nous avons réduits les matrices nucléotidiques à 50 loci, échantillonnés aléatoirement. Le logiciel MCMCtree a été lancé sur 2 millions de générations, dont 300 000 de *burnin*. Lors de l'échantillonnage de ces 50 *loci*, nous nous sommes assurés que le nombre minimal de *loci* pour un échantillon était de 10. Si ce critère n'était pas rempli, un autre set de *loci* fut échantillonné, et ce, jusqu'à que ce critère soit respecté. Si au bout de 100 randomisations, le nombre de *loci* minimal pour un échantillon était toujours inférieur à 10, le *set* de *loci* dont la valeur minimale est la plus élevée est conservé. Si ce nombre est égal à 0, le script s'arrête : il y a un problème pour l'un des échantillons (qui sera indiqué sur les fichiers d'erreur).

Les points de calibration pour estimer les temps de divergence de l'arbre *backbone* sont des fossiles et/ou des calibrations secondaires. Dans ce cas précis, il s'agit de deux fossiles (1 Sphingidae, 1 Saturniidae), ainsi que de cinq calibrations secondaires provenant de la datation de la phylogénie des Lépidoptères par Wahlberg et al. (2013) (voir le Chapitre 1 pour plus de détails). Les calibrations ont été définies à l'aide d'une loi uniforme couvrant, pour les fossiles, l'âge de fossilisation ou, pour les points de datation secondaires, l'intervalle de *confidence* à 95%. Afin d'établir l'influence du set de *loci*

sur l'estimation des temps de divergence, nous avons indépendamment daté le *backbone* avec deux sets de *loci* distincts (*50_loci_random_A* et *50_loci_random_B*). Pour s'assurer de la convergence des analyses, nous avons également lancé la datation deux fois sur chaque set de *loci* (*approx1* et *approx2*).

Ce n'est que dans un second temps que nous avons daté les *subtrees*. Pour cela, nous avons utilisé les estimations des temps de divergence des nœuds communs au *backbone* et aux *subtrees* comme *input* pour dater ces derniers. Contrairement au *backbone*, nous avons utilisé ici des *point calibrations* : nous ne considérons donc pas d'incertitude concernant les âges des nœuds datés. Bien que *MCMCtree* ne permette pas d'effectuer de *point calibration* à proprement parler, nous avons contraint les nœuds concernés dans des intervalles très étroits, ce qui équivaut à une *point calibration*. Par exemple pour le *subtree* Agliinae (ici l'unité de temps est la centaine de million d'année) :

```
(((RROU00153_0101,SPHI00754_0101),RROU00154_0101),(SPHI00413_0101,SARBB257_09))'>0.059897<0.059897';
```

Pour tous les *subtrees*, un seul set de loci a été utilisé et l'analyse n'a été lancée qu'une seule fois. Des tests préliminaires ont montré que, d'un set de loci à l'autre, les temps de divergence étaient relativement similaires et que deux analyses d'un même set de *loci* convergeaient vers des estimations identiques. Egalement, ces tests ont permis de mettre en évidence que les estimations ne variaient pas d'une génération à la suivante. Pour la datation de l'ensemble des *subtrees*, nous avons donc utilisé une seule génération, sans *burnin*.

En plus de l'inférence d'une mégaphylogénie 'principale', le *pipeline* infère une série de réplicats afin de prendre en compte les incertitudes liées à la topologie et aux temps de divergence. Pour cela, nous avons effectué l'étape de datation des *subtrees* puis de l'assemblage de la mégaphylogénie plusieurs fois (nous nous sommes limité à 10 réplicats pour l'instant). Lors de la datation des *subtrees*, nous avons considéré les temps de divergence d'un arbre *backbone* différent pour chaque réplicat, aléatoirement échantillonné dans la distribution postérieure de l'analyse MCMCtree (la topologie reste identique). Egalement, à chaque replicat de la mégaphylogénie inféré, nous avons considéré des topologies différentes pour les *subtrees*, issues des réplicats de bootstrap (partie 3.2).

Quant à la mégaphylogénie 'principale', nous avons utilisé les âges médians estimés par MCMCtree pour le *backbone* ainsi qu'en considérant les topologies *subtrees* estimées à partir des alignements originaux (replicat n°0).

4.1 Datation du backbone et assemblage de l'arbre

4.1.1 Création des fichiers d'entrée

Dans un premier temps, nous avons généré les fichiers d'input pour les 4 datations indépendantes du backbone. La structure des dossiers est la suivante :

à partir du dossier /`work/parnal/SAT_splvl_Phylogeny/backbone/selected_loci_S88` :

```
./DATATION
	./50_loci_random_A
		./Hessian
		./approx1
		./approx2
	./50_loci_random_A
		./Hessian
		./approx1
		./approx2
```

Chaque dossier `./Hessian` ou `./approx[1,2]` comporte les fichiers suivants :

- L'alignement. Il diffère en fonction que l'on se place dans le dossier `./50_loci_random_A` ou `./50_loci_random_B`.

- La topologie, identique dans l'ensemble des dossiers, avec les calibrations.
- Un fichier de configuration de l'analyse. Il diffère en fonction qu'on se place dans un dossier `./Hessian` ou `./approx[1,2]`.

`./Hessian`

```
seed = -1
seqfile = my random based alignment A.phy
treefile = DATATION input backbone.tre
outfile = mcmctree_RDMA_3M_cor.txt
ndata = 2
seqtype = 0
usedata = 3
clock = 3
RootAge = '<1'
model = 4
alpha = 0.5
ncatG = 4
cleandata = 0
BDparas = 1 1 0.1
kappa_gamma = 6 2
alpha_gamma = 1 1
rgene gamma = 2 20 1
sigma2 gamma = 1 10 1
finetune = 1: .1 .1 .1 .1 .1 .1
print = 1
burnin = 300000
sampfreq = 100
nsample = 17000
```

`./approx[1,2]`

```
seed = -1
seqfile = my random based alignment A.phy
treefile = DATATION input backbone.tre
outfile = mcmctree_RDMA_3M_cor.txt
ndata = 2
seqtype = 0
usedata = 2
clock = 3
RootAge = '<1'
model = 4
alpha = 0.5
ncatG = 4
cleandata = 0
BDparas = 1 1 0.1
kappa_gamma = 6 2
alpha_gamma = 1 1
rgene gamma = 2 20 1
sigma2 gamma = 1 10 1
finetune = 1: .1 .1 .1 .1 .1 .1
print = 1
burnin = 300000
sampfreq = 100
nsample = 17000
```

- Un ficher permettant de lancer l'analyse, `run_MCMCtree.sh`.

```
$sbatch run_datation_backbone_generatefiles.sh
```

```
#!/bin/bash
#SBATCH -o datation_backbone_files.o
#SBATCH -e datation_backbone_files.e
#SBATCH -t 01:00:00
#SBATCH --mem=2G

module load system/R-3.4.3
Rscript /work/parnal/SAT_splvl_Phylogeny/datation_backbone_generatefiles.R
```

```
$cat /work/parnal/SAT_splvl_Phylogeny/datation_backbone_generatefiles.R
```

```
library(phytools)
library(phangorn)
library(openxlsx)

setwd("/work/parnal/SAT splvl Phylogeny/backbone/selected loci S88")

#####################################################################
#Import the tree inferred the IQTREE on the merged alignement
iqtree tre=read.tree("./IQTREE/IQTREE backbone alignment.phy.treefile")

#reroot the tree
 my outgroup = "RROU00175 0101" #Endromis versicolora
 index = which(iqtree_tre$tip.label==my_outgroup)
 my_position = 0.5*iqtree_tre$edge.length[which(iqtree_tre$edge[,2]==index)]
 iqtree tre = reroot(iqtree tre, index, position = my position)

#First import tree and prune from outgroup but Sphingidae (needed for datation)
to_prune = c("RROU00175_0101",#Endromis versicolora
             "RROU00224_0101",#Lemonia taraxaci
             "RROU00221_0101"#Acanthobrahmaea europaea
             )

iqtree_tre = drop.tip(iqtree_tre,to_prune)

UCE_samples = readLines("../backbone_UCEs_sample.txt")
UCE samples = UCE samples[-grep(paste(to prune,collapse="|"),UCE samples)]


#####################################################################
#Import the COI alignment
```

```
COI_alignment = read.dna("backbone_COI_alignment.fas.al",
 as.character=T, as.matrix=T, format="fasta")
#Prune for outgroup is in the alignment
if(any(grepl(paste(to_prune,collapse="|"),row.names(COI_alignment)))) {
 COI alignment = COI alignment[-grep(paste(to prune,collapse="|"),row.names(COI alignment)),]
}
#####################################################################
#Import the per loci alignements
alignment_path = list.files("./PMCOA_clean",pattern = "^locus.*\\.phy$",full.names = T)
my alignments=list()
for(i in 1:length(alignment_path)) {
 temp_alignment = read.dna(alignment_path[i], as.character=T, as.matrix=T)

 #prune from outgroup
 if(any(grepl(paste(to_prune,collapse="|"),row.names(temp_alignment)))){
  temp alignment = temp alignment[-grep(paste(to prune,collapse="|"),row.names(temp alignment)),]
 }
 my alignments[[i]] = temp alignment
}

#Alignments names
alignment path names = list.files("./PMCOA clean",pattern = "^locus.*\\.phy$",full.names = F)
my_loci = c()
for(i in 1:length(alignment_path_names)) my_loci = c(my_loci,
gsub("_gbrelaxed.phy","",alignment_path_names[i]))

# Assign names
names(my_alignments) = my_loci

#####################################################################
## First random set of 50 loci
#The while loop allows to relaunch the randomization until the minimum number of loci for a
sample is 10
success <- FALSE

while(!success){

random loci A=sample(my loci,50)

#Count the number of loci per sample
random_loci_spcount_A=data.frame(UCE_samples,rep(0,length(UCE_samples)))
names(random loci spcount A)=c("tip.label","nb")
for(i in 1:length(random_loci_A)){
  index=grep(random loci A[i],my loci)
  temp tr tips=row.names(my alignments[[index]])
  temp_tr_tips = gsub("\t","",temp_tr_tips)
  for(j in 1:length(temp_tr_tips)) {
   my row = random loci spcount A$tip.label==temp tr tips[j]
   random loci spcount A[my row,]$nb=random loci spcount A[my row,]$nb+1
  }
}

success=min(random_loci_spcount_A$nb)>=10
print(min(random_loci_spcount_A$nb))
}

#####################################################################
## Second sed of random loci
success <- FALSE

while(!success){

random_loci_B=sample(my_loci,50)

#Count the number of loci per sample
random loci spcount B=data.frame(UCE samples,rep(0,length(UCE samples)))
names(random_loci_spcount_B)=c("tip.label","nb")
for(i in 1:length(random_loci_B)){
  index=grep(random_loci_B[i],my_loci)
  temp tr tips=row.names(my alignments[[index]])
  temp tr tips = gsub("\t","",temp tr tips)
  for(j in 1:length(temp_tr_tips)) {
   my_row = random_loci_spcount_B$tip.label==temp_tr_tips[j]
   random_loci_spcount_B[my_row,]$nb=random_loci_spcount_B[my_row,]$nb+1
  }
}
```

```
success=min(random_loci_spcount_B$nb)>=10
print(min(random_loci_spcount_B$nb))
}

######################################################################
# Create dataset
if(!dir.exists("DATATION")) dir.create("DATATION")
setwd("DATATION")

######################################################################
# Statistics

#random A
if(!dir.exists("50_loci_random_A")) dir.create("50_loci_random_A")
write.table(random loci A,file="./50 loci random A/random selcted UCEloci A.txt",
 row.names = F,col.names = F,quote = F,sep="\t")
write.table(random_loci_spcount_A,file="./50_loci_random_A/species_count_in_selcted_UCEloci.txt",
 row.names = F,col.names = F,quote = F,sep="\t")

#random B
if(!dir.exists("50_loci_random_B")) dir.create("50_loci_random_B")
write.table(random_loci_B,file="./50_loci_random_B/random_selcted_UCEloci_B.txt",
 row.names = F,col.names = F,quote = F,sep="\t")
write.table(random_loci_spcount_B,file="./50_loci_random_B/species_count_in_selcted_UCEloci.txt",
 row.names = F,col.names = F,quote = F,sep="\t")

######################################################################
# Create the alignments
#random A
random_selcted_loci_A = read.table("./50_loci_random_A/random_selcted_UCEloci_A.txt")
random_selcted_loci_A = as.character(random_selcted_loci_A[,1])

#Get the loci we want
random_loci_A=list()
for (i in 1:length(random_selcted_loci_A)){
 random_loci_A[i]=my_alignments[grep(random_selcted_loci_A[i],names(my_alignments))]
}
my loci names=c()
my_charset=c()
for(i in 1:length(random_loci_A)){
  if(i==1){
    my random alignment A=random loci A[[i]]
    my loci names[i]=random selcted loci A[i]
    my charset[i]=paste(1,length(my random alignment A[1,]))
  }
  if(i!=1){
    temp=random loci A[[i]]

    temp not align= setdiff(row.names(temp),row.names(my random alignment A))
    align not temp= setdiff(row.names(my random alignment A),row.names(temp))

    #Add the new species to the alignment. Put "-" gap
    to add align=matrix("-",length(temp not align),length(my random alignment A[1,]))
    row.names(to add align)=temp not align
    my_random_alignment_A=rbind(my_random_alignment_A,to_add_align)

    #Add the new species to the temp. Put "-" gap
    to add temp=matrix("-",length(align not temp),length(temp[1,]))
    row.names(to add temp)=align not temp
    temp=rbind(temp,to add temp)

    if(nrow(temp) == nrow(my_random_alignment_A)){
      #Match same order
      temp=temp[match(row.names(my random alignment A),row.names(temp)),]

      if(length((grep(FALSE,row.names(temp) == row.names(my_random_alignment_A))))>0){
       print("row.names not same order")
      }
      if(length((grep(FALSE,row.names(temp) ==  row.names(my_random_alignment_A))))==0) {
        my charset[i]=paste(length(my random alignment A[1,])+1,
                            length(my random alignment A[1,])+length(temp[1,]))
        my random alignment A=cbind(my random alignment A,temp)
        print(paste(random_selcted_loci_A[i],"-",length(temp[1,]), "pb"))
        my_loci_names[i]=random_selcted_loci_A[i]
      }
```

```
    }

    if(nrow(temp) != nrow(my_random_alignment_A)) print("not same dimension")

  }
}
write.table(data.frame(my_loci_names,my_charset),
 file="./50_loci_random_A/my_random_based_charset_A.txt", sep="\t",
 col.names = F,row.names = F,quote = F)
my random alignment A = my random alignment A[paste(UCE samples,"\t",sep=""),]

my_random_alignment_A_list = list()
my_random_alignment_A_list[[1]] =
paste(nrow(my_random_alignment_A),ncol(my_random_alignment_A),sep="\t")
for(i in 1:nrow(my_random_alignment_A)){
 my random alignment A list[[i+1]] = paste(gsub("\t","",row.names(my random alignment A)[i]),
                                           paste(my random alignment A[i,], collapse = ""),sep="
")
}
#Add the COI alignment
my random alignment A list[[length(my random alignment A list)+1]] = "\n"
my_random_alignment_A_list[[length(my_random_alignment_A_list)+1]] = paste(nrow(COI_alignment),

ncol(COI_alignment),sep="\t")
for(i in 1:nrow(COI_alignment)){
 end of file=length(my random alignment A list)+1
 my random alignment A list[[end of file]] = paste(gsub("\t","",row.names(COI alignment)[i]),
                                                   paste(COI_alignment[i,], collapse = ""),sep="
")
}
#Write "End of file"
my random alignment A list[[length(my random alignment A list)+1]] = "//end of file"

if(!dir.exists("./50_loci_random_A/Hessian")) dir.create("./50_loci_random_A/Hessian")
if(!dir.exists("./50_loci_random_A/approx1")) dir.create("./50_loci_random_A/approx1")
if(!dir.exists("./50_loci_random_A/approx2")) dir.create("./50_loci_random_A/approx2")

write(unlist(my random alignment A list),file =
"./50 loci random A/Hessian/my random based alignment A.phy")
write(unlist(my_random_alignment_A_list),file =
"./50_loci_random_A/approx1/my_random_based_alignment_A.phy")
write(unlist(my random alignment A list),file =
"./50 loci random A/approx2/my random based alignment A.phy")


#_____________________________

#random based B
random selcted loci B=read.table("./50 loci random B/random selcted UCEloci B.txt")
random_selcted_loci_B=as.character(random_selcted_loci_B[,1])

#Get the loci we want
random loci B=list()
for (i in 1:length(random_selcted_loci_B))
random loci B[i]=my alignments[grep(random selcted loci B[i],names(my alignments))]

my_loci_names=c()
my_charset=c()
for(i in 1:length(random_loci_B)){
  if(i==1){
    my random alignment B=random loci B[[i]]
    my_loci_names[i]=random_selcted_loci_B[i]
    my_charset[i]=paste(1,length(my_random_alignment_B[1,]))
  }
  if(i!=1){

    temp=random_loci_B[[i]]

    temp_not_align= setdiff(row.names(temp),row.names(my_random_alignment_B))
    align not temp= setdiff(row.names(my random alignment B),row.names(temp))

    #Add the new species to the alignment. Put "-" gap

write.table(data.frame(my_loci_names,my_charset),
 file="./50_loci_random_B/my_random_based_charset_B.txt",
```

```
            sep="\t",col.names = F,row.names = F,quote = F)
my_random_alignment_B = my_random_alignment_B[paste(UCE_samples,"\t",sep=""),]
my_random_alignment_B_list = list()
my random alignment B list[[1]] =
paste(nrow(my random alignment B),ncol(my random alignment B),sep="\t")
for(i in 1:nrow(my_random_alignment_B)){
 my_random_alignment_B_list[[i+1]] = paste(gsub("\t","",row.names(my_random_alignment_B)[i]),
      paste(my_random_alignment_B[i,], collapse = ""),sep="     ")
}
#Add the COI alignment
my random alignment B list[[length(my random alignment B list)+1]] = "\n"
my random alignment B list[[length(my random alignment B list)+1]] =
paste(nrow(COI_alignment),ncol(COI_alignment),sep="\t")
for(i in 1:nrow(COI_alignment)){
 my_random_alignment_B_list[[length(my_random_alignment_B_list)+1]] =
paste(gsub("\t","",row.names(COI alignment)[i]),

paste(COI alignment[i,], collapse = ""),sep="     ")
}
#Write "End of file"
my random alignment B list[[length(my random alignment B list)+1]] = "//end of file"

if(!dir.exists("./50_loci_random_B/Hessian")) dir.create("./50_loci_random_B/Hessian")
if(!dir.exists("./50_loci_random_B/approx1")) dir.create("./50_loci_random_B/approx1")
if(!dir.exists("./50_loci_random_B/approx2")) dir.create("./50_loci_random_B/approx2")

write(unlist(my random alignment B list),file =
"./50 loci random B/Hessian/my random based alignment B.phy")
write(unlist(my_random_alignment_B_list),file =
"./50_loci_random_B/approx1/my_random_based_alignment_B.phy")
write(unlist(my random alignment B list),file =
"./50 loci random B/approx2/my random based alignment B.phy")

######################################################################
# TREE input
tre_clean = iqtree_tre
tre clean$edge.length<-NULL

#rename the node labels to the node number
for(i in 1:length(tre_clean$node.label))  tre_clean$node.label[i]= length(tre_clean$tip.label)+i

## SPHINGIDAE/Saturniidae CALIBRATION ##
node=findMRCA(tre clean, tips=tre clean$tip.label, type=c("node","height"))
tre_clean$node.label[tre_clean$node.label==as.character(node)]= "'>.64128<.847602'"

## SPHINGIDAE CALIBRATION ##
node=findMRCA(tre_clean, tips=c("SPHI00232_0101","RROU00478_0101","RROU00591_0101",
                                "RROU00566 0101","RROU00254 0101","SPHI00206 0101",
                                "RROU00277 0101","SPHI00163 0101","SPHI00171 0101",
                                "RROU00209 0101","JRAS05755 0101"), type=c("node","height"))
tre_clean$node.label[tre_clean$node.label==as.character(node)]= "'>.265833<.555616'"

## SATURNIIDAE CALIBRATION ##
node=findMRCA(tre clean, tips=c("SPHI00639 0101","SPHI00895 0101"), type=c("node","height"))
tre_clean$node.label[tre_clean$node.label==as.character(node)]= "'>.4847<.69875'"

## Cercophaninae/Other Saturniidae CALIBRATION ##
node=findMRCA(tre_clean, tips=c("SPHI00639_0101","SPHI00895_0101"), type=c("node","height"))
tre_clean$node.label[tre_clean$node.label==as.character(node)]= "'>.429783<.641207'"

## Ceratocampinae/Hemileucinae CALIBRATION ##
node=findMRCA(tre clean, tips=c("SPHI00488 0101","SPHI00895 0101"), type=c("node","height"))
tre_clean$node.label[tre_clean$node.label==as.character(node)]= "'>.269071<.496079'"

## Salassinae/Saturniinae CALIBRATION ##
node=findMRCA(tre clean, tips=c("SPHI00460 0101","SPHI00423 0101"), type=c("node","height"))
tre_clean$node.label[tre_clean$node.label==as.character(node)]= "'>.238774<.454925'"

### FOSSILS ###
## Smerinthini : Mioclanis ##
node=findMRCA(tre clean, tips=c("RROU00277 0101","SPHI00163 0101","SPHI00171 0101",
 "RROU00209 0101","JRAS05755 0101"), type=c("node","height"))
tre_clean$node.label[tre_clean$node.label==as.character(node)]= "'>.152'"

## Base Bunaeini ##
node=findMRCA(tre_clean, tips=c("SPHI00357_0101","SPHI00777_0101"), type=c("node","height"))
```

```
tre_clean$node.label[tre_clean$node.label==as.character(node)]= "'>.0366'"

for(i in 1:length(tre_clean$node.label)){
 if(tre_clean$node.label[i]==as.character(i+length(tre_clean$tip.label)))
tre clean$node.label[i]=""
}
write.tree(tre_clean,file="./DATATION_input_backbone.tre")
text_tree=readLines("./DATATION_input_backbone.tre")

mcmc text tree=list()
mcmc text tree[[2]]=text tree
mcmc text tree[[1]]=paste(length(tre clean$tip.label),1,sep=" ")
mcmc_text_tree[[3]]="//end of file"

write(unlist(mcmc_text_tree),"./DATATION_input_backbone.tre")

file.copy("./DATATION input backbone.tre", "./50 loci random A/Hessian/")
file.copy("./DATATION input backbone.tre", "./50 loci random A/approx1/")
file.copy("./DATATION_input_backbone.tre", "./50_loci_random_A/approx2/")

file.copy("./DATATION input backbone.tre", "./50 loci random B/Hessian/")
file.copy("./DATATION input backbone.tre", "./50 loci random B/approx1/")
file.copy("./DATATION_input_backbone.tre", "./50_loci_random_B/approx2/")

######################################################################
# MCMCtree control file and run.sh

#MCMCtree control file
Control_Hessian_A = list()
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'seed = -1'
Control Hessian A[[length(Control Hessian A)+1]] = 'seqfile = my random based alignment A.phy'
Control Hessian A[[length(Control Hessian A)+1]] = 'treefile = DATATION input backbone.tre'
Control Hessian A[[length(Control Hessian A)+1]] = 'outfile = mcmctree RDMA 3M cor.txt'
Control Hessian A[[length(Control Hessian A)+1]] = 'ndata = 2'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'seqtype = 0'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'usedata = 3'
Control Hessian A[[length(Control Hessian A)+1]] = 'clock = 3'
Control Hessian A[[length(Control Hessian A)+1]] = "RootAge = '<1'"
Control Hessian A[[length(Control Hessian A)+1]] = 'model = 4'
Control Hessian A[[length(Control Hessian A)+1]] = 'alpha = 0.5'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'ncatG = 4'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'cleandata = 0'
Control Hessian A[[length(Control Hessian A)+1]] = 'BDparas = 1 1 0.1'
Control Hessian A[[length(Control Hessian A)+1]] = 'kappa gamma = 6 2'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'alpha_gamma = 1 1'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'rgene_gamma = 2 20 1'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'sigma2_gamma = 1 10 1'
Control Hessian A[[length(Control Hessian A)+1]] = 'finetune = 1: .1 .1 .1 .1 .1 .1'
Control Hessian A[[length(Control Hessian A)+1]] = 'print = 1'
Control Hessian A[[length(Control Hessian A)+1]] = 'burnin = 300000'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'sampfreq = 100'
Control_Hessian_A[[length(Control_Hessian_A)+1]] = 'nsample = 17000'

Control approx A = Control Hessian A
Control approx A[[7]] = 'usedata = 2'

write(unlist(Control_Hessian_A), file = "./50_loci_random_A/Hessian/mcmctree.ctl")
write(unlist(Control_approx_A), file = "./50_loci_random_A/approx1/mcmctree.ctl")
write(unlist(Control_approx_A), file = "./50_loci_random_A/approx2/mcmctree.ctl")

Control Hessian B = Control Hessian A
Control Hessian B[[2]]  = 'seqfile = my random based alignment B.phy'
Control_Hessian_B[[4]]  = 'outfile = mcmctree_RDMB_3M_cor.txt'

Control_approx_B = Control_Hessian_B
Control approx B[[7]] = 'usedata = 2'

write(unlist(Control_Hessian_B), file = "./50_loci_random_B/Hessian/mcmctree.ctl")
write(unlist(Control_approx_B), file = "./50_loci_random_B/approx1/mcmctree.ctl")
write(unlist(Control_approx_B), file = "./50_loci_random_B/approx2/mcmctree.ctl")

# run MCMCtree.sh files
my runMCMCtree Hessian = list()
my_runMCMCtree_Hessian[[length(my_runMCMCtree_Hessian)+1]] = '#!/bin/sh'
my_runMCMCtree_Hessian[[length(my_runMCMCtree_Hessian)+1]] = '#SBATCH -p workq'
my_runMCMCtree_Hessian[[length(my_runMCMCtree_Hessian)+1]] = '#Load modules'
```

```
my_runMCMCtree_Hessian[[length(my_runMCMCtree_Hessian)+1]] = 'module load bioinfo/paml4.9h'
my_runMCMCtree_Hessian[[length(my_runMCMCtree_Hessian)+1]] = 'mcmctree mcmctree.ctl'

my runMCMCtree approx = list()
my runMCMCtree approx[[length(my runMCMCtree approx)+1]] = '#!/bin/sh'
my runMCMCtree approx[[length(my runMCMCtree approx)+1]] = '#SBATCH -p unlimitq'
my_runMCMCtree_approx[[length(my_runMCMCtree_approx)+1]] = '#Load modules'
my_runMCMCtree_approx[[length(my_runMCMCtree_approx)+1]] = 'module load bioinfo/paml4.9h'
my_runMCMCtree_approx[[length(my_runMCMCtree_approx)+1]] = 'mcmctree mcmctree.ctl'

write(unlist(my runMCMCtree Hessian), file = "./50 loci random A/Hessian/run MCMCtree.sh")
write(unlist(my runMCMCtree approx), file = "./50 loci random A/approx1/run MCMCtree.sh")
write(unlist(my_runMCMCtree_approx), file = "./50_loci_random_A/approx2/run_MCMCtree.sh")

write(unlist(my_runMCMCtree_Hessian), file = "./50_loci_random_B/Hessian/run_MCMCtree.sh")
write(unlist(my runMCMCtree approx), file = "./50 loci random B/approx1/run MCMCtree.sh")
write(unlist(my_runMCMCtree_approx), file = "./50_loci_random_B/approx2/run_MCMCtree.sh")
```

4.1.2 Approximation de la surface de vraisemblance

Une fois les alignements ainsi que les fichiers d'entrée générés, nous avons lancé l'approximation par maximum de vraisemblance. Cette étape créé un fichier `out.BV`, qui est ensuite copié dans les dossiers `./approx[1,2]`.

```
$sbatch launch_backbone_Hessian.sh
```

```
#!/bin/bash
#SBATCH -t 00:10:00

for i in $(ls -d backbone/selected_loci_S88/DATATION/*/*/); do
 cd $i
 dos2unix *
 cd /work/parnal/SAT_splvl_Phylogeny
done

for i in $(ls -d ./backbone/selected_loci_S88/DATATION/*/Hessian/); do
 cd $i
 sbatch run_MCMCtree.sh
 cd /work/parnal/SAT_splvl_Phylogeny
done
```

4.1.3 Inférence des temps de divergence

Une fois cette approximation terminée, nous avons pu lancer les chaines MCMC, c'est-à-dire l'estimation des temps de divergence à proprement parler.

```
$./launch_backbone_Hessian.sh
```

```
#!/bin/bash
#SBATCH -t 00:10:00

for i in $(ls -d backbone/selected_loci_S88/DATATION/*/approx*/); do
 cd $i
 cp ../Hessian/out.BV ./in.BV
 sbatch run_MCMCtree.sh
 cd /work/parnal/SAT_splvl_Phylogeny
done
```

4.1.4 Combinaison des *runs*

Après avoir vérifié que les *runs* indépendants avaient convergé à l'aide de Tracer (Rambaut et al. 2018, tous les Effective Sample Size - ESS>200), ainsi que les temps de divergences obtenus étaient cohérents (script R), nous avons combiné les deux *runs* indépendants à l'aide de Logcombiner (Drummond & Rambaut 2007).

```
$sbatch create_backbone_combined_dir_and_files.sh
```

```
#!/bin/bash

my_dirs=$(ls -d backbone/selected_loci_S88/DATATION/50_loci_random*)

for i in $my_dirs; do
 cd $i || continue
```

```
 mkdir combined
 cd combined

 cp ../approx1/my_random_based_alignment_*.phy .
 cp ../approx1/mcmctree_RDM* .
 cp ../approx1/DATATION_input_backbone.tre .
 cp ../approx1/mcmctree.ctl .
 cp ../approx1/in.BV .
 cp ../approx1/run_MCMCtree.sh .

 sbatch /work/parnal/SAT splvl Phylogeny/run logcombiner.sh

 sleep 1m

 sed -i 's/print = 1/print = -1/g' mcmctree.ctl

 sbatch run_MCMCtree.sh

 cd /work/parnal/SAT_splvl_Phylogeny
done
```

4.2 Datation des *subtrees* et assemblage

Nous avons rassemblé l'ensemble des étapes de la datation des *subtrees* ainsi que l'étape d'assemblage dans un même script. Ici nous avons crée 1 mégaphylogénie et 10 réplicats pour lesquels nous avons utilisé les topologies de bootstrap. Ce nombre devra être augmenté mais il faudra modifier le script pour limiter les temps de calcul (~5h pour n=11 arbres). Le nombre dans la commande suivant indique le nombre de réplicats à générer.

```
$sbatch launch_all_datation_subtrees_files.sh 10
```

```
#!/bin/sh
#SBATCH -t 08:00:00

date

#First get $1 backbones trees from the posterior distribution and store it in a temp folder
module load system/R-3.4.3 ; Rscript get_backbone_trees.R $1

my dirs=$(ls -d /work/parnal/SAT splvl Phylogeny/back*/sel*/DA*/*B/comb*/samp*/*)
for i in $my_dirs;do
  cd $i
  module load bioinfo/paml4.9h ; mcmctree mcmctree.ctl
  cd /work/parnal/SAT_splvl_Phylogeny
done

my_dirs=$(ls -d /work/parnal/SAT_splvl_Phylogeny/*/sel*/ | grep -vw "backbone")
iterations=$(seq 0 $1)

for j in $iterations;do
 echo "----- Iteration "$j" -----"

 for i in $my_dirs;do
  echo $i" - Started!"
  sleep 5s

  cd $i
  rm -R DATATION/

  echo Start creation files

  module load system/R-3.4.3 ; Rscript
/work/parnal/SAT splvl Phylogeny/datation subtrees generatefiles.R $j

  echo Creation files Done!

  approx_dir=$(ls -d ./DA*/*/approx*/)
  cd $approx_dir
  dos2unix *
  module load bioinfo/paml4.9h ; mcmctree mcmctree.ctl

  cd /work/parnal/SAT_splvl_Phylogeny
```

```
  echo $i" - Done!"
 done

 my files=$(ls /work/parnal/SAT splvl Phylogeny/*/sel*/DA*/*/ap*/FigTree* | grep -vw
"backbone")

 for i in $my_files;do
  sed -i 's/ #1.000000//g' $i
  sed -i 's/ #2.000000//g' $i
 done

 Rscript merge_TREE.R $j

done
```

Il repose sur un premier script R qui échantillonne n arbres *backbone* dans la distribution postérieure de MCMCtree. Le script crée également tous les fichiers nécessaires pour que MCMCtree puisse générer les fichiers *nexus* FigTree.tre.

```
$cat /work/parnal/SAT_splvl_Phylogeny/get_backbone_trees.R
library(phytools)

trees nb = commandArgs(trailingOnly=TRUE)

sample_mcmc_path = "backbone/selected_loci_S88/DATATION/50_loci_random_B/combined/mcmc.txt"
sample_mcmc = read.table(sample_mcmc_path,header=T)

print(dim(sample_mcmc))

my dir="backbone/selected loci S88/DATATION/50 loci random B/combined/samples"
if(!dir.exists(my_dir)){
 dir.create(my_dir)
}

for(i in 1:trees_nb){
 my sample = sample mcmc[sample(1:nrow(sample mcmc),1),]
 my_dir = paste("backbone/selected_loci_S88/DATATION/50_loci_random_B/combined/samples/",i,sep="")
 if(!dir.exists(my_dir)){
  dir.create(my dir)
 }
 write.table(my sample,

paste("backbone/selected_loci_S88/DATATION/50_loci_random_B/combined/samples/",i,"/mcmc.txt",sep=""),
  quote=F, sep="\t", row.names=F)


file.copy("backbone/selected_loci_S88/DATATION/50_loci_random_B/combined/DATATION_input_backbone.tre",
  paste("backbone/selected_loci_S88/DATATION/50_loci_random_B/combined/samples/",i,sep=""))

 file.copy("backbone/selected loci S88/DATATION/50 loci random B/combined/mcmctree.ctl",
  paste("backbone/selected loci S88/DATATION/50 loci random B/combined/samples/",i,sep=""))

 file.copy("backbone/selected_loci_S88/DATATION/50_loci_random_B/combined/run_MCMCtree.sh",
  paste("backbone/selected_loci_S88/DATATION/50_loci_random_B/combined/samples/",i,sep=""))

 path = paste("backbone/selected loci S88/DATATION/50 loci random B/combined/samples/",
 i,"/mcmctree.ctl",sep="")
 control file = readLines(path)
 index = grep("seqfile", control_file)
 control_file[index]="seqfile = ../../my_random_based_alignment_B.phy"
 writeLines(control_file,path)

}
```

Il repose également sur un autre script R qui crée les fichiers d'entrée pour dater les *subtrees*.

```
$cat /work/parnal/SAT_splvl_Phylogeny/datation_subtrees_generatefiles.R
library(phytools)
library(phangorn)
library(openxlsx)
source("/work/parnal/software/readMCMCtree_output.R")

iteration = commandArgs(trailingOnly=TRUE)
iteration = as.numeric(iteration)
```

```
print(iteration)

#############################################################################
# import the backbone dated tree
if(iteration==0){
backbone = readMCMCTree(paste("/work/parnal/SAT splvl Phylogeny/backbone/",
 "selected_loci_S88/DATATION/50_loci_random_B/combined/FigTree.tre",sep=""))
}
if(iteration!=0){
backbone = readMCMCTree(paste("/work/parnal/SAT splvl Phylogeny/backbone/",

"selected loci S88/DATATION/50 loci random B/combined/samples/",iteration,"/FigTree.tre",sep=""))
}


#############################################################################
#For rerooting the tree
#reroot the tree
my_outgroup_path = list.files("..", pattern= "outgroup.txt$", full.names = T)
#Get the name of the subtree
my subtree = gsub("\\.\\.\\/","",gsub(" outgroup.txt","", my outgroup path))
my outgroup = readLines(my outgroup path)
#############################################################################

if(iteration==0){
 #Import the tree inferred the IQTREE on the merged alignement
 iqtree path = list.files("./IQTREE/",pattern=".treefile$", full.names = T)
 iqtree tre=read.tree(iqtree path)
 index = which(iqtree_tre$tip.label==my_outgroup)
 iqtree_tre = reroot(iqtree_tre, index,
  position = 0.5*iqtree tre$edge.length[which(iqtree tre$edge[,2]==index)])
 #discard outgroup
 iqtree tre = drop.tip(iqtree tre, my outgroup)

}

if(iteration!=0){
 check=FALSE
 bootstrap index=1
 print("Start While loop")
 #Import a tree from the bootstrap replicates
 bootstrap_path = list.files("./IQTREE/",pattern=".ufboot$", full.names = T)
 bootstrapLines = readLines(bootstrap path)

 while(!check){
  print(bootstrap_index)
  bootstrap_sampled = bootstrapLines[sample(1:length(bootstrapLines),1)]
  write(bootstrap sampled, "./IQTREE/temp.tre")

  iqtree tre=read.tree("./IQTREE/temp.tre")
  file.remove("./IQTREE/temp.tre")

  index = which(iqtree_tre$tip.label==my_outgroup)
  iqtree tre = reroot(iqtree tre, index,
   position = 0.5*iqtree tre$edge.length[which(iqtree tre$edge[,2]==index)])

  #discard outgroup
  iqtree_tre = drop.tip(iqtree_tre, my_outgroup)

  common tips = intersect(backbone$apePhy$tip.label, iqtree tre$tip.label)
  root = length(iqtree tre$tip.label)+1
  cladertight=iqtree tre$edge[which(iqtree tre$edge[,1]==root),2][1]
  cladeleft=iqtree_tre$edge[which(iqtree_tre$edge[,1]==root),2][2]

  tipsright=iqtree_tre$tip.label[Descendants(iqtree_tre,cladertight)[[1]]]
  tipsleft=iqtree tre$tip.label[Descendants(iqtree tre,cladeleft)[[1]]]

  checkright=any(grepl(paste(common_tips,collapse="|"), tipsright))
  checkleft=any(grepl(paste(common_tips,collapse="|"), tipsleft))

  bootstrap index=bootstrap index+1
  if(all(checkright, checkleft)) check=TRUE
  if(bootstrap_index==200){
   check=TRUE
   print("No bootstrap trees with the root we infered in the backbone")
   print("We thus considered the best tree inferred by IQTREE")
```

```
   iqtree_path = list.files("./IQTREE/",pattern=".treefile$", full.names = T)
   iqtree_tre=read.tree(iqtree_path)
   index = which(iqtree_tre$tip.label==my_outgroup)
   iqtree tre = reroot(iqtree tre, index,
   position = 0.5*iqtree tre$edge.length[which(iqtree tre$edge[,2]==index)])
   #discard outgroup
   iqtree_tre = drop.tip(iqtree_tre, my_outgroup)
  }
 }
 print("END of While loop")
}

#################################################################################
#Get the number sample ID
my_UCElist_path = list.files("../", pattern="UCEs_sample.txt$", full.names = T)
UCE samples = readLines(my UCElist path)
#print(UCE samples)
UCE samples = UCE samples[-grep(my outgroup, UCE samples)]

#################################################################################
#Import the COI alignment
COI alignment path = list.files(pattern =  paste("^",my subtree,".*\\.al$",
  sep=""))
COI_alignment = read.dna(COI_alignment_path, as.character=T,
  as.matrix=T, format="fasta")
#Prune for outgroup
COI alignment = COI alignment[-grep(my outgroup,row.names(COI alignment)),]


#################################################################################
#Import the per loci alignements
alignment path = list.files("./PMCOA clean",pattern = "^locus.*\\.phy$",
 full.names = T)
my alignments=list()
for(i in 1:length(alignment_path)) {
 temp_alignment = read.dna(alignment_path[i], as.character=T, as.matrix=T)

 #prune from outgroup
 if(any(grepl(my_outgroup, row.names(temp_alignment)))){
  temp alignment = temp alignment[-grep(my outgroup,row.names(temp alignment)),]
 }
 ########################
 # We discard OTHORENE in that version because of datation issue.
 # HAVE TO INCLUDE in the next versions
 if(any(grepl("SPHI00488_0101", row.names(temp_alignment)))){
  temp_alignment = temp_alignment[-grep("SPHI00488_0101",row.names(temp_alignment)),]
 }
 ########################

 my alignments[[i]] = temp alignment
}

#Alignments names
alignment path names = list.files("./PMCOA clean",
 pattern = "^locus.*\\.phy$",full.names = F)
my loci = c()
for(i in 1:length(alignment_path_names)){
  my_loci = c(my_loci, gsub("_gbrelaxed.phy","",alignment_path_names[i]))
}
# Assign names
names(my alignments) = my loci


#################################################################################
## First random set of 50 loci
#The while loop allows to relaunch the randomization
#until the minimum number of loci for a sample is 10
success <- FALSE

rep=list()
rep min nb = c()
k = 0

while(!(success|(length(rep)==100))){
 k = k + 1
```

```
 random_loci_A=sample(my_loci,50)

 #Count the number of loci per sample
 random loci spcount A=data.frame(UCE samples,rep(0,length(UCE samples)))
 names(random loci spcount A)=c("tip.label","nb")

 for(i in 1:length(random_loci_A)){
   index=grep(random_loci_A[i],my_loci)
   temp_tr_tips=row.names(my_alignments[[index]])
   temp tr tips = gsub("\t","",temp tr tips)
   for(j in 1:length(temp_tr_tips)){
    my row = random loci spcount A$tip.label==temp tr tips[j]
    random_loci_spcount_A[my_row,]$nb=random_loci_spcount_A[my_row,]$nb+1
   }
 }

 rep[[k]] = random loci A
 rep min nb = c(rep min nb, min(random loci spcount A$nb))

 success=min(random_loci_spcount_A$nb)>=10
#print(min(random loci spcount A$nb))

}


if(!success) random_loci_A = rep[[which.max(rep_min_nb)]]

if(min(random_loci_spcount_A$nb)==0) {
  focus_sample = random_loci_spcount_A[random_loci_spcount_A$nb==0,]$tip.label
  stop(paste("Problem with the following sample(s) : ",focus_sample,
       " - Not found in the alignement",sep=""))
}


print("Random set of Loci A - OK !")

################################################################################
# Create dataset
if(!dir.exists("DATATION")) dir.create("DATATION")
setwd("DATATION")

# Statistics

#random A
if(!dir.exists("50_loci_random_A")) dir.create("50_loci_random_A")
write.table(random_loci_A,file="./50_loci_random_A/random_selcted_UCEloci_A.txt",
 row.names = F,col.names = F,quote = F,sep="\t")
write.table(random loci spcount A,
 file="./50 loci random A/species count in selcted UCEloci.txt",
 row.names = F,col.names = F,quote = F,sep="\t")

################################################################################
# Create the alignments

#random A
random selcted loci A=read.table("./50 loci random A/random selcted UCEloci A.txt")
random_selcted_loci_A=as.character(random_selcted_loci_A[,1])

#Get the loci we want
random loci A=list()
for (i in 1:length(random_selcted_loci_A)){
  random loci A[i]=my alignments[grep(random selcted loci A[i],names(my alignments))]
}
my_loci_names=c()
my_charset=c()

for(i in 1:length(random_loci_A)){
  if(i==1){
    my_random_alignment_A=random_loci_A[[i]]
    my_loci_names[i]=random_selcted_loci_A[i]
    my charset[i]=paste(1,length(my random alignment A[1,]))
  }
  if(i!=1){

    temp=random_loci_A[[i]]
```

```
    temp_not_align= setdiff(row.names(temp),row.names(my_random_alignment_A))
    align_not_temp= setdiff(row.names(my_random_alignment_A),row.names(temp))

    #Add the new species to the alignment. Put "-" gap
    to add align=matrix("-",length(temp not align),length(my random alignment A[1,]))
    row.names(to add align)=temp not align
    my_random_alignment_A=rbind(my_random_alignment_A,to_add_align)

    #Add the new species to the temp. Put "-" gap
    to add temp=matrix("-",length(align not temp),length(temp[1,]))
    row.names(to add temp)=align not temp
    temp=rbind(temp,to add temp)

    if(nrow(temp) == nrow(my_random_alignment_A)){
      #Match same order
      temp=temp[match(row.names(my random alignment A),row.names(temp)),]

      if(length((grep(FALSE,row.names(temp) == row.names(my_random_alignment_A))))>0){
       print("row.names not same order")
      }
      if(length((grep(FALSE,row.names(temp) ==  row.names(my_random_alignment_A))))==0) {
        my charset[i]=paste(length(my random alignment A[1,])+1,
                            length(my_random_alignment_A[1,])+length(temp[1,]))
        my_random_alignment_A=cbind(my_random_alignment_A,temp)
        #print(paste(random_selcted_loci_A[i],"-",length(temp[1,]), "pb"))
        my_loci_names[i]=random_selcted_loci_A[i]
      }

    }

    if(nrow(temp) != nrow(my_random_alignment_A)) print("not same dimension")

  }
}

write.table(data.frame(my_loci_names,my_charset),
 file="./50 loci random A/my random based charset A.txt",
 sep="\t",col.names = F,row.names = F,quote = F)
my random alignment A = my random alignment A[paste(UCE samples,"\t",sep=""),]

my_random_alignment_A_list = list()
my_random_alignment_A_list[[1]] = paste(nrow(my_random_alignment_A),
 ncol(my random alignment A),sep="\t")
for(i in 1:nrow(my_random_alignment_A)){
 my_random_alignment_A_list[[i+1]] = paste(gsub("\t","",
                                           row.names(my_random_alignment_A)[i]),
                                           paste(my_random_alignment_A[i,],
                                                 collapse = ""),sep="     ")
}
#Add the COI alignment
my_random_alignment_A_list[[length(my_random_alignment_A_list)+1]] = "\n"
my_random_alignment_A_list[[length(my_random_alignment_A_list)+1]] = paste(nrow(COI_alignment),
 ncol(COI_alignment),sep="\t")
for(i in 1:nrow(COI_alignment)){
 my random alignment A list[[length(my random alignment A list)+1]] = paste(gsub("\t","",
  row.names(COI alignment)[i]), paste(COI alignment[i,], collapse = ""),sep="     ")
}

#Write "End of file"
my random alignment A list[[length(my random alignment A list)+1]] = "//end of file"

if(!dir.exists("./50_loci_random_A/approx1")) dir.create("./50_loci_random_A/approx1")
#if(!dir.exists("./50_loci_random_A/approx2")) dir.create("./50_loci_random_A/approx2")

write(unlist(my_random_alignment_A_list),
 file = paste("./50 loci random A/approx1/my random based ",
 my subtree," alignment A.phy",sep=""))



#############################################################################
# TREE input
tre clean = iqtree tre
tre_clean$edge.length<-NULL

#rename the node labels to the node number
```

```
#for(i in 1:length(tre_clean$node.label))  tre_clean$node.label[i]= length(tre_clean$tip.label)+i

tre_clean$node.label =
(length(tre clean$tip.label)+1):(length(tre clean$tip.label)+tre clean$Nnode)

tree calib = drop.tip(tre clean, setdiff(tre clean$tip.label, backbone$apePhy$tip.label))

# Get all the clades and assign calibration

for(i in 1:tree_calib$Nnode) {
 my node = length(tree calib$tip.label)+i
 tips = tree calib$tip.label[Descendants(tree calib, my node)[[1]]]


#######
#These lines ares added in order to discard Dirphiopsis from the analysis and
#thus consider monophyly at the root of the Automerina C subtree
# TO REMOVE WHEN NO PB with Dirphiopsis : when UCEs are availables !
backbonebis = drop.tip(backbone$apePhy, c("SAFMA214_08", "SACMA283_09", "SAMPA596_17"))
monophyl = is.monophyletic(backbonebis, tips)

######

#monophyl = is.monophyletic(backbone$apePhy, tips) # TO UNCOMMENT in next versions

 if(monophyl){
  backbone node = getMRCA(backbone$apePhy, tips)
  calib = backbone$nodeAges[as.character(backbone_node),"mean"]

  node_bis = getMRCA(tre_clean, tips)
  temp_index = tre_clean$node.label==as.character(node_bis)
  tre clean$node.label[temp index]=paste("'>",calib,"<",calib,"'",sep="")

 }
 if(!monophyl) print(paste(paste(tips,collapse=" | "),
  " - Non Monophyletic in the backbone tree !!! Different Topology !!!", sep=""))
}

for(i in 1:length(tre_clean$node.label)){
 if(tre_clean$node.label[i]==as.character(i+length(tre_clean$tip.label)))
tre_clean$node.label[i]=""
}
write.tree(tre clean,file="./DATATION input backbone.tre")
text tree=readLines("./DATATION input backbone.tre")

mcmc_text_tree=list()
mcmc_text_tree[[2]]=text_tree
mcmc text tree[[1]]=paste(length(tre clean$tip.label),1,sep=" ")
mcmc text tree[[3]]="//end of file"

write(unlist(mcmc_text_tree),"./DATATION_input_backbone.tre")

file.copy("./DATATION_input_backbone.tre", "./50_loci_random_A/approx1/")
#file.copy("./DATATION input backbone.tre", "./50 loci random A/approx2/")

#file.copy("./DATATION input backbone.tre", "./50 loci random B/approx1/")
#file.copy("./DATATION_input_backbone.tre", "./50_loci_random_B/approx2/")

file.remove("./DATATION_input_backbone.tre")

##############################################################################
# MCMCtree control file and run.sh

#MCMCtree control file
Control_A = list()
Control A[[length(Control A)+1]] = 'seed = -1'
Control A[[length(Control A)+1]] = paste('seqfile = my random based ',
 my_subtree,'_alignment_A.phy',sep="")
Control_A[[length(Control_A)+1]] = 'treefile = DATATION_input_backbone.tre'
Control_A[[length(Control_A)+1]] = 'outfile = mcmctree_RDMA_3M_cor.txt'
Control A[[length(Control A)+1]] = 'ndata = 2'
Control A[[length(Control A)+1]] = 'seqtype = 0'
Control A[[length(Control A)+1]] = 'usedata = 1'
Control_A[[length(Control_A)+1]] = 'clock = 3'
Control_A[[length(Control_A)+1]] = "RootAge = '<1'"
Control_A[[length(Control_A)+1]] = 'model = 4'
```

```
Control_A[[length(Control_A)+1]] = 'alpha = 0.5'
Control_A[[length(Control_A)+1]] = 'ncatG = 4'
Control_A[[length(Control_A)+1]] = 'cleandata = 0'
Control_A[[length(Control_A)+1]] = 'BDparas = 1 1 0.1'
Control_A[[length(Control_A)+1]] = 'kappa gamma = 6 2'
Control_A[[length(Control_A)+1]] = 'alpha gamma = 1 1'
Control_A[[length(Control_A)+1]] = 'rgene_gamma = 2 20 1'
Control_A[[length(Control_A)+1]] = 'sigma2_gamma = 1 10 1'
Control_A[[length(Control_A)+1]] = 'finetune = 1: .1 .1 .1 .1 .1 .1'
Control_A[[length(Control_A)+1]] = 'print = 1'
Control_A[[length(Control_A)+1]] = 'burnin = 0'
Control_A[[length(Control_A)+1]] = 'sampfreq = 1'
Control_A[[length(Control_A)+1]] = 'nsample = 1'

write(unlist(Control_A), file = "./50_loci_random_A/approx1/mcmctree.ctl")
#write(unlist(Control_A), file = "./50_loci_random_A/approx2/mcmctree.ctl")

# run MCMCtree.sh files
my_runMCMCtree = list()
my_runMCMCtree[[length(my_runMCMCtree)+1]] = '#!/bin/sh'
my_runMCMCtree[[length(my_runMCMCtree)+1]] = '#SBATCH -p workq'
my_runMCMCtree[[length(my_runMCMCtree)+1]] = '#Load modules'
my_runMCMCtree[[length(my_runMCMCtree)+1]] = 'module load bioinfo/paml4.9h'
my_runMCMCtree[[length(my_runMCMCtree)+1]] = 'mcmctree mcmctree.ctl'

write(unlist(my_runMCMCtree), file = "./50_loci_random_A/approx1/run_MCMCtree.sh")
```

Enfin, il repose sur un dernier script R qui va assembler le *backbone* avec les *subtrees* à l'aide de la fonction *bind.tree* du package *ape* (Paradis & Schliep 2019). A noter qu'il en résulte un arbre qui a de très légers défauts en cela qu'il n'est pas ultramétrique à proprement parler. Cela est dû aux marges d'erreurs des datations avec *MCMCtree*. Nous avons donc utilisé la fonction *force.ultrametric* du package *phytools* (Revell 2012).

```
$cat merge_TREE.R
library(phytools)
source("/work/parnal/software/readMCMCtree_output.R")

iteration = commandArgs(trailingOnly=TRUE)
iteration = as.numeric(iteration)

#Import the backbone
if(iteration==0){
 path = paste("/work/parnal/SAT splvl Phylogeny/backbone/",
  "selected loci S88/DATATION/50 loci random B/combined/FigTree.tre",sep="")
}
if(iteration!=0){
 path = paste("/work/parnal/SAT_splvl_Phylogeny/backbone/",
  "selected loci S88/DATATION/50 loci random B/combined/samples/",
  iteration,"/FigTree.tre",sep="")
}


backbone = readMCMCTree(path)
backbone = backbone$apePhy

#discard outgroups
backbone = drop.tip(backbone,c("SPHI00232_0101","RROU00478_0101",
 "RROU00591_0101","RROU00566_0101","RROU00254_0101","RROU00277_0101",
 "SPHI00206_0101","RROU00209_0101","SPHI00163_0101","SPHI00171_0101",
 "JRAS05755_0101"))

tr_paths = list.files("/work/parnal/SAT_splvl_Phylogeny",
 recursive = T, full.names=T, pattern = "FigTree.tre")
tr_paths = grep("approx",tr_paths, value=T)
tr_paths = tr_paths[-grep("backbone",tr_paths)]

print(tr_paths)

full_tr = backbone

for(i in 1:length(tr_paths)){

 subtree = strsplit(tr_paths[i],"\\/")[[1]][5]
```

```
 print(paste(subtree, "- started"))

 tr = readMCMCTree(tr_paths[i])
 tr = tr$apePhy

 tips = intersect(tr$tip.label,full tr$tip.label)
 node = getMRCA(full_tr, tips)

 for(j in 1:length(tips)){
  index = grep(tips[j],full tr$tip.label)
  if(length(index)==1){
   full tr$tip.label[index] = paste(full tr$tip.label[index],
                                    "to_remove",sep="")
  }
  if(length(index)!=1) stop(paste("Error: ",tips[j],
                            " matched",length(index)," tips!",sep=""))
 }

 full_tr = bind.tree(full_tr,tr, where = node)

 if(length(grep("to_remove", full_tr$tip.label,value = T))!=length(tips)){
  stop("!! Common tips and removed tips not same length !!")
 }

 full_tr = drop.tip(full_tr, grep("to_remove", full_tr$tip.label,value = T))

 print(paste("Tree length: ",length(full_tr$tip.label),"tips!"))

 print(paste(subtree, "- Done!"))

}

if(!dir.exists("/work/parnal/SAT_splvl_Phylogeny/RESULTS")){
 dir.create("/work/parnal/SAT splvl Phylogeny/RESULTS")
}

full tr = force.ultrametric(full tr)

path=paste("/work/parnal/SAT splvl Phylogeny/RESULTS/Saturniidae",iteration,".tre",sep="")
write.tree(full_tr,path)
```

Les arbres complets sont écrits dans le dossier '*./RESULTS*'.

Références

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution, 17(4), 540-552.

de Vienne, D. M., Ollier, S., & Aguileta, G. (2012). Phylo-MCOA: a fast and efficient method to detect outlier genes and species in phylogenomics using multiple co-inertia analysis. Molecular Biology and Evolution, 29(6), 1587-1598.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology, 7(1), 1-8.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14(6), 587-589.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution, 32(1), 268-274.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526-528.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology, 67(5), 901.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in ecology and evolution, 3(2), 217-223.

Tagliacollo, V. A., & Lanfear, R. (2018). Estimating improved partitioning schemes for ultraconserved elements. Molecular Biology and Evolution, 35(7), 1798-1811.

Wahlberg, N., Wheat, C. W., & Peña, C. (2013). Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLOS one, 8(11), e80875.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences, 13(5), 555-556.

Conclusion générale

I] Avancée des connaissances systématiques sur la famille des Saturniidae & utilisation des marqueurs UCEs

Les Saturniidae Boisduval 1837 (superfamille des Bombycoidea) forment une famille de Lépidoptères particulièrement diversifiée et populaire, suscitant un intérêt fort des entomologistes. Pourtant, aucune hypothèse phylogénétique basée sur un échantillonnage conséquent n'avait été inférée jusqu'à maintenant (les positions relatives de différentes sous-familles étaient en revanche reconnues ; Regier *et al.* 2008 ; Barber *et al.* 2015 ; Hamilton *et al.* 2019). Établir les relations phylogénétiques de rang supérieur au sein de la famille des Saturniidae a donc été le premier objectif de ma thèse. Dans le Chapitre 1, j'ai généré un jeu de données, comprenant 1381 *loci* très conservés (*Ultra Conserved Elements* - UCEs) pour l'ensemble des 180 genres de Saturniidae décrits, à partir duquel j'ai inféré une hypothèse phylogénétique particulièrement soutenue. Il faut souligner qu'un tel échantillonnage exhaustif au niveau du genre est exceptionnel, car il inclut des représentants de genres monospécifiques ou très peu diversifiés, particulièrement rares (par ex. *Microdulia, Pseudoludia, Eosia, Pararhodia, Jaiba, Arias*, etc.). La phylogénie produite m'a permis de proposer une nouvelle classification des Saturniidae, présentée dans le Chapitre 1. Cette classification fera l'objet d'une publication dans laquelle les résultats phylogénétiques seront plus amplement discutés et corroborés à des caractères morphologiques.

Les résultats phylogénétiques obtenus au cours de ma thèse démontrent la pertinence de l'utilisation des marqueurs UCE dans l'inférence de relations phylogénétiques de profondeurs variées : entre les différentes sous-familles, tribus ou genres (Chapitres 1 et 3), à l'intérieur des différents genres (Chapitres 2 et 3) ou groupant des espèces très proches (Chapitre 3, au sein du genre *Eacles* par ex.). L'intérêt des UCEs s'explique par le fait que ceux-ci peuvent être découpés en 3 régions évoluant à des rythmes distincts : une région très conservée et deux régions flanquantes qui la juxtaposent. Depuis près d'une demi-douzaine d'année, l'utilisation de tels marqueurs s'est généralisée (par ex. Streicher & Wiens 2016 ; Branstetter *et al.* ; Prebus 2017 ; Zhang et al. 2020) et est très prometteuse pour les inférences phylogénétiques d'une multitude d'organismes : en octobre 2020, 11 ensemble de *loci* UCE (notamment 7 pour les invertébrés ; Faircloth 2017) étaient disponibles en téléchargement sur le site www.ultraconserved.org. Plus globalement, ce sont les méthodes de capture par hybridation (Voir Introduction générale, Figure 3 ; Lemmon & Lemmon 2012) qui révolutionnent aujourd'hui notre façon de générer des données génomiques (notamment en systématique ; par ex. Branstetter et al. 2017 ;

Espeland et al. 2018 ; Kawahara et al. 2019). Ces méthodes, qui reposent sur le lien streptavidine-biotine afin d'isoler les séquences ADN d'intérêt, peuvent être adaptées sur un panel de marqueurs particulièrement larges ; le seul prérequis étant la disponibilité de génomes de référence dont la divergence doit correspondre à l'échelle étudiée (par ex. phylogénies *higher-level* ou intra-spécifiques). Il ne fait pas de doute que l'utilisation de ces méthodes continuera à se généraliser dans les prochaines années et permettra de répondre à une grande diversité de questions allant de l'étude des processus de spéciation (par ex. Dömel et al. 2019) aux signatures génétiques de la sélection (par ex. Yang et al. 2019).

II] L'influence des traits d'histoire de vie sur la dynamique spatiale et temporelle des Saturniidae

Les différentes analyses présentées dans les deux premiers Chapitres soulignent l'importance des facteurs biotiques dans la diversification spatiale des Saturniidae (Chapitres 1, 2), bien qu'elle soit également conditionnée par des facteurs abiotiques tels que les conditions climatiques. Nos résultats suggèrent qu'effectivement la capacité de tisser des cocons pleins et denses, qui constituent une couche protectrice à la fois contre les parasites ou les prédateurs mais aussi contre le froid, permettraient la colonisation de nouvelles entités géographiques (notamment à travers des corridors aux climats tempérés). La polyphagie apparaît également comme un facteur ayant influencé positivement la diversification spatiale du groupe : les lignées polyphages ayant plus de facilités à trouver des plantes hôtes dans un nouvel environnement que les lignées spécialistes. L'influence des traits biotiques sur la dynamique spatiale d'un clade de Saturniidae a aussi été démontrée dans le Chapitre 2 : la niche climatique des *Copaxa*, héritée d'un ancêtre distribué dans la région Holarctique, a effectivement façonné le patron de diversification spatiale « Into the tropics » que nous avons établi : la majorité des espèces de *Copaxa* volent dans les zones montagneuses et les deux colonisations indépendantes de la chaine andine ont impliqué des shifts positifs des taux de diversification (évènements de *dispersification*).

D'une manière similaire à ce que nous avons inféré pour la diversification spatiale, l'évolution de traits d'histoire de vie caractéristiques des *capital breeders* a influencé de manière significative la diversification temporelle des Saturniidae (Chapitre 1). Ces résultats seraient particulièrement intéressant à comparer aux patrons de diversification au sein des Sphingidae, le groupe frère des Saturniidae qui, au contraire de ces derniers, se nourrissent à l'état adulte et peuvent être qualifiés de *income breeders*. L'inférence de la phylogénie des Sphingidae et de leur diversification est justement en cours d'investigation par les membres du groupe de travail de l'ANR SPHINX. En plus de l'étude réalisée sur le genre *Copaxa*, nous prévoyons également de caractériser les dynamiques de

diversification d'autres clades de Saturniidae (par ex. *Lonomia* (Diaz Diaz *et al.* in prep.), *Dirphia* (Reboud *et al.* in prep.), *Eacles*, Attacini, *Cricula* et *Actias* (Rougerie *et al.* in prep.)) dans d'autres contextes géographiques (Afrique, Asie), afin d'identifier de potentiels patrons récurrents de diversification ou de mieux en comprendre les contrastes. Au sein des Attacini, par exemple, nous nous intéresserons tout particulièrement au genre *Rothschildia* qui lui aussi a une origine Holarctique et s'est diversifié dans les Néotropiques. La comparaison de leurs préférences altitudinales, de la chronologie des évènements de colonisation et des taux de diversification avec ceux des *Copaxa* nous permettra de comprendre si le patron « Into the tropics » est également façonné par le conservatisme phylogénétique de niche au sein du genre *Rothschildia*. La génération de phylogénies nous permettra également d'inférer l'influence des traits biotiques sur l'histoire évolutive de ces groupes et nous permettra ainsi d'établir des patrons récurrents de diversification ou de ou dévoiler une diversité de modalités d'évolution régies par des caractéristiques propres aux lignées.

III] Une avancée significative vers la génération d'une phylogénie complète de la famille des Saturniidae

Les méthodes macroécologiques et macroévolutives actuelles reposent aujourd'hui largement sur les phylogénies (par ex. Tonini *et al.* 2016 ; Jetz & Pyron 2018 ; Chazot *et al.* 2020) et l'étude des patrons globaux de la biodiversité nécessite des supports phylogénétiques particulièrement larges, au gand pouvoir statistique : les mégaphylogénies (Chapitre 3). De nombreuses mégaphylogénies ont été générées pour les Vertébrés (par ex. Jetz *et al.* 2012 ; Upham *et al.* 2019), mais aucune phylogénie ayant une complétion similaire n'a jusqu'alors été estimée chez les insectes, entravant de fait notre capacité à expliquer les patrons de diversité globaux chez ces organismes dont la diversité est pourtant considérable. Dans ma thèse, je présente les résultats provisoires de la construction de la première mégaphylogénie inférée pour un clade d'insecte. Cette phylogénie constituera le support de futures analyses qui nous permettront, par exemple, de comprendre si les tendances évolutives de certains traits biotiques sont les mêmes au sein des différents genres que celles pouvant être inférées au niveau de divergences beaucoup plus anciennes (au sein de la phylogénie *higher-level*). Par exemple, dans le Chapitre 1 nous avons estimé que les lignées de Saturniidae devenaient globalement plus polyphages au fil de leur évolution. Cette tendance évolutive se retrouvera-t-elle à l'intérieur de genres ? Ou est-ce que plusieurs lignées indépendantes évolueraient vers la spécialisation, comme prédit par la théorie des oscillations (O*scillations hypothesis* ; Janz & Nylin 2008) ? La taille est également un trait particulièrement hétérogène (au sein par exemple des *Automeris*), dont l'étude des processus de diversification et d'évolution à partir d'une mégaphylogénie permettrait l'identification de convergences morphologiques et biologiques (par ex. présence de queues, d'ocelles, utilisation de plantes hôtes

similaires) qui nous permettrait de progresser dans notre compréhension de la diversité des Saturniidae et, plus généralement, des Lépidoptères.

Dans leur ensemble, les travaux que j'ai effectués dans le cadre de ma thèse permettent une meilleure compréhension de la systématique et de l'évolution de la famille des Saturniidae. Les résultats obtenus ouvrent la voix à de futures recherches permettant de mieux comprendre les particularités de l'évolution de la famille vis-à-dis des autres groupes de Lépidoptères ainsi qu'à une meilleure compréhension des patrons d'évolution des insectes.

Références

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten years of progress following the 'revolution'. Italian Journal of Zoology, 71(1), 5-16.

Alfaro, M. E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D. L., ... & Harmon, L. J. (2009). Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences, 106(32), 13410-13414.

Barber, J. R., Leavell, B. C., Keener, A. L., Breinholt, J. W., Chadwell, B. A., McClure, C. J., ... & Kawahara, A. Y. (2015). Moth tails divert bat attack: evolution of acoustic deflection. Proceedings of the National Academy of Sciences, 112(9), 2812-2816.

Barthlott, W., Hostert, A., Kier, G., Küper, W., Kreft, H., Mutke, J., ... & Sommer, J. H. (2007). Geographic patterns of vascular plant diversity at continental to global scales (Geographische Muster der Gefäßpflanzenvielfalt im kontinentalen und globalen Maßstab). Erdkunde, 305-315.

Beeravolu, C. R., & Condamine, F. L. (2016). An extended maximum likelihood inference of geographic range evolution by dispersal, local extinction and cladogenesis. BioRxiv.

Bénéluz, F. (1986). Description d'un *Copaxa* inédit du Costa Rica (Lepidoptera, Saturniidae). Revue française d'entomologie (1979), 8(2), 88-90.

Bininda-Emonds, O. R., Cardillo, M., Jones, K. E., MacPhee, R. D., Beck, R. M., Grenyer, R., ... & Purvis, A. (2007). The delayed rise of present-day mammals. Nature, 446(7135), 507-512.

Branstetter, M. G., Danforth, B. N., Pitts, J. P., Faircloth, B. C., Ward, P. S., Buffington, M. L., ... & Brady, S. G. (2017). Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. Current Biology, 27(7), 1019-1025.

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., ... & Gušić, I. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science, 325(5938), 318-321.

Brito, D. (2010). Overcoming the Linnean shortfall: data deficiency and biological survey priorities. Basic and Applied Ecology, 11(8), 709-713.

Bouvier, E. L. (1927). Les Saturniens du genre *Aurivillius*. Bulletin du Muséum National d'Histoire Naturelle, 33, 71-75.

Cardillo, M., Orme, C. D. L., & Owens, I. P. (2005). Testing for latitudinal bias in diversification rates: an example using New World birds. Ecology, 86(9), 2278-2287.

Chaw, S. M., Parkinson, C. L., Cheng, Y., Vincent, T. M., & Palmer, J. D. (2000). Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proceedings of the National Academy of Sciences, 97(8), 4086-4091.

Chazot, N., Condamine, F., Dudas, G., Peña, C., Matos-Maraví, P., Freitas, A. V., ... & Lohman, D. J. (2020). The latitudinal diversity gradient in brush-footed butterflies (Nymphalidae): conserved ancestral tropical niche but different continental histories. bioRxiv.

Chen, F. C., & Li, W. H. (2001). Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. The American Journal of Human Genetics, 68(2), 444-456.

Church, G. (2006). The race for the $1000 genome. Science, 311, 1544-1546.

Condamine, F. L., Rolland, J., & Morlon, H. (2013). Macroevolutionary perspectives to environmental change. Ecology letters, 16, 72-85.

Costello, M. J., May, R. M., & Stork, N. E. (2013). Can we name Earth's species before they go extinct? Science, 339(6118), 413-416.

Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., ... & Rasplus, J. Y. (2019). Optimized DNA extraction and library preparation for minute arthropods: application to target enrichment in chalcid wasps used for biocontrol. Molecular Ecology Resources, 19(3), 702-710.

Darwin, C. R. & Wallace, A. R. 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. Journal of the Proceedings of the Linnean Society of London, Zoology, 3(9), 45-62.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics, 6(5), 361-375.

D'Erchia, A. M., Gissi, C., Pesole, G., Saccone, C., & Arnason, U. (1996). The guinea-pig is not a rodent. Nature, 381(6583), 597-600.

Diniz-Filho, J. A. F., De Marco, P., & Hawkins, B. A. (2010). Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. Insect Conservation and Diversity, 3, 172–179.

Diniz-Filho, J. A. F., Loyola, R. D., Raia, P., Mooers, A. O., & Bini, L. M. (2013). Darwinian shortfalls in biodiversity conservation. Trends in Ecology & Evolution, 28(12), 689-695.

Dömel, J. S., Macher, T. H., Dietz, L., Duncan, S., Mayer, C., Rozenberg, A., ... & Melzer, R. R. (2019). Combining morphological and genomic evidence to resolve species diversity and study speciation processes of the Pallenopsis patagonica (Pycnogonida) species complex. Frontiers in Zoology, 16(1), 36.

Dupuis, J. R., Peigler, R. S., Geib, S. M., & Rubinoff, D. (2018). Phylogenomics supports incongruence between ecological specialization and taxonomy in a charismatic clade of buck moths. Molecular Ecology, 27(22), 4417-4429.

Economo, E. P., Narula, N., Friedman, N. R., Weiser, M. D., & Guénard, B. (2018). Macroecology and macroevolution of the latitudinal diversity gradient in ants. Nature communications, 9(1), 1-8.

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences, 107(37), 16196-16200.

Etienne, R. S., & Rosindell, J. (2012). Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. Systematic Biology, 61(2), 204.

Espeland, M., Breinholt, J., Willmott, K. R., Warren, A. D., Vila, R., Toussaint, E. F., ... & Jarzyna, M. A. (2018). A comprehensive and dated phylogenomic analysis of butterflies. Current Biology, 28(5), 770-778.

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. Methods in Ecology and Evolution, 8(9), 1103-1112.

Friedman, W. E., & Carmichael, J. S. (1996). Double fertilization in Gnetales: implications for understanding reproductive diversification among seed plants. International Journal of Plant Sciences, 157(S6), S77-S94.

Gilbert, P. S., Wu, J., Simon, M. W., Sinsheimer, J. S., & Alfaro, M. E. (2018). Filtering nucleotide sites by phylogenetic signal to noise ratio increases confidence in the Neoaves phylogeny generated from ultraconserved elements. Molecular Phylogenetics and Evolution, 126, 116-128.

Hamilton, C. A., St Laurent, R. A., Dexter, K., Kitching, I. J., Breinholt, J. W., Zwick, A., ... & Kawahara, A. Y. (2019). Phylogenomics resolves major relationships and reveals significant diversification rate shifts in the evolution of silk moths and relatives. BMC Evolutionary Biology, 19(1), 1-13.

Hebert, P. D., Cywinska, A., Ball, S. L., & Dewaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1512), 313-321.

Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. PLoS Biology, 2(10), e312.

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. The American Naturalist, 163(2), 192-211.

Höhna, S., Freyman, W. A., Nolen, Z., Huelsenbeck, J. P., May, M. R., & Moore, B. R. (2019). A Bayesian approach for estimating branch-specific speciation and extinction rates. bioRxiv.

IISE (2011). State of Observed Species. Tempe, AZ. International Institute for Species Exploration.

Janz, N. & Nylin S. (2008). The oscillation hypothesis of host-plant range and speciation. In The evolutionary biology of herbivorous insects: specialization, speciation and radiation. Tilman D. (eds.), University of California Press, 203-215.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. Nature, 491(7424), 444-448.

Jetz, W., & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nature Ecology & Evolution, 2(5), 850-858.

Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F., Donath, A., ... & Barber, J. R. (2019). Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. Proceedings of the National Academy of Sciences, 116(45), 22657-22663.

Kitching, I. J., Rougerie, R., Zwick, A., Hamilton, C. A., St Laurent, R. A., Naumann, S., ... & Kawahara, A. Y. (2018). A global checklist of the Bombycoidea (Insecta: Lepidoptera). Biodiversity Data Journal, (6).

Klaus, K. V., & Matzke, N. J. (2020). Statistical comparison of trait-dependent biogeographical models indicates that Podocarpaceae dispersal is influenced by both seed cone traits and geographical distance. Systematic Biology, 69(1), 61-75.

Lagomarsino, L. P., Condamine, F. L., Antonelli, A., Mulch, A., & Davis, C. C. (2016). The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). New Phytologist, 210(4), 1430-1442.

Lemaire, C. (1978). Les Attacidae Américains : Attacinae. C. Lemaire (eds.), Neuilly-sur-Seine, 238pp.

Lemaire, C. (1980). Les Attacidae américains : Arsenurinae. C. Lemaire (eds.), Neuilly-sur-Seine, 199pp.

Lemaire, C. (1988). Les Saturniidae Américains : Ceratocampinae, Museo Nacional de Costa Rica (eds.), San José, 480pp.

Lemaire, C. (2002). The Saturniidae of America. Les Saturniidae américains (= Attacidae). Hemileucinae. Goecke & Evers, Keltern, Germany, 1388 pp., 140 pls.

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. Systematic biology, 61(5), 727-744.

Lewis, Z. A., Shiver, A. L., Stiffler, N., Miller, M. R., Johnson, E. A., & Selker, E. U. (2007). High-density detection of restriction-site-associated DNA markers for rapid mapping of mutated loci in Neurospora. Genetics, 177(2), 1163-1171.

Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. Systematic Biology, 56(5), 701-710.

Maliet, O., Hartig, F., & Morlon, H. (2019). A model with many small shifts for estimating species-specific diversification rates. Nature Ecology & Evolution, 3(7), 1086-1092.

Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Frontiers of Biogeography, 5(4).

McCormack, J. E., Maley, J. M., Hird, S. M., Derryberry, E. P., Graves, G. R., & Brumfield, R. T. (2012). Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Molecular Phylogenetics and Evolution, 62(1), 397-406.

Metzker, M. L. (2010). Sequencing technologies—the next generation. Nature reviews genetics, 11(1), 31-46.

Michener, C. D. (1952). The Saturniidae (Lepidoptera) of the Western Hemisphere: morphology, phylogeny, and classification. Bulletin of the AMNH, v. 98, article 5.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome research, 17(2), 240-248.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. (2011). How many species are there on Earth and in the ocean? PLoS Biol, 9(8), e1001127.

Morlon, H., Parsons, T. L., & Plotkin, J. B. (2011). Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences, 108(39), 16327-16332.

Murphy, W. J., Eizirik, E., Johnson, W. E., Zhang, Y. P., Ryder, O. A., & O'Brien, S. J. (2001). Molecular phylogenetics and the origins of placental mammals. Nature, 409(6820), 614-618.

Naumann, S., & Peigler, R. S. (2001). Four new species of the silkmoth genus *Samia* (Lepidoptera: Saturniidae). Nachrichten des Entomologischen Vereins Apollo, 22, 75-83.

Nee, S., May, R. M., & Harvey, P. H. (1994). The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 344(1309), 305-311.

Newman, M. E. (1997). A model of mass extinction. Journal of Theoretical Biology, 189(3), 235-252.

Packard, A. S. (1895). Monograph of the Bombycine Moths of America North of Mexico: including their transformations and origin of the larval markings and armature, vol. 7,. US Government Printing Office.

Peigler, R. S., & Maldonado, M. (2005). Uses of cocoons of *Eupackardia calleta* and *Rothschildia cincta* (Lepidoptera: Saturniidae) by Yaqui Indians in Arizona and Mexico. Nachrichten des Entomologischen Vereins Apollo, 26(3), 111-119.

Prebus, M. (2017). Insights into the evolution, biogeography and natural history of the acorn ants, genus *Temnothorax* Mayr (hymenoptera: Formicidae). BMC Evolutionary Biology, 17(1), 250.

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature communications, 4(1), 1-8.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... & Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. Nature, 559(7714), 392-395.

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Molecular Ecology Notes, 7(3), 355-364.

Raup, D. M. (1986). Biological extinction in earth history. Science, 231(4745), 1528-1533.

Ree, R. H. (2005). Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. Evolution, 59(2), 257-265.

Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology, 57(1), 4-14.

Regier, J. C., Mitter, C., Peigler, R. & Friedlander T. P. (2002). Monophyly, composition, and relationships within Saturniinae (Lepidoptera: Saturniidae): Evidence from two nuclear genes. Insect Systematics and Evolution, 33: 9-21.

Regier, J. C., Cook, C. P., Mitter, C., & Hussey, A. (2008). A phylogenetic study of the 'bombycoid complex'(Lepidoptera) using five protein-coding nuclear genes, with comments on the problem of macrolepidopteran phylogeny. Systematic Entomology, 33(1), 175-189.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L., & Rasnitsyn, A. P. (2012). A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology, 61(6), 973-999.

Rougeot, P. (1955). Les Attacides (Saturnidae) de l'Equateur africain français. Encyclopédie Entomologique, 34, 1-116, xii pls.

Rubin, J. J., Hamilton, C. A., McClure, C. J., Chadwell, B. A., Kawahara, A. Y., & Barber, J. R. (2018). The evolution of anti-bat sensory illusions in moths. Science Advances, 4(7), eaar7428.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, 74(12), 5463-5467.

Sibley, C. G., & Ahlquist, J. E. (1984). The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. Journal of Molecular Evolution, 20(1), 2-15.

Sibley, C. G., & Ahlquist, J. E. (1990). Phylogeny and classification of birds: a study in molecular evolution. Yale University Press.

Stadler, T. (2011). Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences, 108(15), 6187-6192.

Streicher, J. W., & Wiens, J. J. (2016). Phylogenomic analyses reveal novel relationships among snake families. Molecular Phylogenetics and Evolution, 100, 160-169.

Sukumaran, J., Economo, E. P., & Lacey Knowles, L. (2016). Machine learning biogeographic processes from biotic patterns: a new trait-dependent dispersal and diversification model with model choice by simulation-trained discriminant analysis. Systematic Biology, 65(3), 525-545.

Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., & Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biological Conservation, 204, 23-31.

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biology, 17(12), e3000494.

Valentine, J. W. (1968). Climatic regulation of species diversification and extinction. Geological Society of America Bulletin, 79(2), 273-276.

Von Bubnoff, A. (2008). Next-generation sequencing: the race is on. Cell, 132(5), 721-723.

Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology, Evolution, and Systematics, 34(1), 273-309.

Yang, X., Song, J., Todd, J., Peng, Z., Paudel, D., Luo, Z., ... & Zhao, Y. (2019). Target enrichment sequencing of 307 germplasm accessions identified ancestry of ancient and modern hybrids and signatures of adaptation and selection in sugarcane (Saccharum spp.), a 'sweet'crop with 'bitter'genomes. Plant Biotechnology Journal, 17(2), 488-498.

Yeates, D. K., Zwick, A., & Mikheyev, A. S. (2016). Museums are biobanks: unlocking the genetic potential of the three billion specimens in the world's biological collections. Current opinion in insect science, 18, 83-88.

Zhang, Y. M., Buffington, M. L., Looney, C., László, Z., Shorthouse, J. D., Ide, T., & Lucky, A. (2020). UCE data reveal multiple origins of rose gallers in North America: Global phylogeny of Diplolepis Geoffroy (Hymenoptera: Cynipidae). Molecular Phylogenetics and Evolution, 153, 106949.

Abstract

Phylogenomic approach of spatial and temporal dynamic of diversification in Saturniidae moths (Lepidoptera)

Understanding the evolutionary and ecological mechanisms governing the global patterns of Biodiversity is a central question in the fields of Ecology and Evolution. Phylogenies, as representations of the evolutionary relationships between lineages of the living world, are a fundamental support to identify the processes at the origin of these patterns. This thesis presents my work on the diversity and evolution of moths in family Saturniidae Boisduval 1837 (Lepidoptera: Bombycoidea). This biologically and morphologically diverse family includes nearly 3,500 species distributed on all continents. Using phylogenomic approaches, I inferred the phylogenetic relationships between all knwown genera – from which I introduce a new classification – and I proposed a phylogeny for all species in the Neotropical genus *Copaxa*. I also designed a phylogenomic pipeline allowing the generation of megaphylogenies – i.e. dated phylogenies with more than a thousand tips and whose completion is >50% of all species – which I applied on a data set combining ultraconserved elements and DNA barcodes to generate a phylogeny including between 88 and 100% of all described saturniid species. The phylogenies thus inferred were used in order to characterize spatial and temporal diversification of Saturniidae. Altogether, my results demonstrate the importance of biotic factors in the spatial and temporal diversification of the family. In particular, I identified that the ability to spin plain and dense cocoons as well as high degrees of polyphagy have been the keys to the biogeographical success of Saturniidae, and that the heterogeneity of the diversification rates within the family was explained by the evolution of traits linked to the capital breeding strategy of these moths: increases in body size and in polyphagy level. I also inferred that the climatic niche of *Copaxa* moths, inherited from an ancestor distributed in the Holarctic region, shaped their diversification within the Neotropical region: the majority of species fly in mountainous areas, the climate of which is similar to those found in temperate areas, and the two independent colonizations of the Andean chain implied positive shifts in diversification rates (dispersification events). Taken together, these results represent a major advance in understanding the evolution of Saturniidae, Lepidoptera and more generally constitute a set of materials allowing a better understanding of the evolutionary processes which have generated the incredible diversity of insects.

Résumé

Approche phylogénomique de la dynamique spatiale et temporelle de diversification chez les Lépidoptères Saturniidae

La compréhension des mécanismes évolutifs et écologiques à l'origine des patrons globaux de Biodiversité est une question centrale dans les domaines de l'Écologie et de l'Évolution. Les phylogénies, représentations des liens évolutifs entre les lignées du monde vivant, constituent le socle incontournable permettant d'identifier les processus à l'origine de ces patrons. Cette thèse présente les travaux de recherche que j'ai entrepris sur l'évolution de la diversité de papillons de la famille des Saturniidae Boisduval 1837 (Lepidoptera : Bombycoidea). Cette famille, particulièrement diversifiée biologiquement et morphologiquement, comprend près de 3500 espèces distribuées sur l'ensemble des continents. A l'aide d'approches phylogénomiques, j'ai inféré les relations phylogénétiques entre tous les genres décrits – à partir desquelles j'introduis une nouvelle classification des Saturniidae - et j'ai proposé une phylogénie incluant toutes les espèces du genre Néotropical *Copaxa*. J'ai également conçu un *pipeline* phylogénomique permettant la génération de mégaphylogénies – phylogénies datées de plus de mille feuilles dont la complétion est >50% des espèces – que j'ai appliqué sur un jeu de données combinant des éléments ultra-conservés du génome et des codes-barres ADN pour générer une phylogénie représentant entre 88 et 100% des espèces connues des Saturniidae. Les phylogénies ainsi inférées ont ensuite été utiliser afin d'examiner les dynamiques spatio-temporelles de la diversification des Saturniidae. Dans leur ensemble, les résultats obtenus au cours de ma thèse démontrent l'importance des facteurs biotiques dans la diversification spatiale et temporelle de la famille. J'ai notamment identifié que la capacité de tisser des cocons pleins et denses ainsi qu'un fort degré de polyphagie ont été les clés du succès biogéographique des Saturniidae et que l'hétérogénéité des taux de diversification au sein de la famille s'explique par l'évolution de traits en lien avec la stratégie dite de « *capital breeding* » de ces papillons : augmentation de la taille et de la polyphagie. J'ai également inféré que la niche climatique des *Copaxa*, héritée d'un ancêtre distribué dans la région Holarctique, avait façonné leur patron de diversification au sein de la région Néotropicale : la majorité des espèces volent dans les zones montagneuses, aux climats proches de ceux des zones tempérées, et les deux colonisations indépendantes de la chaine andine ont impliqué des *shifts* positifs des taux de diversification (évènements de *dispersification*). Dans leur ensemble, ces résultats représentent une avancée majeure dans la compréhension de la phylogénie et de l'évolution des Saturniidae, des Lépidoptères et plus généralement constituent un ensemble de supports permettant de mieux comprendre les processus évolutifs qui ont généré l'incroyable diversité des insectes.