Alma Mater Studiorum – Università di Bologna

Dottorato di ricerca in

Scienze della Terra, della Vita e dell'Ambiente

Ciclo 33

Settore Concorsuale: 05/B1 - ZOOLOGIA E ANTROPOLOGIA

Settore Scientifico Disciplinare: BIO/05 - ZOOLOGIA

Molecular Systematics and Traits Evolution in Phasmatodea Leach, 1815 (Hexapoda; Insecta)

Presentata da:
Giobbe Forni

Coordinatore Dottorato:
Giulio Viola

Supervisore:
Andrea Luchetti

Esame finale anno 2021

Abstract

Phasmatodea Leach, 1815 (Hexapoda; Insecta) is a polyneopteran order which counts approximately 3000 described species, often known for their remarkable forms of mimicry. Most of Phasmatodea species richness lies within the suborder Euphasmatodea: several hypotheses on the systematic relationship and diversification times of its major clades have been proposed, but they are conflicting and contributed to the difficulty in interpreting phasmids extant diversity and evolution. In this thesis, I provide a comprehensive systematic framework which includes over 180 species never considered in a phylogenetic framework: the latter can facilitate a better understanding of the processes underlying phasmids evolutionary history. The clade represents in fact an incredible testing ground to study trait evolution and its striking disparity of reproductive strategies and wing morphologies have been of great interest to the evolutionary biology community. Throughout the thesis, these traits are used to unravel different processes related to the evolution of complex phenotypes. Phasmids wings represent one of the first and most notable rejection of Dollo's law and they played a central role in initiating a long-standing debate on the irreversibility of complex traits loss. Macroevolutionary analyses presented here confirm that wings evolution in phasmids is a reversible process even when possible biases - such as systematic uncertainty and trait-dependent diversification rates - are considered. These findings remark how complex traits can evolve in a dynamic, reversible manner and imply that their molecular groundplan can be preserved despite its phenotypical absence. This concept has been further tested with phylogenetic and transcriptomic approaches in two phasmids parthenogenetic lineages and a bisexual congeneric of the European *Bacillus* species complex. Leveraging a gene co-expression network approach, male gonad associated genes were retrieved in the bisexual species and then their modifications in the parthenogens were charachterized. Pleiotropy appears to constrain gene modifications associated to male reproductive structures after their loss in parthenogens, so that the lost trait molecular groundplan can be largely preserved in both transcription patterns and sequence evolution. Yet, parthenogenesis is not just the loss of males and their reproductive machinery but instead requires several adaptations to overcome the obstacles of reproducing in the absence of fertilization and other clues provided by males. Parthenogen gonads transcriptional programme is largely assembled from genes that were present before the shifts in reproductive strategies. The convergent expression reprogramming of hundreds of genes is observed in parthenogens reproductive tissues compared to sexual ones, while convergent changes played a marginal role at the sequence evolution level. Overall, the results presented in this thesis contribute to shape our understanding of the interplay between the phenotypic and molecular levels in trait evolution.

1. Introduction. 05

1.1 Systematics and evolutionary history of Phasmatodea. 06

1.2 Dollo's law of irreversibility and the evolution of wings in stick insects. 08

1.3 Trait loss and establishment in parthenogenetic species. 10

2. Systematics and Evolutionary History of Phasmatodea. 18

2.1 Phylomitogenomics provides new perspectives on the Euphasmatodea radiation (Insecta: Phasmatodea). 19

2.2 Molecular systematics and phylogenetics of the Australian stick insect genus *Candovia* Stål, 1875 (Phasmida, Necrosciinae). 53

2.3 Molecular systematics of four Euphasmatodea sub-families and the incongruence with morphological taxonomy. 80

3. Trait Evolution in Phasmids. 148

3.1 Macroevolutionary analyses provide new evidences of ohasmids wings evolution as a reversible process. 149

3.2 Evolution of automixis through different cytological mechanisms is associated with convergent transcriptional reprogramming in *Bacillus* insects. 203

3.3 Lost but not forgotten: pleiotropic interactions constrain gene expression and sequence evolution of male gonad genes after the shift to parthenogenesis. 255

4. Methods Development. 293

4.1 Complete mitochondrial genomes from transcriptomes: assessing pros and cons of data mining for assembling new mitogenomes. 294

4.2 BASE: a novel workflow to integrate non-ubiquitous genes in comparative genomics analyses for selection. 336

5. Conclusions. 349

1. Introduction.

1.1 Systematics and evolutionary history of Phasmatodea.

Phasmatodea Leach, 1815 (Hexapoda; Insecta) is a order of polyneopteran insects, counting over three thousands described species (Bradler and Buckley, 2018). The clade diversity is distributed across all major landmasses - with the exception of Antarctica and the Patagonia region - and a predominantly tropical distribution: the most species-rich regions are Southeast Asia, Southern and Central America and Australia. The majority of extant phasmatodeans show an elongated and tubular body with slender legs (*i.e.* stick insects), while a few taxa show a wider body with lateral expansions (*i.e.* leaf insects). Phasmids are strictly herbivores and predominantly nocturnal animals and, as their name suggests - *phasma* being the greek word for ghost - they include some of the most remarkable examples of cryptic mimicry. With the exception of a few diurnal species which show aposematic colorations, most species exhibit extreme forms of morphological and behavioral mimicry of branches, twigs, leaves, bark and lichens, (Bedford 1978). Phasmids are adapted to their arboreal environment during all life stages: phytomimesis is present in the adult as well as in the nymphal stages and also their hard-shelled eggs show a strong resemblance to plant seeds. Even many behaviors are related to mimicry, such as the typical rocking and rolling movements.

Phasmatodea species are traditionally subdivided in ~473 genera (Brock et al, 2020), but their actual species richness may be substantially underestimated. Stick and leaf insects appear to have a huge degree of morphological disparity, with more than 35% of the genera being monotypic; this phenomenon could result from a plethora of phenomena such as a highly fragmented species sampling or rapid morphological divergence. In comparison with other insect taxa, they are difficult to sample due to their mimicry and collection efforts need to cover extremely vast areas, due to the species reduced dispersal abilities. Species-rich geographical areas, such as south America and Africa, have been often overlooked in taxonomic investigation, contrary to other less rich areas such as the Mediterranean basin and the south-east Asia (Bradler and Buckley, 2018). It also has to be considered that when molecular species delimitation approach was applied, they recovered more species than what was expected from morphological observation (Velonà et al, 2015; Glaw et al, 2019; Cumming et al, 2020), suggesting instances of cryptic species richness.

Phasmatodea molecular systematics has received an increasing attention in the last twenty years: this large body of work has greatly contributed to our understanding of the clade evolutionary history but, in some instances, phylogenetic relationships among higher taxonomic ranks still lack support or reproducibility. Yet, all these findings consistently show that the traditional classification of the order is

not congruent in a phylogenetic framework. Nonetheless, there are some established facts about phasmids phylogeny, such as their position inside Polyneoptera: recent phylogenomic studies corroborate the hypothesis of Embioptera as their sister group, to form Eukinolabia (Misof et al, 2014; Wipfler et al, 2019; Simon et al, 2019). This sister relationship was initially proposed by Rähle (1970), based on a shared paraglossae flexor muscle and subsequently other Eukinolabia apomorphies have been proposed, such as largely reduced ovipositor and a unique head morphology (Friedemann et al, 2012). Phasmatodea monophyly is also well established and it has been demonstrated both by molecular (*e.g.* Whiting et al, 2003; Buckley et al, 2009; Robertson et al, 2018; Bradler et al, 2014; Simon et al, 2019) and morphological means (Hennemann et al, 2008). Within the order, the sister relationship of the Timematodea suborder - counting just 21 species distributed only in the Western Nearctic - with Euphasmatodea (all other phasmids) is solid, although relationships within Euphasmatodea are still debated. The systematics of Euphasmatodea (families and subfamilies) has been originally proposed by Günther (1953) and subsequently adopted by other authors (Bradley & Galil, 1977; Kevan, 1977 and 1982). Günther himself was aware of the difficulties of this effort, due to the presence of several irreconcilable morphological characters. In more recent years, the phylogenetic analyses on morphological (*e.g.* Tilgner, 2002) and molecular data (*e.g.* Whiting et al, 2003; Buckley et al, 2009a; Bradler et al, 2015; Robertson et al, 2018) rejected the traditional classification and highlighted the necessity of an extensive taxonomic revision. As an example, Phasmatodea have traditionally been subdivided into Areolatae and Anareolatae on the basis of the presence or absence of the area apicalis - a triangular area located ventrally on the apex of the tibiae - but this division has never been supported by any molecular phylogenetic study. Yet, the different molecular phylogenetic hypotheses which have been proposed are conflicting in nature and contributed to increment the confusion in interpreting morphological observations. Nonetheless, it has been possible to identify some monophyletic groups with high confidence and to explore novel hypotheses, such as the subdivision of the Neophasmatodea (*i.e.* all Euphasmatodea except the Aschiphasmatidae) in two major clades that may reflect biogeography: Occidophasma (“Western Phasmids”; New World) and Oriophasmata (“Eastern phasmids”; Old World) (Simon et al, 2019).

This complex and dynamic scenario appears to be due to several concurring factors: at the morphological level, convergent evolution and intraspecific variability appeared to be the two main reasons for which the traditional classification of the Phasmatodea (Günther, 1953; Bradley & Galil, 1977) is inconsistent in a phylogenetic framework. At the molecular level, the rapid radiation of Euphasmatodea (Buckley et al, 2009; Bradler et al, 2014; Robertson et al, 2018; Simon et al, 2019)

causes molecular phylogenetics to incur in an extremely scarce phylogenetic signal, with topology being very sensitive to model misspecification and traditional nodal support metrics concealing uncertainty.

In addition to the uncertainty of phasmids phylogenetic relationships, conflicting results can be found in the literature also regarding the timing of Phasmatodea origin and diversification. The outcome of molecular divergence time analyses can be roughly divided into two kinds, based on nodes calibration strategies and taxon sampling: (1) analyses with insect orders other than Phasmatodea as their main focus, which have a scarce taxon sampling for phasmids and use most fossil calibrations outside of the clade (*e.g.* Evangelista et al, 2019; Tong et al, 2015; Bourguignon et al, 2018) or (2) analyses carried out using a high taxon sampling for phasmids but relying mostly - or exclusively - on calibrations internal to the clade (Buckley et al, 2009; Bradler et al, 2014; Robertson et al, 2018). The first kind of studies date the divergence between Euphasmatodea and *Timema* in a time frame ranging from ~190 mya (late Jurassic; Tong et al, 2015) to ~270 mya (early Permian; Bourguignon et al, 2018). The second group of studies, instead, find the same split at more recent times, ranging from a minimum of ~51.9 mya (Paleogene; Buckley et al, 2009) to a maximum of ~121.8 mya (early Cretaceous; Simon et al, 2019). As also acknowledged by the same authors, the latter kind of studies were suspected to underestimate the divergence times. Understanding the timing of phasmids cladogenetic events is also important to understand the evolutionary radiation of Euphasmatodea. Beside key innovations, such as hardened egg (Robertson et al, 2018) or the acquisition of pectinase by horizontal gene transfer (Shelomi et al, 2016), concurrent events could have contributed to the radiation process. For example, hypotheses sustaining a recent evolutionary radiation connect it to the radiation of angiosperms (Peris et al, 2017) and explain the extant phasmid distribution as a process mainly driven by active dispersal, with vicariance playing a marginal role.

Even if these shortcomings have partially impeded the understanding of evolutionary patterns in this group, phasmids have proven themselves as an incredible testing ground to study trait evolution. The striking disparity of reproductive strategies and wing morphologies have been of great interest to the scientific community and are here used as case studies to explore different evolutionary processes underlying complex phenotypes evolution.

1.2 Dollo's law of irreversibility and the evolution of wings in stick insects

The concept that complex phenotypes, after being lost down the line of evolution, cannot be acquired

again derived from Simpson's (Simpson, 1953) modern formulation of Dollo's law. If a trait is lost at the phenotypical level, it is expected that the evolutionary constraints acting on its molecular underpinning are removed: as a consequence, negative selection will no longer act on them and mutations will accumulate, making a reversion unlikely. Despite a broader original formulation (Dollo, 1893), in which reversals of complex ancestral traits were considered possible as secondary convergence events, several intuitive examples reinforced the concept that lost trait can never come back: for example, whales have never reacquired legs or teeth never reappeared in birds. Although this evolutionary principle is still very popular in the scientific community, an increasing number of exceptions are described in literature, mainly exposed through the means of macroevolutionary analyses. These include shell coiling in limpets (Collin and Cipriani, 2003), compound eyes in ostracods (Syme and Oakley, 2012), sex and parasitism in mites (Klimov and O'Connor, 2013; Domes et al. 2007), mandibular teeth in frogs (Wiens, 2011), limb evolution (Kohlsdorf and Wagner, 2006), eggshells and oviparity (Lynch and Wagner, 2010; Recknagel et al. 2018; Esquerré et al. 2020) in squamata. Some authors have hypothesized that a re-acquisition of complex traits is possible only within a short time frame after their loss (Porter and Crandall, 2003; Teotónio and Rose, 2001); however, in other instances ancestral lost traits appear to have reverted back even several million years later (Chippindale et al, 2004; Collin and Cipriani, 2003).

Phasmids have played an important role in rethinking of Dollo's law. Wings emerged early in the diversification of insects with subsequent modifications and frequent losses occurring during their evolutionary history and leading to the extant diversity (Engel, 2015; Wipfler et al, 2019). In 2003, Whiting and co-workers proposed that phasmids have initially diversified as wingless and then wings appeared later during their evolutionary history, in multiple and independent instances (Whiting et al, 2003). In insects, partial reduction or complete loss of wings may have an adaptive value, for example related to cryptic capacity (Whiting et al, 2003). It is possible to hypothesize that during the early evolutionary history of phasmids a selective pressure towards increased crypticity resulted in wingless forms, while, later on, a greater capacity of dispersal and defense against predators may have been selected in some lineages, leading to a reversion to the ancestral winged forms. Shortly after this claim several comments followed (Stone and French, 2003; Trueman et al, 2004) exposing possible methodological flaws, while at the same time new approaches have been developed which allow to avoid incorrect rejection of the Dollo's law. Root prior probability (Sauquet et al, 2017; Goldberg and Igić, 2008), trait-dependent diversification rates (Goldberg and Igić, 2008; Holland et al, 2020) and tree uncertainty (Rangel et al, 2015; Bollback, 2006) have all to be considered when testing the irreversible evolution of a trait.

Transcriptomic and genomic approaches are contributing to better understanding possible processes of complex traits reversible evolution, taking into account the complex interaction between genotype and phenotype which underlie the development of organisms. Many genes that drive the development of organisms are necessary at different times or in different tissues, sometimes guiding the development of apparently uncorrelated traits (Collin and Miglietta, 2008). Genes can, in fact, have different levels of pleiotropy - *i.e.* the potential of a single gene to affect multiple phenotypical traits - which seems to be the rule rather than the exception in biological systems (Stearns, 2010). This phenomenon appears to be a direct consequence of the evolution acting as a molecular "tinkerer", with existing assets being often co-opted for evolutionary novelties (Jacob, 1977). In such a prospective, even if a character is lost in a lineage, the underlying blueprint guiding its development can persist - at least partially - when it is associated to other traits. In many insects the imaginal discs of wings and legs share the same origin and the genetic pathway that guide their development appears to be largely shared. Moreover, it has been observed that in stick insects the neural structures and their functional connectivity necessary to sustain flight capability are conserved also in wingless forms (Kutsch and Kittmann, 1991). If, after the wings' loss, their molecular blueprint could remain under selection because involved in the development of other traits - such as legs - the re-acquisition scenario does not appear as unlikely as it might be expected. In this view, it is interesting to note that all known fossil species of winged stick insects found before mid-Cretaceous had long tegmina (*i.e.* modified leathery front wing) and there are no paleontological evidences of wings with shorter tegmina, which seem to have decreased during the evolution of the order. This modification of ancestral and derived structures could imply that if a reversal of the lost character occurred, it was restored in a slightly different form.

1.3 Trait loss and establishment in parthenogens.

Reproduction may be considered the most fundamental characteristics of living beings and occurs *via* a plethora of different mechanisms. Although being mostly bisexual, many phasmids species adopt different reproductive strategies, including parthenogenesis, hybridogenesis and androgenesis. About 10% of the order species, scattered across different families, is found capable of parthenogenesis (Scali et al. 2009). These mechanisms have been studied and described in detail in only a few genera, such as *Timema*, *Clonopsis*, *Pijnackeria* and *Bacillus,* where a vast literature on the reproductive strategies is available. Parthenogens can be roughly divided into apomictic and automictic: the main

difference lies in the retention or loss of the meiotic process. In apomixis, meiosis is replaced by a mitotic division and a genetically-invariant offspring is produced, while in automixis meiosis is carried out but, in the long run, it is known to cause severe heterozygosity erosion (Scali et al. 2009). Much attention has been paid to apomictic species while automictic ones are rather overlooked, both in general and specifically for phasmids.

In this thesis, I explored causes and consequences of automixis in the Mediterranean stick-insects genus *Bacillus*, an outstanding experimental system to explore the causes and consequences of meiotic parthenogenesis, which comprises the following: (1) the strictly bisexual *B. grandii* (2*n*=33/34, XO-XX), endemic to tiny areas in Sicily and formally split into three subspecies; (2) the western Mediterranean facultative parthenogen, *B. rossius* 2*n* = 35/36, XO-XX, with two Italian subspecies; (3) the eastern Mediterranean *B. atticus*, at present an all-female obligate parthenogen, forming a complex of three different karyological (2*n*=32; 2*n*=34; 3*n*=48–51) and allozymic races. These species are the ancestors of parthenogenetic thelytokous hybrids: *Bacillus whitei* (=rossius/grandii) and *Bacillus lynceorum* (= rossius/grandii/atticus). Despite the challenges in demonstrating the lack of sexual reproduction in a lineage (Schurko et al., 2009), no males have been ever observed in our captive bred populations of *B. atticus* and parthenogenetic populations of *B. rossius*. On the whole, *Bacillus* constitute an example of reticulate evolution (*i.e.* they experience an array of reproductive interactions realizing a complex network of phyletic relationships) and provide an outstanding experimental system to analyse the multifaceted links among sexual and derived clonal taxa.

My initial focus has been to understand the relative contribution of novel and pre-existing genes in the establishment of automixis (Tautz et al, 2011; McLysaght and Guerzoni 2015). The two mechanisms can concur to the evolution of novel traits, but the latter can present different signatures, with some traits seemingly coupled with the *de-novo* origin of genes (Babonis et al, 2016; Sim et al, 2016) and others more strongly associated with the co-option of pre-existing ones (Almundi et al, 2020). Changes in function involving pre-existing genes can result either from modification in a *cis*-regulatory element that changes gene expression (McGirr et al, 2020) and/or from a modification in the protein sequence that alters its function (Casewell et al, 2019; Jebb et al, 2020). Another important aspect is the degree to which convergent evolution of traits involves modifications in the same set of genes (Stern, 2013; Sackton and Clark, 2019): while the establishment of some traits seems to have happened through similar trajectories, in other evidences are lacking, implying that similar outcomes are achieved through different paths.

Subsequently I focused on the consequences of the absence of males (and their reproductive

structures) in such lineages. When selective pressures change, traits may become useless or disadvantageous, decay and be lost. This is a rather common phenomenon across the tree of life and can have adaptive values as strong as the establishment of novel ones. Trait loss is predicted to have evolutionary consequences for the genes that control its development, either via drift or selection (van der Kooi & Schwander, 2014). Several kinds of decay have been associated to trait-specific genes after its loss, including mutations in coding and regulatory parts of genes, changes in their expression patterns and/or complete gene losses. On the other side, trait loss is not necessarily coupled with the decay of the molecular pathway behind its development. For example, parasitic wasps which have lost lipogenesis do not show any gene sequence degradation across the pathway (Lammers et al., 2019); the expression of functional opsins has been found in cave crustacean with reduced eyes (Perez-Moreno et al., 2018); and genes underlying photosynthesis are found to be under a strong purifying selection in a parasitic plant (McNeal et al., 2007). However, these observations on maintenance and decay can be reconciled: phenotypes are the product of large networks of interactions and different evolutionary trajectories can be expected for genes underlying a trait after its loss. Genes concurring to the establishment of a trait can, in fact, have different properties, including variable levels of pleiotropy (He & Zhang, 2006; Paaby & Rockman, 20013). Some genes can maintain their functionality over time - due to selection on functions that are not related to the lost trait - while those specifically associated with the trait are expected to decay. Genes that are involved in multiple biological processes are less likely to degenerate, after the selective constraints are removed from a specific trait; thus, few genes could be really dispensable after trait loss, either because they are already involved in the making of other phenotypes or because they are co-opted for novel purposes (Hunt et al, 2011; Smith et al, 2015). As pleiotropy is known to modulate selection (Fraïsse et al., 2019) and constrain both gene expression (Papakostas, et al., 2014) and sequence evolution (Mack et al., 2014), we hypothesized that the different fate of genes related to a trait after its loss could be explained by their different degrees of pleiotropy. Genes with many interactions in a gene regulatory network are in fact more pleiotropic and more likely to be essential than the others (Jeong et al, 2001; MacNeil and Walhoutm, 2011).

References:

Babonis, L.S., Martindale, M.Q. and Ryan, J.F., 2016. Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone Nematostella vectensis. BMC evolutionary biology, 16(1), pp.1-22.

Bedford, G.O., 1978. Biology and ecology of the Phasmatodea. *Annual review of entomology*, *23*(1),pp.125-149.

Bollback, J.P., 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC bioinformatics*, *7*(1), p.88.

Bourguignon, T., Tang, Q., Ho, S.Y., Juna, F., Wang, Z., Arab, D.A., Cameron, S.L., Walker, J., Rentz, D., Evans, T.A. and Lo, N., 2018. Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Molecular Biology and Evolution*, *35*(4), pp.970-983.

Bradler, S. and Buckley, T.R., 2018. Biodiversity of phasmatodea. *Insect Biodiversity: Science and Society*, *2*, pp.281-313.

Bradler, S., Robertson, J.A., Whiting, M.F., 2014. A molecular phylogeny of Phasmatodea with emphasis on Necrosciinae, the most species-rich subfamily of stick insects. Systematic entomology 39, 205– 222.

Bradley, J.C. & Galil, B.S. 1977. The taxonomic arrangement of the Phasmatodea with keys to the subfamilies and tribes. Proceedings of the Entomological Society of Washington, 79, 176-208.

Brock, P.D., Büscher, T. & Baker, E. Phasmida Species File Online. Version 5.0/5.0. [12-12-2020]. <http://Phasmida.SpeciesFile.org>

Buckley, Thomas R, Attanayake, D., Bradler, S., 2009. Extreme convergence in stick insect evolution: phylogenetic placement of the Lord Howe Island tree lobster. Proceedings of the Royal Society B: Biological Sciences 276, 1055–1062.

Casewell, N.R., Petras, D., Card, D.C., Suranse, V., Mychajliw, A.M., Richards, D., Koludarov, I., Albulescu, L.O., Slagboom, J., Hempel, B.F. and Ngum, N.M., 2019. Solenodon genome reveals convergent evolution of venom in eulipotyphlan mammals. Proceedings of the National Academy of Sciences, 116(51), pp.25745-25755.

Chippindale, P.T., Bonett, R.M., Baldwin, A.S., Wiens, J.J., 2004. Phylogenetic Evidence for a Major Reversal of Life-History Evolution in Plethodontid Salamanders. Evolution 58, 2809–2822.

Collin, R., Cipriani, R., 2003. Dollo's law and the re–evolution of shell coiling. Proceedings of the Royal Society of London. Series B: Biological Sciences 270, 2551–2555.

Collin, R., Miglietta, M.P., 2008. Reversing opinions on Dollo's Law. Trends in Ecology & Evolution 23, 602–609.

Cumming, R.T., Bank, S., Le Tirant, S. and Bradler, S., 2020. Notes on the leaf insects of the genus Phyllium of Sumatra and Java, Indonesia, including the description of two new species with purple coxae (Phasmatodea, Phylliidae). *ZooKeys*, *913*, p.89.

Dollo, L., 1893. Lesloisdel'évolution. *Bul. Soc. Belge Géol. Pal. Hydr*, *7*, pp.164-166.

Domes, K., Norton, R.A., Maraun, M., Scheu, S., 2007. Reevolution of sexuality breaks Dollo's law. Proceedings of the National Academy of Sciences 104, 7139–7144.

Engel, M.S., 2015. Insect evolution. Current Biology 25, R868–R872.

Esquerré, D., Brennan, I.G., Catullo, R.A., Torres-Pérez, F., Keogh, J.S., 2019. How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America's most species-rich lizard radiation (Squamata: Liolaemidae). Evolution 73, 214–230.

Evangelista, D.A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M.K., Legendre, F., Liu, S., Machida, R., Misof, B., Peters, R.S., Podsiadlowski, L., Rust, J., Schuette, K., Tollenaar, W., Ware, J.L., Wappler, T., Zhou, X., Meusemann, K., Simon, S., 2019. An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites

(Blattodea). Proceedings of the Royal Society B: Biological Sciences 286, 20182076.

Fraïsse, C., Puixeu Sala, G. and Vicoso, B., 2019. Pleiotropy modulates the efficacy of selection in Drosophila melanogaster. *Molecular biology and evolution*, *36*(3), pp.500-515.

Friedemann, K., Wipfler, B., Bradler, S., Beutel, R.G., 2012. On the head morphology of Phyllium and the phylogenetic relationships of Phasmatodea (Insecta). Acta Zoologica 93, 184–199.

Glaw, F., Hawlitschek, O., Dunz, A., Goldberg, J. and Bradler, S., 2019. When giant stick insects play with colors: Molecular phylogeny of the Achriopterini and description of two new splendid species (Phasmatodea: Achrioptera) from Madagascar. *Frontiers in Ecology and Evolution*, *7*, p.105.

Goldberg, E.E. and Igić, B., 2008. On phylogenetic tests of irreversible evolution. *Evolution: International Journal of Organic Evolution*, *62*(11), pp.2727-2741.

Günther, K. 1953. Über die taxonomische Gliederung und geographische Verbreitung der Insektenordnung der Phasmatodea. Beiträge zur Entomologie, 3, 541-563.

Hennemann, F.H. & Conle, O.V. 2008. Revision of Oriental Phasmatodea: The tribe Pharnaciini Günther, 1953, including the description of the world's longest insect, and a survey of the family Phasmatidae Gray, 1835 with keys to the subfamilies and tribes (Phasmatode: “Anareolatae”: Phasmatidae). Zootaxa, 1906, 1-316.

Holland, B.R., Ketelaar-Jones, S., o’Mara, A.R., Woodhams, M.D. and Jordan, G.J., 2020. Accuracy of ancestral state reconstruction for non-neutral traits. *Scientific reports*, *10*(1), pp.1-10.

Hunt, B.G., Ometto, L., Wurm, Y., Shoemaker, D., Soojin, V.Y., Keller, L. and Goodisman, M.A., 2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity. Proceedings of the National Academy of Sciences, 108(38), pp.15936-15941.

Jacob, F., 1977. Evolution and tinkering. Science 196, 1161–1166.

Jebb, D., Huang, Z., Pippel, M., Hughes, G.M., Lavrichenko, K., Devanna, P., Winkler, S., Jermiin, L.S., Skirmuntt, E.C., Katzourakis, A. and Burkitt-Gray, L., 2020. Six reference-quality genomes reveal evolution of bat adaptations. Nature, 583(7817), pp.578-584.

Jeong, H., Mason, S.P., Barabási, A.L. and Oltvai, Z.N., 2001. Lethality and centrality in protein networks. *Nature*, *411*(6833), pp.41-42.

Kevan, D.K. McE. (1977). The higher classification of orthopteroid insects: a general view. Lyman Entomological Museum Research Laboraatory Memoirs 4, 1-31.

Kevan, D.K. McE. (1982). “Phasmatoptera”, in: Parker, S.F. [Ed]: Synopsis and Classification of Living Organisms, Vol. 2. New York, NY: McGraw-Hill, 379-383.

Klimov, P.B. and OConnor, B., 2013. Is permanent parasitism reversible?—Critical evidence from early evolution of house dust mites. *Systematic biology*, *62*(3), pp.411-423.

Kohlsdorf, T., Wagner, G.P., 2006. Evidence for the reversibility of digit loss: a phylogenetic study of limb evolution in Bachia (Gymnophthalmidae: Squamata). Evolution 60, 1896–1912.

Kutsch, W., Kittmann, R., 1991. Flight motor pattern in flying and non-flying Phasmida. J Comp Physiol A 168, 483–490.

Lammers, M., Kraaijeveld, K., Mariën, J. and Ellers, J., 2019. Gene expression changes associated with the evolutionary loss of a metabolic trait: lack of lipogenesis in parasitoids. *BMC genomics*, *20*(1), pp.1-14.

Lynch, V.J. and Wagner, G.P., 2010. Did egg-laying boas break Dollo's law? Phylogenetic evidence for reversal to oviparity in sand boas (Eryx: Boidae). *Evolution: International Journal of Organic*

Evolution, *64*(1), pp.207-216.

MacNeil, L.T. and Walhout, A.J., 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome research*, *21*(5), pp.645-657.

McGirr, J.A. and Martin, C.H., 2020. Few fixed variants between trophic specialist pupfish species reveal candidate cis-regulatory alleles underlying rapid craniofacial divergence. Molecular Biology and Evolution.

McLysaght, A. and Guerzoni, D., 2015. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1678), p.20140332.

McNeal, J.R., Kuehl, J.V., Boore, J.L. and de Pamphilis, C.W., 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus Cuscuta. *BMC Plant Biology*, *7*(1), p.57.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G. and Niehuis, O., 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science*, *346*(6210), pp.763-767.

Papakostas, S., Vøllestad, L.A., Bruneaux, M., Aykanat, T., Vanoverbeke, J., Ning, M., Primmer, C.R. and Leder, E.H., 2014. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nature Communications*, *5*(1), pp.1-9.

Pérez-Moreno, J.L., Balázs, G. and Bracken-Grissom, H.D., 2018. Transcriptomic insights into the loss of vision in Molnár János Cave's crustaceans. *Integrative and Comparative Biology*, *58*(3), pp.452-464.

Peris, D., Pérez-de la Fuente, R., Peñalver, E., Delclòs, X., Barrón, E., Labandeira, C.C., 2017. False Blister Beetles and the Expansion of Gymnosperm-Insect Pollination Modes before Angiosperm Dominance. Current Biology 27, 897–904.

Porter, M.L., Crandall, K.A., 2003. Lost along the way: the significance of evolution in reverse. Trends in Ecology & Evolution 18, 541–547.

Rähle, W., 1970. Untersuchungen an Kopf und Prothorax von Embia ramburi Rimsky-Korsakow 1906 (Embioptera, Embiidae).

Rangel, T.F., Colwell, R.K., Graves, G.R., Fučíková, K., Rahbek, C. and Diniz-Filho, J.A.F., 2015. Phylogenetic uncertainty revisited: Implications for ecological analyses. *Evolution*, *69*(5), pp.1301-1312.

Recknagel, H., Kamenos, N.A. and Elmer, K.R., 2018. Common lizards break Dollo's law of irreversibility: genome-wide phylogenomics support a single origin of viviparity and re-evolution of oviparity. *Molecular Phylogenetics and Evolution*, *127*, pp.579-588.

Robertson, J.A., Bradler, S., Whiting, M.F., 2018. Evolution of Oviposition Techniques in Stick and Leaf Insects (Phasmatodea). Front. Ecol. Evol. 6.

Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J.S., Wheeler, N.E., Gardner, P.P., Clarke, J.A., Baker, A.J., Clamp, M. and Edwards, S.V., 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. Science, 364(6435), pp.74-78.

Sauquet, H., von Balthazar, M., Magallón, S., Doyle, J.A., Endress, P.K., Bailes, E.J., de Morais, E.B., Bull-Hereñu, K., Carrive, L., Chartier, M. and Chomicki, G., 2017. The ancestral flower of angiosperms and its early diversification. *Nature communications*, *8*(1), pp.1-10.

Scali, V., 2009. Metasexual stick insects: model pathways to losing sex and bringing it back. In *Lost sex* (pp. 317-345). Springer, Dordrecht.

Shelomi, M., Danchin, E.G., Heckel, D., Wipfler, B., Bradler, S., Zhou, X. and Pauchet, Y., 2016. Horizontal gene transfer of pectinases from bacteria preceded the diversification of stick and leaf insects. *Scientific reports*, *6*(1), pp.1-9.

Sim, A.D. and Wheeler, D., 2016. The venom gland transcriptome of the parasitoid wasp Nasonia vitripennis highlights the importance of novel genes in venom function. BMC genomics, 17(1), p.571.

Simon, S., Letsch, H., Bank, S., Buckley, T., Donath, A., Liu, S., Machida, R., Meusemann, K., Misof, B., Podsiadlowski, L. and Zhuo, X., 2019. Old world and new world phasmatodea: phylogenomics resolve the evolutionary history of stick and leaf insects. *Frontiers in Ecology and Evolution*, *7*, p.345.

Simpson, G.G., 1953. *The major features of evolution* (No. 575 S55).

Smith, C.R., Helms Cahan, S., Kemena, C., Brady, S.G., Yang, W., Bornberg-Bauer, E., Eriksson, T., Gadau, J., Helmkampf, M., Gotzek, D. and Okamoto Miyakawa, M., 2015. How do genomes create novel phenotypes? Insights from the loss of the worker caste in ant social parasites. Molecular biology and evolution, 32(11), pp.2919-2931.

Stearns, F.W., 2010. One hundred years of pleiotropy: a retrospective. *Genetics*, *186*(3), pp.767-773.

Stern, D.L., 2013. The genetic causes of convergent evolution. Nature Reviews Genetics, 14(11), pp.751-764.

Stone, G. and French, V., 2003. Evolution: have wings come, gone and come again?. *Current Biology*, *13*(11), pp.R436-R438.

Syme, A.E. and Oakley, T.H., 2012. Dispersal between shallow and abyssal seas and evolutionary loss and regain of compound eyes in cylindroleberidid ostracods: conflicting conclusions from different comparative methods. *Systematic Biology*, *61*(2), p.314.

Tautz, D. and Domazet-Lošo, T., 2011. The evolutionary origin of orphan genes. Nature Reviews Genetics, 12(10), pp.692-702.

Teotónio, H., Rose, M.R., 2001. Perspective: Reverse Evolution. Evolution 55, 653–660.

Terasawa, M., Shimokawa, R., Terashima, T., Ohya, K., Takagi, Y., Shimokawa, H., 2004. Expression of dentin matrix protein 1 (DMP1) in nonmineralized tissues. Journal of bone and mineral metabolism 22, 430–438.

Tilgner, E.H., 2009. Chapter 194 - Phasmida: Stick and Leaf Insects, in: Resh, V.H., Cardé, R.T. (Eds.), Encyclopedia of Insects (Second Edition). Academic Press

Tong, K.J., Duchêne, S., Ho, S.Y.W., Lo, N., 2015. Comment on “Phylogenomics resolves the timing and pattern of insect evolution.” Science 349, 487–487.

Trueman, J.W.H., Pfeil, B.E., Kelchner, S.A. and Yeates, D.K., 2004. Did stick insects really regain their wings?. *Systematic Entomology*, *29*(2), pp.138-139.

van der Kooi, C.J. and Schwander, T., 2014. On the fate of sexual traits under asexuality. *Biological Reviews*, *89*(4), pp.805-819.

Velona, A., Brock, P.D., Hasenpusch, J. and Mantovani, B., 2015. Cryptic diversity in Australian stick insects (Insecta; Phasmida) uncovered by the DNA barcoding approach. *Zootaxa*, *3957*(4), pp.455-466.

Whiting, M.F., Bradler, S., Maxwell, T., 2003. Loss and recovery of wings in stick insects. Nature 421,

264– 267.

Wiens, J.J., Kuczynski, C.A., Duellman, W.E., Reeder, T.W., 2007. Loss and re-evolution of complex life cycles in marsupial frogs: does ancestral trait reconstruction mislead? Evolution: International Journal of Organic Evolution 61, 1886–1899.

Wipfler, B., Letsch, H., Frandsen, P.B., Kapli, P., Mayer, C., Bartel, D., Buckley, T.R., Donath, A., Edgerly-Rooks, J.S., Fujita, M. and Liu, S., 2019. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. *Proceedings of the National Academy of Sciences*, *116*(8), pp.3024-3029.

2. Systematics and Evolutionary History of Phasmatodea.

Phylomitogenomics provide new perspectives on the Euphasmatodea radiation (Insecta: Phasmatodea).

Giobbe Forni, Federico Plazzi, Alex Cussigh, Oskar Conle, Frank Hennemann, Andrea Luchetti and Barbara Mantovani.

Abstract - Phasmatodea species diversity lies almost entirely within its suborder Euphasmatodea, which exhibits a pantropical distribution and is considered to derive from a recent and rapid evolutionary radiation. To shed light on Euphasmatodea origins and diversification, we assembled the mitogenomes of 17 species from transcriptomic sequencing data and analysed them along with 22 already available Phasmatodea mitogenomes and 33 mitogenomes representing most of the Polyneoptera lineages. Maximum Likelihood and Bayesian Inference approaches retrieved consistent topologies, both showing the widespread conflict between phylogenetic approaches and traditional systematics. We performed a divergence time analysis leveraging ten fossil specimens representative of most polyneopteran lineages: the time tree obtained supports an older radiation of the clade with respect to previous hypotheses. Euphasmatodea diversification is inferred to have started ~187 million years ago, suggesting that the Triassic-Jurassic mass extinction and the breakup of Pangea could have contributed to the process. We also investigated Euphasmatodea mitogenomes patterns of dN, dS and dN/dS ratio throughout our time-tree, trying to characterize the selective regime which may have shaped the clade evolution.

Keywords: Divergence time; Evolutionary radiation; Mitogenomics; Phasmatodea.

Introduction

Phasmatodea is a a small order of polyneopteran insects, with ~3000 described species present across all major landmasses and with a predominantly tropical distribution. The clades is known to include some of the most outstanding examples of cryptic mimicry, with lineages which resemble leaves, twigs and mosses (Bradler and Buckley, 2018). Several studies are contributing to the understanding of the systematic relationships within this clade, suggesting that present-day distribution is the result of rather recent dispersal events (Simon et al., 2019; Robertson et al., 2018). Nonetheless, many of the hypotheses which have been proposed throughout the years are conflicting. From a morphological perspective, the lack of resolved relationships at deeper nodes can be attributed to the convergent evolution of several morphological features, which blur the boundaries between taxa. At the same time, molecular phylogenetic approaches lack substantial support and consistency for deep relationships, while their outcome is often in disagreement with those found using morphological approaches (Buckley et al., 2010; Bradler et al., 2014; Bradler et al., 2015). Nonetheless, having a solid phylogenetic framework and deciphering the timing of lineages diversification is crucial for our understanding of phasmids evolution and their major adaptations.

The lack of fully resolved relationships in molecular phylogenies can result from a variety of factors, including: (i) uninformative or (ii) conflicting data, (iii) model violations, (iv) high homoplasy and (v) incomplete lineage sorting. A possible biological explanation for weak phylogenetic signal may also lie in ancient and rapid radiations (Whitfield et al., 2007), which are often associated to low supports at deeper nodes and the lack of consistency when different kind of data and analytical approaches are employed. This uncertainty can sometimes be resolved by either increasing the amount of data or by having a more extensive taxon sampling (Bragg et al., 2018).

Several processes can be involved in radiations, including a wide range of biotic and abiotic factors (Simões et al., 2016). Adaptive radiations can be triggered by key evolutionary innovations and/or exaptation. This has been proposed, for example, for the antifreeze glycoproteins of notothenioid fishes living in the ice-cold waters of Antarctica (Matschiner et al., 2011) or for wing patterns, pollen feeding and the expansion of olfactory receptor gene families that are related to a burst of diversification in Heliconius butterflies (Kozak et al., 2015). On the other hand, the increase in diversifications within a clade can also be driven by a wide range of abiotic factors which affect physical isolation, such as climate shifts or physical barriers (Lopèz-Estrada et al., 2018). Evolutionary radiations can also occur concurrently across different clades, as described by the Turnover-Pulse theory, which postulates that physical changes trigger biotic changes eventually resulting in a process of lineages turnover across a wide range of clades (Vrba, 1993).

We aimed to investigate the processes that shaped the initial divergence of Euhasmatodea lineages using mitochondrial genomes, which are an increasingly common tool for molecular systematics and phylogenetics in metazoan, particularly in insects (Cameron, 2014). Their wide use stems from the ease of getting new sequence data for a set of clearly orthologous genes, which can be readily compared to an increasing amount of available mitogenomic

sequences. Mitogenomes have been used many times to solve controversial systematic relationships at inter-ordinal (Rota-Stabelli and Telford, 2008; Song et al., 2016; Rutschmann et al., 2017) or intra-ordinal (Bourguignon et al., 2018; Li et al., 2017; Timmermans et al., 2014) taxonomic levels.

Although sequence variation in mitochondrial DNA was traditionally considered selectively neutral (Dowling et al., 2008), there is an increasing evidence that mitogenomes can experience episodes of positive selection as a consequence of shifts in physiological or environmental conditions. Those modifications underlie metabolic adaptations such as the accommodation of hematophagy and the origin of flight in vampire bats (Botero-Castro et al., 2018; Shen et al., 2010), lung reduction and venom evolution in snakes (Castoe et al., 2008), or the evolution of bioelectrogenesis in fishes (Elbassiouny et al., 2019). Also environmental conditions may affect mitochondrial genomes evolution, as pointed out in several animal taxa like marine and terrestrial mammals (Hassanin et al., 2009; Foote et al., 2011; Tomasco and Lessa, 2011), birds (Zhou et al., 2014), fishes (Wang et al., 2016), bivalve molluscs (Plazzi et al., 2017) and hexapods (Jiang et al., 2018; Yuan et al., 2018; Carapelli et al., 2019).

In the present paper we reconstructed new mitogenome sequences for 17 phasmid species from transcriptomic data deposited in the NCBI SRA database and analysed them in a wide phylogenetic framework along with previously published data of phasmid and other representatives of the major polyneopteran orders. Our aim was to investigate the evolutionary radiation of its suborder Euphasmatodea: we leveraged several well characterized fossils - mainly external to the Phasmatodea clade - to perform a divergence time analysis and obtained a timeframe for phasmid evolution which is found older from previous estimates. Moreover, we investigated the possibility of selective pressures which may have shaped the mitogenomes genetic variation.

Materials and Methods

Mitogenomes assembly and annotation.

We downloaded the RNA-Seq reads of all the 17 euphasmatodean species which did not have a corresponding mitogenome already assembled (last checked in March 2019) (Table 1). Reads were quality trimmed with a phred score threshold of 33, with a 25 bp-sliding window using Trimmomatic (Bolger et al., 2014) and quality checked with FastQC (Andrews, 2010). The Python script mitoRNA was then used to assemble mitogenomes from transcriptomic data (Forni et al., 2019; available at https://github.com/mozoo/ MitoRNA): this consists of a target assembly process, based on an iterative reference mapping and de-novo assembly process. This straightforward approach is useful to recover mitochondrial genomes from RNA-Seq experiments, given a reference. For each transcriptome analysed, all available phasmid mitogenomes (last checked in March 2019) were used as starting references for mitoRNA. On the obtained contigs, the 13 protein coding genes (PCGs) and the two ribosomal RNA (rRNAs) genes were manually annotated,

following the closest reference sequence. Control regions and tRNAs where omitted as they presented a low coverage, due to the transcriptional architecture of mitochondria (Forni et al., 2019).

Dataset, saturation and base composition analyses.

We carried out phylogenetic analyses on the newly obtained 17 nearly complete molecules combined with the 22 mitogenomes already available for Euphasmatodea (Table 1), bringing the total number up to 39. Additionally, we included 35 Polyneoptera mitogenomes for tree rooting and time tree calibration (Suppl. Table S1). We excluded representatives from the two orders Zoraptera and Embioptera due to phylogenetic artefacts, especially long branch attraction, that those mitogenomes are known to generate (Song et al., 2016).

Each gene was separately aligned using MAFFT with the option --auto for PCGs and X-INS-i for rRNAs (Katoh et al., 2013). PCG alignments were also inspected using AliView (Larsson, 2014) to select the correct reading frame and to check for stop codons. All poorly aligned regions were removed using Gblocks, v.0.91b (Castresana, 2000) with default settings and the codon parameter. Saturation and composition analyses were then performed using DAMBE v. 7.0.28 (Xia, 2018) on each single gene and on the concatenated alignment, with and without the third codon position. Saturation analyses were also carried out on a subset including only the Phasmatodea spp. We also used Aligroove to test for data heterogeneity on the concatenated alignment, with and without the third codon position.

Alignments, including the sequences of annotated genes of the new 17 mitochondrial genomes generated in this study, were deposited in Figshare (doi: https://doi.org/10.6084/m9.figshare.9885176.v3). The final concatenation was partitioned into three subsets: (i) first codon positions of PCGs; (ii) second codon positions of PCGs; (iii) small and large ribosomal subunit rRNAs.

Model selection, phylogenetic inference and divergence time analysis.

Model selection, Bayesian Inference and Maximum Likelihood analyses were carried out through the CIPRES Science Gateway (www.phylo.org).

For Bayesian Inference, the best-fit nucleotide substitution model and the optimal partitioning scheme were determined using PartitionFinder2, based on the corrected Akaike Information Criterion (Suppl. Table S2; Lanfear et al., 2016). The concatenated sequence alignment was analysed using BEAST 2.4.8 (Bouckaert et al., 2014) jointly estimating topology and node age. An uncorrelated lognormal relaxed clock model of rate variation across branches was implemented as a time prior and both Yule and Birth-Death speciation processes were used for the tree prior. Node calibrations were set using 10 fossil taxa with justified phylogenetic placement (Table 2). The minimum age constraints based on fossil records were implemented as exponential priors, with soft maximum priors (95% probability) taken from Bourguignon et al., 2018 or fixed at 411 Million years ago (Mya), which is the estimated age of the oldest hexapod fossil (Rhyniella praecursor Hirst & Maulik, 1926; Wolfe et al., 2016). Four replicates of the MCMC analysis were run with trees and

parameter values sampled every 5,000 steps over a total of 700 million generations. A 10% burn-in was discarded and the estimated sample size of parameters were assessed (ESS>200) using Tracer 1.6 (Rambaut et al., 2018). Run convergence was estimated with Tracer 1.6 and by visually comparing tree topologies.

In order to test the reliability of our inferred topology, we also analysed the concatenated sequence alignment of nucleotides and amino acids using the Maximum Likelihood approach of IQ-TREE v1.6.1 (Trifinopoulos et al., 2016). The best-fit models of nucleotide substitution were identified using IQ-TREE Model Selection (Suppl. Tab. S4; Kalyaanamoorthy et al., 2017) using the FreeRate model and the edge-linked parameter. Ten ML searches were run with 1,000 ultrafast bootstraps replicates; the run with the best likelihood was selected as the most reliable.

We downloaded all fossil occurrences of arthropods from the Fossilworks database (http://fossilworks.org) and generated three-timer (3T) origination and extinction rates, which were plotted through time. These metrics implement a sliding window approach, which is considered to minimize the artefacts deriving from the variation in sampling completeness (Alroy et al., 2008).

dN/dS analysis

The dN, dS and dN/dS ratios for each gene and for the concatenated matrix were calculated using the branch model of the CODEML program implemented in PAML 4.8a (Yang, 2007). The IQ-TREE-computed tree with fixed branch lengths was set as the user tree. Two models were used for dN/dS estimation: i) a single dN/dS ratio along the entire tree (model 0); ii) a specific dN/dS ratio for each given branch in the tree (model 1). We then used the likelihood ratio test (LRT) to determine the best-fitting model (Yang et al., 2007). Before all subsequent analyses, we excluded dN/dS estimates derived from values of dS<0.01, as they could lead to biased dN/dS estimates, as well as values of dN>2 and dS>2, which could indicate saturation of substitutions (Villanueva-Cañas et al., 2013). The timespan covered by the inferred time-tree was split into time bins of 10 million years and for each one we gathered the median of dN/dS values relative to all contemporary branches, for each PCG and their concatenation (as described in Plazzi et al., 2017). To check for possible biases in our analyses on selective pressures, linear regression were carried out a) between median dN, dS, dN/dS values and temporal midpoint of each time bin in the complete time-tree, b) between median dN, dS, dN/dS values and temporal midpoint of each time bin in the Phasmatodea sub-tree, c) between dN, dS, dN/dS values and the branch lengths found in the Maximum Likelihood inference, d) between dN, dS and dN/dS. We also tested for significant differences in dN, dS and dN/dS ratio between the branch leading to Timema californicum Scudder, 1895 and the one leading to Euphasmatodea by means of the two-tailed Wilcoxon signed rank test.

Results

Dataset, saturation and base composition analyses

When analyzed using Xia's method, all alignments showed little to no saturation (ISS < ISS.C; P<0.001), but ISS values were found smaller when the 3rd codon positions were excluded from PCGs (Suppl. Table S3a). The same method was applied to the sequences of phasmatodean species only, giving similar results as for the complete dataset (Suppl. Table S3b). When considering only the third codon positions of PCGs, the majority of the genes showed signs of saturation in the complete dataset, while it was present to a lesser extent when only phasmid species were considered (Suppl. Table S3c). Several genes and the concatenated alignment were found to have a heterogeneous nucleotide composition, even when excluding third codon positions (Chi-square test, P<0.01; Suppl. Table S4). However, values of Cramer's V, which measure the strength of the observed heterogeneity (i.e. the strength of association between nucleotide composition and taxa in the Chi-square test; Cramer, 1946), were low (<0.15), and were even lower when excluding 3rd codon positions. This, therefore, suggested that the observed nucleotide frequency differences among the taxa was not expected to substantially affect the phylogenetic inference under time-reversible models of evolution. Moreover, AliGROOVE showed a general lack of confounding signals related to sequence heterogeneity (Suppl. Fig. S1). Phylogenetic inferences were performed on a final matrix which included 72 species and 13,547 nucleotide positions, resulting from the concatenation of all the genes after the use of GBlocks and the removal of the 3rd codon position from PCGs.

Mitochondrial phylogeny and divergence times

Maximum Likelihood (-lnL = -215066.9305) analysis and Bayesian inferences, using both the Birth-Death (-lnL = -215655.2673) and Yule (lnL = -215655.8712) tree priors, produced identical topologies (Fig. 1; Fig. 2; Suppl. Fig. S2A-C). Since no Embioptera mitogenome was included in the analysis due to the phylogenetic artefacts associated to them (Song et al., 2016), the Notoptera clade (which includes the two orders Grylloblattodea and Mantophasmatodea) was recovered as phasmids sister lineage. Within the Phasmatodea clade, the first split was represented by the divergence between Timema californicum Scudder, 1895 (Timematodea) and the highly supported, monophyletic clade of Euphasmatodea (posterior probability, PP=1.0; bootstrap proportion, BP=100). While a few euphasmatodean families, like Phylliidae, formed highly supported clades, most of the families appeared either paraphyletic (Diapheromeridae) or polyphyletic (Phasmatidae and Lonchodidae). At the subfamily level, the supported clades were Phylliinae (Phyllium giganteum Hausleithner, 1984, Ph. tibetense Liu, 1993 and Ph. siccifolium (Linné, 1758)) (PP=1, BP=100), Necrosciinae (Micadina phluctainoides (Rehn, 1904), Neohirasea japonica (Haan, 1842), Sipyloidea sipylus (Westwood, 1859) and Calvisia medogensis Bi, 1993) (PP=1, BP=100) and Lonchodinae (Phraortes elongatus (Thunberg, 1815) [= illepidus Brunner v. Wattenwyl, 1907], 1907, Phraortes sp. Iriomote Island, Phraortes sp. Miyako Island, Carausius morosus

(Sinéty, 1901), Megalophasma granulatum Bi, 1995) (PP=1, BP=100). It has to be noted, though, that another available Phraortes sp. sample do not cluster with the other Phraortes species; no information is available about the specimen voucher and it could be likely a species misidentification. While several genera, like Micrarchus Carl, 1913, Bacillus Berthold, 1827, Phraortes Stål, 1875 and Phyllium Illiger, 1789 were recovered as monophyletic, Ramulus Saussure, 1862 was found paraphyletic and Tectarchus Salmon, 1954 appeared polyphyletic. As defined by Buckley et al. (2010), the New Zealand clade (labelled in Fig. 2) was here confirmed as monophyletic, with maximum support (PP=1, BP=100). The ML inference on the amino acid dataset (-lnL = -131128.4724) presented a slightly different topology regarding the Euphasmatodea clade with respect to the ML and BI inferences based on the nucleotide dataset. However, when collapsing poorly supported branches (BP<85) the two trees are fully compatible (Suppl. Fig. S2).

Divergence time estimates (Fig. 2) indicated that Phasmatodea diversification dates back to the Mid-Permian, 273.8 million years ago (95% HPD = 233.4-320.6 Mya), when the divergence between the Timema lineage and the Euphasmatodea clade took place. The Euphasmatodea radiation started in the Mid-Jurassic (187.5 Mya; 95% HPD = 145.3-205.8 Mya) and by the End Jurassic-Early Cretaceous all family-level lineages were already established. The New Zealand clade would have diverged later, the estimated age being 71.3 Mya (95% HPD = 53.8-89.5 Mya).

dN/dS analysis

For all single genes and the concatenated dataset, the null hypothesis that a single dN/dS applies to all the tree branches was rejected by the LRT test (P<0.001; Suppl. Table S5) and the alternative hypothesis of a dN/dS value for each branch was, therefore, accepted.

We found that dN/dS ratios showed a significantly positive correlation with time, both with the complete dataset and with the Phasmatodea sub-tree only (Fig. 3; Suppl. Fig. S3A-3B). To tentatively check whether these dN/dS estimates might have resulted from biased dN or dS values, although potentially problematic values were already filtered out (Villanueva-Cañas et al., 2013), we investigated their possible dependency from time and branch length. Both dN and dS values do not correlate with time in the complete dataset or correlate negatively in the Phasmatodea sub-tree only (Suppl. Fig. S3A-3B). Moreover, as expected, dN and dS correlated significantly with branch lengths, while dN/dS do not (Suppl. Fig. S3C). Finally, no correlation was found between dS and dN/dS both in the complete dataset and when considering only the Phasmatodea sub-tree (Suppl. Fig. S3D).

Interestingly, values of dN, dS and dN/dS calculated over the 13 PCGs for the T. californicum and Euphasmatodea stem branches resulted significantly different (Wilcoxon signed ranked test, P<0.001) (Table 3). Overall, comparing dN/dS values between the two branches, the difference varies between 1.4-fold to 8.6-fold, depending on the gene considered (Table 3).

Discussion

Relationships among Euphasmatodea lineages

New mitochondrial DNA data have been recovered for 17 phasmid species, by mining transcriptome sequencing, and are analysed in a phylomitogenomic framework for the first time. Overall, our analysis includes representative of eight out of the ten recognised polyneopteran orders and six Euphasmatodea families out of the eleven described. In our analysis, the systematic relationships of the polyneopteran orders are congruent with previous findings based on both molecular markers and phylogenomics with either mitochondrial (Song et al., 2016) or nuclear genes (Misof et al., 2014). The placement of all the euphasmatodean species included in the analysis and the relationships between families and subfamilies are generally consistent with previous phylogenetic analysis on this clade both using whole mitogenomes (Kômoto et al., 2010; Zhou et al., 2017) and molecular markers (Tilgner, 2002; Whiting et al., 2003; Bradler, 2009; Kômoto et al., 2011; Bradler et al., 2014). Despite recently proposed as a new family (Robertson et al., 2018), being supported also by different studies (Bradler et al., 2014; Bradler et al., 2015; Goldberg et al., 2015), we recovered Lonchodidae as a polyphyletic group, consistently with other analyses that utilized mitogenomes (Kômoto et al., 2011; Zhou et al., 2017) and molecular markers (Buckley et al., 2010; Glaw et al., 2019). In our analysis Megacrania alpheus (Westwood, 1859) and Phobaeticus serratipes (Gray, 1835), both belonging to the family Phasmatidae, form a monophyletic cluster, consistently with Kômoto et al. (2011). On the contrary, in Zhou et al. (2017), M. alpheus was instead recovered as the sister taxon of Extatosoma tiaratum (Macleay, 1826) and Phobaeticus serratipes. The Necrosciinae clade presents a topology consistent with previous works dealing with the same species (Kômoto et al., 2011; Zhou et al., 2017) and the placement of Neohirasea japonica in this clade validates the change of the taxonomic status from Lonchodinae to Necrosciinae suggested by Bradler et al. (2014). The New Zealand clade presents a topology largely consistent with the previous literature based on molecular markers; the only difference we found in our inferred topology is that while in the previous literature (Trewick et al., 2008; Buckley et al., 2010; Bradler et al., 2015; Dennis et al., 2015) Niveaphasma annulatum (Hutton, 1898) was in a sister relationship with Micrarchus hystriculeus (Westwood, 1859) and Asteliaphasma jucundum (Salmon, 1991) was external to this group, in our work we recover a sister relationship between Asteliaphasma jucundum and Niveaphasma annulatum, and Micrarchus hystriculeus clustered with the congeneric Mirarchus spp. (PP=1, BP=100).

Timing of Euphasmatodea evolution

In order to provide a temporal framework for the evolution of phasmids, we calculated a time tree using fossils calibration. Fossil calibrations on phasmids have been already used in previous studies, recovering the origin of Phasmatodea between 90 and 122 Mya and the one of Euphasmatodea between 50 and 61 Mya ago (Buckley et al., 2010; Bradler et al. 2015; Robertson et al., 2018, Simon et al., 2019). However, as acknowledged by the same authors,

the use of those fossil calibrations could have potentially led to a substantial underestimation of cladogenetic events (Buckley et al., 2010; Bradler et al., 2015). Therefore, in order to avoid such a possible bias, we included calibration points based on eight polyneopteran fossils and only two phasmid fossils. Our analysis dates the appearance of major Polyneoptera orders consistently with previous studies which used molecular markers, mitogenomics and nuclear phylogenomics (Misof et al., 2014; Tong et al., 2015; Legendre et al., 2105; Bourguignon et al., 2018, Evangelista et al., 2019; Montagna et al., 2019) (Suppl. Fig. S4). In our time tree, the Phasmatodea clade is estimated ~275 Million years old and the Euphasmatodea crown group is dated shortly after 190 Mya: these dates are somehow consistent with estimates obtained in recent phylogenetic analyses based on whole mitogenomes (Bourguignon et al., 2018) and transcriptomes (Evangelista et al., 2019). Our results are also congruent with the literature on phasmids fossils: a recent paper by Yang et al. (2019) provides support for a diversification of major clades of stick insects happening during or before the mid-Cretaceous. Moreover, the presently obtained estimates of Phasmatodea origin are consistent with the oldest fossils attributed to this taxon, such as Arachnophasma scurra (279.5-272.5 Mya; Aristov and Rasnitsyn, 2015) or Isadyphasma bashkuevi (259-254 Mya; Gorochov et al., 2013) which were not used as calibrations points for the divergence time analysis.

The outcome of our dating analysis suggests that, the radiation of Euphasmatodea most likely occurred shortly after the Triassic-Jurassic boundary (~200 Mya). There are mainly two events that could have played key roles in the evolutionary radiation and diversification of this order of insects: the concurrent mass extinction event and the beginning of the Pangea break up.

The coincidence of the Euphasmatodea radiation and the breakup of Pangea which started around 180 Mya (McIntyre et al., 2017) supports a diversification scenario where vicariance has contributed significantly to shape phasmid distribution and diversity. Yet, instances of long-range dispersal, cannot be ruled out: for example, it has been shown that eggs of Megacrania tsudai Shiraki, 1933 can be dispersed by floating on the marine surface (Kobayashi et al., 2014) and those of Ramulus mikado (Rehn, 1904) can be dispersed by birds (Suetsugu et al., 2018, where it is referred to as Ramulus irregulariterdentatus). However, it is still unclear whether passive, long-range dispersal capacities are common among stick insects and they probably cannot be generalized at the level of the entire order. Previous hypotheses on the divergence of New Zealand phasmids dated the event at 26 Mya (16.8 – 35.5 Mya; Buckley et al. 2010), supporting a long-range dispersal from New Caledonia subsequent to the putative re-emergence of New Zealand (22 Mya). On the contrary, our divergence time estimate for the clade is retrieved at 73 Mya (95% HPD = 54.8 – 91.7 Mya), which is consistent with a vicariance scenario and a long-term survival of phasmatodean lineages in New Zealand.

The ecological modifications and the increase in habitat diversity associated with the Pangea fragmentation have been linked to an increase of metazoan diversity (Jordan et al., 2016) and could have played a key role in the diversification of other polyneopteran (Svenson, et al., 2009) and non-polyneopteran (Tang, et al., 2019) insect clades. Environmental

perturbations, such as the Triassic-Jurassic extinction, have been also associated with high lineage turnover in different clades, such as amphibians (Roelants et al., 2007), reptiles (Toljagic et al., 2013) and insects as well (Tang et al., 2019). This turnover dynamic can be also observed for most arthropods, indicating a phenomenon that affected several clades at the same time (Fig. 3). Having found Euphasmatodea radiation shortly after the Triassic-Jurassic boundary suggests that their diversification could have happened in the context of a wider process of lineages turnover.

Molecular evolution of Euphasmatodea mitogenomes

We investigated dN/dS ratios throughout the inferred time-tree in order to better understand the evolution of euphasmatodean mitogenomes. The dN/dS ratio is widely used to infer changes in selection regimes, but it is known that its estimation can be affected by substitution saturation, which may result in an underestimation of the real number of synonymous substitutions and eventually leading to inflated dN/dS values. We tried to correct estimates for saturation by excluding values of of dN and dS >2, as they may actually imply undetected multiple substitutions (Villanueva-Cañas et al., 2013). Moreover, specific tests indicated that saturation is a marginal feature of our dataset, being only observed in some genes 3ª codon positions. We further analyzed the dN and dS behaviour with respect to time, branch length and dN/dS, in order to have a better understanding of possible biases. First of all, both dN and dS showed the same trend when correlated with time, suggesting that both are similarly affected. Second, dS values well correlated with branch length, indicating that longer branches do not lead to dS underestimation. Third, dN/dS values do not correlate with dS, implying that higher dN/dS values are not resulting by low, potentially underestimated dS values. Overall, these results suggested that obtained dN/dS values are reliable.

Two observations are worth to be considered, that may have implications with phasmid evolution: 1) Euphasmatodea stem branch presented significantly higher dN/dS values compared to the stem branch of T. californicum; 2) higher dN/dS values were found in the early-diverging branches of the Phasmatodea sub-tree compared to the more recent ones.

The difference observed between euphasmatodean and T. californicum stem branches is difficult to explain, although we may speculate about possible adaptive processes acting on mitogenomes. In fact, a generalized reduction of selective pressures should impact similarly all the molecule, the observed differential pattern across genes is expected in case of an adaptive process (Tomasco and Lessa, 2011). However, it is interesting to note that branches with higher dN/dS values were those corresponding to early divergences: this may indicate a possible implication in the subsequent euphasmatodean radiation. It is, therefore, possible to hypothesize that a relaxation of selective pressures may have helped during the early steps of the radiation, when environmental condition changed because of the Pangea fragmentation, eventually facilitating phasmid diversification. This is, in fact, consistent with the lower dN/dS value of T. californicum stem branch, and the limited diversity and restricted geographic distribution of Timematodea species (Vickery and Sandoval, 2001). Alternatively, it can be considered that a rapid radiation may have concurred to shape

dN/dS values: it is known in fact, that mildly deleterious, non-synonymous mutations can be randomly brought to fixation as a consequence of low effective population size (Ne) and the subsequent relaxation of purifying selection (Otha, 1993; Woolfit and Bromhan, 2005; Hughes 2007; Brandvain and Wright, 2016). A decrease in Ne can be caused by the occurrences of narrow and repeated colonization bottlenecks which happen during a radiation. When main lineages emerged and the radiation process slowed down a purifying selection regime was restored. However, this cannot explain the higher dN/dS value on the Euphasmatodea stem branch occurring before the radiation.

Conclusions

Mitogenomics provided a novel perspective on Euphasmatodea diversification. Our analyses supported a more ancient radiation of Euphasmatodea than previously thought: this new perspective suggested several contemporary events which could have acted as drivers of the clade fast diversification. Euphasmatodea radiation appeared coincident with the Pangea fragmentation and, therefore, vicariance could have played a prominent role in shaping the diversity and distribution of leaf and stick insects. Their diversification, then, could have taken place in the context of a wider dynamic of lineages turnover that occurred subsequently to the Triassic-Jurassic mass extinction. Our analyses also suggested that a possible relaxation of selective pressures may have played a role in facilitating the Euphasmatodea radiation.

Acknowledgements

The present work has been supported by Canziani funding to AL and BM. We wish to thank the two anonymous Reviewers and the Editors for their useful and constructive comments that greatly improved the manuscript.

References

Alroy J. 2008. Dynamics of origination and extinction in the marine fossil record. Proc. Natl. Acad. Sci. USA. 105, 11536–11542.

Andrews, S. 2010. FastQC. http://www.bioinformatics.babraham.ac.uk/ projects/fastqc/.

Aristov, D.S. and Rasnitsyn, A.P. 2015. New insects from the Kungurian of Tshekarda fossil site in Permian territory of Russia. Russian Entomol. J. 24:17-35

Bethoux, O., Nel, A., Lapeyrie, J., Gand, G., Galtier, J. 2002. *Raphogla rubra* gen. n., sp. n., the oldest representative of the clade of modern Ensifera (Orthoptera: Tettigoniidae, Gryllidea). Eur. J. Entomol. 99, 111–116.

Bolger, A.M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Botero-Castro, F., Tilak, M.K., Justy, F., Catzeflis, F., Delsuc, F., Douzery, E.J.P. 2018. In cold blood: compositional bias and positive selection drive the high evolutionary rate of vampire bats mitochondrial genomes. Genome Biol. Evol. 10, 2218-2239.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard, MA., Rambaut, A., Drummond, A. J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PloS Comput. Biol., 10, e1003537.

Bourguignon, T., Tang, Q., Ho, S.Y.W., Juna, F., Wang, Z., Arab, D. A., et al. 2018. Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. Mol. Biol. Evol. 35, 970–983.

Bradler, S., 2009. Die Phylogenie der Stab-und Gespenstschrecken (Insecta: Phasmatodea), Thesis, University of Göttingen.

Bradler, S., Buckley, T. R. 2018. Biodiversity of phasmatodea. Insect Biodiversity: Science and Society, 2, 281-313.

Bradler, S., Robertson, J.A., Whiting, M.F. 2014. A molecular phylogeny of Phasmatodea with emphasis on Necrosciinae, the most species rich subfamily of stick insects. Syst. Entomol. 39, 205-222.

Bradler, S., Cliquennois, N., Buckley, T.R. 2015. Single origin of the Mascarene stick insects: ancient radiation on sunken islands? BMC Evol. Biol. 15, 196.

Bragg, J.G., Potter, S., Silva, A.C.A., Hoskin, C.J., Bai, B.Y.H., Moritz, C. 2018. Phylogenomics of a rapid radiation: the Australian rainbow skinks. BMC Evol. Biol. 18, 1–12.

Brandvain, Y., Wright, S.I. 2016. The limits of natural selection in a nonequilibrium world. Trends Genet. 32, 201–210.

Brongniart, C. 1885. Les Insectes fossiles des terrains primaires: coup d'oeil rapide sur la faune entomologique des terrains paléozoïques. Rouen,impr. J. Lecerf.

Buckley, T.R, Attanayake, D., Nylander, J.A.A., Bradler, S. 2010: The phylogenetic placement and biogeographical origins of the New Zealand stick insects (Phasmatodea). Syst. Entomol. 35, 207-225.

Cameron, S.L. 2014. insect mitochondrial genomics: implications for evolution and phylogeny. Annu. Rev. Entomol. 59, 95–117.

Carapelli, A., Fanciulli, P.P., Frati, F., Leo, C., 2019. Mitogenomic data to study the taxonomy of Antarctic springtail species (Hexapoda: Collembola) and their adaptation to extreme environments. Polar Biol. 42, 715–732.

Castoe, T.A., Jiang, Z.J., Gu, W., Wang, Z.O., Pollock, D.D., 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. PLoS ONE. 3, e2201–2214.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540-552.

Cramer, H. 1946. Mathematical methods of statistics. Princeton, Princeton University Press.

Dennis, A. B., Dunning, L. T., Sinclair, B. J., Buckley, T. R. 2015. Parallel molecular routes to cold adaptation in eight genera of New Zealand stick insects. Sci. Rep. 5, 13965.

Dowling, D., Friberg, U., Lindell, J., 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. Trends Ecol. Evol. 23, 546–554.

Engel, M.S., Grimaldi, D.A., Krishna, K. 2007. Primitive termites from the Early Cretaceous. Stuttgarter Beitr Naturk. 371, 1–32.

Elbassiouny, A.A., Lovejoy, N.R., Chang, B.S.W., 2019. Convergent patterns of evolution of mitochondrial oxidative phosphorylation (OXPHOS) genes in electric fishes. Phil. Trans. R. Soc. B 375, 20190179–9.

Engel, M.S., Ortega-Blanco, J., Azar, D. 2011. The earliest earwigs in amber (Dermaptera): a new genus and species from the Early Cretaceous of Lebanon. Insect Syst. Evol. 42, 139–148.

Engel, M.S., Wang, B., Alqarni, A.S. 2016. A thorny, ‘anareolate’ stick-insect (Phasmatidae s.l.) in Upper Cretaceous amber from Myanmar, with remarks on diversification times among Phasmatodea. Cretac. Res. 63, 45–53.

Evangelista, D.A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M.K., et al. 2019. An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). Proc. Roy. Soc. B. 286, 20182076–20182079.

Foote, A.D., Morin, P.A., Durban, J.W., Pitman, R.L., Wade, P., Willerslev, E., Gilbert, M.T.P., da Fonseca, R.R. 2011. Positive selection on the killer whale mitogenome. Biol. Lett. 7, 116–118.

Forni, G., Puccio, G., Bourguignon, T., Evans, T., Mantovani, B., Rota-Stabelli, O., Luchetti A. 2019. Complete mitochondrial genomes from transcriptomes: assessing pros and cons of data mining for assembling new mitogenomes. Sci. Rep. 9, 14806.

Glaw, F., Hawlitschek, O., Dunz, A., Goldberg, J., Bradler, S., 2019. When giant stick insects play with colors: molecular phylogeny of the Achriopterini and description of two new splendid species (Phasmatodea: Achrioptera) from Madagascar. Front. Ecol. Evol. 7, 125–18.

Goldberg, J., Bresseel, J., Constant, J., Kneubühler, B., Leubner, F., Michalik, P., Bradler, S. 2015. Extreme convergence in egg-laying strategy across insect orders. Sci. Rep. 5, 485–487.

Gorochov, A.V. 2013. New taxa of the superorder Orthopteroidea from the latter half of the Permian of European Russia. *Paleontol. J.* 47, 782-793.

Hassanin, A., Ropiquet, A., Couloux, A., Cruaud, C. 2009. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). J. Mol. Evol. 68, 293–310.

Huang, D.Y., Nel, A., Zompro, O., Waller, A. 2008. Mantophasmatodea now in the Jurassic. Naturwissenschaften. 95, 947–52.

Hughes, A.L. 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity, 99, 364–373.

Jiang, G.-F., 2018. Positive selection drove the adaptation of mitochondrial genes to the demands of flight and high-altitude environments in grasshoppers. Front. Genet. 9:605

Jordan, S.M.R., Barraclough, T.G., Rosindell, J. 2016. Quantifying the effects of the break up of Pangaea on global terrestrial diversification with neutral theory. Phil. Trans. R. Soc. B 371, 20150221–9.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., Jermiin, L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14, 587.

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

Kobayashi, S., Usui, R., Nomoto, K., Ushirokita, M., Denda, T., Izawa, M., 2014. Does egg dispersal occur via the ocean in the stick insect *Megacrania tsudai* (Phasmida: Phasmatidae)? Ecol. Res. 29, 1025-1032.

Kômoto, N., Yukuhiro, K., Ueda, K., Tomita, S. 2011. Exploring the molecular phylogeny of phasmids with whole mitochondrial genome sequences. Mol. Phylogenet. Evol. 58, 43-52.

Kozak, K.M., Wahlberg, N., Neild, A.F.E., Dasmahapatra, K.K., Mallet, J., Jiggins, C.D. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. Syst. Biol. 64, 505–24.

Krishna, K., Grimaldi, D.A., Krishna, V., Engel, M.S., 2013. Treatise on the Isoptera of the World: volume 1 Introduction. Bull. Am. Mus. Nat. Hist. 377, 1–200.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773.

Larsson, A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 30, 3276-3278.

Legendre, F., Nel, A., Svenson, G. J., Robillard, T., Pellens, R., Grandcolas, P. 2015. Phylogeny of Dictyoptera: dating the origin of cockroaches, praying mantises and termites with molecular data and controlled fossil evidence. PloS ONE. 10, e0130127.

Li, H. et al., 2017. Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. Proc. R. Soc. B. 284, 20171223.

López-Estrada, E.K., Sanmartín, I., García-París, M., Zaldívar-Riverón, A. 2018. High extinction rates and non-adaptive radiation explains patterns of low diversity and extreme morphological disparity in North American blister beetles (Coleoptera, Meloidae). Mol. Phylogenet. Evol.130, 156–68.

Matschiner, M., Hanel, R., Salzburger, W. 2011. On the origin and trigger of the notothenioid adaptive radiation. PloS ONE. 6, e18911.

McIntyre, S.R.N., Lineweaver, C.H., Groves, C.P., Chopra, A. 2017. Global biogeography since Pangaea. Proc. R. Soc. B. 284, 20170716.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science. 346, 763-767.

Montagna, M., Tong, K.J., Magoga, G., Strada, L., Tintori, A., Ho, S.Y.W., Lo, N., 2019. Recalibration of the insect evolutionary time scale using Monte San Giorgio fossils suggests survival of key lineages through the End-Permian Extinction. Proc. Roy. Soc. B 286, 20191854–20191859.

Ohta T., 1993. Amino acid substitution at the *Adh* locus of Drosophila is facilitated by small population size. Proc Natl Acad Sci USA, 90, 4548-4551.

Plazzi, F., Puccio, G., Passamonti, M., 2017. Burrowers from the past: mitochondrial signatures of Ordovician bivalve infaunalization. Genome Biol. Evol. 9, 956–967.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901-904.

Robertson, J.A., Bradler, S., Whiting, M.F. 2018. Evolution of oviposition techniques in stick and leaf insects (Phasmatodea). Front. Ecol. Evol. 6, 125–15.

Roelants, K., Gower, D.J., Wilkinson, M., Loader, S.P., Biju, S.D., Guillaume, K., Moriau, L., Bossuyt, F. 2007. Global patterns of diversification in the history of modern amphibians. Proc. Natl. Acad. Sci. USA 104, 887–892.

Rota-Stabelli, O., Telford, M.J. 2008. A multi criterion approach for the selection of optimal outgroups in phylogeny: recovering some support for Mandibulata over Myriochelata using mitogenomics. Mol. Phylogenet. Evol. 48,103–111.

Rutschmann, S., Chen, P., Zhou, C., Monaghan, M.T. 2017. Using mitochondrial genomes to infer phylogenetic relationships among the oldest extant winged insects (Palaeoptera). BioRxiv 1–33, https://doi.org/10.1101/164459

Sharov, A.G., 1961. Otryad Plecoptera (Order Plecoptera). Paleozojskoe nasekomye Kuznetskovo bassejna [Paleozoic insects from the Kuznetsk basin]. Tr. Paleontol. Inst. Akad. Nauk SSSR 85, 225–234.

Suetsugu, K., Funaki, S., Takahashi, A., Ito, K., Yokoyama, T. 2018. Potential role of bird predation in the dispersal of otherwise flightless stick insects. Ecology. 99, 1504–1506.

Shen, Y.Y., Liang, L., Zhu, Z.H., Zhou, W.P., Irwin, D.M., Zhang, Y.P. 2010. Adaptive evolution of energy metabolism genes and the origin of flight in bats. Proc. Natl. Acad. Sci. USA. 107, 8666-8671.

Simões, M., Breitkreuz, L., Alvarado, M., Baca, S., Cooper, J.C., Heins, L., Lieberman, B.S., 2016. The evolving theory of evolutionary radiations. Trends Ecol. Evol. 31, 27-34.

Simon, S., Letsch, H., Bank, S., Buckley, T.R., Donath, A., Liu, S., Machida, R., Meusemann, K., Misof, B., Podsiadlowski, L., Zhou, X., Wipfler, B., Bradler, S., 2019. Old World and New World Phasmatodea: phylogenomics resolve the evolutionary history of stick and leaf insects. Front. Ecol. Evol. 7, 744–14.

Song, H., Amédégnato, C., Cigliano, M.M., Desutter-Grandcolas, L., Heads, S.W., Huang, Y., et al. 2015. 300 million years of diversification: elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. Cladistics. 31, 621–651.

Svenson, G.J., Whiting, M.F. 2009. Reconstructing the origins of praying mantises (Dictyoptera, Mantodea): the roles of Gondwanan vicariance and morphological convergence. Cladistics. 25, 468–514.

Tang, P., Zhu, J.C., Zheng, B.Y., Wei, S.J., Sharkey, M., Chen, X.X., et al. 2019. Mitochondrial phylogenomics of the Hymenoptera. Mol. Phylogenet. Evol. 131, 8–18.

Tilgner, E.H., 2002. Systematics of phasmida, Unpublished PhD. Thesis. University of Georgia.

Timmermans, M.J.T.N., Lees, D.C., Simonsen, T.J. 2014. Towards a mitogenomic phylogeny of Lepidoptera. Mol. Phylogenet. Evol. 79, 169–78.

Toljagic, O., Butler, R. J., 2013. Triassic-Jurassic mass extinction as trigger for the Mesozoic radiation of crocodylomorphs. Biol. Lett. 9, 20130095–20130095.

Tomasco, I.H., Lessa, E.P. 2011. The evolution of mitochondrial genomes in subterranean caviomorph rodents: adaptation against a background of purifying selection. Mol. Phylogenet. Evol. 61, 64–70.

Tong, K.J., Duchene, S., Ho, S.Y.W., Lo, N., 2015. Comment on “Phylogenomics resolves the timing and pattern of insect evolution”. Science. 349, 487-487.

Trewick, S.A., Morgan-Richards M., Collins J.L. 2008. Are you my mother? Phylogenetic analysis reveals orphan hybrid stick insect genus is part of a monophyletic New Zealand clade. Mol. Phylogenet. Evol. 48, 799-808.

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., Minh, B. Q. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucl. Acids Res. 44(W1), W232-W235.

Vickery, V. R., Sandoval, C. P. 2001. Descriptions of three new species of *Timema* (Phasmatoptera: Timematodea: Timematidae) and notes on three other species. J Orthoptera Res. 10(1), 53-61.

Villanueva-Cañas, J. L., Laurie, S., Albà, M. M. 2013. Improving genome-wide scans of positive selection by using protein isoforms of similar length. *Genome Biol Evol*. *5*(2), 457–467.

Vrba, E. S., 1993. Turnover-pulses, the Red Queen, and related topics. Am. J. Sci. 293, 418.

Vrsansky, P., 2002. Origin and the early evolution of mantises. Amba Proj. 6, 1-16.

Wang, Y., Shen, Y., Feng, C., Zhao, K., Song, Z., Zhang, Y., Yang, L., He, S. 2016. Mitogenomic perspectives on the origin of Tibetan loaches and their adaptation to high altitude. Sci. Rep. 6, 2969.

Wedmann, S., Bradler, S., Rust, J., 2007. The first fossil leaf insect: 47 million years of specialized cryptic morphology and behavior. Proc. Natl. Acad. Sci. USA. 104, 565–569.

Whitfield, J.B., Lockhart, P.J. 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22, 258–265.

Whiting, M. F., Bradler, S., Maxwell, T., 2003. Loss and recovery of wings in stick insects. Nature. 421, 264.

Wolfe, J.M., Daley, A.C., Legg, D.A., Edgecombe, G.D. 2016. Fossil calibrations for the arthropod Tree of Life. Earth Sci. Rev. 160, 43–110.

Woolfit, M., Bromham, L. 2005. Population size and molecular evolution on islands. Proc Biol Sci, 272(1578), 2277-2282.

Xia X. 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. Mol. Biol. Evol. 35, 1550–1552

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24, 1586-1591.

Yang H, Yin X, Lin X, Wang C, Shih C, Zhang W, Ren D, Gao T., 2019. Cretaceous winged stick insects clarify the early evolution of Phasmatodea. Proc. Biol. Sci. 286(1909), 20191085.

Yuan, M.L., Zhang, Q.-L., Zhang, L., Jia, C.L., Li, X.P., Yang, X.Z., et al. 2018. Mitochondrial phylogeny, divergence history and high-altitude adaptation of grassland caterpillars (Lepidoptera Lymantriinae Gynaephora) inhabiting the Tibetan Plateau. Mol. Phylogenet. Evol. 122, 116–24.

Zhou, T., Shen, X., Irwin, D.M., Shen, Y., Zhang, Y., 2014. Mitogenomic analyses propose positive selection in mitochondrial genes for high-altitude adaptation in galliform birds. Mitochondrion. 18, 70–75.

Zhou, Z., Guan, B., Chai, J., Che, X. 2017. Next-generation sequencing data used to determine the mitochondrial genomes and a preliminary phylogeny of Verophasmatodea insects. J. Asia-Pac. Entomol. 20, 713-719.

Species	Genbank acc. no.	SRA
Acanthoxyla sp.		SRR2089887
Aretaon asperrimus		SRR1172270
Argosarchus horridus		SRR2089902
Asteliaphasma jucundum		SRR2089913
Carausius morosus		SRR3211828
Clitarchus hookeri		SRR3080266
Medauroidea extradentata		SRR1172394
Micrarchus sp.		SRR1054193
Micrarchus hystriculeus		SRR1054191
Niveaphasma anulata		SRR2089878
Peruphasma schultei		SRR1002984
Phraortes sp. YW-2014		SRR1189755
Phyllium siccifolium		ERR392012
Ramulus artemis		SRR1172422
Spinotectarchus acornutus		SRR2089897
Tectarchus salebrosus		SRR2089908
Tectarchus ovobessus		SRR2089893
Bacillus atticus	GU001955	
Bacillus rossius	GU001956	
Calvisia medogensis	KY124330	
Dryococelus australis	AP018522	
Extatosoma tiaratum	AB642680	
Entoria okinawaensis	AB477459	
Heteropteryx dilatata	AB477468	
Megacrania alpheus	AB477471	
Megalophasma granulatum	KY124331	
Micadina phluctainoides	AB477466	
Neohirasea japonica	AB477469	
Orestes mouhotii	AB477462	
Phobaeticus serratipes	AB477467	
Phraortes illepidus	AB477460	
Phraortes sp. Iriomote Island	AB477464	
Phraortes sp. Miyako Island	AB477465	
Phyllium giganteum	AB477461	
Phyllium tibetense	KY124332	
Ramulus irregulariterdentatus	AB477463	
Ramulus hainanense	FJ156750	
Sipyloidea sipylus	AB477470	
Timema californicum	DQ241799	

Table 1 - Genbank accession numbers for the 17 RNA-Seq experiments and the 21 assembled mitogenomes utilized.

Species	Minimum age (Mya)	Maximum age (Mya)	Calibration group	Reference
Rhadinolabis phoenicica	129.4	411	crown Dermaptera	Engel et al., 2011
Palaeotaeniopteryx elegans	268.3	411	crown Plecoptera	Sharov et al., 1961
Oedischia williamsoni	299.0	411	stem Orthoptera	Brongniart et al., 1885
Raphogla rubra	271.8	411	crown Orthoptera	Bethoux et al., 2002
Juramantis initialis	145	316	crown Dyctioptera	Vrsansky et al., 2002
Valditermes brennenae	130.3	325	crown Isoptera	Krishna et al., 2013
Reticulitermes antiquus	33.9	95	crown Coptotermitinae + Heterotermitinae	Engel et al., 2007
Juramantophasma sinica	158.1	411	stem Mantophasmatodea - Grylloblattodea	Huang et al., 2008
Echinosomiscus primoticus	98.2	411	crown Phasmatodea	Engel et al., 2016
Eophyllium messelensis	47.0	411	stem *Phyllium*	Wedeman et al., 2007

Table 2 - Fossil records, dating and calibration points for the divergence time analysis.

	ATP6	ATP8	CO1	CO2	CO3	CYTB	ND1	ND2	ND3	ND4	ND4L	ND5	ND6	Concatenated
T. californicum dN/dS	0.099	0.481	0.043	0.085	0.089	0.063	0.107	0.188	0.217	0.106	0.110	0.094	0.262	0.113
T. californicum dN	0.034	0.064	0.017	0.031	0.029	0.024	0.042	0.047	0.050	0.039	0.042	0.038	0.055	0.036
T. californicum dS	0.341	0.132	0.389	0.361	0.329	0.373	0.390	0.248	0.231	0.372	0.383	0.409	0.210	0.319
Euphasmatodea dN/dS	0.285	NA	0.216	0.263	0.770	0.267	0.246	0.497	0.295	0.246	0.821	0.443	0.731	0.546
Euphasmatodea dN	0.052	0.075	0.044	0.051	0.065	0.049	0.054	0.060	0.053	0.053	0.066	0.061	0.065	0.061
Euphasmatodea dS	0.183	0.000	0.205	0.192	0.084	0.185	0.219	0.121	0.181	0.214	0.081	0.137	0.089	0.112
Fold change	2.9	NA	5.1	3.1	8.6	4.2	2.3	2.6	1.4	2.3	7.4	4.7	2.8	4.8

Table 3 - Values of dN/dS. dN and dS obtained for *T. californicum* and Euphasmatodea stem branches

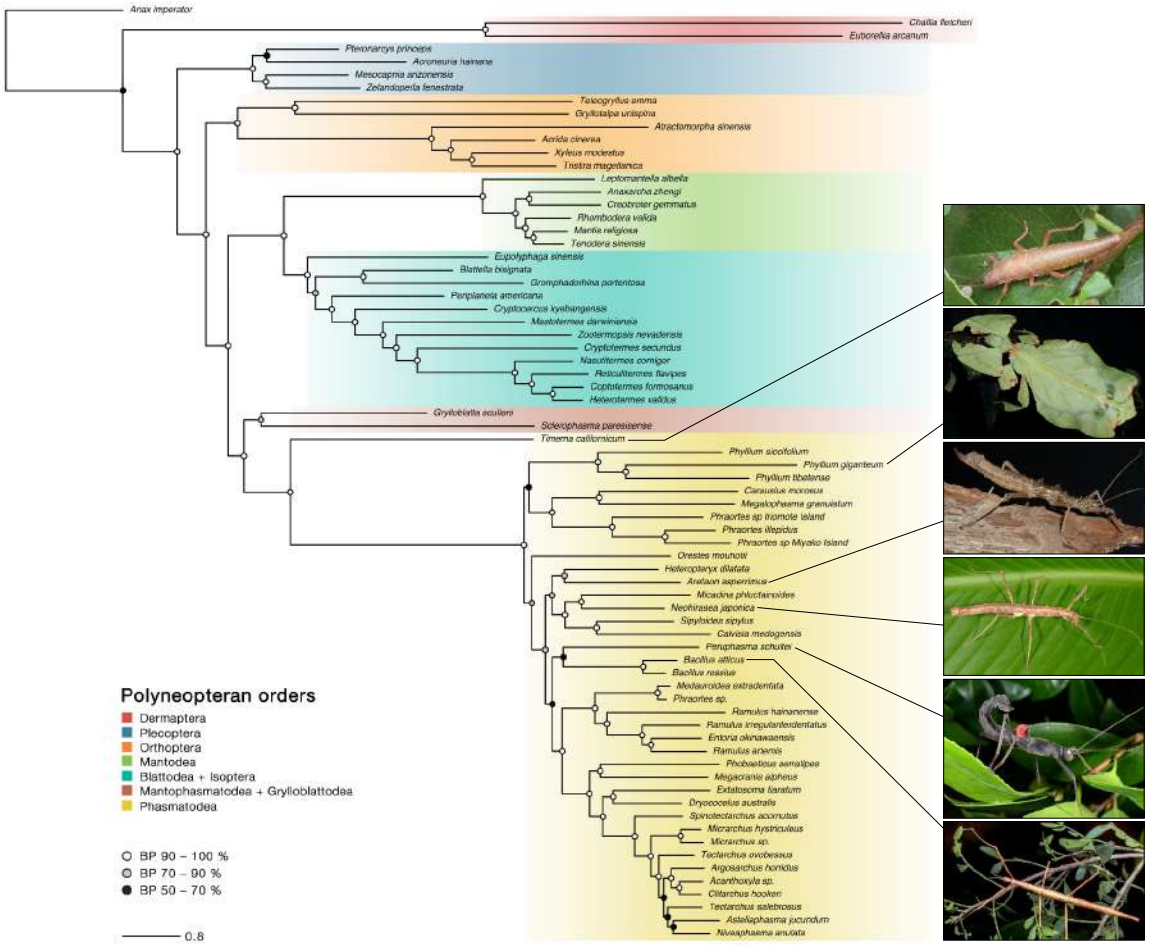


Figure 1 – Maximum likelihood tree obtained from the full dataset, with pictures illustrating some of the stick and leaf insects analysed in the present paper.

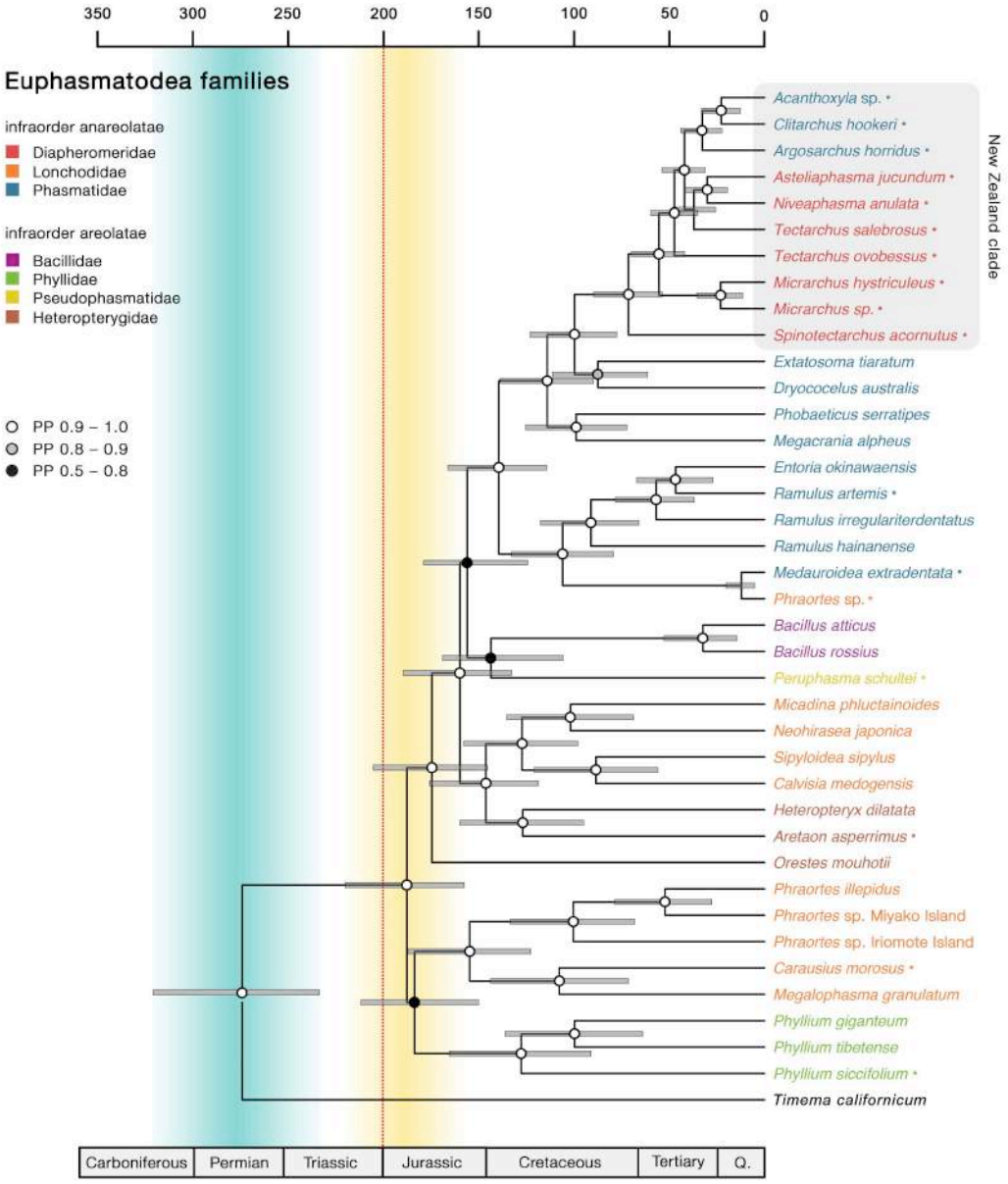


Figure 2 – Bayesian phylogeny and estimates of divergence time for Phasmatodea lineages. Asterisks highlight newly assembled mitogenomes; the scale on the top indicates Million years.

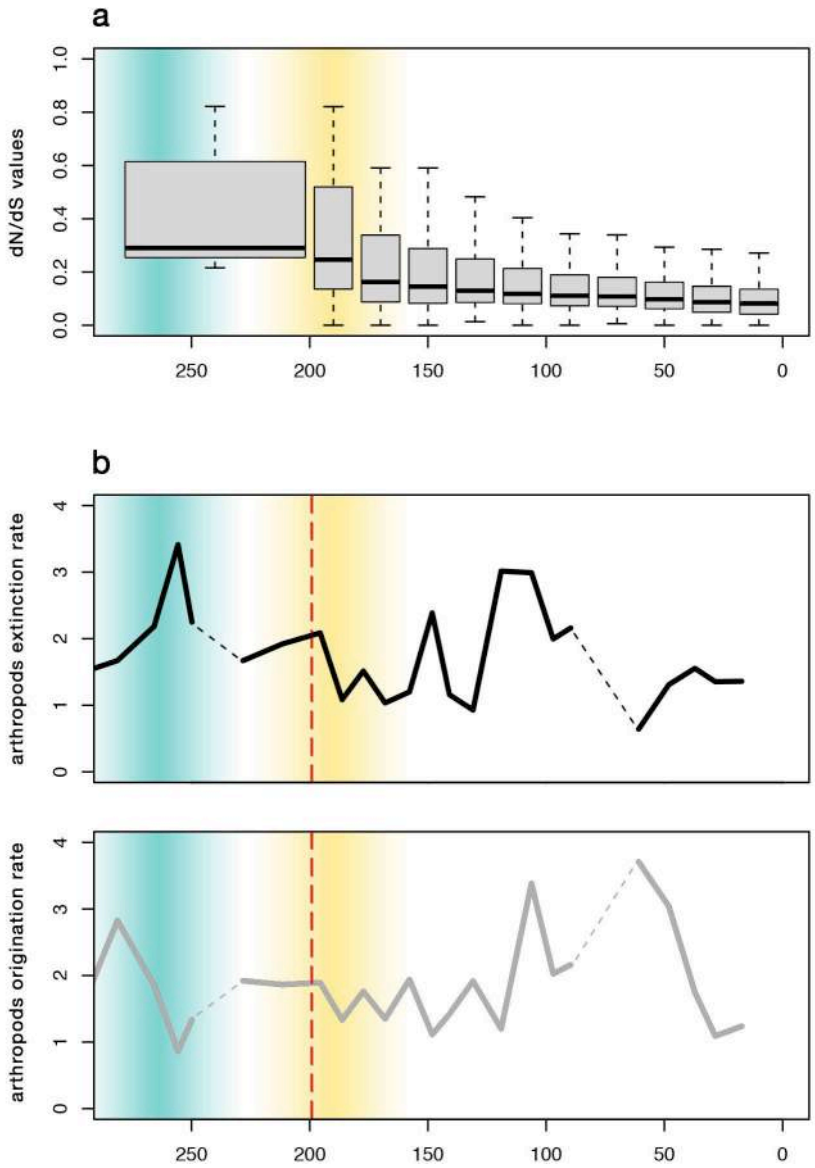


Figure 3 – Comparison of divergence dates of Euphasmatodean stem (blue) and crown (yellow) nodes relatively to (a) variation in median dN/dS values thourgh time and (b) to three-timer (3T) origination and extinction rates of arthropods obtained by the Fossilworks portal.

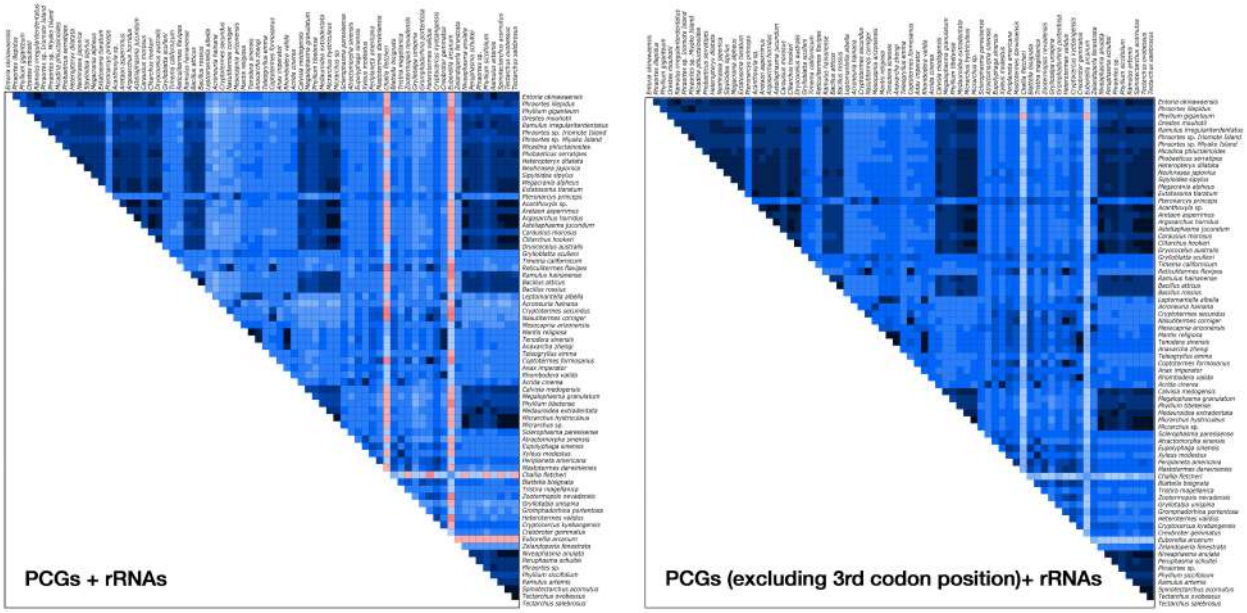


Figure S1 – Outcome of the AliGROOVE analysis, including and excluding 3rd codon position.

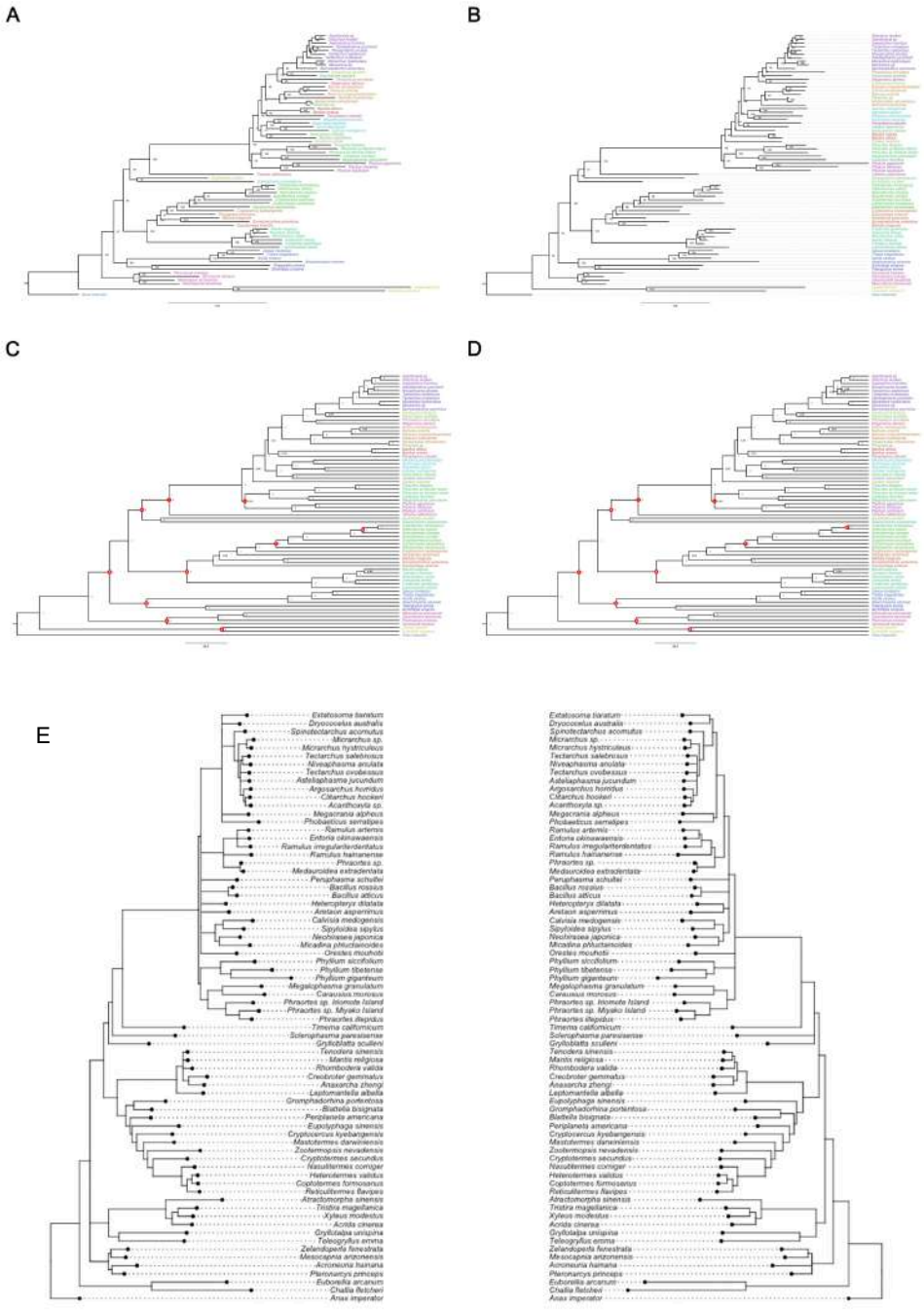


Figure S2 - A: ML inference on the nucleotide dataset (13 PCGs with third codon position excluded and 2 rRNA). B: ML inference on the amino acid dataset (13 PCGs). C: Estimate of divergence on the nucleotide dataset (13 PCGs with third codon position excluded and 2 rRNA) using a yule tree prior. D: Estimate of divergence time on the nucleotide dataset (13 PCGs with third codon position excluded and 2 rRNA) using a Birth-Death tree prior. For A-D: posterior probabilities and rapid bootstrap values are reported at nodes only if they are < 1 and < 100 respectively. For C - D: calibration points described in Table 2 are marked in red with the letters from A to L. E: ML inference on the amino acid dataset (left panel, 13 PCGs) and on the nucleotide dataset (right panel, 13 PCGs with third codon position excluded and 2 rRNA).

Figure S3 - Plot of median values of dN, dS and dN/dS through time and their linear regression with time for each PCG of the mitochondrion and their concatenation for the whole dataset (A) and phasmatodea only (B). In the C panel, the plot of median values of dN, dS and dN/dS against branch length and their linear regression. In the D panel dN/dS ratios were plotted against corresponding dS values in both complete dataset (left) and in Phasmatodea only (right)

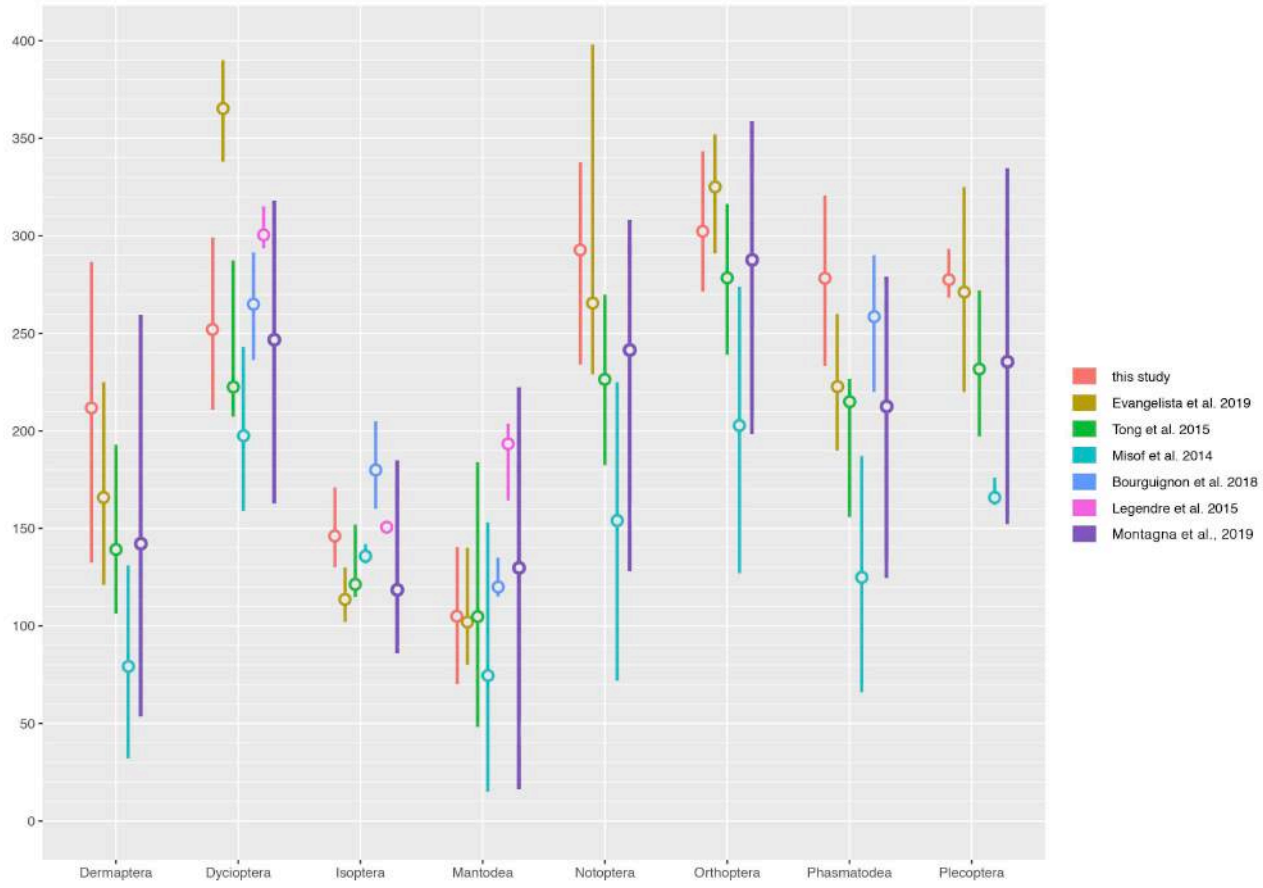


Figure S4 – Comparison of the dating of the main polyneopteran lineages with relevant literature.

Genbank acc. no.	Species	Order
JQ910991	*Zorotypus medoensis*	Zoraptera
AB639034	*Aposthonia japonica*	Embioptera
NC_018538	*Challia fletcheri*	Dermaptera
NC_032075	*Euborellia arcanum*	Dermaptera
KM199685	*Acroneuria hainana*	Plecoptera
KP642637	*Mesocapnia arizonensis*	Plecoptera
NC_034997	*Zelandoperla fenestrata*	Plecoptera
AY687866	*Pteronarcys princeps*	Plecoptera
KX673195	*Acrida cinerea*	Orthoptera - Caelifera
NC_014490	*Xyleus modestus*	Orthoptera - Caelifera
NC_011824	*Atractomorpha sinensis*	Orthoptera - Caelifera
NC_020773	*Tristira magellanica*	Orthoptera - Caelifera
NC_029148	*Gryllotalpa unispina*	Orthoptera – Ensifera
KU562917	*Teleogryllus emma*	Orthoptera – Ensifera
NC_007701	*Sclerophasma paresisense*	Mantophasmatodea
DQ241796	*Grylloblatta sculleni*	Grylloblattodea
KU201317	*Mantis religiosa*	Mantodea
KX611804	*Rhombodera valida*	Mantodea
NC_030267	*Creobroter gemmatus*	Mantodea
KU201320	*Anaxarcha zhengi*	Mantodea
KJ463364	*Leptomantella albella*	Mantodea
KU201318	*Tenodera sinesi*	Mantodea
NC_030191	*Cryptocercus kyebangensis*	Blattodea
NC_016956	*Periplaneta americana*	Blattodea
NC_030001	*Gromphadorhina portentosa*	Blattodea
NC_018549	*Blattella bisignata*	Blattodea
NC_014274	*Eupolyphaga sinensis*	Blattodea
NC_018120	*Mastotermes darwiniensis*	Isoptera
KP026283	*Cryptotermes secundus*	Isoptera
NC_024658	*Zootermopsis nevadensis*	Isoptera
KU925203	*Coptotermes formosanus*	Isoptera
NC_030034	*Heterotermes validus*	Isoptera
EF206314	*Reticulitermes flavipes*	Isoptera
KP091691	*Nasutitermes corniger*	Isoptera
KX161841	*Anax imperator*	Odonata

Table S1. Polyneopteran species included in the analysis.

Subset	Best Model	lnL	No. sites	Partition names
1	GTR+I+G+X		3869	PCG_1st
2	GTR+I+G+X		3869	PCG_2nd
3	GTR+I+G+X		1353	rRNA

Table S2b. IQ-Tree Model Selection and partition scheme for nucleotides

Subset	Best Model	lnL	No. sites	Partition names
1	GTR+F+R6	-115129.4979	3869	PCG_1st
2	GTR+F+R5	-66861.5838	3869	PCG_2nd
3	GTR+F+R5	-33275.8469	1353	rRNA

Table S2c. IQ-Tree Model Selection and partition scheme for amino acids

Subset	Best Model	lnL	No. sites	Partition names
1	mtZOA+F+R6	-41194.575	897	ATP6, ATP8, CO3, ND2, ND3, ND6
2	mtZOA+F+R6	-30868.934	1073	CO1, CO2, CYTB
3	mtZOA+F+R7	-59287.817	1303	ND1, ND4, ND4L, ND5

Table S2a. PartitionFinder2 model selection and partition scheme for BEAST v2.4.8 analysis

Gene	I	I_Sym	T	DF	P	I_Asym	T	DF	P
atp6	0.410	0.719	13.435	662	0.0000	0.392	0.747	662	0.4550
atp6_no	0.292	0.696	14.500	441	0.0000	0.368	2.750	441	0.0062
atp8	0.475	0.818	5.951	95	0.0000	0.609	2.329	95	0.0220
atp8_no	0.374	0.981	9.093	65	0.0000	0.895	7.796	65	0.0000
co1	0.287	0.775	32.853	1511	0.0000	0.492	13.826	1511	0.0000
co1_no	0.139	0.748	42.029	1007	0.0000	0.443	20.982	1007	0.0000
co2	0.383	0.719	14.749	662	0.0000	0.393	0.424	662	0.6716
co2_no	0.253	0.696	17.134	441	0.0000	0.368	4.467	441	0.0000
co3	0.349	0.729	18.108	770	0.0000	0.409	2.867	770	0.0043
co3_no	0.221	0.704	20.756	515	0.0000	0.378	6.757	515	0.0000
cytB	0.339	0.756	24.038	1124	0.0000	0.457	6.824	1124	0.0000
cytB_no	0.206	0.727	27.317	749	0.0000	0.406	10.487	749	0.0000
nd1	0.362	0.741	20.541	908	0.0000	0.430	3.673	908	0.0003
nd1_no	0.241	0.713	22.405	605	0.0000	0.385	6.834	605	0.0000
nd2	0.521	0.740	11.064	893	0.0000	0.428	4.748	893	0.0000
nd2_no	0.402	0.712	12.827	595	0.0000	0.384	0.725	595	0.4689
nd3	0.434	0.686	8.045	350	0.0000	0.356	2.493	350	0.0131
nd3_no	0.323	0.684	9.478	233	0.0000	0.362	1.020	233	0.3090
nd4	0.424	0.766	21.453	1313	0.0000	0.476	3.259	1313	0.0011
nd4_no	0.309	0.738	22.298	875	0.0000	0.425	6.041	875	0.0000
nd4l	0.499	0.682	5.505	269	0.0000	0.355	4.317	269	0.0000
nd4l_no	0.389	0.697	7.859	181	0.0000	0.390	0.015	181	0.9881
nd5	0.444	0.781	23.115	1679	0.0000	0.504	4.110	1679	0.0000
nd5_no	0.331	0.756	23.568	1119	0.0000	0.457	6.986	1119	0.0000
nd6	0.483	0.685	6.754	338	0.0000	0.355	4.270	338	0.0000
nd6_no	0.382	0.685	8.234	225	0.0000	0.364	0.477	225	0.6336
concat	0.380	0.818	81.705	11939	0.0000	0.527	35.829	11939	0.0000
concat_no	0.276	0.815	87.896	8416	0.0000	0.571	48.017	8416	0.0000

Table S3a. Saturation analyses performed using DAMBE v. 7.0.28 on the whole 72 spp. dataset.

Gene	I_{ss}	I_{ss} Sym	T	DF	P	I_{ss} Asym	T	DF	P
atp6	0.345	0.719	17.866	662	0.0000	0.393	2.302	662	0.0216
atp8	0.363	0.818	9.073	95	0.0000	0.609	4.916	95	0.0000
co1	0.219	0.775	42.782	1511	0.0000	0.492	20.997	1511	0.0000
co2	0.272	0.719	22.092	662	0.0000	0.393	5.971	662	0.0000
co3	0.279	0.729	24.097	770	0.0000	0.409	6.982	770	0.0000
cytB	0.279	0.756	30.538	1124	0.0000	0.457	11.418	1124	0.0000
nd1	0.268	0.741	29.108	908	0.0000	0.430	3.673	908	0.0000
nd2	0.381	0.740	20.089	893	0.0000	0.428	2.600	893	0.0095
nd3	0.346	0.686	12.319	350	0.0000	0.356	0.373	350	0.7097
nd4	0.322	0.766	30.647	1313	0.0000	0.476	10.645	1313	0.0000
nd4l	0.324	0.682	11.279	269	0.0000	0.355	0.981	269	0.3276
nd5	0.344	0.781	32.987	1679	0.0000	0.504	12.065	1679	0.0000
nd6	0.356	0.685	12.040	338	0.0000	0.355	0.008	338	0.9938

Table S3b. Saturation analyses performed using DAMBE v. 7.0.28 on the Phasmatodea clade.

Table S3c. Saturation analyses performed using DAMBE v. 7.0.28 on third codon positions.

Gene	I_ss	I_ss Sym	T	DF	P	I_ss Asym	T	DF	P
atp6	0.767	0.686	2.970	221	0.0033	0.366	14.732	221	0.0000
atp6_phasmatodea	0.673	0.686	0.438	219	0.6620	0.366	10.320	219	0.0000
atp8	0.820	1.156	4.625	50	0.0000	1.195	5.174	50	0.0000
atp8_phasmatodea	0.756	1.156	5.825	49	0.0000	1.195	6.404	49	0.0000
co1	0.704	0.703	0.077	506	0.9389	0.378	18.102	506	0.0000
co1_phasmatodea	0.618	0.703	4.454	500	0.0000	0.378	12.535	500	0.0000
co2	0.751	0.686	2.325	220	0.0210	0.366	13.822	220	0.0000
co2_phasmatodea	0.635	0.686	1.644	218	0.1017	0.366	8.677	218	0.0000
co3	0.730	0.682	1.869	257	0.0627	0.356	14.770	257	0.0000
co3_phasmatodea	0.646	0.682	1.405	254	0.1614	0.356	11.034	254	0.0000
cytB	0.709	0.688	0.993	377	0.3213	0.359	17.178	374	0.0000
cytB_phasmatodea	0.625	0.688	2.984	370	0.0030	0.359	12.479	370	0.0000
nd1	0.716	0.683	1.322	305	0.1870	0.354	14.515	305	0.0000
nd1_phasmatodea	0.613	0.683	2.594	296	0.0100	0.354	9.570	296	0.0000
nd2	0.846	0.683	7.268	294	0.0000	0.354	21.865	294	0.0000
nd2_phasmatodea	0.724	0.683	1.604	290	0.1098	0.354	14.420	290	0.0000
nd3	0.767	0.763	0.105	116	0.9168	0.512	6.699	116	0.0000
nd3_phasmatodea	0.682	0.763	2.135	114	0.0349	0.521	4.429	114	0.0000
nd4	0.784	0.840	1.065	89	0.2897	0.648	2.635	89	0.0099
nd4_phasmatodea	0.618	0.840	3.998	86	0.0001	0.648	0.542	86	0.5891
nd4l	0.789	0.840	0.988	89	0.3258	0.648	2.744	89	0.0073
nd4l_phasmatodea	0.616	0.840	4.042	86	0.0001	0.648	0.577	86	0.5658
nd5	0.757	0.708	2.665	559	0.0079	0.380	20.511	559	0.0000
nd5_phasmatodea	0.668	0.708	2.018	546	0.0441	0.380	14.342	546	0.0000
nd6	0.552	0.689	4.986	366	0.0000	0.359	7.012	366	0.0000
nd6_phasmatodea	0.673	0.747	1.933	124	0.0555	0.484	4.895	124	0.0000

Table S4. Compositional bias analysis performed using DAMBE v. 7.0.28 ("_yes" and "_no" indicate presence and absence of the 3rd codon position, respectively).

Gene	Statistic	Value	d.f.	P
Concat_yes	Chi-square	7064.24	213	0.0000
	Likelihood ratio chi-square	6864.45	213	0.0000
	Cramer's V	0.0524		
Concat_no	Chi-square	2393.00	213	0.0000
	Likelihood ratio chi-square	2363.23	213	0.0000
	Cramer's V	0.0363		
12s	Chi-square	328.94	213	0.0000
	Likelihood ratio chi-square	349.14	213	0.0000
	Cramer's V	0.0561		
16s	Chi-square	594.00	213	0.0000
	Likelihood ratio chi-square	624.07	213	0.0000
	Cramer's V	0.0566		
ATP6_yes	Chi-square	1101.22	213	0.0000
	Likelihood ratio chi-square	1088.05	213	0.0000
	Cramer's V	0.0877		
ATP6_no	Chi-square	333.58	213	0.0000
	Likelihood ratio chi-square	338.99	213	0.0000
	Cramer's V	0.0591		
ATP8_yes	Chi-square	216.03	213	0.4292
	Likelihood ratio chi-square	204.53	213	0.6493
	Cramer's V	0.1022		
ATP8_no	Chi-square	82.95	210	1.0000
	Likelihood ratio chi-square	85.50	210	1.0000
	Cramer's V	0.0769		
CO1_yes	Chi-square	1209.87	210	0.0000
	Likelihood ratio chi-square	1218.96	210	0.0000
	Cramer's V	0.0613		
CO1_no	Chi-square	178.52	210	0.9438
	Likelihood ratio chi-square	175.97	210	0.9579
	Cramer's V	0.0288		
CO2_yes	Chi-square	785.80	213	0.0000
	Likelihood ratio chi-square	782.58	213	0.0000
	Cramer's V	0.0742		
CO2_no	Chi-square	192.48	213	0.8402
	Likelihood ratio chi-square	209.03	213	0.5640
	Cramer's V	0.0450		
CO3_yes	Chi-square	960.05	213	0.0000
	Likelihood ratio chi-square	956.92	213	0.0000
	Cramer's V	0.0759		
CO3_no	Chi-square	263.27	213	0.0108
	Likelihood ratio chi-square	257.75	213	0.0195

	Cramer's V	0.0486		
CYTB_yes	Chi-square	1641.81	213	0.0000
	Likelihood ratio chi-square	1636.52	213	0.0000
	Cramer's V	0.0822		
CYTB_no	Chi-square	355.62	213	0.0000
	Likelihood ratio chi-square	355.20	213	0.0000
	Cramer's V	0.0469		
ND1_yes	Chi-square	1219.14	213	0.0000
	Likelihood ratio chi-square	1198.59	213	0.0000
	Cramer's V	0.0788		
ND1_no	Chi-square	268.47	213	0.0059
	Likelihood ratio chi-square	255.84	213	0.0237
	Cramer's V	0.0453		
ND2_yes	Chi-square	2099.13	213	0.0000
	Likelihood ratio chi-square	2025.75	213	0.0000
	Cramer's V	0.1044		
ND2_no	Chi-square	730.71	213	0.0000
	Likelihood ratio chi-square	731.06	213	0.0000
	Cramer's V	0.0754		
ND3_yes	Chi-square	613.23	213	0.0000
	Likelihood ratio chi-square	619.19	213	0.0000
	Cramer's V	0.0913		
ND3_no	Chi-square	216.99	213	0.4112
	Likelihood ratio chi-square	232.11	213	0.1757
	Cramer's V	0.0656		
ND4_yes	Chi-square	2135.63	213	0.0000
	Likelihood ratio chi-square	2062.82	213	0.0000
	Cramer's V	0.0868		
ND4_no	Chi-square	589.87	213	0.0000
	Likelihood ratio chi-square	581.99	213	0.0000
	Cramer's V	0.0559		
ND4L_yes	Chi-square	400.09	213	0.0000
	Likelihood ratio chi-square	385.97	213	0.0000
	Cramer's V	0.0831		
ND4L_no	Chi-square	122.55	213	1.0000
	Likelihood ratio chi-square	151.01	213	0.9996
	Cramer's V	0.0560		
ND5_yes	Chi-square	2589.90	213	0.0000
	Likelihood ratio chi-square	2470.48	213	0.0000
	Cramer's V	0.0846		
ND5_no	Chi-square	658.96	213	0.0000
	Likelihood ratio chi-square	654.21	213	0.0000
	Cramer's V	0.0523		
ND6_yes	Chi-square	810.88	213	0.0000
	Likelihood ratio chi-square	777.48	213	0.0000
	Cramer's V	0.1054		
ND6_no	Chi-square	331.88	213	0.0000

Likelihood ratio chi-square	343.49	213	0.0000	
Cramer's V	0.0826			

Gene	d.f. model 0	*lnL* model 0	d.f. model 1	*lnL* model 1	d.f._tot	LRT	P
atp6	2	-30360.43335	143	-30033.46768	141	653.931324	0
atp8	2	-10103.91239	143	-9982.624533	141	242.575714	2.26412E-07
co1	2	-51212.94269	143	-50753.72117	141	918.443024	0
co2	2	-28220.83637	143	-27922.88866	141	595.895428	0
co3	2	-31835.20386	143	-31532.90308	141	604.601564	0
cytb	2	-45323.86456	143	-45013.20347	141	621.322182	0
nd1	2	-36423.6979	143	-36175.91101	141	495.573778	0
nd2	2	-59519.80984	143	-59143.74421	141	752.131254	0
nd3	2	-17129.03471	143	-16991.73711	141	274.595188	1.2167E-10
nd4	2	-57913.6413	143	-57489.46682	141	848.348968	0
nd4l	2	-12144.9913	143	11940.47297	141	48170.92853	0
nd5	2	-78137.29453	143	-77618.92948	141	1036.730102	0
nd6	2	-29077.1655	143	-28889.7924	141	374.746206	0
concatenated	2	-490670.2616	143	-487814.6644	141	5711.194468	0

Table S5. Likelihood ratio test (LRT) to determine the best-fitting model between (0) a single dN/dS ratio along the entire tree (1) a specific dN/dS ratio for each given branch in the tree.

Molecular systematics and phylogenetics of the Australian stick insect genus *Candovia* Stål, 1875 (Phasmida, Necrosciinae).

Giobbe Forni, Alex Cussigh, Paul D. Brock, Braxton R. Jones, Filippo Nicolini, Jacopo Martelossi, Andrea Luchetti and Barbara Mantovani.

Abstract - The phasmid genus *Candovia* counts nine recognised species endemic to Australia. However, much of its species diversity has gone undetected, mainly due to the highly convergent morphological evolution related to crypsis in phasmids. In this study, we unravel the diversity of the *Candovia* genus through molecular species delimitation approaches using the Folmer region of Cytochrome Oxidase subunit I, along with phylogenetic analyses on seven additional mitochondrial and nuclear loci. Our analysis confirmed the already described species and led to the delineation of twelve putative new taxa. The phylogenetic analyses reinforced our understanding of the systematic relationships among *Candovia* species, and suggest the monophyly of the taxon. Moreover, mapping the presence of egg capitulum on the tree suggests the possible multiple origin of this structure in *Candovia*.

Keywords: Australian Fauna; Species Delimitation; Molecular phylogenetics; Stick insects; Phasmids.

Introduction

The order Phasmida Leach, 1815 (Hexapoda; Insecta) consists of 3348 valid species (Brock *et al*., 2020), mostly with a tropical or subtropical distribution. It contains insects that are well known for their remarkable mimicry, from crypsis to background matching, or disruptive coloration to masquerading (Merilaita and Lind, 2005; Skelhorn, 2015). Their camouflage influences the understanding of their taxonomy and systematic relationships. This is because their morphology presents instances of both convergent evolution and high intraspecific variability, but also they bear high phenotypic plasticity (Gutiérrez-Valencia *et al*., 2017). It is, therefore, important to rely on the combined outcomes of different approaches to unravel their diversity. A possible solution to the problem lies in implementing a molecular approach to provide species hypotheses and compare them with morphological observations. Despite several shortcomings having been highlighted, concerning the usage of a single locus mitochondrial gene for this purpose (Moritz and Cicero 2004), and just after the barcoding identification system was proposed by Hebert *et al*. (2003), molecular approaches been considered a quick and efficient exploratory strategy for the delimitation and validation of species boundaries, and integrative taxonomy. Many empirical studies have in fact shown that some species delimitation methods can lead to an under- or over-splitting of species, so that the results of molecular species delimitation could be better interpreted along with other information, such as comparative morphology or ecology (Dellicour and Flot, 2018; Jacobs *et al*., 2018; Luo *et al.,* 2018).

The Australian Phasmida fauna lists, at present, 103 species distributed in 11 subfamilies ascribed to four families; three within the infraorder Anaereolate: Diapheromeridae Kirby, 1904, Lonchodidae Brunner, 1893 and Phasmatidae Leach, 1815, consisting of stick insects, and one family in the Areolatae: Phylliidae Brunner, 1893, which are true leaf insects. Diapheromeridae and Lonchodidae are small- to medium-sized species and Phasmatidae are generally very large and winged, while Phylllidae have a lateralized leaf-shaped body (Brock and Hasenpush, 2009). Velonà *et al*. (2015) successfully applied, for the first time, the DNA barcoding methodology on 16 Australian stick insect taxa to verify its suitability as a tool to both identify species and uncover possible instances of undetected diversity. Among others, it emerged a high differentiation of the three putative morphospecies of the genus *Candovia*, *Candovia* spp. A, B and C, and their divergence from *C. annulata* (Brunner, 1907). The genus *Candovia* Stål, 1875 (Lonchodidae: Necrosciinae Brunner, 1893: Necrosciini Brunner, 1893) consists of nine recognized species, all of which are present in Australia, feeding

on a wide range of trees, shrubs and ferns. The species *C. evoneobertii* Zompro and Adis, 2001, usually listed as the tenth taxon, is endemic of South America and seems to be more related to the genus *Paracalynda* Zompro, 2001, sharing just a superficial similarity to *Candovia*. It is, therefore, to be removed from this genus (Brock and Hasenpush, 2009).

This study aims to explore the molecular diversity of the *Candovia* genus across different populations. Moreover, recent sampling resulted in the observation of morphologically diverse specimens, not recognizable as known taxa. Therefore, to support their formal descriptions as possible new taxa and highlight their phylogenetic relationships, they were included in this study and, along with data from previous investigations (Velonà *et al.,* 2015), they were analyzed with species delimitation approaches and phylogenetic methods.

Material and methods

Taxon sampling

Adult specimens were collected between 2009 and 2017 across 17 locations in Southern and Eastern Australia (Figure 1 and Table 1). After morphological observations were taken, voucher specimens were deposited at the Brisbane Museum (BM), Australian National Insect Museum, Canberra (ANIC) and Natural History Museum, London (NHMUK). We referred to new specimens which could not be reliably assigned to any known species as *Candovia* spp. D to L, following the species delimitation analysis (see below). Sixty-seven *Candovia* individuals (Table 1) were selected for molecular analyses; their leg tissues were preserved in high grade ethanol (99%) and stored at 4 °C until use. For phylogenetic analyses, 18 publicly available sequences of Lonchodidae species (17 Necrosciinae and one Lonchodinae Brunner, 1893 species) were drawn from GenBank and *Extatosoma tiaratum* (Macleay, 1826) (Phasmatidae, Extatosomatinae) was utilized as outgroup (Suppl. Table 1).

DNA extraction, amplification and sequencing

Genomic DNA was isolated for the 67 newly collected specimens using the kit Smarter Nucleic Acid Preparation (Stratec) following the manufacturer's standard protocol.

The Folmer region of Cytochrome Oxidase subunit I for DNA barcoding (Folmer *et al*., 1994; henceforth referred to as *cox1*$_{Fol}$) was PCR amplified from all samples using primers and the condition is given in Suppl. Table 2.

On the basis of the species delimitation results, we selected 39 specimens for which, in addition to the *cox1*$_{Fol}$, we amplified further seven loci: four mitochondrial (a *cox1* fragment downstream the *cox1*$_{Fol}$ region, and indicated as *cox1*; Cytochrome Oxidase subunit II – *cox2*; 12S and 16S rDNAs) and three nuclear markers (Histone subunit 3, H3; 18S and 28S rDNAs). Amplified fragment length, primers and thermal cycling conditions can be found in Supp. Table 2. All PCR products were visualized by 1% agarose gel electrophoresis, subsequently purified using ExoSAP-IT PCR Product Cleanup Reagent (Thermofisher) and sequenced by Macrogen Inc. Chromatograms were inspected using SeqTrace 0.9.0 (Stucky, 2012) and the resulting sequences were manually checked using Aliview 1.26 (Larsson, 2014). We then compared the sequences obtained to the NCBI Genbank database using BLAST (with *blastn* algorithm: Altschul *et al*., 1990) to identify potential contaminants. New sequences were submitted to GenBank under the accession numbers MT077516-MT077845.

Cox1$_{Fol}$ *analysis and species delimitation*

Cox1$_{Fol}$ fragments have been aligned as amino acids using MAFFT v. 7 with *--auto* parameter setting (Katoh and Standley, 2013) and then retro-translated to nucleotides. The dataset was partitioned in the three codon positions and analyzed by both Maximum Likelihood (ML) and Bayesian Inference (BI) methods. All phylogenetic analyses were carried out with XSEDE (eXtreme Science and Engineering Discovery Environment, https://www.xsede.org) through CIPRES Science Gateway (www.phylo.org; Miller, Pfeiffer and Schwartz, 2012). ML tree was obtained with IQ-TREE v1.6.1 (Nguyen *et al*., 2014). The best-fit models of nucleotide substitution were identified using ModelFinder (Kalyaanamoorthy *et al.,* 2017; Suppl. Table 3.1); then, 10 ML searches were run with 1,000 ultrafast bootstrap replicates and the run with the best likelihood was selected as the most reliable. For the BI, we determined the best-fit nucleotide substitution model using PartitionFinder2, based on the corrected Bayesian Information Criterion using the edge-linked parameter and the greedy strategy (Lanfear *et al*., 2016; Suppl. Table 3.1). But the partition scheme and the substitution model selected by PartitionFinder led to poor mixing, most likely due to over-parametrization. We thus modified the partition scheme to use the less parametrized model HKY.

The *cox1*$_{Fal}$ dataset was, then, analyzed using several species delimitation approaches to explore the cryptic diversity of the genus and to explore the species hypotheses. For this purpose we used both distance and evolutionary model-based methods, without *a priori* species hypothesis: (i) Automatic Barcode Gap Discovery (ABGD; Puillandre *et al*., 2012) analysis was performed using Jukes-Cantor distances, a relative gap width of 1 and 10 bins; (ii) the different Poisson Tree Processes approaches (PTP and mPTP; Zhang *et al*., 2013; Kapli *et al*., 2017) were carried out using the IQ-TREE best ML tree as input: two runs of 1,000,000 generations with sampling every 100 and a burn-in of 20% as parameters were carried out and compared for convergence; (iii) the General Mixed Yule Coalescent (GMYC; Pons *et al.*, 2006) was run on ultrametric trees, using the single threshold parameter. The concatenated sequence alignment (see below) was analyzed with BEAST 2.5 (Bouckaert *et al*., 2019) to generate ultrametric trees for the GMYC species delimitation method. We used four combinations of different clock and speciation model priors: strict or relaxed clock model and Yule or coalescent process. Each analysis was run with trees and parameter values sampled every 5,000 steps over a total of 60 million generations. A burn-in of 20% steps was discarded and adequate sample sizes (ESS>200) were checked using Tracer 1.6 (Rambaut *et al.*, 2018). Trees were summarized in the maximum clade credibility tree from the posterior distribution in TreeAnnotator v1.4.7 (Drummond *et al.*, 2012). GMYC and ABGD were used with their online implementation (available at https://species.h-its.org/gmyc/ and https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html, respectively).

Phylogenetic analyses

Each gene was aligned separately using MAFFT v. 7 with the option *--auto* for protein coding genes (PCGs) and with the option *--X-INS-i* for the rDNA genes (Katoh and Standley, 2013). We then manually inspected alignments using AliView (Larsson, 2014) to select the correct reading frame and to check for stop codons. Ambiguously aligned regions were removed from the single gene alignments with Gblocks (Talavera and Castresana, 2007) with options for a less stringent selection. We then concatenated gene sequences to form a single character matrix using Phyutility (Smith and Dunn, 2008).

Model selection, ML and BI analyses were carried out with XSEDE (eXtreme Science and Engineering Discovery Environment, https://www.xsede.org) through CIPRES Science Gateway (www.phylo.org; Miller *et al*., 2012). For the ML approach, the best-fit partitioning scheme and models of nucleotide substitution were identified using IQ-TREE v1.6.12 ModelFinder (Kalyaanamoorthy *et al*., 2017) (Suppl. Table 3.2) using the

edge-proportional parameter. We then inferred 10 trees using IQ-TREE v1.6.12 (Nguyen *et al*., 2014) with 1,000 ultrafast bootstraps, and the run with the best likelihood was selected. For what concerns the BI, we determined the best-fit nucleotide substitution model and the optimal partitioning scheme using PartitionFinder2, based on the corrected Bayesian Information Criterion, using the edge-linked parameter and the greedy strategy (Lanfear *et al*., 2016), with each rDNA and each codon position of the four PCGs as separate initial partition (Suppl. Table 3.2). The MCMC analysis was run with the tree and parameter values sampled every 5,000 steps over a total of 50 million generations. A burn-in of 10% steps was discarded, and adequate sample sizes (ESS>200) were checked using Tracer 1.7.7 (Rambaut *et al*., 2018).

Results

Molecular species delimitations

The *cox1*$_{aa}$ sequences obtained were 666 bp long, and correctly translate for 222 amino acids sequences without any stop codon. BLAST searches against NCBI Genbank database consistently resulted in congeneric *Candovia* species best-hits. The final dataset consisted in 67 newly generated sequences and 29 previously published sequences, bringing the total number of operational taxonomic units (OTUs) to 96. The obtained ML and BI trees showed a fully compatible topologies (Figure 2; Suppl. Figure 1).

Using the different species delimitation approaches, the number of hypothetical species ranged from 19 to 24 for PTP, which proved to be the most conservative approach, while GMYC was the method identifying the highest number of hypothetical species (24 spp.; Figure 2).

A full agreement between some of the known morphospecies and the hypothetical species identified by the different species delimitation approaches can be observed for *C. aberrata* (Brunner, 1907)*, C. granulosa* (Brunner, 1907)*, C. pallida* (Sjöstedt, 1918)*, C. robinsoni* Brock and Hasenpusch, 2007*, C. strumosa* (Redtenbacher, 1908). This holds also for previously recognized putative taxa *Candovia* spp. B and C. In other instances, morphospecies are split in different taxa or merged in a single entity. *C. peridromes* (Westwood, 1859) splits in all approaches in two groups, according to a geographical pattern, while this occurs for *C. annulata* in the GMYC analyses only. The specimen PB-0178 identified as *C. peridromes* is further found as a separate taxon by sPTP and the two GMYC approaches using the strict molecular clock. *C. coenosa* (Gray, 1833) is consistently recognized as composed by two putative species, separating the PB-0164 specimen

from the other two congeneric samples. On the other hand, the mPTP method merges all *C. coenosa* specimens with those of *C. spurcata* (Brunner, 1907) in a single hypothetical species. Finally, the specimen JH-0027 of *Candovia* sp. A is recognized as divergent by two GMYC analyses (Figure 2).

Regarding the newly collected specimens, in both ML and BI trees they have been partitioned into well-defined clusters that the species delimitation analysis identified as new putative taxonomic entities (henceforth named *Candovia* spp. D, E, F, G, H, I and L; Figure 2). Though, for *Candovia* sp. F, two specimens were recognized as a separate taxon by two GMYC analyses (Figure 2).

Phylogenetic Analysis

The final dataset used for phylogenetic analyses, obtained concatenating the eight analysed mitochondrial and nuclear markers, consisted in 4540 sites and 58 specimens representing 30 species. The ML and BI analyses provided identical tree topologies, with most nodes more strongly supported in the BI tree than in the ML one. Both phylogenetic analyses agreed in identifying the genus as monophyletic with respect to the other Necroscinae taxa included in the analysis (Figure 3; Suppl. Figure 2). Within the *Candovia* clade, the observed clustering pattern appeared more resolved, in term of phyletic relationships and nodal supports, than in the analysis of *cox1*$_{Fa}$ dataset alone.

The first split within the clade identified a cluster including *C. peridromes*, *C. pallida* and three putative species (*Candovia* spp. G, I and L) in sister relationship to all remaining *Candovia* species. Samples of *C. spurcata* and *C. coenosa* [the latter type species of the genus] are then clustered in a monophyletic clade, with the latter taxon being paraphyletic. Further well-defined clusters are given by: *C. annulata* + *Candovia* spp. B and E; *C. strumosa* + *Candovia* spp. C and F; *C. aberrata* + *Candovia* sp. D; *C. robinsoni* and *C. granulosa* + *Candovia* spp. A and H (Figure 3).

The pattern observed in the species delimitation analysis for *C. coenosa* appears here explained by the *C. coenosa* PB-164 sequence which is more related to the *C. spurcata* clade than to the conspecific specimens.

Discussion

The present analysis aimed to investigate the molecular phylogenetic diversity within the stick insect genus *Candovia* and to help in better defining the diversity observed in specimens not identified as previously known *Candovia* species.

A general consensus on which is the more robust method for molecular species delimitation is lacking, and different methods can outperform depending on the specific characteristics of a dataset (Luo *et al*., 2018). Therefore, the combination of different species delimitation approaches and phylogenetic reconstruction methods (Tang *et al*., 2014) gave us a stronger confidence while exploring the diversity of the *Candovia* genus. Although all methods, whether distance- (ABGD) or tree-based (PTP and GMYC), were able to detect most interspecific boundaries, it is clear that the use of a single approach could have been misleading in some instances.

Generally speaking, the different species delimitation analyses consistently recognize previously described species and yet undescribed taxa (*Candovia* spp. A, B and C; Velonà *et al.*, 2015). Moreover, analyses are generally concordant in identifying new putative species (*Candovia* spp. D - L). For some lineages, the GMYC method highlighted some degree of intra-specific divergence for the specimen JH-0027 of *Candovia* sp. A, specimens PB-0261 and PB-0271 of *Candovia* sp. F, and five specimens of *C. annulata*. This can be due to either general properties of our dataset, including unbalanced geographical range sampling, skewed species abundances and the availability for analyses of a single specimen (Talavera *et al.*, 2013; Ahrens *et al*., 2016), or because of paraphyly/polyphyly of some clades (Hendrich *et al*., 2010; Scicchitano *et al*., 2018). Interestingly, the intraspecific differentiation observed in *C. annulata* corresponds to samplings at different geographic areas (Montville and Apple Tree Park - Springbrook). This could also be the case for *C. peridromes*, where all methods consistently recognize two groups of specimens that have been, in fact, collected in two distant locations (Wilmington and Mount Gambier). Therefore, the geographical ground of the observed pattern suggests the presence of structured populations that may represent different subspecific entities. A peculiar situation is given by the specimens PB-164 of the species *C. coenosa*, recognized as a different entity by all species delimitation methods but the mPTP one. Although this could be due to the paraphyly of *C. coenosa*, which is also confirmed by the phylogenetic analyses with high nodal supports (see below), it is possible that the specimen PB-164 may represent a more differentiated taxon. Despite a substantial proportion of species level-diversity is represented by species described on a single specimen (Lim *et al*., 2012), in our opinion, only a wider sampling of specimens from different geographical locations will allow to properly describe the aforementioned lineage.

Overall, data observed here do not seem to suggest significant deviations from the known taxonomy, especially considering that inferences drawn from species delimitation approaches should rely on a

conservative consensus (Carstens *et al*., 2013). However, the possibility of intraspecific differentiation on a geographical ground should be taken into account in a few instances. Moreover, regarding taxonomically unassigned specimens, our analysis suggests the presence of new putative species (*Candovia* spp. D - L). Further sampling and formal descriptions of these taxa are in progress, also including egg morphology which in some lineages appears to be a much more reliable character than the traditional body morphology.

The phylogenetic analysis suggests *Candovia* is as a monophyletic genus, although the nodal support in the Maximum Likelihood tree is not 100%. In this regard, however, it is to be noted that no other Australian Necroscinae species were available for comparison (although the *Sipyloidea* genus includes some Australian species, these are not among the two here considered). Only further analyses will help in increasing the confidence in the present results. The phyletic relationships within the *Candovia* clade appear quite well-defined, even if a general decrease of nodal support at the deepest nodes emerges. At variance of some intra-specific divergences scored in the species delimitation analysis, no clear geographical pattern emerges from the phylogenetic analysis: in fact, most of the taxa are distributed along eastern Australian coasts, with overlapping ranges. Only *C. pallida* and *C. peridromes* are known to have a different geographic distribution, the former within the central-western Australia and the latter in the south-east of the continent (Brock and Hasenpusch, 2009). Accordingly, they fall in the same cluster which is in sister relationship with all remaining species. The same cluster also include *Candovia* sp. L, collected from central Australia, and Candovia spp. I and G which, on the other hand, have been collected along the east coast.

The phylogenetic analysis presented here provides a strong framework for further morphological investigations in order to better describe the taxonomy and systematics of the genus. For example, an important part of phasmid taxonomy relies on egg morphological character, like the capitulum (Clark, 1976). This structure is a lipid-rich extension of the operculum whose function was not understood until recently, where it was shown to serve as a reward to promote ant-mediated dispersal, or myrmecochory (Compton and Ware, 1991). This adaptation represents one of the most extraordinary examples of convergent evolution across different kingdoms of life, where the capitulum in phasmid eggs are analogous to elaiosomes of angiosperm seeds (Stanton *et al*., 2015). Both structures are used to exploit ant behaviour: this mutualistic relationship allows for both plant seeds and phasmid eggs to be buried in the soil, protected from environmental changes, predators and parasites, also greatly enhancing the dispersal abilities of the species (Hughes and Westoby, 1992). A study on an African phasmid species also suggests the possibility to escape

recurrent bushfire when ants drag the egg into their nest (Compton and Ware, 1991). Some *Candovia* species, such as *C. aberrata, C. spurcata and C. coenosa*, lay eggs with capitulum, while others do not (Brock and Hasenpusch, 2009; Brock, personal communication: see Suppl. Figure S3). Mapping this character into our phylogenetic tree (Figure 3) clearly suggests the non-monophyly of the capitulum, possibly pointing out its multiple origin. At the order level, it has been shown that capitulum likely evolved independently in several lineages across the phasmid tree of life (Robertson *et al*., 2018 and references therein), but to our knowledge this is the first documented case among congeneric species. It is interesting to note, although belonging to different phylogenetic clusters, the three aforementioned *Candovia* species all contain capitulum's and have strictly overlapping distribution in the southern part of the Australian east coast. At present, however, we can only speculate if the presence of capitulum can be linked to local adaptations or to the presence of particular ant species performing myrmecochory.

Altogether, analyses presented here have helped us better understanding the intra-generic diversity of the *Candovia* stick insect genus, suggested new putative taxa, and clarified instances of intraspecific diversity. This data might also provide a reliable framework to further refine taxonomy and systematics, reconsider some diagnostic character for species description and identification, and to better understand the drivers behind the evolution and diversification of this overlooked - yet remarkable - clade.

Acknowledgements

Authors are grateful to all people that kindly helped in obtaining material, sometimes with PDB and/or BRJ. In particular Authors want to thank Noelene Tweed, Steve Cross, Kirsten Dalgleish, Alexander Karas, John Koens and Andreas Urban. Geoff Monteith and Helen Brock assisted in the Springbrook area and Geoff organised permits. Ross Coupland and Beth Ripper supplied photographs of species in parts of Queensland. Oskar Conle is also thanked for his assistance. Export and collecting permits were otherwise obtained thanks to the Authorities (ABRS). Jack Hasenpusch helped in rearing some stock sent by collectors. This work has been supported by Canziani funding to BM.

References

Ahrens, D., Fujisawa, T., Krammer, H.-J., Eberle, J., Fabrizi, S., and Vogler, A. P. (2016). Rarity and Incomplete Sampling in DNA-Based Species Delimitation. *Systematic Biology* **65**, 478–494. https://doi.org/10.1093/sysbio/syw002

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bouckaert, R, Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M. Gavryushkina A, ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **15**(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Brock, P. D., and Hasenpush, J. W. (2009). 'The complete field guide to stick and leaf insects of Australia.' CSIRO PUBLISHING.

Brock, P. D., Büscher, T. and Baker, E. *Phasmida Species File Online*. Version 5.0/5.0. [2020]. http://phasmida.speciesfile.org/HomePage/Phasmida/HomePage.aspx

Campbell, K. G. (1960). Preliminary studies in population estimation of two species of stick insects (Phasmatidae: Phasmatodea) occurring in plague numbers in highland forest areas of south-eastern Australia. *Proceedings of the Linnean Society of New South Wales* **85**, 121-141.

Carstens, B. C., Pelletier, T. A., Reid, N. M., and Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology* **22**, 4369–4383. https://doi.org/10.1111/mec.12413

Clark, J. T. (1976). The capitulum of phasmid eggs (Insecta: Phasmida). *Zoological Journal of the Linnean Society* **59**, 365–375.

Compton, S. G., and Ware, A. B. (1991). Ants disperse the elaiosome-bearing eggs of an African stick insect. *Psyche* **98**, 207-213.

Dellicour, S., and Flot, J. F. (2018). The hitchhiker's guide to single-locus species delimitation. *Molecular Ecology Resources* **18**(6), 1234-1246. https://doi.org/10.1111/1755-0998.12908

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969–1973. https://doi.org/10.1093/molbev/mss075

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**(5), 294–299.

Gutiérrez-Valencia, J., Gutiérrez, Y., and Gomes Dias, L. (2017). Species delimitation in the crypsis-defended and polymorphic stick insects of the genus *Libethra* (Phasmatodea, Diapheromeridae). *Zoologica Scripta* **46**, 693-705. https://doi.org/10.1111/zsc.12242

Hebert, P. D. N., Cywinska, A., Balland, S. L., and deWaard J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the RoyalSociety London B* **270**, 313-321. https://doi.org/10.1098/rspb.2002.2218

Hendrich, L., Pons, J., Ribera, I., and Balke, M. (2010). Mitochondrial cox1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. *PLoS ONE* **5**, e14448.

https://doi.org/10.1371/journal.pone.0014448

Hughes, L., and Westoby M. (1992). Capitula on stick insect eggs and elaiosomes on seeds: convergent adaptations for burial by ants. *Functional Ecology* **6**, 642-648.

Jacobs, S. J., Kristofferson, C., Uribe-Convers, S., Latvis, M., and Tank ,D. C. (2018). Incongruence in molecular species delimitation schemes: What to do when adding more data is difficult. *Molecular Ecology* **27**(10), 2397-2413. https://doi.org/:10.1111/mec.14590

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–589. https://doi.org/10.1038/nmeth.4285

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., and Flouri, T. (2017). Multi-rate Poisson tree processes for single- locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* **33**,1630–1638. https://doi.org/10.1093/bioinformatics/btx025

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780. https://doi.org/10.1093/molbev/mst010

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**, 772-773. https://doi.org/10.1093/molbev/msw260

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lim, G. S., Balke, M., and Meier, R. (2012). Determining Species Boundaries in a World Full of Rarity: Singletons, Species Delimitation Methods. *Systematic Biology* **61**, 165–169. https://doi.org/10.1093/sysbio/syr030

Luo, A., Ling, C., Ho, S. Y. W, and Zhu, C. D. (2018). Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology* **67**(5), 830-846. https://doi.org/10.1093/sysbio/syy011

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2012**).** The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources**.** Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the Campus and Beyond. ACM, Chicago, Illinois, pp. 1-8.

Merilaita, S., and Lind, J. (2005). Background-matching and disruptive coloration, and the evolution of cryptic coloration. *Proceedings of the Royal Society London B* **272**(1563), 665–670. https://doi.org/10.1098/rspb.2004.3000

Moritz, C., Cicero, C., (2004). DNA barcoding: promise and pitfalls. *PLoS Biology* **2**, e354. https://doi.org/10.1371/journal.pbio.0020354

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q., (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–274. https://doi.org/10.1093/molbev/msu300

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., and Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**, 595–609. https://doi.org/10.1080/10635150600852011

Puillandre, N., Lambert, A., Brouillet, S., and Achaz, G. (2012). ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology* **21**, 1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M.A. (2018). posterior summarization in bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**, 901–904. https://doi.org/10.1093/sysbio/syy032

Robertson, J. A., Bradler, S., and Whiting, M. F. (2018). Evolution of oviposition techniques in stick and leaf insects (Phasmatodea). *Frontiers in Ecology and Evolution* **6**, 216. https://doi.org/10.3389/fevo.2018.00216

Skelhorn, J. (2015). Masquerade. *Current Biology* **25**, R643-R644 https://doi.org/10.1016/j.cub.2015.02.069

Smith, S. A., and Dunn, C. W. (2008). Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24**, 715–716. https://doi.org/10.1093/bioinformatics/btm619

Stanton, A. O., Dias, D. A., and O'Hanlon, J. C. (2015). Egg dispersal in the Phasmatodea: convergence in chemical signaling strategies between plants and animals? *Journal of Chemical Ecology* **41**, 689–695. https://doi.org/10.1007/s10886-015-0604-8

Stucky, B. J. (2012). SeqTrace: A graphical tool for rapidly processing DNA sequencing chromatograms. *Journal of Biomolecular Techniques* **23**, 90–93. https://dx.doi.org/10.7171%2Fjbt.12-2303-004

Svenson, G. J., Medellin, C., and Sarmiento, C. E. (2016). Re-evolution of a morphological precursor of crypsis investment in the newly revised horned praying mantises (Insecta, Mantodea, Vatinae). *Systematic Entomology* **41**, 229–255. https://doi.org/10.1111/syen.12151

Syme, A. E., and Oakley, T. H., (2012). Dispersal between shallow and abyssal seas and evolutionary loss and regain of compound eyes in cylindroleberidid ostracods: conflicting conclusions from different comparative methods. *Systematic Biology* **61**, 314–23. https://doi.org/10.1093/sysbio/syr085

Scicchitano, V., Dedeine, F., Mantovani, B., and Luchetti, A. (2018). Molecular systematics, biogeography and colony fusion in the European dry-wood termites *Kalotermes* spp. (Blattodea, Termitoidae, Kalotermitidae). *Bulletin of Entomological Research* **108**, 523-531. https://doi.org/10.1017/S0007485317001080

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**, 564–577. https://doi.org/10.1080/10635150701472164

Talavera, G., Dincă, V., and Vila, R. (2013). Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution* **4**, 1101–1110. https://doi.org/10.1111/2041-210X.12107

Tang, C. Q., Humphreys, A. M., Fontaneto, D., and Barraclough, T. G. (2014). Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution* **5**, 1086–1094. https://doi.org/10.1111/2041-210X.12246

Trueman, J. W. H., Pfeil, B. E., Kelchner, S. A., and Yeates, D. K. (2004). Did stick insects really regain their wings? *Systematic Entomology* **29**, 138–139. https://doi.org/10.1111/j.0307-6970.2004.00251.x

Velonà, A., Brock, P., Hasenpusch, J., and Mantovani, B. (2015). Cryptic diversity in Australian stick insects (Insecta; Phasmida) uncovered by the DNA barcoding approach. *Zootaxa* **3957**(4): 455–466. https://doi.org/10.11646/zootaxa.3957.4.6

Whiting, M. F., Bradler, S., and Maxwell, T. (2003). Loss and recovery of wings in stick insects. *Nature* **421**, 264-267. https://doi.org/10.1038/nature01313

Zhang, J., Kapli, P., Pavlidis, P., and Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869–2876. https://doi.org/10.1093/bioinformatics/btt499

Map ID	Sampling site	Morphospecies	*N*	Latitude	Longitude	Sample ID
1	Fountain Springs Rest Area (QLD)	*Candovia* sp. L	1	-20.800	139.996	BJ-(95)
2	Eungella N.P., Broken River (QLD)	*Candovia* sp. B	4	-21.162	148.512	PB-0(18 -21)
3	Byfield (QLD)	*Candovia* sp. F	3	-22.847	150.650	PB-(325, 327, 328)
4	Rossmoya Road, The Caves (QLD)	*Candovia* sp. F	4	-23.159	150.457	PB-(258,261,271,272)
5	Blackdown Tableland (QLD)	*Candovia* sp. D	4	-23.582	149.063	PB-(188-191)
5	Blackdown Tableland (QLD)	*Candovia* sp. E	4	-23.582	149.063	PB-(184-187)
6	Charlevue Road (QLD)	*Candovia* sp. I	4	-23.638	149.110	BJ-(91-94)
7	Ellery Creek Big Hole, West Macdonnell NP (NT)	*C. pallida*	4	-23.778	133.073	BJ-(30,31,32,34)
8	Baralaba, Duaninga Road (QLD)	*Candovia* sp. G	1	-24.136	149.830	PB-295
9	Narrows Road, Montville (QLD)	*C. annulata*	10	-26.698	152.866	JH-0(37-42); PB-0(66,67,77,98)
9	Narrows Road Montville (QLD)	*Candovia* sp. A	9	-26.698	152.866	JH-0(26-31,43,44); PB-060
9	Narrows Road, Montville (QLD)	*Candovia* sp. C	5	-26.698	152.866	JH-0(66-70)
10	D`Aguilar Range, Boombana (QLD)	*Candovia* sp. A	1	-27.404	152.794	PB-058
11	Springbrook, Apple Tree Park (QLD)	*C. annulata*	5	-28.166	153.259	PB-(207-211)
11	Queensland, Springbrook, Apple Tree Park (QLD)	*C. strumosa*	5	-28.166	153.259	PB-(214-218)
12	Springbrook, Repeater Station Rd (QLD)	*C. granulosa*	4	-28.234	153.267	PB-(220-223)
13	Mt Warning, Wollumbin (NSW)	*Candovia* sp. H	2	-28.400	153.282	PB-(173, 174)
14	Evans Head (NSW)	*C. aberrata*	5	-29.108	153.431	PB-(157-161)
15	Pilliga (NSW)	*C. robinsoni*	1	-30.350	148.890	PB-162
16	Wilmington, Park road (SA)	*C. peridromes*	5	-32.840	138.036	BJ-(12-16)
17	Colo Heights (NSW)	*C. coenosa*	1	-33.369	150.722	PB-164
18	Little bay, Harvey Street (NSW)	*C. coenosa*	2	-33.983	151.244	BJ-(41-42)
19	Bulli (NSW)	*C. robinsoni*	7	-34.330	150.901	PB-(107, 108, 112, 115, 118, 119, 121)
20	Kosciuszko National Park, Guthega Road (NSW)	*C. spurcata*	2	-36.500	148.266	PB-(155, 156)
21	Mount Gambier (SA)	*C. peridromes*	3	-37.843	140.765	PB-(178-180)

*from Velonà et al., 2015

Table 1 - Sampling information about analyzed *Candovia* samples. Map IDs refer to Figure 1.

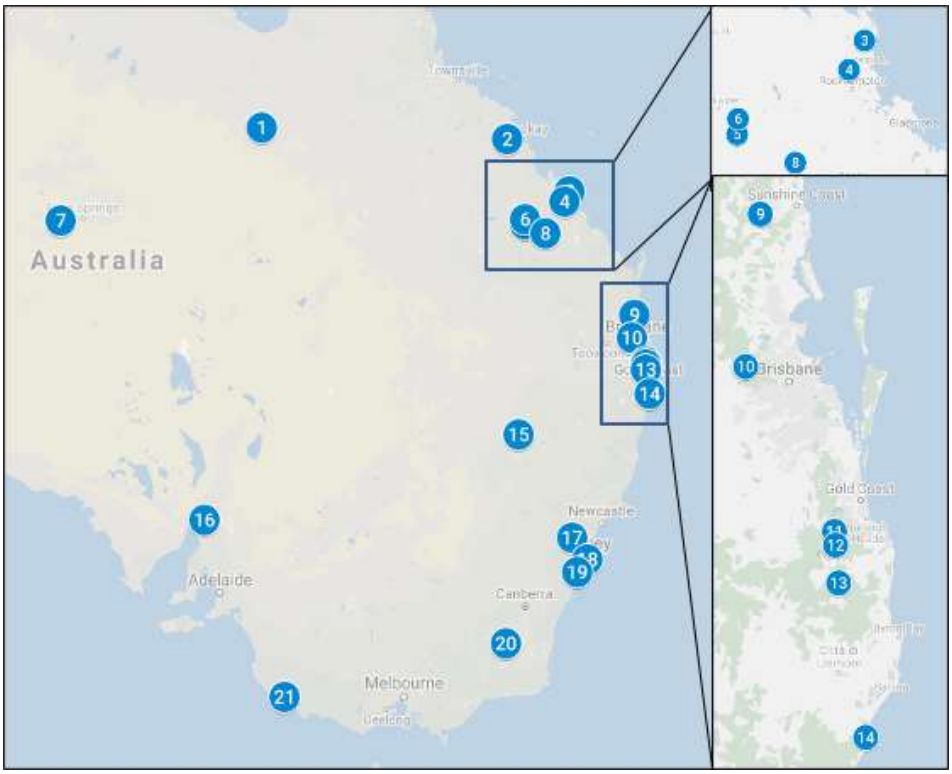


Fig. 1. Geographical distribution of the sampled populations from Australia (1-Fountain Springs Rest Area (QLD)*;* 2-Eungella N.P., Broken River (QLD)*;* 3-Byfield (QLD)*;* 4-Rossmoya Road, The Caves (QLD)*;* 5-Blackdown Tableland (QLD)*;* 5-Blackdown Tableland (QLD)*;* 6-Charlevue Road (QLD)*;* 7-Ellery Creek Big Hole, West Macdonnell NP (NT)*;* 8-Baralaba, Duaninga Road (QLD)*;* 9-Narrows Road, Montville (QLD)*;* 9-Narrows Road Montville (QLD)*;* 9-Narrows Road, Montville (QLD)*;* 10-D`Aguilar Range, Boombana (QLD)*;* 11-Springbrook, Apple Tree Park (QLD)*;* 11-Queensland, Springbrook, Apple Tree Park (QLD)*;* 12-Springbrook, Repeater Station Rd (QLD)*;* 13-Mt Warning, Wollumbin (NSW) *;* 14-Evans Head (NSW)*;* 15-Pilliga (NSW)*;* 16-Wilmington, Park road (SA)*;* 17-Colo Heights (NSW)*;* 18-Little bay, Harvey Street (NSW)*;* 19-Bulli (NSW)*;* 20-Kosciuszko National Park, Guthega Road (NSW)*;* 21-Mount Gambier (SA).

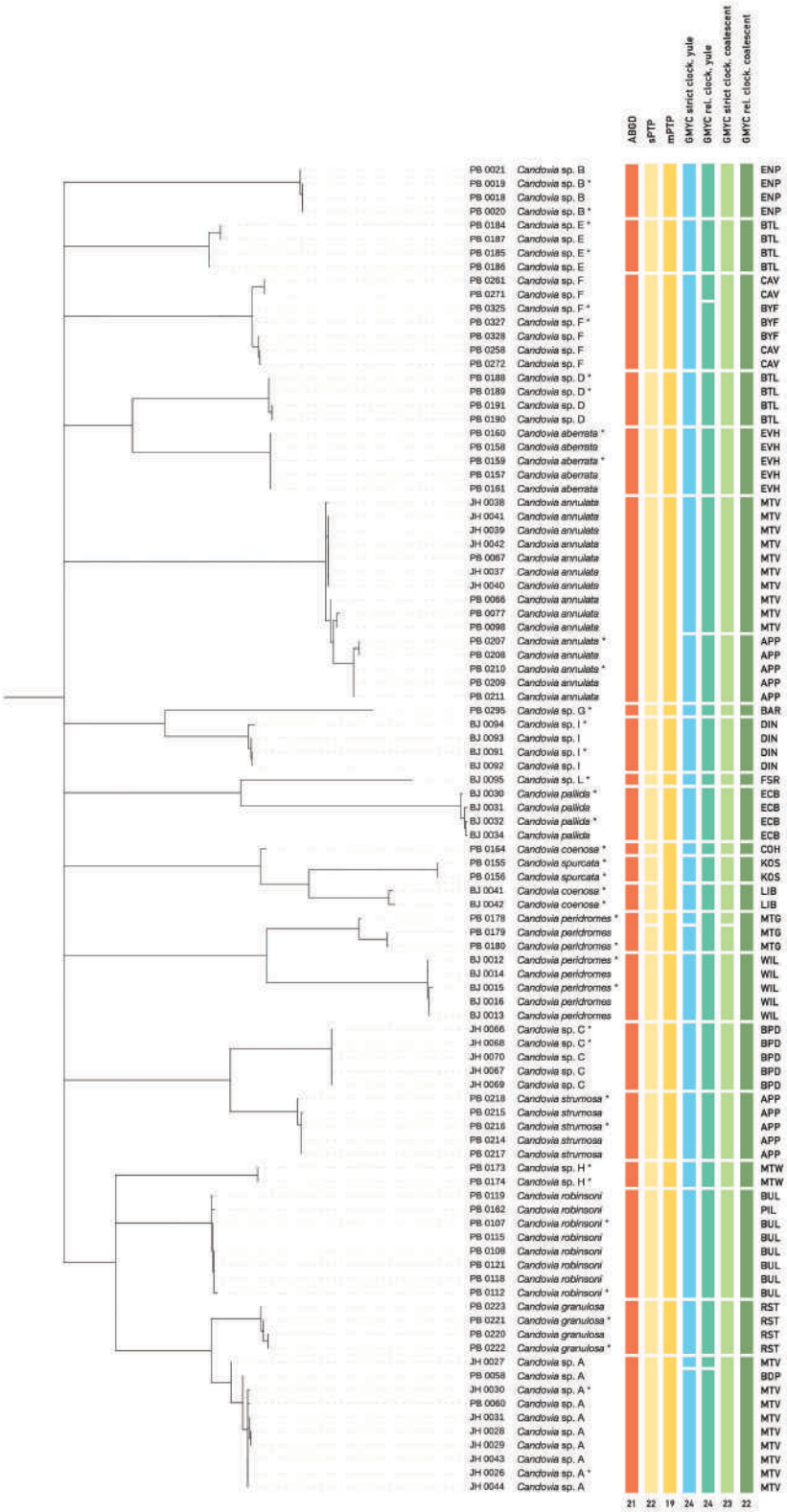

ABGD
sPTP
mPTP
GMYC strict clock, yule
GMYC rel. clock, yule
GMYC strict clock, coalescent
GMYC rel. clock, coalescent
PB 0021 Candowia sp. B ENP
PB 0019 Candowia sp. B * ENP
PB 0018 Candowia sp. B ENP
PB 0020 Candowia sp. B * ENP
PB 0184 Candowia sp. E * BTL
PB 0187 Candowia sp. E BTL
PB 0185 Candowia sp. E * BTL
PB 0186 Candowia sp. E BTL
PB 0261 Candowia sp. F CAV
PB 0271 Candowia sp. F CAV
PB 0325 Candowia sp. F * BYF
PB 0327 Candowia sp. F * BYF
PB 0328 Candowia sp. F BYF
PB 0258 Candowia sp. F CAV
PB 0272 Candowia sp. F CAV
PB 0188 Candowia sp. D * BTL
PB 0189 Candowia sp. D * BTL
PB 0191 Candowia sp. D BTL
PB 0190 Candowia sp. D BTL
PB 0160 Candowia aberrata * EVH
PB 0158 Candowia aberrata EVH
PB 0159 Candowia aberrata * EVH
PB 0157 Candowia aberrata EVH
PB 0161 Candowia aberrata EVH
JH 0038 Candowia annulata MTV
JH 0041 Candowia annulata MTV
JH 0039 Candowia annulata MTV
JH 0042 Candowia annulata MTV
PB 0067 Candowia annulata MTV
JH 0037 Candowia annulata MTV
JH 0040 Candowia annulata MTV
PB 0066 Candowia annulata MTV
PB 0077 Candowia annulata MTV
PB 0098 Candowia annulata MTV
PB 0207 Candowia annulata * APP
PB 0208 Candowia annulata APP
PB 0210 Candowia annulata * APP
PB 0209 Candowia annulata APP
PB 0211 Candowia annulata APP
PB 0295 Candowia sp. G * BAR
BJ 0094 Candowia sp. I * DIN
BJ 0093 Candowia sp. I DIN
BJ 0091 Candowia sp. I * DIN
BJ 0092 Candowia sp. I DIN
BJ 0095 Candowia sp. L * FSR
BJ 0030 Candowia pallida * ECB
BJ 0031 Candowia pallida ECB
BJ 0032 Candowia pallida * ECB
BJ 0034 Candowia pallida ECB
PB 0164 Candowia coenose * COH
PB 0155 Candowia spurcata * KOS
PB 0156 Candowia spurcata * KOS
BJ 0041 Candowia coenose * LIB
BJ 0042 Candowia coenose * LIB
PB 0178 Candowia peridromes * MTG
PB 0179 Candowia peridromes MTG
PB 0180 Candowia peridromes * MTG
BJ 0012 Candowia peridromes * WIL
BJ 0014 Candowia peridromes WIL
BJ 0015 Candowia peridromes * WIL
BJ 0016 Candowia peridromes WIL
BJ 0013 Candowia peridromes WIL
JH 0066 Candowia sp. C * BPD
JH 0068 Candowia sp. C * BPD
JH 0070 Candowia sp. C BPD
JH 0067 Candowia sp. C BPD
JH 0069 Candowia sp. C BPD
PB 0218 Candowia strumosa * APP
PB 0215 Candowia strumosa APP
PB 0216 Candowia strumosa * APP
PB 0214 Candowia strumosa APP
PB 0217 Candowia strumosa APP
PB 0173 Candowia sp. H * MTW
PB 0174 Candowia sp. H * MTW
PB 0119 Candowia robinsoni BUL
PB 0162 Candowia robinsoni PIL
PB 0107 Candowia robinsoni * BUL
PB 0115 Candowia robinsoni BUL
PB 0108 Candowia robinsoni BUL
PB 0121 Candowia robinsoni BUL
PB 0118 Candowia robinsoni BUL
PB 0112 Candowia robinsoni * BUL
PB 0223 Candowia granulosa RST
PB 0221 Candowia granulosa * RST
PB 0220 Candowia granulosa RST
PB 0222 Candowia granulosa * RST
JH 0027 Candowia sp. A MTV
PB 0058 Candowia sp. A BDP
JH 0030 Candowia sp. A * MTV
PB 0060 Candowia sp. A MTV
JH 0031 Candowia sp. A MTV
JH 0028 Candowia sp. A MTV
JH 0029 Candowia sp. A MTV
JH 0043 Candowia sp. A MTV
JH 0026 Candowia sp. A * MTV
JH 0044 Candowia sp. A MTV
21 22 19 24 24 23 22
MTV Queensland, Narrows Road, Montville
KOS NSW, Guthega Road, Kosciuszko National Park
BUL New South Wales, Bulli
EVH New South Wales, Evans Head
LIB New South Wales, Little bay, Harvey street
COH New South West, Colo Heights
MTW New South West, Mt Warning (Wollumbin)
PIL New South West, Pilliga
ECB Northern territory, Ellery Creek Big Hole, West Macdonnel NP
GST Queensalnd, Gladstone
BAR Queensland, Barralaba, Duaringa Road
BTL Queensland, Blackdown Tableland
BYF Queensland, Byfield
DIN Queensland, Charlevevu Road, 18km West of Dingo
DAR Queensland, D'Aguilar Range, Boombana
ENP Queensland, Eungella N.P, Broken River
FSR Queensland, Fountain Springs Rest Area, 60km East of Mount Isa
BPD Queensland, Narrow Road, Nr Barcoo Pocket Dam, Montville
CAV Queensland, Rosemoya Road, The Caves
APP Queensland, Springbrook, Apple Tree Park, 580m.
RST Queensland, Springbrook, Repeater Station Road
MTG South Australia, Mount Gambier
WIL South Australia, Wilmington, Park road

Fig. 2. Species delimitations summarised on the Maximum Likelihood tree with a bootstrap cut off of 80%. Coloured bars represent hypothetical species identified by the different approaches. Acronyms on the right are relative to geographic collection points, as reported at the bottom of the figure.

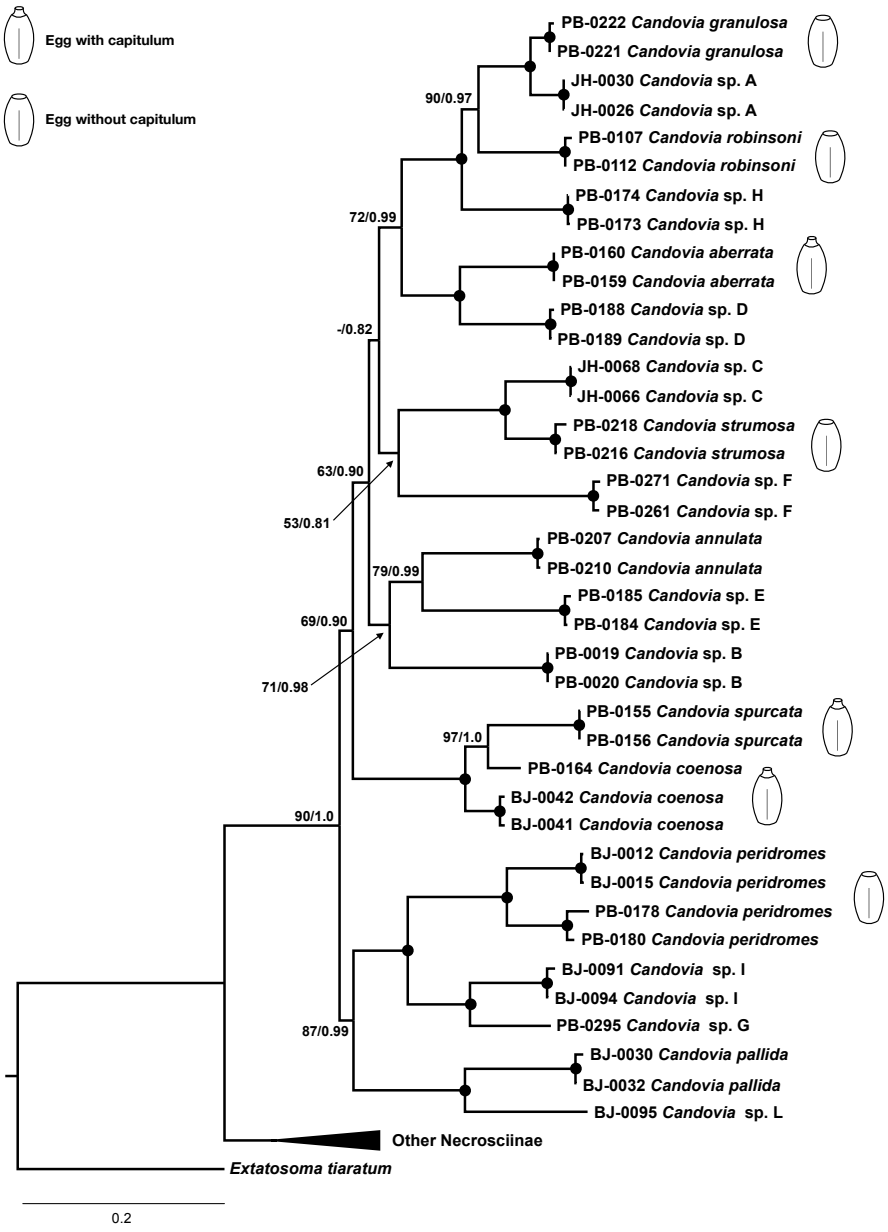
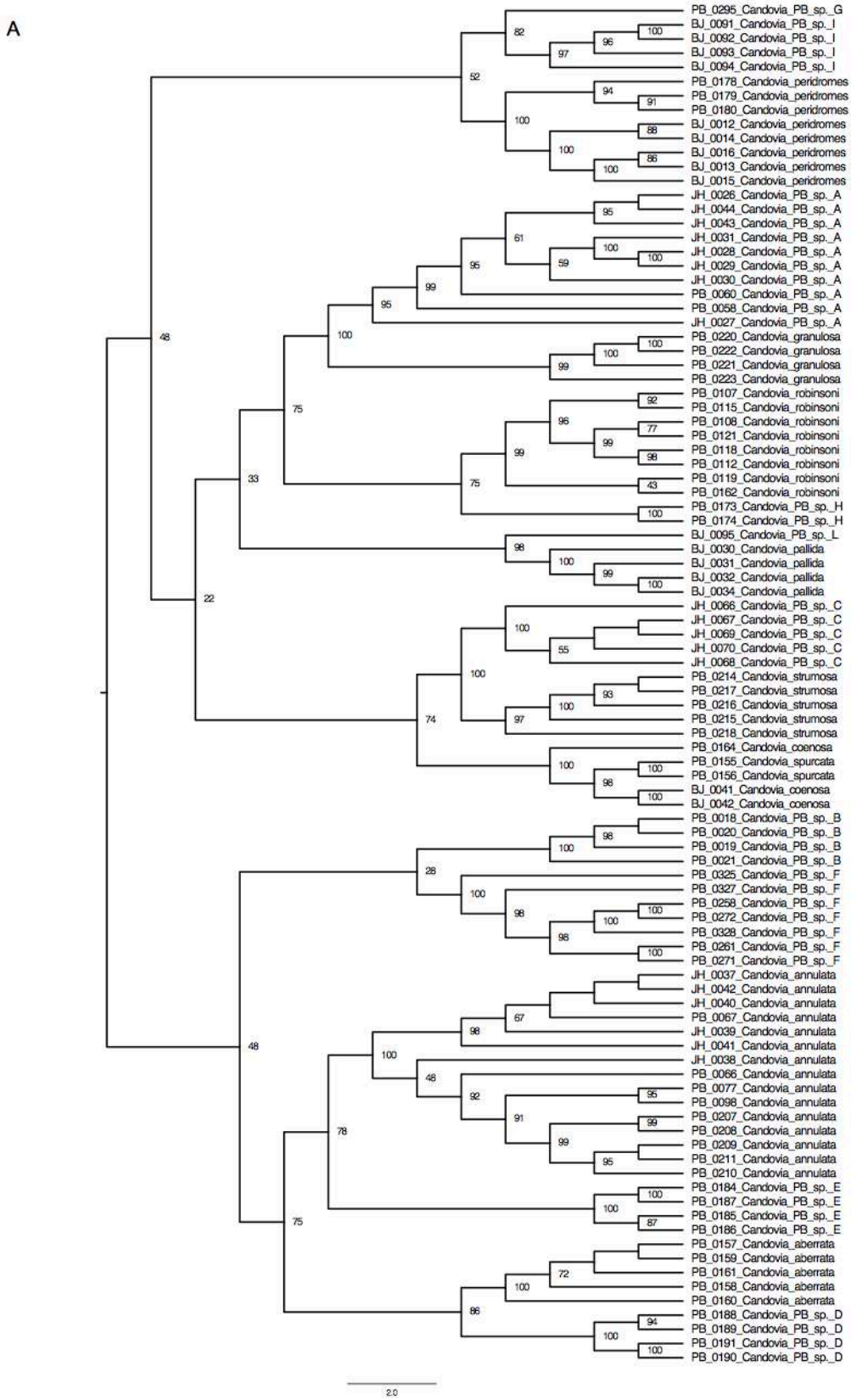
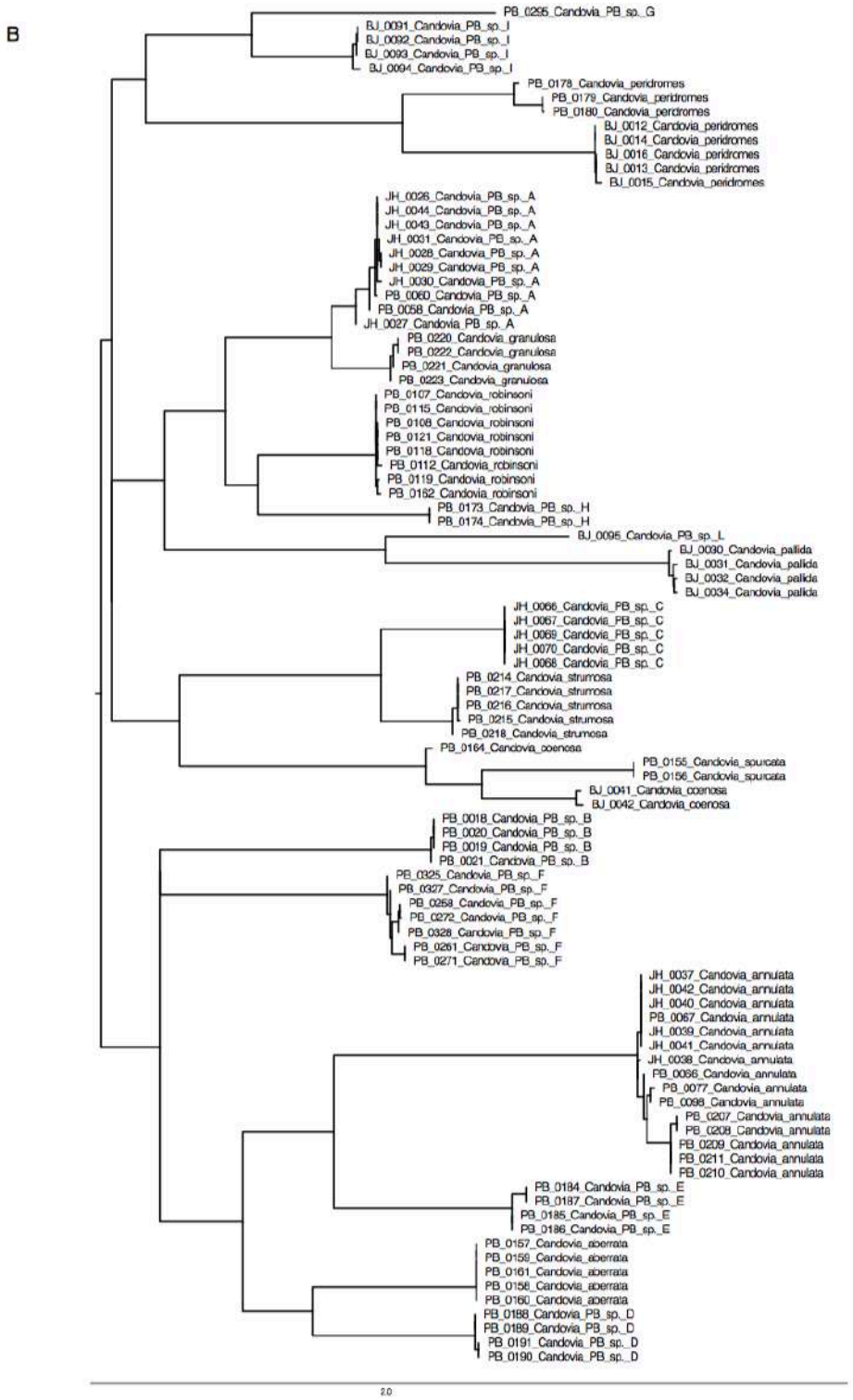
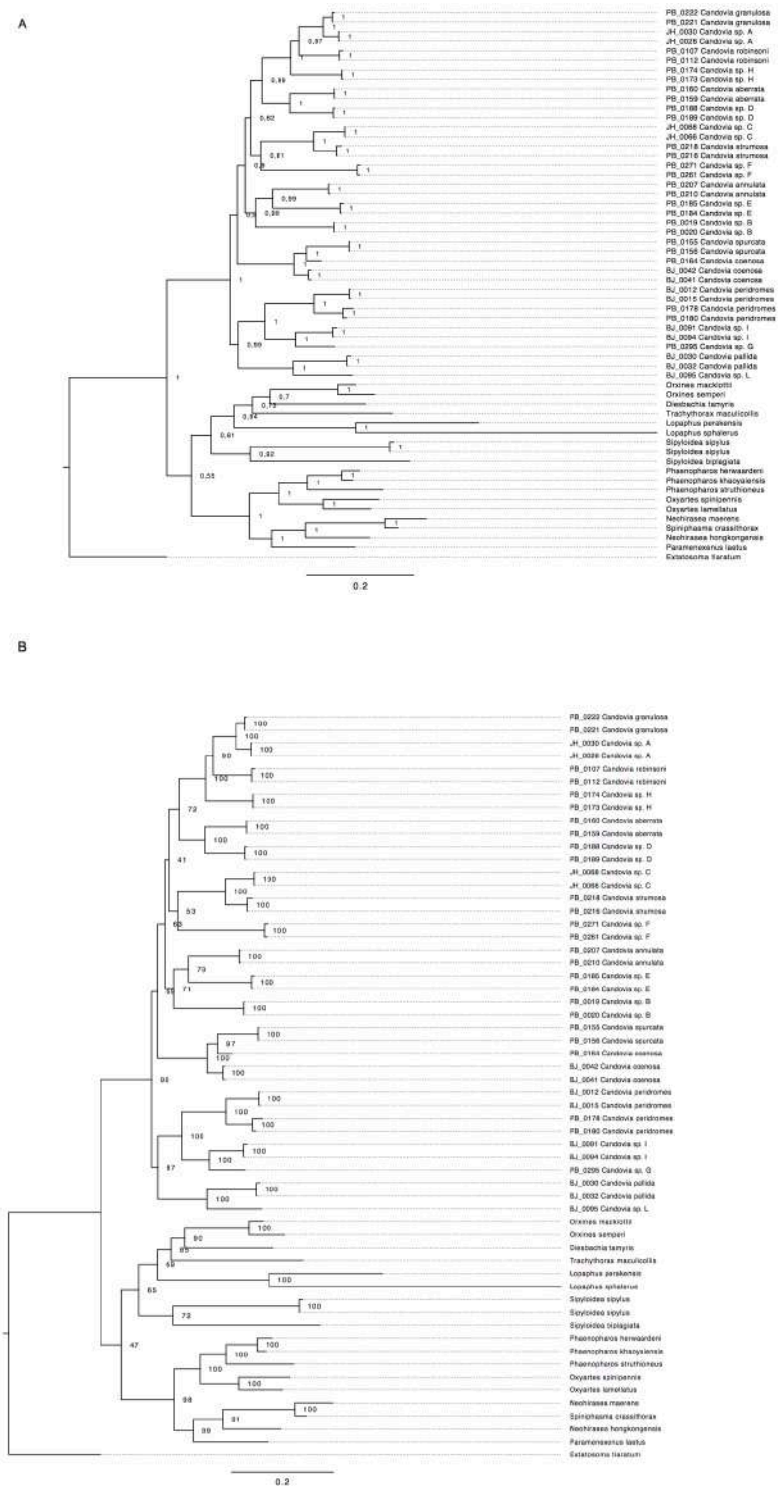


Fig. 3. Combined Maximum Likelihood and Bayesian Inference trees. Maximum nodal support (bootstrap=100%; posterior probability=1.0) are indicated with a black dot on the node; when support values are lower than the maximum, actual values are shown (dashes indicate no support at all). Eggs with or without capitulum, when known, are drawn on the right of the corresponding species.

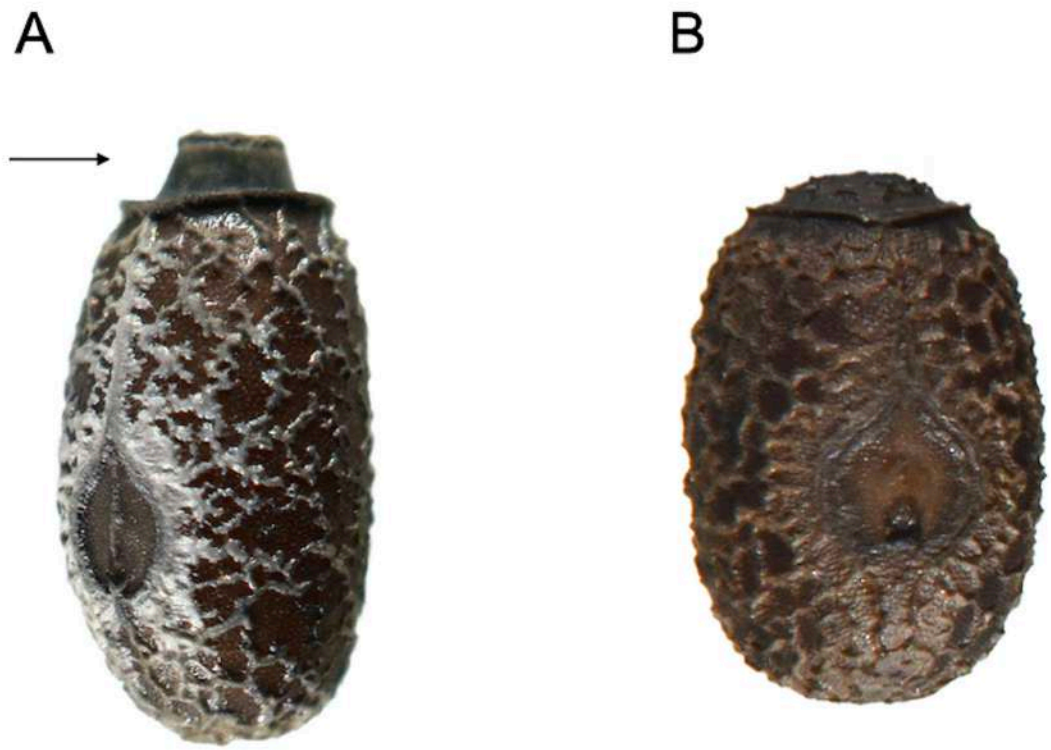

A
PB_0295_Candovia_PB_sp._G
BJ_0091_Candovia_PB_sp._I
BJ_0092_Candovia_PB_sp._I
BJ_0093_Candovia_PB_sp._I
BJ_0094_Candovia_PB_sp._I
PB_0178_Candovia_peridromes
PB_0179_Candovia_peridromes
PB_0180_Candovia_peridromes
BJ_0012_Candovia_peridromes
BJ_0014_Candovia_peridromes
BJ_0016_Candovia_peridromes
BJ_0013_Candovia_peridromes
BJ_0015_Candovia_peridromes
JH_0026_Candovia_PB_sp._A
JH_0044_Candovia_PB_sp._A
JH_0043_Candovia_PB_sp._A
JH_0031_Candovia_PB_sp._A
JH_0028_Candovia_PB_sp._A
JH_0029_Candovia_PB_sp._A
JH_0030_Candovia_PB_sp._A
PB_0060_Candovia_PB_sp._A
PB_0058_Candovia_PB_sp._A
JH_0027_Candovia_PB_sp._A
PB_0220_Candovia_granulosa
PB_0222_Candovia_granulosa
PB_0221_Candovia_granulosa
PB_0223_Candovia_granulosa
PB_0107_Candovia_robinsoni
PB_0115_Candovia_robinsoni
PB_0108_Candovia_robinsoni
PB_0121_Candovia_robinsoni
PB_0118_Candovia_robinsoni
PB_0112_Candovia_robinsoni
PB_0119_Candovia_robinsoni
PB_0162_Candovia_robinsoni
PB_0173_Candovia_PB_sp._H
PB_0174_Candovia_PB_sp._H
BJ_0095_Candovia_PB_sp._L
BJ_0030_Candovia_pallida
BJ_0031_Candovia_pallida
BJ_0032_Candovia_pallida
BJ_0034_Candovia_pallida
JH_0066_Candovia_PB_sp._C
JH_0067_Candovia_PB_sp._C
JH_0069_Candovia_PB_sp._C
JH_0070_Candovia_PB_sp._C
JH_0068_Candovia_PB_sp._C
PB_0214_Candovia_strumosa
PB_0217_Candovia_strumosa
PB_0216_Candovia_strumosa
PB_0215_Candovia_strumosa
PB_0218_Candovia_strumosa
PB_0164_Candovia_coenosa
PB_0155_Candovia_spurcata
PB_0156_Candovia_spurcata
BJ_0041_Candovia_coenosa
BJ_0042_Candovia_coenosa
PB_0018_Candovia_PB_sp._B
PB_0020_Candovia_PB_sp._B
PB_0019_Candovia_PB_sp._B
PB_0021_Candovia_PB_sp._B
PB_0325_Candovia_PB_sp._F
PB_0327_Candovia_PB_sp._F
PB_0258_Candovia_PB_sp._F
PB_0272_Candovia_PB_sp._F
PB_0328_Candovia_PB_sp._F
PB_0261_Candovia_PB_sp._F
PB_0271_Candovia_PB_sp._F
JH_0037_Candovia_annulata
JH_0042_Candovia_annulata
JH_0040_Candovia_annulata
PB_0067_Candovia_annulata
JH_0039_Candovia_annulata
JH_0041_Candovia_annulata
JH_0038_Candovia_annulata
PB_0066_Candovia_annulata
PB_0077_Candovia_annulata
PB_0098_Candovia_annulata
PB_0207_Candovia_annulata
PB_0208_Candovia_annulata
PB_0209_Candovia_annulata
PB_0211_Candovia_annulata
PB_0210_Candovia_annulata
PB_0184_Candovia_PB_sp._E
PB_0187_Candovia_PB_sp._E
PB_0185_Candovia_PB_sp._E
PB_0186_Candovia_PB_sp._E
PB_0157_Candovia_aberrata
PB_0159_Candovia_aberrata
PB_0161_Candovia_aberrata
PB_0158_Candovia_aberrata
PB_0160_Candovia_aberrata
PB_0188_Candovia_PB_sp._D
PB_0189_Candovia_PB_sp._D
PB_0191_Candovia_PB_sp._D
PB_0190_Candovia_PB_sp._D
2.0



Supplementary Fig. S1. Maximum Likelihood tree used for the species delimitation, presented as a cladogram with bootstrap proportion (A) and with branch lengths (B).



Supplementary Fig. S2. Trees obtained using (A) Maximum Likelihood inference and (B) Bayesian Inference on the phylogenetic dataset.



Supplementary Fig. S3. Exemplificative picture of egg with capitulum (indicated by the arrow) in *Candovia aberrata* (A) and without capitulum in *C. annulata* (B).

	12S	16S	18S	28S	*cox1*	*cox2*	H3
Lonchodidae							
Necrosciinae							
Diesbachia_tamyris	KJ024494.1			KJ024389.1		KJ024558.1	KJ024526.1
Lopaphus_perakensis	KJ024485.1	KJ024439.1	AY121171.1	AY125311.1			AY125254.1
Lopaphus_sphalerus			AY121182.1	AY125322.1			AY125265.1
Neohirasea_hongkongensis	KJ024484.1		KJ024403.1	KJ024380.1			KJ024517.1
Neohirasea_maerens	KJ024480.1	KJ024432.1	AY121168.1	AY125308.1		KJ024546.1	AY125251.1
Orxines_macklottii		KJ024438.1	AY121153.1	AY125293.1			AY125237.1
Orxines semperi				FJ474150.1	FJ474306.1		FJ474228.1
Oxyartes lamellatus				FJ474151.1	FJ474307.1	FJ474380.1	FJ474229.1
Oxyartes_spinipennis	KJ024489.1	KJ024443.1	KJ024407.1	KJ024384.1			KJ024521.1
Paramenexenus_laetus	KJ024490.1	KJ024444.1	KJ024408.1	KJ024385.1		KJ024554.1	KJ024522.1
Phaenopharos_herwaardeni	KJ024498.1	KJ024451.1	KJ024414.1	KJ024393.1			KJ024530.1
Phaenopharos khaoyaiensis				FJ474157.1	FJ474313.1	FJ474383.1	FJ474235.1
Phaenopharos_struthioneus	KJ024486.1	KJ024440.1	KJ024404.1	KJ024381.1		KJ024551.1	KJ024518.1
Sipyloidea biplagiata	MK291676.1			MK291676.1			
Sipyloidea_sipylus				FJ474168.1		KJ024550.1	AY125264.1
Sipyloidea sipylus				FJ474168.1	FJ474324.1	FJ474393.1	FJ474246.1
Trachythorax_maculicollis	KJ024495.1	J024448.1		KJ024448.1		KJ024559.1	KJ024527.1
Lonchodinae							
Spiniphasma crassithorax				FJ474169.1	FJ474325.1	FJ474394.1	FJ474247.1
Phasmatidae							
Extatosomatinae							
Extatosoma_tiaratum	KJ024511.1	KJ024468.1			FJ474284.1	KJ024574.1	

Suppl. Table 1. NCBI Genbank accession numbers of all the non-Candovia Necroscinae species included in the phylogenetics dataset.

		SEQUENCE	%GC	AMPLICON SIZE (bp)	ANNEALING TIME AND TEMPERATURE
12S	F	CTGCACCTTGACYTGAAATA	42,5	399	56°C, 30s
	R	CAGCATACGCGGTTATACAA	45,0		
16S	F	CGCCTGTTTATCAAAAACAT	35,0	548	56°C, 30s
	R	CTCCGGTTTGAACTCAGATCA	47,6		
COX1	F	TTGATTTTTTGGTCATCCAGAAGT	33,3	819	61°C, 60s
	R	TCCAATGCACTAATCTGCCATATTA	36,0		
COX2	F	CAGATAAGTGCATTGGATTT	35,0	1000	50°C, 60s
	R	GTTTAAGAGACCAGTACTTG	40,0		
$COX1_{FOL}$	F	GTCAACAAATCATAAAGATATTGG	55,0	708	55°C, 45s
	R	TAAACTTCAGGGTGACCAAAAAATCA	41,3		
18S	F	CCTGAGAAACGGCTACCACATC	54,5	800	57°C, 60S
	R	GAGTCTCGTTCGTTATCGGA	50,0		
28S	F	TCAAGACGGGTCGGGAGA	61,0	564	58°C, 30s
	R	AGAACTTTGAAGAGAGAGTTCAAGA	36,0		
H3	F	ATGGCTCGTACCAAGCAGAC	55,0	500	58°C, 30s
	R	ATATCCTTRGGCATRATRGTGAC	41,3		

Suppl. Table 2. Gene target, sequence, GC%, amplicon size (bp), temperature and time of annealing of the primers used in the present study.

1) Cox1$_{Fol}$ dataset (species delimitation analyses)

IQTree model selection with ModelFinder

PARTITION NAMES	ID	MODEL	LogL	AIC	w-AIC	AICc	w-AICc	BIC	w-BIC
FOLst	1	TN+F+I+G4	-1033.5600	2083.1200	+0.0000	2083.7960	+0.0000	2110.3414	+0.0000
FOLnd	2	F81+F+I	-421.4696	852.9393	+0.0000	853.2170	+0.0000	869.9526	+0.0000
FOLrd	3	TIM2+F+G4	-3602.3608	7222.7216	+0.0000	7223.5706	+0.0000	7187.6429	+0.0000

BEAST 2.5 model selection with PartitionFinder2

BEST MODEL	# SITES	PARTITION NAMES
GTR+G+X	222	FOLst
HKY+I+X	222	FOLnd
GTR+G+X	222	FOLrd

2) Species phylogenetic analyses

IQTree model selection with ModelFinder

PARTITION NAME	BEST MODEL	LogL	AIC	w-AIC	AICc	w-AICc	BIC	w-BIC
12S	GTR+F+G4		4373.988529	0	4374.909032	0	4409.203139	0
16S	GTR+F+I+G4		7614.369048	0	7614.86343	0	7661.697857	0
18S+H03st+H03nd	TNe+I		5105.018321	0	5105.183762	0	5130.650312	0
28S	TIM2e+G4		8950.841332	0	8951.33479	0	8992.206915	0
H03rd	TVM+F+G4		3386.588787	0	3388.444457	0	3410.644247	0
CO1st+CO2st+FOLst	TIM2+F+G4		7247.319122	0	7247.656029	0	7285.468161	0
CO1nd+CO2nd+FOLnd	HKY+F+R2		3635.414644	0	3635.751551	0	3671.506054	0
CO1rd+CO2rd+FOLrd	GTR+F+R3		22471.54156	0	22472.10156	0	22530.01922	0

MrBayes model selection with PartitionFinder2.

BEST MODEL	# SITES	PARTITION NAMES
GTR+G	250	12S
GTR+I+G	546	16S
SYM+I	1098	H03st, H03nd, 18S
GTR+G	547	28S
GTR+G	664	CO1st, CO2st, FOLst
HKY+I+G	664	FOLnd, CO2nd, CO1nd
GTR+G	664	CO2rd, CO1rd, FOLrd
GTR+G	107	28Srd

Suppl. Table 3. Results of the model selection performed with ModelFinder and PartitionFinder2 on (1) the species delimitation dataset (2) the phylogenetics dataset.

Reconciling Molecular Systematics and Morphological Taxonomy across four major Euphasmatodea Clades (Cladomorphinae; Diapheromerinae; Lonchodinae; Necrosciinae).

Giobbe Forni, Jacopi Martelossi, Frank Hennemann, Oskar Conle, Alex Cussigh, Pablo Valero, Andrea Luchetti, and Barbara Mantovani.

Abstract. During the two past decades, molecular systematics continuously challenged the traditional taxonomy of Euphasmatodea. Yet, the resulting phylogenies are conflicting and contributed to the difficulty in interpreting morphology-based taxonomical classification. Here, we explore the phenomena underlying molecoular phylogenetic uncertainty and try to reconcile it with the current taxonomy in four Euphasmatodea major clades: Cladomorphinae, Diapheromerinae, Necrosciinae and Lonchodinae. We generated molecular data from seven nuclear and mitochondrial loci for 89 species and analyzed them along with 55 species from previous studies, building the most comprehensive taxon sampling for these clades so far. Our results clearly show how both molecular and morphological approaches have specific limitations. Molecular phylogenetics is affected by a scarce phylogenetic signal, so that topology results very sensitive to different approaches to model specification and data processing; moreover, traditional nodal support metrics can bring a misleading confidence to specific hypotheses. On the other hand, morphology is complicated by the convergent evolution and extreme plasticity of several features. Throughout the paper, we propose changes to current taxonomy and systematics when strongly supported and necessary; we also suggest corrections for discordances that have occurred in clades naming and definitions, applying strict taxonomic rules and the use of type-taxa.

keywords: Cladomorphinae; Diapheromerinae; Euphasmatodea; Lonchodinae; Necrosciinae; phasmids.

Introduction

Phasmatodea is a mesodiverse order of polyneopteran insects subdivided in the two suborders Euphasmatodea and Timematodea. While the latter consists of 21 species distributed across California, Euphasmatodea counts over 3300 species currently divided into about 570 genera and with a predominantly tropical distribution. Phasmids are strictly phytophagous and predominantly nocturnal animals. As their name suggests - phasma being the Greek word for ghost - they exhibit extreme forms of morphological and behavioural mimicry (Figure 1). Their systematics has received increasing attention with numerous efforts either based on morphological (Whiting *et al.*, 2003; Bradler, 2009) or molecular data (Whiting *et al.*, 2003; Buckley *et al.,* 2009; Kômoto *et al*., 2011; Bradler *et al*., 2014; Bradler & Buckley, 2018; Robertson *et al*., 2018; Simon *et al.,* 2019). Nonetheless, this body of literature presents a plethora of conflicting hypotheses, which are also clashing with the traditional taxonomical classification proposed by Günther (1953) and enforced by subsequent authors (Bradley & Galil, 1977; Kevan, 1977, 1982).

Convergent evolution (Buckley *et al.,* 2009; Bradler *et al*., 2015) and intraspecific variability (Hennemann & Conle, 1997) of several features appear to be the main factors that challenged the establishment of a solid taxonomy able to reflect the systematic relationships among lineages, based on morphology alone. Phasmids phenotipic features are often strikingly alike when species live in similar habitats and/or undergo comparable evolutionary pressures, a phenomenon strongly tied to their mimicry of host plants or other elements of their natural environment *(e.g.* mosses, lichens, bark or dry leaf litter on the forest floor). As such, discerning between anatomical traits shaped by convergent adaptations and those valuable for taxonomical classification is a complex and cumbersome process. An striking example of convergent evolution is represented by robust and ground-dwelling phasmids known as “tree lobsters”, which were traditionally placed in the subfamily Eurycanthinae (Günther, 1953; Bradley & Galil, 1977). This subfamily comprised different taxa from New Guinea and New Caledonia and the Lord Howe Island tree lobster, *Dryococelus australis*. Although traditional Eurycanthinae were considered morphologically highly consistent, molecular studies and subsequent morphological re-evaluations showed that *Dryococelus* and New Caledonia species form two monophyletic clades within Phasmatidae *s. str.* (=Lanceocercata), while the New Guinea ones form a monophyletic clade (indicated as the tribe Eurycanthini = Eurycanthomorpha *sensu* Bradler, 2009; Buckley *et al.,* 2009; Bradler *et al.,* 2014) in the subfamily Lonchodinae. Euphasmatodea species also show a high intraspecific variability of morphological characters associated to camouflage. For example, the Indonesian species *Hermagoras foliopeda* (subfamily Lonchodinae) shows a remarkable variability of body sculpturing, shape, size and coloration of protibiae expansions that resulted in the description of at least eight synonyms (Hennemann & Conle, 1997).

Molecular approaches have not been fully capable to resolve the instances where morphology hasn't been up to the task: for the clades on which we focus in this paper, different analyses retrieved a plethora of phylogenetic hypotheses (Figure 2; Bradler *et al*., 2015; Goldberg *et al.,* 2015; Glaw *et al.,* 2019). Also the mopnophily of traditional subfamiles is sometimes questioned, such as for Diapheromerinae, which is retrieved as monophyletic in some studies (Buckley *et al.,* 2009; Robertson *et al*., 2018; Simon *et al*., 2019 and Glaw *et al*., 2019) but polyphyletic in others (e.g., Bradler *et al*., 2015), also depending on the tree inference method (Bradler *et al.,* 2014). Uncertainty in tree topology is a common phenomenon across molecular phylogenies and instances of unresolved or conflicting relationships were found all across the tree of life, including algae (Silberfeld *et al*., 2010), plants (Geuten *et al.,* 2004), fishes (Meyer *et al*., 2015), birds (Suh, 2016), reptiles (Stanley *et al.,* 2011), mammals (Veeramah *et al.,* 2015), and insect clades like Hymenoptera (Romiguier *et al*., 2016), Coleoptera (Timmermans *et al.,* 2016) and Lepidoptera (Bazinet *et al.,* 2017). Moreover, uncertainty can be found at all taxonomic levels, from the genus level (Dillenberger *et al.,* 2017) up to the root of the animal phylogeny (King & Rokas, 2017). Although many of these problematic issues were solved with the use of phylogenomic approaches, the use of a large amount of data is not by itself a solution to phylogenetic problems (Philippe, 2011; Giarla *et al.,* 2015). This appears especially true when it happens at expense of a wide taxonomic sampling, which may have a strong impact on phylogenies as well as limiting assesments on clades monophyly (Heath *et al*., 2008). Uncertainty can be observed from weak nodal support, polytomies and/or sensitivity to the use of independent sampling of species and analytical frameworks (Yuan *et al.,* 2016). Moreover, standard measures of clade support, such as posterior probabilities and bootstrap proportions, can support several conflicting hypotheses with high apparent confidence (Lewis *et al.,* 2005; Simmons & Norton, 2014). Many phylogenetic controversies are centred around branches close to evolutionary radiations, with short and deep branches being highly susceptible to analytical artefacts (Whitfield & Kjer, 2008), that are often related to scarce phylogenetic signal (Rokas & Carrol, 2006; Suh *et al*., 2016), substitution saturation (Zhang *et al.,* 2020), nucleotide compositional heterogeneity (Romiguier *et al.,* 2016; Timmermans *et al.,* 2016), variation of substitution rates among lineages (López-López *et al*., 2017) and incomplete lineage sorting (Alda *et al.,* 2018). The differential impact of these phenomena, together with stochastic errors, can explain conflicting tree topologies in the absence of other biological interpretation (Nosenko *et al.,* 2013). Therefore, some phylogenetic uncertainties have been solved taking into account technical aspects or with the usage of large amounts of data. On the other hand, uncertainties caused by nearly simultaneous cladogenetic events which prevent the reconstruction of the correct branching order are much more difficult to deal with (Tarver *et al.,* 2016). As a phylogenetic uncertainty may reflect biological processes, though, it should not be simply dismissed as a failure but it can instead be considered valuable information on a clade evolutionary history (Rokas & Carrol, 2006).

Yet, even when molecular phylogenetics retrieves strong and consistent support of taxonomic groupings, this does not always result in an appropriate adjustment of traditional taxonomy. Further discrepancies between molecular phylogenetics and morphological studies have arisen when conclusions on higher ordinate taxa and clades were proposed without referring to the eponymous type-species and type-genera of analyzed taxa. An example of incorrect use of a traditional taxon name caused by ignoring the type-genus concerns to Cladomorphinae sensu Robertson *et al.* (2018), where the subfamily name was assigned to a clade that actually does not include the type-genus *Cladomorphus* Gray, 1835. Although the practical use of traditional taxonomic ranks is often replaced by rank-free taxonomy due to its immediacy (*e.g.* Ax, 1984; Willmann, 1989; Dubois, 2007), the consideration of taxonomic type-taxa is still of utmost importance for correct and meaningful assignments of names to clades. Recently several rank-free names have been introduced to define supposedly monophyletic clades in Euphasmatodea, arguing that the traditional classification differed from phylogenetic result to such an extent that saving traditional names in terms of Linnean categories was virtually impossible (e.g., Bradler, 2009). However, in most cases, it is easily possible to retain old names if corresponding family-group names are available and type-taxa are strictly considered according to the code of the International Commission on Zoological Nomenclature (ICZN, 1999). A clear-cut example is the rank-free Lanceocercata Bradler, 2001, which includes the well-known genus *Phasma* Lichtenstein, 1796, the type-genus of the tribe Phasmatini Leach, 1815, of the subfamily Phasmatinae Leach, 1815 and of the family Phasmatidae Leach, 1851. The delimitation of Lanceocercata corresponds to that of the traditional family Phasmatidae *sensu stricto* (Hennemann & Conle, 2008), therefore it should actually be treated as a junior synonym of Phasmatidae. Another similar example is represented by the rank-free Sermyleformia Bradler, 2009 and the subordinate Eusermyleformia Bradler, 2009, both of which include *Diapheromera* Gray, 1835, the type-genus of the traditional and available subfamily name Diapheromerinae Kirby, 1904 and subordinate tribe name Diapheromerini Kirby, 1904. Currently, the Phasmida Species File database (Brock *et al.,* 2017) lists Eusermyleformia as a synonym of Diapheromerini, without morphological or molecular supported delimitation. Discrepancies between the results of molecular phylogenetic and morphological taxonomic studies have also been generated by adjustments to the classification based on misidentified taxa. This, besides overlooking the type-species of the concerned genera, has for example led to a misinterpretation of the genus *Otocrania* (Robertson *et al.*, 2018).

The present analysis focuses on Cladomorphinae, Diapheromerinae, Lonchodinae and Necrosciinae due to their uncertain status. Combining molecular previously published and new data - which have undergone a thorough taxonomic evaluation - we built a 143 species dataset through which we explored the features responsible for inconsistencies a) among different phylogenetic hypotheses based on molecular data and b) between

morphological taxonomy and molecular phylogenetics. We argue that traditional metrics for nodal support can conceal uncertainty in Euphasmatodea molecular phylogenies, which is instead revealed by additional support metrics, inconsistency to different approaches to model specifications and data processing. While comparing our findings with traditional taxonomy, we highlight how convergent evolution and extreme plasticity can explain the inconsistency of certain morphological characters in a phylogenetic framework.

Materials and Methods

Taxon sampling, DNA extraction and amplification

Samples were collected, morphologically identified and preserved dry or in 95% ethanol until molecular analyses were carried out. A total of 89 individuals (Table S1) were selected for molecular analyses. Genomic DNA was isolated using the Smarter Nucleic Acid Preparation kit (Stratec), following the manufacturer's standard protocol. Eight molecular markers were PCR amplified: two mitochondrial protein coding genes (PCGs; two distinct fragments of the Cytochrome Oxidase subunit 1 - the so-called *Folmer* region (the first 600 bp), indicated as $CO1_{fol}$, and the central region, indicated with $CO1_{mid}$ - and Cytochrome Oxidase subunit 2), two mitochondrial rDNAs (12S and 16S), one nuclear PCG (Histone subunit 3) and two nuclear rDNAs (18S and 28S). Primers and their relative thermal cycling conditions can be found in Table S2. All PCR products were purified using ExoSAP-IT PCR Product Cleanup Reagent (Thermofisher). The Sanger sequencing has been carried out by Macrogen Europe Lab. Chromatograms were inspected using SeqTrace 0.9.0 (Stucky, 2012); the resulting sequences were manually checked using Aliview and compared with BLAST (Altschul *et al*., 1997) to the NCBI Genbank nt database to identify potential contaminants. Sequences were submitted to GenBank under the accession numbers MN449491 - MN449962.

Dataset alignment and sequence filtering

Obtained sequences were then merged with additional data, drawn from NCBI Genbank database, relative to 54 species for our groups of interest: they included one species of Cladomorphinae, five species of Diapheromerinae, 26 species of Lonchodinae, 12 species of Necrosciinae, six species of Phylliinae, four species of Aschiphasmatinae and three species of Timematinae along with one Embioptera outgroup (Table S1). All subsequent steps of analysis are summarized in Figure S1. PCGs have been aligned as amino acids using MAFFT v. 7 with auto parameter settings (Katoh & Standley, 2013) and subsequently back translated to nucleotides, while rDNAs were aligned using the "--X-INS-i" algorithm. We cleaned each gene alignment with Gblock (Talavera & Castresana, 2007) using default parameters, with the nucleotide flag for rRNAs and the codon flag for PCG. Alignments were then concatenated using Phyutility (Smith & Dunn, 2008).

Compositional heterogeneity, saturation and phylogenetic signal

Sequence heterogeneity was tested within all the gene alignments using AliGROOVE (Kück *et al.,* 2014), with the default sliding window size and the gaps treated as fifth character. To further check for heterogeneity in nucleotide frequencies among OTUs we used the Yates correction implemented in DAMBE v. 7.0.28 (Xia *et al.,* 2003; Xia, 2018). Substitution saturation was evaluated by generating plots of transitions and transversions against the F84 distances and with Xia test for substitution saturation, both implemented in DAMBE v. 7.0.28. The phylogenetic signal of our dataset was assessed in two ways: (1) by performing a likelihood mapping analysis (Strimmer & Haeseler, 1997), as implemented in IQ-TREE 1.6.10 (Nguyen *et al.,* 2015), with the construction of 15,000 quartets for each gene; (2) by reconstructing gene trees, with the aim to investigate the presence of conflicting signals between topologies derived by different loci. Gene trees were generated with a maximum likelihood (ML) analysis using IQ-TREE 1.6.10, with 1000 ultra-fast bootstrap replicates (Hoang *et al.,* 2018). Best-fit substitution models were selected according to the BIC score using ModelFinder (Kalyaanamoorthy *et al.,* 2017). Three different metrics, with different statistics and discriminative power (Bogdanowicz *et al.,* 2017), were taken as an index of discordance among gene trees topologies: (1) the commonly used Robinson-Foulds distances (Robinson & Foulds, 1981), (2) the quartet distances and (3) matching split distances, which have been demonstrated to be less subject to overestimation (Bryant *et al.,* 2000; Bogdanowicz & Giaro, 2012). Each metric has been computed on the TreeCmp web application (Bogdanowicz *et al.,* 2012), which allows to normalize the distances, indicating whether two trees are more dissimilar than two random trees with the same number of tips, according to a given metric. Values of 1 indicate that the two trees are no more similar than two random trees, while values of 0 indicate the same topology.

Model selection and phylogenetic inference

In the light of the previous analyses (See Results), we built two concatenated datasets: a complete nucleotide dataset, including all sequence's position (henceforth named CMP), and a reduced dataset, excluding the third codon position of mitochondrial PCGs (henceforth named RED). Both datasets had as initial partitions (prior to model selection) genes and codon positions separately. The best-fit partitioning scheme and substitutions model for ML tree searches were obtained using ModelFinder, as implemented in IQ-TREE 1.6.10, with the -TESTNEWMERGE option, which tries to merge partition sharing similar model parameters, decreasing the risk of over-parameterization (Chernomor *et al.,* 2016). For both dataset (CMP and RED) we carried out two model selections: one unlinking branch lengths (-sp option) and one linking them proportionally (-spp option), allowing each partition to have its own evolution rate (Table S3a). The best-fitting branch length's set and the

corresponding tree were chosen according to the BIC scores (Table 1), and we arbitrarly carried out 42 independents runs and subsequently selected the one with the best likelihood value (Table S4). To improve the scanning of the tree space and to try to avoid local optima, the number of unsuccessful iterations to stop was modified from 100 to 400 (-nstop 400). Branch support was estimated using the parametric ultrafast bootstrap approximation (BS; Hoang *et al.,* 2018) with 1000 replicates. To ensure bootstrap convergence, the maximum number of iterations was set to 1500 (-nm option) and then we checked that bootstrap correlation coefficient of split occurrence frequencies reached a value >0.95.

We also used a Bayesian Inference (BI) approach on the two datasets: we determined the best-fit substitutions model and the optimal partitioning scheme using PartitionFinder2 (Lanfear *et al.,* 2016), linking branch lengths (Table S3b). Two independent MCMC analyses were run, with trees and parameters sampled every 2000 steps over a total of 120 million generations in MrBayes 3.2.7a (Ronquist et al., 2012) on the CIPRES gateway (Miller et al., 2010). A conservative burn-in of 25% steps was applied; adequate sample sizes of parameters (>200), average standard deviation of split occurrence frequency (<0.01) and PSRF (Potential Scale Reduction Factor; Gelma & Rubin, 1992) approaching to 1.0 were used as convergence diagnostic. The resulting trees were summarized collapsing nodes showing posterior probability (PP) <0.5, via the sumt command.

Topologies found using ML and BI approaches with linked branch lengths on the CMP and the RED datasets were compared in the same way as described previously for the gene trees. The six clades were then extracted from the CMP tree obtained with the ML approach (see Results and Discussion for justification) using the "plot.phylo" function of the "ape 5" package (Paradis & Schliep, 2019).

Phylogenetic signal of concatenation, additional nodal support metrics and topology testing

The phylogenetic signal in the CMP and in the RED datasets was compared by computing two likelihood mapping analyses by building 14,900 quartets (100-fold the number of sequences) using the best-fit partitioning scheme, substitutions model and branch lengths found previously. For the tree obtained with the CMP dataset and linked branch lengths (see Results for justification) support was also estimated by additional metrics: (a) 1000 replicates of Shimodaira-Hasegawa approximate likelihood ratio test (shALRT; Guindon *et al.,* 2010) (b) the consensus support, i.e. the frequency with which a specific split was present throughout the 42 independents runs carried out with IQ-TREE 1.6.10, (c) the recent implementation in IQ-TREE 2 of gene and site concordance factors (sCF and gCF, respectively; Minh *et al.,* 2020). Values of shALRT, sCF and gCF were tested for correlation with bootstrap using the Pearson's correlation test. Alternative phylogenetic hypotheses were comprehensively evaluated through several topology tests: bootstrap proportion (Kishino et al., 1990), Kishino-Hasegawa test (Kishino & Hasegawa, 1989), Shimodaira-Hasegawa test (Shimodaira & Hasegawa, 1999),

expected likelihood weights (Strimmer & Rambaut, 2002) and approximately unbiased test (Shimodaira, 2002), as implemented in IQ-TREE 1.6.10 with a number of RELL replicates set to 10.000 (-zb option). The first topology tested was constrained to be congruent with the results of Robertson et al. (2018): this has been done because of the similarities in taxon and gene sampling of the two analyses. Further constrained topologies tested were: (1) Aschiphasmatinae as the sister group of Phylliinae or sister to all other Euphasmatodea; (2) the sister group relationship between Lonchodinae and Necrosciinae; (3) the monophyly of Diapheromerinae.

Results

Dataset and sequence analyses

Altogether, we had sequences for 89 individuals: 18 species of Cladomorphinae, 23 species of Diapheromerinae, 24 species of Lonchodinae, 23 species Necrosciinae, and one Phylliinae species. These were analyzed here for the first time in a phylogenetic framework. Moreover, 54 additional species were drawn from NCBI Genbank to complete the dataset. The concatenated alignments count a total of 143 taxa, with a final length of 4701 positions for the CMP dataset and 4042 for the RED one. The percentage of missing data - calculated as the number of available sequences for genes/number of taxa multiplied by the number of gene analyzed - is equal to 40,85% (number of taxa for each marker: $CO1_{fac}$ = 54; $CO1_{mid}$ = 73; CO2 = 82; 12S = 93; 16S = 112; 18S = 53; 28S = 97; H3 = 118).

No substantial difference in sequence composition emerged within whole gene alignments: the mean similarity score calculated by AliGROOVE tends to +1, showing similarities to all other comparisons (Figure S2) and the analyses on compositional bias performed with DAMBE (Table S5) showed a homogeneous composition of genes. However, when analyzing separately each mitochondrial PCGs codon position, third codon positions were found significantly heterogeneous ($P< 0.05$) while the first position, the second position and both positions combined resulted as homogenous (Table S5).

Concerning the saturation of the phylogenetic signal, only third codon positions of mitochondrial PCGs were found saturated by Xia's test (Table S6). Though, transversions outnumbered transitions when analyzing the $CO1_{mid}$ fragment and the CO2 genes, indicating possible substitution saturations (Figure S3). Moreover, Xia's test highlighted some degree of saturation in the two ribosomal genes 16S and 28S that could also be observed in the Ts/Tv plots (Figure S3).

Likelihood mapping analyses (Table S7) recovered the 18S, the 16S and the CO2 as respectively the first, the second and the third most informative markers (number of unresolved quartets = 20.30%; 25.70% and 29.6%, respectively), while for the other sequences the percentage of unresolved quartets were always>30%. The tree

distance analyses (Table S8a) showed a high level of discordance among gene trees by all three evaluated metrics (mean Robinson-Foulds = 0.714; mean Quartet = 0.499; mean Matching Split = 0.525). Maximum distances were found between 12S and 18S for Robinson-Foulds and Matching Split distances (0.874 and 0.723, respectively) and between CO2 and H3 Quartet distances (0.628).

Model selection, phylogenetic inference and nodal support metrics

For both datasets (CMP and RED), unlinking branch lengths in the model selection carried out in ModelFinder caused the a priori partitions to be merged into four and three a posteriori partitions, respectively (Table S3a). The resulting partitioning scheme may have been influenced by the high parametrization of the models - which had respectively 906 and 789 free parameters (Table 1) - so that the attempt to reduce the number of parameters resulted into partitions that are highly heterogeneous and have a poor biological meaning. Instead, using the proportional branch lengths option in ModelFinder, more homogeneous a priori partitions were generated (Table S3a), reflecting their different origin and evolutionary rate (e.g., mitochondrial PCGs in distinct partitions with respect to ribosomal and/or nuclear markers). Between linked and unlinked branch lengths, the mean bootstrap value was approximately equal for the CMP dataset (respectively 91.2 and 91.3) and slightly decreased for the RED dataset (Table S3a). AICc and BIC scores were lower when proportional branch lengths were used, for both the datasets, showing a better fit of the corresponding partitioning scheme (Table 1). For this reason, in ML analyses, we focused further investigation on trees obtained by linking branch lengths proportionally, while for the Bayesian Inference we have chosen the linked parameter of PartitionFinder, which, similarly to the -spp option of ModelFinder, estimates only one underlying set of branch lengths but with different scaling parameters for each partition (i.e accommodating for subset-specific rates). The partitioning scheme and model selection obtained using PartitionFinder2 (Table S3b) was generally similar to the one obtained with ModelFinder.

Concerning relationships between subfamilies, ML trees obtained with the CMP and RED concatenations showed identical topologies (Figure 3; Figure S4), as also pointed out by the tree distances (normalized Robinson-Foulds = 0.103; normalized Quartet distance = 0.004; normalized Matching split = 0.047; Table S8b). Discordance between the two phylogenies was found mainly at lower taxonomic levels. As for the ML inference, the BI recovered stable deep relationships with respect to the dataset used (normalized Robinson-Foulds = 0.099; normalized Quartet distance = 0.011; normalized Matching split = 0.061 in Table S8b).

Mean bootstrap values and posterior probabilities do not varied substantially when third codon positions of mitochondrial PCGs were removed (Table S3a). This result was only partially consistent with the likelihood mapping analyses, where the number of unresolved quartets is equal to 30.2% for the CMP and to 36% for the

RED (Table S3a). For both datasets, the inference's analytical framework did not seem to impact the phylogenetic reconstruction: tree distance values between ML and BI trees were low (Table S8c) and subfamilies relationships appeared fully congruent (Figure 3; Figure S4).

We selected the phylogenetic tree obtained by analyzing the CMP dataset as the main tree of this analysis, because with substantially similar bootstrap values and posterior probabilities, it showed the smaller number of unresolved quartets.

Generally speaking, concordance factors reflected bootstrap values, with more recent nodes receiving relatively higher support (Figure S5a). Both gCFs and sCFs gave significantly similar estimates (Person's correlation r=0.67, p<0.0001; Figure S5b). Although showing a significant correlation with bootstrap values, discrepancies emerged: beside nodes which were highly supported by concordance factors and by bootstrap values, also nodes with high bootstrap values but low concordance factors can be observed throughout the phylogeny (Figure S5c,d). shALRT values result generally higher than CFs, even if some discordances with bootstrap were still present (Figure S5e). It is interesting to note that all replicates carried out to find the best ML tree (Table S3) showed the same topology, so that the consensus tree showed nodes with frequency of 100%.

Major clades systematic relationships and topology tests

In the reconstructed tree, phylogenetic relationships among main clades were moderately to weakly supported, in some instances with wide differences among different metrics used for node support. However, main clades are generally well-supported (Figure 3).

The sub-family Diapheromerinae resulted paraphyletic, the clade including the tribe Diapheromerini as the sister group of a cluster formed by the other tribe, Oreophoetini, sister to all remaining analyzed taxa (Figure 4). Interestingly, the species *Cranidium gibbosus*, ascribed to the tribe Cranidiini, sub-family Cladomorphinae (see below), resulted as the sister clade of Diapheromerini. Moreover, two species of the genus *Hirtuleius*, which belong to the Cladomorphini tribe (subfamily Cladomorphinae), was included into a monophyletic clade including the Diapheromerini species of the genera *Phantasca, Phanocles, Phanocloidea, Trychopeplus* and *Alienobostra*.

The two subfamilies Aschiphasmatinae and Phylliinae resulted both monophyletic and in sister relationship (Figure 5) even if the tree topology constraining Aschiphasmatinae as a sister group of the cluster *Phyllium* + all remaining Euphasmatodea is supported by the topology test (Table 2). This is in agreement with Engel *et al.* (2015), who regarded this lineage as Aschiphasmatodea and the remaining Euphasmatodea as Neophasmatodea, hypothesis recovered also in numerous previous studies (Tilgner, 2002; Buckley et al. 2009; Bradler *et al.,* 2015, Robertson *et al.,* 2018). In contrast, two recent phylogenomics by Simon *et al.,* (2019) and

Tihelka *et al.,* (2020), carried on with the same transcriptomic dataset but with different analytical methods, gave the Phylliinae in a nested position inside the remaining Euphasmatodea, as sister to Bacillinae and to a clade comprising mainly Lonchodidae, Clitumninae, Lanceocercata and Cladomorphinae. Within Aschiphasmatinae*,* *Dajaca napolovi* (tribe Dajacini) is nested among all other species that have been ascribed to the tribe Aschiphasmatini (Figure 5). The seven Phylliinae species considered, all belong to the tribe Phylliini and include four species of the genus *Phyllium* (subgenera *Phyllium* and *Pulchriphyllium*) and three species of the genus *Chitoniscus*. Although Phylliini is strongly supported as a monophyletic clade, neither genus was found monophyletic, as one species of *Chitoniscus* (the New Caledonian *C. brachysoma*) resulted nested within *Phyllium* with strong support (Figure 5).

The sub-family Lonchodinae resulted monophyletic (Figure 6). It comprises the tribe Lonchodini, presently appearing as paraphyletic, and the monophyletic Eurycanthini (=Eurycanthomorpha sensu Bradler, 2009; Buckley *et al.,* 2009a; Bradler *et al.,* 2014). Within Lonchodini, interestingly, the two congeneric species *Hyrtacus tuberculatus* and *H. procerus* appeared distantly related, the latter resulting in sister relationship with Eurycanthini (Figure 6). In this latter tribe the genus *Neopromachus* appears polyphyletic, because it actually splits into two distantly related lineages.

New Caledonia tree lobsters, that have already been recognized as Phasmatidae *s. str.* (=Lanceocercata) (Buckley *et al.,* 2009), form a monophyletic cluster in sister relationship with a clade including the overall majority of Cladomorphinae (excluding *Cranidium gibbosum* and *Hirtuleius gracilis* clustering with or within Diapheromerinae; see Figure 4) (Figure 7). This clade groups the three tribes Haplopodini, Hesperophasmatini and Pterinoxylini. The monophyly of the tribe Haplopodini, here represented by the genera *Haplopus* and *Diapherodes*, is strongly supported and this holds also for the monophyly of the tribe Pterinoxylini. On the other hand, Hesperophasmatini clearly appear as paraphyletic, their cluster including the Pterinoxylini clade (Figure 7).

The subfamily Necrosciinae was found monophyletic with strong nodal support (Figure 8). The clade is further divided into two main groups: one including the "*Neohirasea* clade" (sensu Buckley *et al.,* 2009; Bradler *et al.,* 2014) and a monophyletic group comprising the genera *Rhamphosipyloidea* and *Hemisosibia*, the other including all the remaining Necrosciinae.

The overall congruence of the presently obtained ML topology with that of Robertson *et al.* (2008) is rejected by all topology tests implemented in our analysis (Table 2). This occurs also when constraining Lonchodinae (except the New Caledonia tree lobsters) as sister clade of Necrosciinae. On the other hand, the constrained monophyly of Diapheromerinae is not rejected as well as the tree topology constraining Aschiphasmatinae as sister group of the cluster Phyllinae + all other Euphasmatodea (Table 2).

Discussion

Uncertainty in Euphasmatodea molecular phylogenetics

We investigated three features that usually lead to artefactual topology reconstruction when analyzing molecular phylogenetics datasets: substitution saturation, compositional heterogeneity and the strength of the phylogenetic signal. Our aim was to understand to which extent these phenomena were affecting our phylogenetic inferences and how to deal with their effects. Overall, we found a clear signature of a reduced phylogenetic signal: likelihood mapping analyses carried on both single genes and the concatenated alignment showed high percentages of unresolved quartets. Moreover, the high topological distances recovered between different gene trees highlight the presence of unclear phylogenetic signals. This phenomenon does not seem to be caused by a possible mito-nuclear conflict, as topological discordance is also present between markers with the same origin (either mitochondrial or nuclear genomes), but most likely results from the arbitrary resolution of topologies in the absence of a strong phylogenetic signal. The third codon positions of mitochondrial PCGs were found saturated both by Xia's test and Ts/Tv plots and the same positions resulted also significantly heterogeneous in nucleotide composition. Nonetheless, the gain in terms of phylogenetic noise reduction when removing these positions does not seem to exceed the signal loss due to their exclusion. In fact, considering only the 1st and the 2nd codon positions led to a marked decrease in both the number of resolved likelihood mapping quartets and in the average nodal support values of the resulting trees, although inferred phylogenetic relationships between sub-families do not change when including 3rd codon positions. As already suggested for Annelida (Struck *et al.,* 2008), it is likely that in Phasmatodea a higher number of aligned positions, even if saturated and with heterogenous nucleotide composition, may actually improve the tree inference when analyzed with a proper model specification. For the same reason we have not considered approaches based on amino acids, as they convey a lower amount of information compared to nucleotides.

When analyzing recovered topologies, we used bootstrapping and posterior probabilities along with different kinds of support metrics (i.e., gene and sites concordance factors, SH-aLRT, frequency of splits throughout independent ML runs), and tested results that were in contrast with previous phylogenetic hypotheses. An increasing number of studies is highlighting the possible overestimation of bootstrap and posterior probability in case of rapid radiation (Suh, 2016) and/or model misspecification (e.g., Roycroft *et al.,* 2019; Minh *et al.,* 2013), which may also lead to a strong confidence towards otherwise uncertain topologies. This confidence is generally lost when other metrics for support and reproducibility are taken into account (Suh, 2016). Main phylogenetic relationships within Euphasmatodea appeared to be similarly affected by this phenomenon: sub-optimal model specifications (unlinked branch lengths) resulted in highly supported relationships among sub-families, which are

found different – still with high support – when best-fitting models were used (linked branch lengths). Even if concordance factors positively correlated with bootstrap values (even though with low correlation coefficient), in most instances CFs are weak and several nodes showing a high bootstrap value present a very poor corresponding gCF and sCF. With the exception of the Timematodea/Euphasmatodea split, most nodes in the tree showed low CF values. This highlights the uncertainty regarding several relationships, as the inclusion or exclusion of some genes - or part of them - may affect the inferred topology. In this respect, another interesting outcome concerns the phylogenetic placement of paraphyletic Diapheromerinae (see below for further taxonomic implications) with respect to the remaining Euphasmatodea: this cannot be considered a novel finding, since it has already been suggested by Whiting *et al.* (2003) and Bradler *et al.* (2014), but the weak phylogenetic signal highlighted by CFs could provide a possible explanation for the conflicting results reported in literature (Whiting *et al.,* 2003; Buckley *et al.,* 2009; Bradler *et al.,* 2014; Bradler *et al.,* 2015; Robertson *et al.,* 2018; Simon *et al.,* 2019). Even if the support for the non-monophyly of the Lonchodidae (Lonchodinae + Necrosciinae) resulted to be weak, the topology test rejected the hypothesis of monophyly. Despite the sister relationship between Lonchodinae and Necrosciinae having been proposed several times (Bradler *et al.,* 2014; Bradler *et al.,* 2015; Robertson *et al.,* 2018; Simon *et al.,* 2019), our analysis agrees with different hypotheses that are already present in literature (e.g., Glaw *et al.,* 2019).

Generally speaking, though, the scarce phylogenetic signal appeared linked to the fast and ancient evolutionary radiation of Euphasmatodea (Forni *et al.,* 2021). The typical signatures of such a phenomenon in phylogenies are short and deep internodes, which reflect a non-stochastic diversification process throughout the tree, and long terminal branches, which reflect the age of the event (Whitfield & Kjer, 2008). Divergence events associated to the radiation are separated by short times while followed by long time spans, so that few evolutionary changes are hindered by many subsequent ones, thus resulting in the erosion of phylogenetic signal through time. This signature is consistently present in phasmid phylogenies, implying that the radiation is one of the most likely causes for the poor phylogenetic signal at the basis of the Euphasmatodea tree of life.

Reconciling morfological taxonomy and molecular phylogenetics

The presently reported molecular phylogeny does not include samples from each of the major Phasmatodea lineages, but it was limited to the sub-families Cladomorphinae, Diapheromerinae, Lonchodinae, Necrosciinae along with Aschiphasmatinae and Phylliinae. In this regard, our aim was to focus efforts on selected clades that are of uncertain status and subject to conflicting hypotheses, in order to investigate features responsible for the controversial results of different studies and trying to reconcile molecular systematic with morphological taxonomy. Our extensive and more representative taxon sampling among the six aforementioned clades, which

comprises 89 species that have never before been included in molecular phylogenetic studies, allowed for a considerably deeper insight into the phylogeny of these taxa. The results we achieved evidence several significant relationships, support some previously proposed hypotheses and provide support for entirely novel ones. These have been carefully evaluated by comparing present molecular data with morphological taxonomy and new taxonomic considerations are only drawn in cases where we had strong confidence.

From the sub-family Diapheromerinae, we analyzed samples from 20 genera belonging to the two tribes Diapheromerini (16 genera) and Oreophoetini (*sensu* Robertson *et al.,* 2018; four genera). With the exception of the RED dataset with unlinked branch lengths, the subfamily was recovered as paraphyletic in all inferences we carried out, with all other analyzed Euphasmatodea included in a monophyletic clade. All analyzed members of the tribe Diapheromerini are clustered in a monophyletic clade which also includes a species of *Hirtuleius*, a genus traditionally ascribed to the sub-family Cladomorphinae (tribe Cladomorphini). Moreover, another Cladomorphinae species, *Cranidium gibbosus* (tribe Cranidiini) resulted closely related to Diapheromerini in the present analysis. Current morphological analyses (Hennemann & Conle, in prep.), aiming to revise the tribe Cladomorphini as a whole, suggest *Hirtuleius* to be very closely related to the genus *Cladomorphus*, the type-genus of the tribe Cladomorphini and subfamily Cladomorphinae. Moreover, it is interesting to note that a recent phylogenetic analysis indicated the species *Cladomorphus phyllinus* as closely related to *Phanocles costaricensis* (Bank *et al.,* 2021). From a taxonomic point of view, considering the rules of the International Commission on Zoological Nomenclature (ICZN, 1999), the monophyletic lineage that in our tree includes the genera *Alienobostra, Hirtuleius, Paraphanocles, Phanocles, Phanocloidea* and *Trychopeplus*, as well as *Cladomorphus* and *Otocrania* (Hennemann & Conle, in prep.; Bank *et al.,* 2021), must be considered as the tribe Cladomorphini *sensu nov*. (for a definition see below). The sister group to Cladomorphini *sensu nov*. is a lineage comprising all the remaining genera belonging to Diapheromerini *sensu* Robertson *et al.* (2018). Since the latter monophyletic lineage includes the type-genus *Diapheromera,* it must be regarded as the tribe Diapheromerini sensu nov. (for a definition see below). Moreover, we found support for the rank-free taxon Eusermyleformia (Bradler, 2009; Hennemann & Conle, 2012), here represented by the three species *Diapheromera femorata*, *Megaphasma dentricus* and *Pseudosermyle phalangiphora* and found nested within Diapheromerini *sensu nov*.

Within Cladomorphini *sensu nov*. the here revealed placement of *Phantasca quadrilobata* is very interesting and deserves a more detailed explanation. Our results appear, at least partially, in contrast to previous suggestions on its placement and relationships retrieved by means of morphological characters (Hennemann *et al.,* 2018). In our tree *P. quadrilobata*, together with *Hirtuleius gracilis*, is the sister taxon to all other Cladomorphini *sensu nov*. and this placement was found consistently across different phylogenetic inferences. Based on morphological characters, but mostly the anatomy of the male terminalia, Hennemann *et al.* (2018) suggested a close

relationship to the *L. strigiventris* species group of the genus *Libethra* (see Günther, 1932), which is actually rejected by our analysis. There are, though, some morphological characters that may support the position of the genus *Phantasca* within Cladomorphini *sensu nov*. such as i) the morphology of the profemora, which have the dorsal carinae strongly approaching each other with the anterior one being strongly raised; ii) the fairly distinct medioventral carina notably displaced towards the anteroventral carina; iii) the presence of a vomer in males; iv) the distinct gonoplacs of females that are equal in length to slightly longer than gonapophyses IX; v) the elongated gonapophyses VIII.

As already pointed out above, the monotypic tribe Cranidiini is here revealed as sister group to Cladomorphini *sensu nov*. + Diapheromerini *sensu nov.* Basically, this topology is in line with the classification suggested by Hennemann et al. (2016b) on morphological dataset, which placed Cranidiini as sister group to Cladomorphini + Cladoxerini. However, since we have not sampled any members of Cladoxerini for this study, the position and relationships of Cladoxerini must for now be carefully considered. While most morphological characters strongly support the subordinate placement within Cladomorphinae, Cladoxerini in fact differ from Cladomorphini, Diapheromerini and Cranidiini by the short antennae of females, which consist of no more than 30 segments and are no more than half the length of the profemora, and the serrate anteroventral carina of the female profemora (Hennemann *et al.,* 2016b).

Consequently, the monophyletic clade that comprises Cranidiini + (Cladomorphini sensu nov. + Diapheromerini *sensu nov*.) must be properly regarded as the subfamily Cladomorphinae *sensu nov*. for which we present a new diagnosis as follows:

Cladomorphinae Brunner v. Wattenwyl, 1893 *sensu nov*. (Figure 4)

= Diapheromerinae Kirby, 1904 *syn. nov*.

Type genus. *Cladomorphus* Gray, 1835: 15.

Diagnosis. Cladomorphinae *sensu nov*. is well supported by molecular data as are the three tribes currently contained, namely Cladomorphini *sensu nov*., Cranidiini and Diapheromerini. However, as already pointed out by Robertson et al. (2018; referring to Diapheromerini) almost no morphological apomorphies have so far been identified for this clade. Members of the subfamily have a gula, lack an area apicalis on the tibiae and the eggs all have a median line, which sometimes is displaced towards the polar area (Cranidiini). The Cladomorphini is characterized by a) the strongly lamellate medioventral carina of the profemora, which is strongly displaced towards the anteroventral carina; b) presence of a praeopercular organ on abdominal sternum VII; c) the often enormously elongated gonapophyses VIII and enlarged gonoplacs of females. Males may possess conspicuous

specializations of the terminalia, e.g. an apical projection of the poculum (= subgenital plate), a strongly enlarged and tube-like phallus or a vomer with two terminal hooks. The first three characters are shared with Cranidiini; this taxon is however readily separated by the indistinct medioventral carina of the profemora being centric on the ventral surface of the femur, strongly flattened and laterally dilated body of females and spinulose phallus of males (Hennemann et al., 2016b). Males of both tribes are either apterous or winged, and if alae are present these are always plain and transparent. Females are exclusively wingless. They comprise the majority of the New World giant stick insects and some of the largest known Phasmatodea of that region with females of *Cladomorphus, Otocrania* or *Phanocles* achieving body lengths of up to 25 cm if the elongated subgenital plate is included. The tribe Diapheromerini *sensu nov*. differs from the other two tribes by numerous morphological characters and is diagnosed separately below.

Taxa included. This subfamily contains the majority of the species diversity of anareolate New World stick insects of the Occidophasmata and comprises the three tribes Cladomorphini *sensu nov*., Diapheromerini *sensu nov*. and Cranidiini. Cladomorphinae *sensu nov*. corresponds to what Robertson *et al.* (2018) and Simon *et al.* (2019) regarded as Diapheromerini. Because Cladomorphinae *sensu nov*. comprises the eponymous genus *Diapheromera* and *Cladomorphinae* has priority over Diapheromerinae the latter name falls as a junior synonym (*syn. nov*.). Thus, the clade automatically comprises all taxa previously classified as Diapheromerinae including *Otocrania* (see Robertson *et al.,* 2018) but also the genera attributed to the tribe Cladomorphini by Hennemann & Conle (2010) and Cladomorphinae by Hennemann *et al.,* (2016) as well as *Phantasca* and *Trychopeplus*. Only the tribe Cladoxerini was not sampled in this or any previous molecular phylogenetic study, hence its relationships are here considered incertae sedis. Though, Hennemann *et al.,* (2016) have presented morphological support for a subordinate position of Cladoxerini within Cladomorphinae.

Distribution: Cladomorphinae *sensu nov*. is distributed throughout most of the New World, ranging from southern Canada and the United States to the north over Central America and the Caribbean to Paraguay and Argentina to the south.

Cladomorphini Brunner v. Wattenwyl, 1893 *sensu nov.* (Figure 4)

= Diapheromerini Kirby, 1904 (in part)

Type genus. *Cladomorphus* Gray, 1835: 15.

Diagnosis: Cladomorphini *sensu nov.* is characterized by the strongly lamellate medioventral carina of the profemora which is strongly displaced towards the anteroventral carina. Females possess a distinct praeopercular organ on abdominal sternum VII, while the gonapophyses VIII are often enormously elongated and

the gonoplacs are much enlarged. Males often show specializations of the terminalia, e.g. an apical projection of the poculum (= subgenital plate), enlarged and hook-like cerci, a strongly enlarged and tube-like phallus or a vomer with two (sometimes asymmetrical) terminal hooks. The vomer is always present and well developed. While females are exclusively apterous, males show the entire range of wing anatomy ranging from apterous to fully winged but with the tegmina always small, scale-like and notably shorter than the alae. If the alae are developed, the anal fan is monochromatic and hyalinous to transparent grey. Eggs show a wide range of morphological variability and always bear a hollow, either conical or crown-like capitulum. The capsule sculpturing varies from almost smooth to distinctly pitted or rugulose to being all over covered with hairy structures.

Taxa included: Cladomorphini *sensu nov*. comprises all general formerly attributed to the tribe (Hennemann & Conle, 2010; Hennemann *et al.,* 2016b) as well as the "*Phanocles* group" and two genera of the "*Clonistria* group" as defined by Zompro (2001). Namely, these are the genera: *Alienobostra, Aplopocranidium, Bostra, Cladomorphus, Globocalynda, Hirtuleius, Jeremia, Jeremiodes, Laciphorus, Otocrania, Paraphanocles, Phanocles, Phanocloidea, Phantasca, Trychopeplus* and *Xylodus*. Only the seven genera *Alienobostra, Hirtuleius, Paraphanocles, Phanocles, Phanocloidea, Phantasca* and *Trychopeplus* have been sampled in our analysis but the eight genera did not include all key out morphologically as members of Cladomorphini.

Distribution: The tribe is distributed throughout most of the Neotropical region, ranging from Central Mexico to the north over Central America and the Caribbean to Paraguay and Argentina to the south.

Diapheromerini Kirby, 1904 *sensu nov*. (Figure 4)

Type genus. Diapheromera Gray, 1835: 18.

Diagnosis: While Diapheromerini *sensu nov*. is well supported by molecular data, morphological apomorphies have not yet been defined. Females do not have considerably elongated gonapophyses VIII, a praeopercular organ on abdominal sternum VII is mostly lacking and most taxa have a short, often scoop-shaped subgenital plate (exceptions are *Andeocalynda* and *Calynda*). Males show a wide array of terminalia specializations, that include strongly elongated cerci, which may have basal protuberances or a forked apex, projections of abdominal sternum IX, or a ventrally closed and tube-like tergum X. A vomer is either present or absent. All known members are exclusively apterous in both sexes.

Taxa included: Diapheromerini *sensu nov*. includes all genera assigned to Sermyleformia and Eusermyleformia by Bradler (2009), namely *Diapheromera, Litosermyle, Manomera, Megaphasma, Parabacillus, Pseudosermyle* and *Sermyle*, as well as *Bacteria, Bostriana, Calynda, Caribbiopheromera, Oncotophasma* and *Paracalynda*, the

latter six being all sampled in our analyses and represented in our tree (Figure 4). Further genera, which on morphological characters key out as members of Diapheromerini *sensu nov*. are *Andeocalynda, Clonistria, Paraclonistria* and *Pseudoclonistria*. Further work is obviously needed for exactly delimiting Diapheromerini *sensu nov*. and it is hoped that these genera can be included in future phylogenetic analyses to clarify their true relationships.

Distribution: This tribe is distributed from southern Canada and the United States in the north to Peru and Bolivia in the south and also has a considerable number of taxa in the Caribbean. The highest density of species is found throughout Central America from Mexico to Colombia.

The analysed members of the tribe Oreophoetini (sensu Robertson *et al.,* 2018), namely *Lobolibethra carbonelli*, *Libethra strigiventris*, *Oreophoetes topoense* and *Dyme bifrons*, form a separate monophyletic lineage. This is in line with the result of Robertson *et al.* (2018), that recovered Oreophoetini as monophyletic based on sampling of the genera *Lobolibethra, Oreophoetes, Dyme* and *Ocnophiloidea*. Therefore, this lineage can now be delimited to include *Dyme, Libethra, Lobolibethra, Ocnophiloidea, Oreophoetes* and *Oreophoetophasma*. Moreover, based on morphological data, there is good support for including also the genera *Laciniobethra, Libethroidea, Nanolibethra* and *Spinopeplus* (see Conle *et al.,* 2011). It is to be noted though, that Robertson *et al.* (2018) showed this lineage as the sister group of the currently recognized Cladomorphinae *sensu nov*. The future inclusion of members of the tribe Ocnophilini in molecular analyses could help in shedding light on the correct phylogenetic relationships of Oreophoetini. For taxonomy this would mean that if the genus *Ocnophila* turns out to belong to the same lineage, then Ocnophilini (Günther, 1953) would have priority over Oreophoetini (Zompro, 2001). So far, any sampling in phylogenetic studies has included *Ocnophila*, the type-genus of the tribe Ocnophilini.

In our tree (Figure 7), the three tribes Haplopodini, Hesperophasmatini and Pterinoxylini form a monophyletic clade, which is not only in concordance to previous molecular-based results (Robertson *et al.,* 2018; Simon *et al.,* 2019) but also to the morphology-based results presented by Hennemann *et al.* (2016b, fig. 409). Tribes in this clade were previously ascribed to Cladomorphinae (Hennemann *et al.,* 2016) and in the present analysis resulted as the sister group to the New Caledonia tree lobsters (Figure 7), which previous studies (e.g., Buckley *et al.,* 2009a) have shown to belong to Phasmatidae s. str. (=Lanceocercata). This relationship is further supported by a recent phylogenetic analysis (Bank et al., 2021). The tribe Haplopodini, here represented by the genera *Haplopus* and *Diapherodes*, forms a strongly supported monophyletic clade and there is also good support for the monophyly of the tribe Pterinoxylini as established by Hennemann *et al.* (2016b). Within Hesperophasmatini, the genus *Sigaruphasma* is recovered as the sister group to all remaining genera,

supporting the numerous morphological peculiarities of this genus discussed by Hennemann *et al.* (2020), who stated that the genus violates several diagnostic characters of Hesperophasmatini. These characters include the strongly expanded and lamellate posteroventral carina of the female profemora, the strongly laterally compressed meso- and metafemora of both sexes as well as the complete lack of hairy structures on the capsule and operculum of the eggs. The Hispaniolan species *Lamponius bocki* renders *Lamponius* as currently treated polyphyletic. However, morphological characters such as the well-developed sensory-areas of the probasisternum and proforcasternum show *L. bocki* to be misplaced in *Lamponius*, a genus that is well characterized within Hesperophasmatini by lacking these sensory-areas (a feature only shared with *Sigaruphasma*; see Hennemann *et al.,* 2020). Hence, *L. bocki* is here transferred to the genus *Hesperophasma*, as *Hesperophasma bocki comb. nov*.

From a morphological point of view, the monophyletic clade including Haplopodini, Hesperophasmatini and Pterinoxylini was hypothesized as not to belong to Cladomorphinae sensu stricto (Hennemann *et al.,* 2020). Given the above re-description of Cladomorphinae *sensu nov*., in consideration of the code of the International Commission on Zoological Nomenclature (ICZN, 1999), we here introduce the new name Haplopodinae *subfam. nov*. for which we present a diagnosis and a list of included taxa as follows:

Haplopodinae Günther, 1953 *subfam. nov.* (Figure 7)

Type-genus. *Haplopus* Burmeister, 1838: 576.

Diagnosis: the subfamily Haplopodinae *subfam. nov*. is well supported by molecular data (see above). Morphological characters that delimit the clade amongst other New World Taxa and separate it from Cladomorphinae *sensu nov*. are the not elongated gonapophyses VIII and reduced gonoplacs of females as well as the lack of a median line in the eggs (Hennemann *et al.,* 2016b). Moreover, the morphology of the profemora readily separates the clade from Cladomorphinae in having the medioventral carinae centric on the ventral surface of the femur or at best very slightly displaced towards the anteroventral carina (Pterinoxylini). All winged taxa have the anal region of the alae coloured, either being plain or conspicuously tessellated (always plain and transparent in Cladomorphinae), and females of Pterinoxylini possess a stridulatory organ in the costal region of the alae, which enables the insects to produce stridulating sounds when disturbed. Haplopodini and Hesperophasmatini lack a gula; Hesperophasmatini (exceptions are given by *Lamponius* and *Sigaruphasma*) and Pterinoxylini possess rough sensory areas on the probasisternum and profurcasternum (Hennemann *et al.,* 2016).

Taxa included. Haplopodinae *subfam. nov*. comprises all genera of the tribes Haplopodini, Hesperophasmatini and Pterinoxylini as discussed by Hennemann *et al.* (2016) as well as the recently described tribe Teruelphasmini (Yong, 2017). This is the clade that has erroneously been referred to as Cladomorphinae previously (Robertson *et al.*, 2018); Simon *et al.*, 2019) and is the only clade of New World phasmatodeans that belongs to the Old World Oriophasmata. Hence, a position within the family Cladomorphidae, which belongs into the Occidophasmata, can be excluded. An Old-World origin of Haplopodinae *subfam. nov*. and possible sister relationship to Phasmatidae s. str. (= Lanceocercata) has already been shown by previous molecular-based data (Robertson *et al.,* 2018; Simon *et al.,* 2019) and is also supported by our analysis (Figure 7). Considering morphological characters, Hennemann *et al.* (2016b) have hypothesized a close relation to the African subfamily Palophinae. This has subsequently found support in the molecular-based result of Robertson *et al.* (2018), which shows Palophinae (sampled genus *Bactrododema*) as sister to Haplopodinae *subfam. nov*., albeit the placement of *Bactrododema* has varied throughout different studies (Buckley *et al.,* 2009a, 2009b; Bradler *et al.,* 2015; Goldberg *et al.,* 2015; Büscher *et al.,* 2018). Consequently, Haplopodinae *subfam. nov*. must be regarded as subordinate to Phasmatidae s. l. (compare Hennemann & Conle, 2008). However, the exact placement and relationships of Haplopodinae *subfam. nov*. still deserve evaluation by including more African taxa in forthcoming phylogenetic analyses.

Distribution: Haplopodinae *subfam. nov*. is a mainly Caribbean clade with the majority of taxa distributed throughout the Antilles and with only a few representatives found in Central America (Hennemann & Conle, 2012; Hennemann *et al.,* 2016b; Conle *et al.,* 2020). The monophyly of this clade and its distribution mirror the important role of biogeography for the evolutionary history of stick insects (Buckley *et al.,* 2009b; Bradler *et al.,* 2015). The Caribbean subregion is in fact defined to extend through southern Mexico, Central America, the Antilles and northwestern South America (Morrone, 2007) with the true West Indies (= Antillean biogeographic dominion) possessing a very typical fauna with an apparently high degree of endemism.

The very diverse subfamily Lonchodinae (Figure 6) currently counts 53 genera that are divided into the two tribes Lonchodini and Eurycanthini (Phasmida Species File; http://phasmida.speciesfile.org). We have sampled 42 species from 18 genera of Lonchodinae, which is the largest set of taxa ever incorporated in a phylogenetic study of this subfamily so far. In our analysis Lonchodinae resulted as monophyletic, which is in concordance to previous molecular results (Buckley *et al*., 2009a, 2009b; Bradler *at al.,* 2014). The two tribes presently considered, Lonchodini and Eurycanthini (=Eurycanthomorpha *sensu* Bradler, 2009; Buckley *et al.,* 2009a; Bradler *et al.,* 2014), were clearly recognized, although Lonchodini resulted as paraphyletic. The monophyletic Eurycanthini is in our analysis represented by 14 species from five different genera and includes only the New Guinea tree-lobsters (e.g., *Eurycantha, Thaumatobactron, Neopromachus, Erinaceophasma, Eupromachus,*

Brachyrtacus). This confirms their divergence from the New Caledonia tree-lobsters (e.g., *Asprenas*, *Canachus*, *Carlius*, *Labidiophasma*, *Microcanachus*) and *Dryococelus*, which have already been revealed as subordinate to the family Phasmatidae s. str. (=Lanceocercata; Buckley *et al.,* 2009). The true delimitations of both clades of tree lobsters however still deserve clarification by detailed examination of all concerned genera. Based on morphological examination we confirm the placement of *Papuacocelus* in Eurycanthini, because the genus exhibits both key features of this clade, having a large gula and strongly reduced gonoplacs. The beak-like secondary ovipositor present in almost all Eurycanthini (with the exception of *Papuacocelus*, *Pseudopromachus* and *Thaumatobactron*) and formed by the elongation of the abdominal tergum X and the subgenital plate, is however also found in certain Lonchodini genera (e.g., *Manduria*; see Hennemann & Conle, 2007) and must thus be presumed to have been evolved at least two times within Lonchodinae. Autoapomorphies of Eurycanthini distinguishing it from Lonchodini are the reduced membrane between the hemiterga of male abdominal tergum X and the lack of a capitulum or raised opercular structures in the eggs (Bradler, 2009). The straight profemora of certain genera appears to have been evolved secondarily within Eurycanthini and is even hypothesized to have been developed independently in *Eurycantha* and *Erinaceophasma* (Bradler, 2009). The strongly developed sexual dimorphism and the extreme masquerade crypsis of several genera of Lonchodini, that are often attended by an impressive range of intraspecific morphological variability of various external characters of the adult insects, make the delimitation and differentiation of individual genera a difficult task on morphological data alone (e.g., Hennemann & Conle, 2007). The genera *Carausius*, *Lonchodes* and *Mnesilochus* were recovered as polyphyletic in our analyses. Molecular data based on an even more extensive sampling than in our analysis will be necessary for obtaining meaningful conclusion on the intergeneric relationships and delimiting genera within Lonchodinae. Our analysis renders the genera *Carausius*, *Lonchodes* and *Mnesilochus* as polyphyletic in their current taxonomic composition while there is support for the monophhyly of *Hermagoras* and *Staelonchodes*. However, these results can be only preliminary indications because only two species for each genus are included in our analysis. Based on the examination of morphological characters, Hennemann & Conle (2007) already hypothesized *Lonchodes* to be polyphyletic and referred to two species-groups, provisionally referred by the Authors as "*Lonchodes* (*sensu stricto*)" and "*Lonchodes* (*sensu lato*)". In this regard, the close relationship between *Mnesilochus mindanaense* and two species of *Phenacephorus* (Figure 6) recovered in our analysis is concordant with the morphology-based results by Hennemann & Conle (2007). Furthermore, the position of the genera *Baculofractum* and *Leprocaulinus* within Lonchodini is here confirmed, corroborating results by Bradler *et al*. (2014). The genus *Stheneboea* resulted as the sister group of the remaining Lonchodini but other than the lack of opercular structures in the eggs, there appear to be no definite morphological characters that would support this position. Although molecular data indicated Lonchodini as a paraphyletic assemblage, their

monophyly is morphologically supported by the longitudinally split abdominal tergum X of males, which consists of two movable hemi-terga that are connected by a distinct membrane, and the presence of opercular structures (mostly a stalked capitulum) in the eggs. Moreover, the profemora are distinctly curved and compressed basally, triangular in cross-section with the two dorsal carinae strongly approaching each other and the anteroventral carinae notably lamellate, and the extremities in general are never serrate or dentate.

Within Eurycanthini the speciose genus *Neopromachus* is recovered as polyphyletic in our analysis, which is not surprising given the morphological variability of the genus in its current composition. Based on morphological characters, Günther (1929) attempted a subdivision into several species-groups, which are partly reflected by our molecular results. Therefore, elucidating the most likely phylogeny of the genus as a whole will require a much more extensive sampling. True *Neopromachus* are in our tree represented by *N. wallacei* (the type-species of *Neopromachus*), *N. muticus*, *N. dyselius* and *N. pachynotus*. The second lineage in our tree, here represented by the tree species *N. doreyanus*, *N. obrutus* and *N. scharreri* (Figure 6), results as the sister group to *Erinaceophasma*. Our results moreover support the validity of *Erinaceophasma*, in agreement with the morphology-based results presented by Bradler (2009), who revealed the genus to differ from all other Eurycanthini by the plesiomorphic presence of gonoplacs in the females. The genus *Hyrtacus* is recovered as polyphyletic: in our tree the Australian type-species *H. tuberculatus* is nested within Lonchodini while the second sampled species, *H. procerus* from New Guinea, is the sister taxon to the remaining genera of Eurycanthini. Looking at morphological aspects, the polyphyly of *Hyrtacus* is strongly supported by distinctive characters of the adult terminalia. While *H. tuberculatus* (tribe Lonchodini) has a somewhat elongated and apically pointed anal segment (= abdominal tergum X) and a short subgenital plate in females as well as a rather flattened posteromedially incised anal segment in males, the morphology of the terminalia of *H. procerus* (tribe Eurycanthini) corresponds to other members of that tribe. The strongly elongated anal segment and subgenital plate of females form a conspicuous beak-like secondary ovipositor and the anal segment in males is strongly tectiform and ventro-apically protruded into two long digitiform, downward directed processes. Taxonomic splitting of the genus hence appears inevitable.

The very species-rich subfamily Necrosciinae includes 103 genera and one tribe, with our analysis including 33 species belonging to 20 genera. As for Diapheromerinae and Lonchodinae, this is the largest set of species and genes ever incorporated in a phylogenetic study of the subfamily for displaying its phylogeny. However, as for Lonchodinae, it becomes obvious that elucidating the true phylogeny of Necrosciinae as a whole requires even more extensive sampling. Although the monophyly of this subfamily was repeatedly questioned (Sellick, 1997; Bradler, 2009), our data support it and are in line with a previous study (Bradler *et al.,* 2014). Interestingly, this clade includes some of the species that were previously ascribed to the tribe Neohiraseini, a clade originally

established as a tribe within Lonchodinae by Hennemann & Conle (2008). This tribe comprises the closely related genera *Andropromachus, Neohirasea, Pseudocentema, Qiongphasma* and *Spinohirasea*, and could readily be separated from the remaining Lonchodinae by a set of significant morphological features. Moreover, some morphological characters of the male terminalia, which include a well-developed vomer and a non-split abdominal tergum X (Hennemann & Conle, 2008), already suggested Nehoiraseini should be ascribed to Necrosciinae, which is in line with other molecular analyses (Whiting *et al.,* 2003). Bradler *et al.* (2014) recovered the eponymous Neohirasea as nested within Necrosciinae, which corroborates the results of Kômoto *et al.* (2011), and formally transferred *Neohirasea* resp. Neohiraseini to the subfamily Necrosciinae. The subordinate position of *Neohirasea* among Necrosciinae and possible non-monophyly of Neohiraseini, as originally constituted by Hennemann & Conle (2008), prompted Bradler *et al.* (2014) to reject the validity of the tribe and to synonymize it with Necrosciinae. In line with these results, we recover Neohiraseini (*Andropromachus* + *Neohirasea* + *Spinohirasea*) as a paraphyletic assemblage, with *Paramenexenus laetus* as the sister group. A monophyletic lineage comprising *Oxyartes* and *Phaenopharos* is revealed as the sister taxon of Neohiraseini *sensu* Hennemann & Conle (2008) + *Paramenexenus* in our tree. This relationship, further supported by morphological data, has already been suggested previously and the whole cluster was therefore referred to as the "*Neohirasea* clade" (Buckley *et al.,* 2009; Bradler *et al.,* 2014). Consequently, the strong support of the *Neohirasea* clade from both molecular and morphological data suggests the possible re-establishment of the tribe Neohiraseini, although in a new constitution to comprise also the genera *Paramenexenus, Oxyartes* and *Phaenopharos*. Moreover, from a biogeographical point of view, it is also to be noted that this clade, as presently proposed, is restricted to continental South East Asia, mainly in the Indo-Chinese area, with *Phaenopharos* being the only representative that has one species in Peninsular Malaysia and Singapore, which biogeographically belong to Sundaland (i.e. a southeast extension of the continental shelf of Southeast Asia, comprising Malay Peninsula, Sumatra, Borneo, Java, Madura, Bali and their surrounding smaller islands).

Our molecular data reveal *Neohirasea* as paraphyletic because *Andropromachus* and *Spinohirasea* are found nested within it. On the basis of morphological features there is good support for the two distinct lineages shown by the molecular data of our analysis and also *Paramenexenus* differs from Neohiraseini *sensu* Hennemann & Conle, 2008 (*Andropromachus* + *Neohirasea* + *Spinohirasea*) by several traits. *Paramenexenus* differs significantly in the egg morphology by having a conical central protrusion on the operculum and an elongate micropylar plate that is closed internally (Sellick, 1998: 216, fig. 26d-e). These traits of the egg morphology are shared with *Phaenopharos* (Sellick, 1998: 212, fig. 20k-l). In contrast, eggs of Neohiraseini *sensu* Hennemann & Conle, 2008 (*Andropromachus* + *Neohirasea* + *Spinohirasea*) are more or less spherical in shape, lack any kind of opercular protrusions (operculum at best roundly convex), and have a small, shield-shaped to

almost circular micropylar plate that is open internally with a wide posterior notch and a median line (Sellick, 1998: 216, fig. 25b-c; Hennemann & Conle, 2008: 78, fig. 59a). Surprisingly, eggs of *Oxyartes* combine characters of both these groups, having a knob or cone-shaped central protrusion on the operculum and a rather oval micropylar plate that is open internally with a distinct posterior notch and a short median line (Sellick, 1998: 214, fig. 21e-f). At this point it must be noted that our molecular results are in contrast to the hypothesis of Sellick (1998), who stated that subgroups that contain eggs of strikingly different micropylar plate form, particularly if these differ internally in being closed and open or both possessing or lacking a median line, are likely to be polyphyletic. The anatomy of the insects is very diverse, and it is hard to draw any meaningful consequences for taxonomy without a cladistic analysis of a good set of morphological characters. A discussion of this wide range of anatomy would be beyond the scope of our present study but it highlights the problems that can arise for morphology-based taxonomy alone. The "*Neohirasea*-clade" is a good example for the necessity of an integrative approach that combines morphological and molecular data for achieving a satisfactory phylogeny. The current discrepancies between molecular and morphological data here detected also show that any broader discussion on the validity and delimitation of Neohiraseini and phylogeny of the "*Neohirasea*-clade" deserve a much more comprehensive sampling.

In our analysis, a clade comprising *Rhamphosipyloidea* and *Hemisosibia* resulted as the sister group of Neohiraseini. While the phylogenetic placement of *Rhamphosipyloidea* is well explained also by morphological characters of female genitalia, the position of *Hemisosibia* lacks any clear morphological support. The genus *Rhamphosipyloidea* differs from all other Necrosciinae by the beak-like secondary ovipositor of females that is formed by an elongated anal segment (= abdominal tergum X) and a subgenital plate (Bradler *et al.,* 2014). The presence of such an ovipositor in a member of Necrosciinae is a further support for a multiple independent evolution of this character within Lonchodidae, a family that comprises Lonchodinae and Necrosciinae (Robertson *et al.,* 2018).

Lopaphus is well supported in the present study and is revealed as the sister group of the Necrosciinae clade including Orxineformia and all other Necrosciinae but excluding Neohiraseini + (*Rhamphosipyloidea* + *Hemisosibia*). This genus is the so far only known representative of the Necrosciinae in which males lack a vomer; hence its position is supported by anatomical evidence. A more detailed discussion of the position within Necrosciinae and the genital morphology was presented by Bradler *et al.* (2014). *Lopaphus* eggs are ovoid to almost spherical with an elongate subgenital plate that is closed internally (Sellick, 1998).

We also find good support for Orxineformia, which was introduced as a rank-free subordinate clade of Necrosciinae to comprise closely related genera with specialized terminalia in the adult insects and egg deposition behaviour (Bradler, 2009; Bradler *et al.,* 2014). Our sampling of Orxineformia, which included the four

genera *Centrophasma, Diesbachia, Orxines* and *Pseudodiacantha*, reveal the group as being closely related and perhaps the sister group to the remaining Necrosciinae, excluding Neohiraseini + (*Rhamphosipyloidea* + *Hemisosibia*). The very characteristic eggs of Orxineformia are bullet-shaped with a spear-like and carinate polar end and fringes on the opercular collar. The internal micropylar plate is open with a distinct median line (Sellick, 1998: 211, figs. 19a-i).

The phylogeny of the remaining Necrosciinae is still obscure and a much wider array of sampling is necessary for obtaining meaningful data in order to conduct any broader discussion.

During the last decade many studies, mainly based on molecular data (Whiting et al., 2003; Buckley *et al.,* 2009b; Bradler *et al.,* 2014, 2015; Kômoto *et al.,* 2011; Robertson *et al.,* 2018; Glaw *et al.,* 2019; Simon *et al.,* 2019), have recovered numerous inaccuracies within the traditional classification of the order originally presented by Günther (1953) and broadly accepted by subsequent authors (Bradley & Galil, 1977; Kevan, 1977, 1982). The fact that, with the exception of Whiting *et al.* (2003) and Bradler (2009), all these more recent studies were exclusively or mostly based on molecular data suggest actual limits of morphological approaches in the context of the phasmatodean systematics. As already pointed out, one of the fundamental problems in morphology-based approaches is the extreme convergent evolution of morphological characters, that may occur in distantly related lineages within Euphasmatodea and result in striking morphological similarities. Mimicry and camouflage of parts of host plants and/or components of the natural environment (*e.g.*, mosses, lichens, bark or dry leaf litter on the forest floor) found throughout phasmids is the main reason for this convergent evolution, leading to often strikingly alike anatomical adaptions in similar habitats. Beside the aforementioned case of tree lobsters, examples of morphological convergence are taxa living in moist mossy forests that mimic mosses and lichens (*e.g.*, *Trychopeplus* or *Taraxippus* in the Neotropical region and *Pericentrus* or *Parastheneboea* in the Oriental region) or taxa inhabiting dry savannas that mimic grasses (*e.g.*, *Parabacillus* in the Neotropical region and *Clonaria* in the African region). The mostly winged giant stick insects comprised in the Australian tribe Phasmatini (*e.g.*, *Acrophylla, Anchiale or Ctenomorpha* in Australia) are paralleled by the similarly large African Palophinae (*Bactrododema* and *Dematobactron*) in ecologically similar habitats. Rather small and slender, often colorful flying taxa inhabiting the canopy layer of tropical forests are found in the Oriental Region (*e.g.*, *Necroscia, Orthonecroscia, Paranecroscia* of the Oriophasmata: Necrosciinae) but also in the New World (*e.g.*, *Brizoides, Cesaphasma, Parastratocles* of the Occidophasmata: Pseudophasmatinae). Phenotypically very similar looking, rather inconspicuous grey or brown flying taxa include *Sipyloidea* (Necrosciinae) in the Oriental region and *Agrostia* (Pseudophasmatinae) in the New World. The Oriental giant stick insects of the tribe Pharnaciini (*e.g.*, *Pharnacia*, *Phobaeticus*, *Tirachoidea*; Oriophasmata: Clitumninae) have exclusively wingless females but either wingless, brachypterous or winged males, the same

applying to members of the Cladomorphinae: Cladomorphini (*e.g.*, *Cladomorphus, Phanocloidea, Phanocles*) in the Neotropical Region. Examples for strikingly similar trunk-dwellers that mimic bark are *Prisopus* (Occidophasmata: Prisopodinae) in the Neotropical region and *Neoclides* in the Oriental region (Oriophasmata: Necrosciinae).

A good example for the limits of morphology based systematic is the evolution of wings within the Phasmatodea, whose true pathways could not have been fully detected by examination and cladistic analysis of morphological characters alone. A secondary derivation (or re-evolution) of wings was originally recovered based on the analysis of molecular data by Whiting *et al.* (2003) that suggested the ancestral condition to be wingless. Recently, this hypothesis was also confirmed through extensive comparative analyses along the entire order by Forni *et al.* (2021). These results are in sharp contrast to the traditional classification of the order (Günther, 1953; Bradley & Galil, 1977), which was strictly based on the concept that the presence of wings was the ancestral state. As a consequence, this resulted in the false hypothesis that within the Phasmatodea wings have been lost secondarily in many independent occasions and that this was a one-way evolution with a regain of wings being impossible. The re-evolution of wings detected by Whiting *et al.* (2013) however has been supported by studies of morphological characters (e.g., Hennemann *et al.,* 2016a) and has since played an important role in revising the phylogeny of the Phasmatodea with fundamental changes.

These phenotypical similarities of distantly related evolutionary lineages within the Phasmatodea represent the main problem for morphology-based systematics. Yet, several of the more recent conclusions on the phylogeny of certain subgroups based on morphological data (Hennemann & Conle, 2008; Hennemann *et al.,* 2016b) are not only corroborated by our analysis but have also found support by other molecular based studies (Whiting *et al.,* 2013; Robertson *et al.,* 2018; Simon *et al.,* 2019). For instance, our result widely supports the relationships within the traditional subfamily Cladomorphinae assumed by Hennemann *et al.* (2016b), who already detected a monophyletic group that comprised Haplopodini, Hersperophasmatini and Pterinoxylini, here named as Haplopodinae *subfam. nov*. Moreover, we can confirm also the fairly separated placement of Cranidiini, which resulted as the sister group to all remaining Cladomorphinae *sensu nov*. Simon *et al.* (2019) have recently inferred the monophyly of the subfamily Clitumninae as established by Hennemann & Conle (2008) to be comprised of Clitumnini, Medaurini and Pharnaciini, which is in contrast to previous results (Buckley *et al.,* 2009a, 2009b; Bradler *et al.,* 2014, 2015; Robertson *et al.,* 2018).

Summary of changes to taxonomic classification

The results of the present study as well as morphological evidence outlined above and strictly considering taxonomic rules specified by the code of the International Commission on Zoological Nomenclature (ICZN, 1999)

make several nomenclatural and systematic changes to the classification of Phasmatodea warranted and justified. We introduce Haplopodini *subfam. nov.* to accomodate the taxa previously regarded as Cladomorphinae by Robertson *et al.* (2018), namely the three tribes Haplopodini, Hesperophasmatini and Pterinoxylini. We propose a new concept for the subfamily Cladomorphinae *sensu nov.*, to comprise the tribes Cladomorphini *sensu nov.* (including the type-genus *Cladomorphus*), Diapheromerini *sensu nov.* and Cranidiini. Diapheromerinae is shown to be a junior synonym of Cladomorphinae because Cladomorphinae *sensu nov.* includes Diapheromera, the type-genus of Diapheromerinae (*syn. nov.*). *Phantasca* and *Trychopeplus* are assigned to Cladomorphini. On the evidence of morphological characters two species are transferred to other genera: *Bacteria ploiaria* is transferred to *Phanocles* (Cladomorphinae: Cladomorphini), as *Phanocles ploiaria* (*comb. nov*.) and *Lamponius bocki* is transferred to *Hesperophasma* (Haplopdinae *subfam. nov.*: Hesperophasmatini), as *Hesperophasma bocki* (*comb. nov.*). *Libethra strigiventris* is shown to belong to the tribe Oreophoetini (*sensu* Robertson et al., 2018). The genus *Papuacocelus* is confirmed as a member of the tribe Eurycanthini (subfamily Lonchodinae), as it exhibits the morphological key features of that tribe.

Conclusions

In this paper we revise past hypothesis and present novel ones, yet we also recognize uncertain instances in Euphasmatodea systematics. Traditional support metrics can provide missleading confidence to some phylogenetic hypotheses - such as the establishment of Lonchodidae - which are challenged by our results. Phylogenetic relationships among main Euphasmatodea lineages are found to be inconsistent and strongly impacted by model specification and data processing, due to the scarce phylogenetic signal associated to Euphasmatodea evolutionary radiation. Contrary to backbone relationships, the subdivisions recovered at lower taxonomic levels show a stronger support, which allowed us to highlight morphological features that appear to have undergone convergent evolution across distant clades. These can be highly confounding in a phylogenetic framework, largely explaining the conflict between traditional taxonomy and molecular phylogenetics.

In Phasmatodea - as in many other insect clades - molecular and morphological approaches suffer from limitations due to a combination of biological and technical phenomena. Morphological characters require a detailed understanding to address their weight (Borkent, 2018), while molecular analyses shouldn't overlook possible phylogenetic biases. Overall, this work highlights once more how the combination of the two approaches is the most promising way to achieve steady, coherent and meaningfull systematic frameworks and taxonomical classifications.

References

Alda, F., Tagliacollo, V.A., Bernt, M.J., Waltz, B.T., Ludt, W.B., Faircloth, B.C., Alfaro, M.E., Albert, J.S. & Chakrabarty, P. (2019) Resolving deep nodes in an ancient radiation of neotropical fishes in the presence of conflicting signals from incomplete lineage sorting. *Systematic Biology*, **68**, 573-593.

Ax, P. (1984) Das Phylogenetische System. Systematisierung der lebenden Natur aufgrund ihrer Phylogenese. Gustav Fischer Verlag, Stuttgart/New York.

Bank, S., Buckley, T.R., Büscher, T.H., Bresseel, J., Constant, J., de Haan, M., Dittmar, D., Dräger, H., Kahar, R.S., Kang, A., Kneubühler, B., Langton-Myers, S.S. & Bradler, S. (2021) Reconstructing the nonadaptive radiation of an ancient lineage of ground-dwelling stick insects (Phasmatodea: Heteropterygidae). *Systematic Entomology*, in press https://doi.org/10.1111/syen.12472

Bogdanowicz, D. & Giaro, K. (2017) Comparing phylogenetic trees by matching nodes using the transfer distance between partitions. *Journal of Computational Biology*, **24**, 422–435.

Bogdanowicz, D. & Giaro, K. (2012) Matching split distance for unrooted binary phylogenetic trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **9**, 150–160.

Bogdanowicz, D., Giaro, K. & Wróbel, B. (2012) TreeCmp: comparison of trees in polynomial time. *Evolutionary Bioinformatics Online*, **8**, 475–487.

Borkent, A. (2018) The state of phylogenetic analysis: narrow visions and simple answers – examples from the Diptera (flies). *Zootaxa*, **4374**, 107-143.

Bradler, S. (2009) Phylogenie der Stab- und Gespenstschrecken (Phasmatodea). *Species Phylogeny and Evolution*, **2**, 3-139.

Bradler, S. & Buckley, T.R. (2018) Biodiversity of Phasmatodea, *Insect Biodiversity: Science and Society, Vol. II* (ed. by R.G. Foottit and P.H. Adler), pp. 281-313. Wiley-Blackwell Hoboken, N.J, USA.

Bradler, S., Cliquennois, N. & Buckley, T.R. (2015) Single origin of the Mascarene stick insects: ancient radiation on sunken islands? *BMC Evolutionary Biology*, **15**, 196.

Bradler, S., Robertson, J.A. & Whiting, M.F. (2014) A molecular phylogeny of Phasmatodea with emphasis on Necrosciinae, the most species-rich subfamily of stick insects. *Systematic Entomology*, **39**, 205–222.

Bradley, J.C. & Galil, B.S. (1977) The taxonomic arrangement of the Phasmatodea with keys to the subfamilies and tribes, *Proceedings of the Entomological Society of Washington*, **79**, 176-208.

Bryant, D., Tsang, J., Kearney, P. & Li, M. (2000) Computing the quartet distance between evolutionary trees, *Proceedings of ACM-SIAM Symposium on Discrete Algorithms*, pp. 285-286. ACMPress, New York, USA..

Buckley, T.R., Attanayake, D. & Bradler, S. (2009a) Extreme convergence in stick insect evolution: phylogenetic placement of the Lord Howe Island tree lobster. *Proceedings of the Royal Society B*, **276**, 1055-1062.

Buckley, T.R., Attanayake, D., Nylander, J.A.A. & Bradler, S. (2009b) The phylogenetic placement and biogeographical origins of the New Zealand stick insects (Phasmatodea). *Systematic Entomology*, **35**, 207-225.

Büscher, T.H., Buckley, T.R., Grohmann, C., Gorb, S.N., & Bardler, S. (2018) The evolution of tarsal adhesive microstructures in stick and leaf insects (Phasmatodea). *Frontiers in in Ecology and Evolution*, **6**, 69.

Chernomor, O., von Haeseler, A., & Minh, B.Q. (2016) Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, **65**, 997–1008.

Conle, O.V., Hennemann, F.H. and Valero, P. (2020) Studies on Neotropical Phasmatodea XXII: Two new species of *Taraxippus* (Phasmatodea: Cladomorphinae: Hesperophasmatini) and the first record of the genus from Central America. *Journal of Orthoptera Research*, **29**, 101-114.

Dillenberger, M.S. & Kadereit, J.W. (2017) Simultaneous speciation in the European high mountain flowering plant genus *Facchinia* (*Minuartia sl*, Caryophyllaceae) revealed by genotyping-by-sequencing. *Molecular Phylogenetics and Evolution*, **112**, 23-35.

Bazinet, A.L., Mitter, K.T., Davis, D.R., Van Nieukerken, E.J., Cummings, M.P. & Mitter, C. (2017), Phylotranscriptomics resolves ancient divergences in the Lepidoptera. *Systematic Entomology*, **42**, 305-316.

Engel, M.S., Wang, B. & Alqarni, A.S. (2016) A thorny 'anareolate' stick-insect (Phasmatidae s.l.) in Upper Cretaceous amber from Myanmar, with remarks on diversification times among Phasmatodea. *Cretaceous Research*, **63**, 45-53.

Forni, G., Plazzi, F., Cussigh, A., Conle, O., Hennemann, F., Luchetti, A., & Mantovani, B. (2021) Phylomitogenomics provide new perspectives on the Euphasmatodea radiation (Insecta: Phasmatodea). *Molecular Phylogenetics and Evolution*, **155**, 106983.

Forni, G., Martelossi, J., Valero, P., Hennemann, F.H., Conle, O., Luchetti, A. & Mantovani, B. (2020) Macroevolutionary analyses provide new evidences of phasmids wings evolution as a reversible process. *bioRxiv* https://doi.org/10.1101/2020.10.14.336354

Gelman, A. & Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.

Geuten, K., Smets, E., Schols, P., Yuan, Y.M., Janssens, S., Küpfer, P. & Pyck, N. (2004) Conflicting phylogenies of balsaminoid families and the polytomy in Ericales: combining data in a Bayesian framework. *Molecular Phylogenetics and Evolution*, **31**, 711-729.

Giarla, T.C. & Esselstyn, J.A. (2015) The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, **64,** 727-740.

Glaw, F., Hawlitschek, O., Dunz, A., Goldberg, J. & Bradler, S. (2019) When giant stick insects play with colors: molecular phylogeny of the Achriopterini and description of two new splendid species (Phasmatodea: Achrioptera) from Madagascar. *Frontiers in Ecology and Evolution*, **7**, 105.

Goldberg, J., Bresseel, J., Constant, J., Kneubühler, B., Leubner, F., Michalik, P. & Bradler, S. (2015) Extreme convergence in egg-laying strategy across insect orders. *Scientific Reports*, **5**, 7825.

Günther, K. (1929) Die Phasmoïden der Deutschen Kaiserin Augusta-Fluss-Expedition 1912/13. Ein Beitrag zur Kenntnis der Phasmoïdenfaune Neuguineas. *Mitteilungen aus dem Zoologischen Museum, Berlin*, **14**, 60-747.

Günther, K. (1932) Columbianische Phasmoiden aus der Sammlung des Rev. Apolinar Maria. Mit einer Übersicht über das Genus *Libethra* Stål. *Mitteilungen aus dem Zoologischen Museum, Berlin*, **18**, 226-261.

Günther, K. (1953) Über die taxonomische Gliederung und geographische Verbreitung der Insektenordnung der Phasmatodea. *Beiträge zur Entomologie*, **3**, 541-563.

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**, 307-321.

Heath, T.A., Hedtke, S.M. & Hillis, D.M. (2008) Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, **46**, 239-257.

Hennemann, F.H. & Conle, O.V. (1997) Intraspezifische Variabilität bei *Lonchodes femoratus* (Stoll, 1787) nebst einigen Bemerkungen zu ihrer Synonymie (Phasmatodea: Phasmatidae: Lonchodinae). *Entomologische Zeitschrift*, **107**, 30-37.

Hennemann, F.H. & Conle, O.V. (2006) *Papuacocelus papuanus* n. gen., n. sp. – a new Eurycanthinae from Papua New Guinea, with notes on the genus *Dryococelus* Gurney, 1947 and description of the egg (Phasmatodea: Phasmatidae: Eurycanthinae). *Zootaxa*, **1375**, 31-49.

Hennemann, F.H. & Conle, O.V. (2007) Studies on Philippine Lonchodinae, with the descriptions of two new genera and eleven new species (Phasmatodea: Phasmatidae: Lonchodinae). *Mitteilungen der Münchner Entomologischen Gesellschaft*, **97**, 3-88.

Hennemann, F.H. & Conle, O.V. (2008) Revision of Oriental Phasmatodea: the tribe Pharnaciini Günther, 1953, including the description of the world's longest insect, and a survey of the family Phasmatidae Gray, 1835 with keys to the subfamilies and tribes (Phasmatode: “Anareolatae”: Phasmatidae). *Zootaxa*, **1906**, 1-316.

Hennemann, F.H. & Conle, O.V. (2010) Studies on Neotropical Phasmatodea X: Redescriptions of *Aplopocranidium* Zompro, 2004 and Jeremia Redtenbacher, 1908, with a survey of the tribe Cladomorphini Brunner v. wattenwyl, 1893 and keys to the genera (Insecta, Phasmatodea: "Anareolatae": Cladomorphinae). *Journal of Orthoptera Research*, **19**, 101-113.

Hennemann, F.H. & Conle, O.V. (2012) Studies on Neotropical Phasmatodea XIV: Revisions of the Central American genera *Hypocyrtus* Redtenbacher, 1908 and *Rhynhacris* Redtenbacher, 1908 (Phasmatodea: "Anareolatae": Xerosomatinae: Hesperophasmatini). *Journal of Orthoptera Research*, **21**, 65-89.

Hennemann, F.H., Conle, O.V., Brock, P.D. & Seow-Choen, F. (2016a) Revision of the Oriental subfamily Heteropteryginae Kirby, 1896, with a re-arrangement of the family Heteropterygidae and the descriptions of five new species of Haaniella Kirby, 1904 (Phasmatodea: Areolatae: Heteropterygidae). *Zootaxa*, **4159**, 1-219.

Hennemann, F.H., Conle, O.V. & Perez-Gelabert, D. (2016b) Studies on Neotropical Phasmatodea XVI: revision of Haplopodini Günther, 1953 (rev. stat.), with notes on the subfamily Cladomorphinae Bradley & Galil, 1977 and the descriptions of a new tribe, four new genera and nine new species (Phasmatodea: "Anareolatae": Phasmatidae: Cladomorphinae). *Zootaxa*, **4128**, 1-211.

Hennemann, F.H., Conle, O.V., Bellanger, Y., Lelong, P. & Jourdan, T. (2018) Studies on Neotropical Phasmatodea XVII: revision of *Phantasca* Redtenbacher, 1906, with the descriptions of six new species (Phasmatodea: Diapheromeridae: Diapheromerinae). *European Journal of Taxonomy*, **435**, 1-62.

Hennemann, F.H., Conle, O.V., Perez-Gelabert, D. & Valero, P. (2020) Studies on Neotropical Phasmatodea XXI: *Sigaruphasma*, a new genus of Hesperophasmatini Bradley & Galil, 1977, from Hispaniola with the descriptions of two new species (Phasmatodea: Cladomorphinae). *Novitates Caribaea*, **16**, 58-79.

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**, 518–522.

ICZN - International Commission on Zoological Nomenclature (1999). International code of zoological nomenclature. The International Trust for Zoological Nomenclature 1999. The Natural History Museum, London, UK.

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**, 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**, 772–780.

Kevan, D.K.M. (1977) The higher classification of orthopteroid insects: a general view. *Lyman Entomological Museum Research Laboratory Memoirs*, **4**, 1-31.

Kevan, D.K. McE. (1982) Phasmatoptera, *Synopsis and Classification of Living Organisms, Vol. 2* (ed. by S.F. Parker), pp 379-383. McGraw-Hill, New York, NY, USA.

King, N. & Rokas, A. (2017) Embracing uncertainty in reconstructing early animal evolution. *Current Biology*, **27**, R1081-R1088.

Kishino, H. & Hasegawa, M. (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution*, **29**, 170–179.

Kishino, H., Miyata, T. & Hasegawa, M. (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution,* **31**, 151–160.

Kômoto, N., Yukuhiro, K., Ueda, K. & Tomita, S. (2011) Exploring the molecular phylogeny of phasmids with whole mitochondrial genome sequences. *Molecular Phylogenetics and Evolution*, **58**, 43-52.

Kück, P., Meid, S.A., Groß, C., Wägele, J.W. & Misof, B. (2014) AliGROOVE – visualization of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support. *BMC Bioinformatics*, **15**, 294.

Kutschera, V.E., Bidon, T., Hailer, F., Rodi, J.L., Fain, S.R. & Janke, A. (2014) Bears in a forest of gene trees: phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Molecular Biology and Evolution*, **31**, 2004-2017.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) PartitionFinder2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.

Lewis, P.O., Holder, M.T. & Holsinger, K.E. (2005) Polytomies and Bayesian phylogenetic inference. *Systematic Biology*, **54**, 241-253.

López-López, A. & Vogler, A.P. (2017) The mitogenome phylogeny of *Adephaga* (Coleoptera). *Molecular Phylogenetics and Evolution*, **114**, 166-174.

Meyer, B.S., Matschiner, M. & Salzburger, W. (2015) A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. *Molecular Phylogenetics and Evolution,* **83**, 56-71.

Miller, M.A., Pfeiffer, W., Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees, *2010 Gateway Computing Environments Workshop (GCE)*, pp. 1–8. New Orleans, LA, USA.

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**, 1530-1534.

Minh, B.Q., Nguyen, M.A.T. & von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, **30**, 1188–1195.

Morando, M., Olave, M., Avila, L.J., Sites Jr, J.W. & Leaché, A.D. (2020) Phylogenomic data resolve higher-level relationships within South American *Liolaemus* lizards. *Molecular Phylogenetics and Evolution*, **147**, 106781.

Morrone, J.J. (2006) Biogeographic areas and transition zones of Latin America and the Caribbean Islands based on panbiogeographic and cladistic analyses of the entomofauna. *Annual Review of Entomology*, **51**, 467-494.

Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., Müller, W.E., Nickel, M., Schierwater, B. & Vacelet, J. (2013) Deep metazoan phylogeny: when different genes tell different stories. *Molecular Phylogenetics and Evolution*, **67**, 223-233.

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**, 268–274.

Paradis, E. & Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G. & Baurain, D. (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, 9: e1000602.

Plazzi, F., Ricci, A. & Passamonti, M. (2011) The mitochondrial genome of *Bacillus* stick insects (Phasmatodea) and the phylogeny of orthopteroid insects. *Molecular Phylogenetics and Evolution*, **58**, 304–316.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Robertson, J.A., Bradler, S. & Whiting, M.F. (2018) Evolution of oviposition techniques in stick and leaf insects (Phasmatodea). *Frontiers in Ecology and Evolution*, **6**, 219.

Robinson, D.F. & Foulds, L.R. (1981) Comparison of phylogenetic trees. *Mathematical Biosciences*, **53**, 131–147.

Romiguier, J., Cameron, S.A., Woodard, S.H., Fischman, B.J., Keller, L., & Praz, C.J. (2016) Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. *Molecular Biology and Evolution*, **33**, 670-678.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Rokas, A. & Carroll, S.B. (2006) Bushes in the tree of life. *PLoS biology*, **4**, e352.

Roycroft, E.J., Moussalli, A. & Rowe, K.C. (2019) Phylogenomics uncovers confidence and conflict in the rapid radiation of Australo-Papuan rodents. *Systematic Biology*, **69**, 431-444.

Sellick, J.T.C. (1997) The range of egg capsule morphology within the Phasmatodea and its relevance to the taxonomy of the order. *Italian Journal of Zoology*, **64**, 97-104.

Sellick, J.T.C. (1998) The micropylar plate of the eggs of Phasmida, with a survey of the range of plate form within theorder. *Systematic Entomology*, **23**, 203-228.

Shimodaira, H. (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, **51**, 492–508.

Shimodaira, H. & Hasegawa M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inferences. *Molecular Biology and Evolution*, **16**, 1114-1116.

Silberfeld, T., Leigh, J.W., Verbruggen, H., Cruaud, C., De Reviers, B. & Rousseau, F. (2010) A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): investigating the evolutionary nature of the “brown algal crown radiation”. *Molecular Phylogenetics and Evolution*, **56**, 659-674.

Simmons, M.P. & Norton, A.P. (2014) Divergent maximum-likelihood-branch-support values for polytomies. *Molecular Phylogenetics and Evolution*, **73**, 87-96.

Simon, S., Letsch, H., Bank, S., Buckley, T.R., Donath, A., Liu, S., Machida, R., Meusemann, K., Misof, B., Podsiadlowski, L., Zhou, X., Wipfler, B. & Bradler, S. (2019) Old World and New World Phasmatodea: phylogenomics resolve the evolutionary history of stick and leaf insects. *Frontiers in Ecology and Evolution*, **7**, 345.

Smith, S.A. & Dunn, C.W. (2008) Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. *Bioinformatics,* **24**: 715-716]

Stanley, E.L., Bauer, A.M., Jackman, T.R., Branch, W.R. & Mouton, P.L.F.N. (2011) Between a rock and a hard polytomy: rapid radiation in the rupicolous girdled lizards (Squamata: Cordylidae). *Molecular Phylogenetics and Evolution*, **58**, 53-70.

Strimmer, K. & von Haeseler, A. (1997) Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences USA*, 94, 6815–6819.

Strimmer, K. & Rambaut, A. (2002) Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society B: Biological Sciences*, **269**, 137–142.

Struck, T.H., Nesnidal, M.P., Purschke, G. & Halanych, K.M. (2008) Detecting possibly saturated positions in 18S and 28S sequences and their influence on phylogenetic reconstruction of Annelida (Lophotrochozoa). *Molecular Phylogenetics and Evolution*, **48**, 628–645.

Stucky, B.J. (2012) SeqTrace: a graphical tool for rapidly processing DNA sequencing chromatograms. *Journal of Biomolecular Techniques*, **23**, 90–93.

Suh, A. (2016) The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zoologica Scripta*, **45**, 50-62.

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

Tarver, J.E., Dos Reis, M., Mirarab, S., Moran, R.J., Parker, S., O'Reilly, J.E., King, B.L., O'Connell, M.J., Asher, R.J., Warnow, T. & Peterson, K.J. (2016) The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biology and Evolution*, **8**, 330-344.

Tilgner, E.H. (2002) Systematics of Phasmida. Dissertation. University of Georgia, Athens, GA, USA.

Timmermans, M.J., Barton, C., Haran, J., Ahrens, D., Culverwell, C.L., Ollikainen, A., Dodsworth, S., Foster, P.G., Bocak, L. & Vogler, A.P. (2016) Family-level sampling of mitochondrial genomes in Coleoptera: compositional heterogeneity and phylogenetics. *Genome Biology and Evolution*, **8**, 161-175.

Tomita, S., Yukuhiro, K. & Kômoto, N. (2011) The mitochondrial genome of a stick insect *Extatosoma tiaratum* (Phasmatodea) and the phylogeny of polyneopteran insects. *Journal of Insect Biotechnology and Sericology*, **80**, 3_079-3_088.

Veeramah, K.R., Woerner, A.E., Johnstone, L., Gut, I., Gut, M., Marques-Bonet, T., Carbone, L., Wall, J.D. & Hammer, M.F. (2015) Examining phylogenetic relationships among gibbon genera using whole genome sequence data using an approximate bayesian computation approach. *Genetics*, **200**, 295-308.

Whitfield, J.B. & Kjer, K.M. (2008) Ancient rapid radiations of insects: challenges for phylogenetic analysis. *Annual Review of Entomology*, **53**, 449-472.

Whiting, M.F., Bradler, S. & Maxwell, T. (2003) Loss and recovery of wings in stick insects. *Nature*, **421**, 264–267.

Willmann, R. (1989) Palaeontology and the systematization of natural taxa. *Abhandlungen des Naturwissenschaftlichen Vereins in Hamburg (NF)*, **28**, 267-291.

Xia, X. (2018) DAMBE7: new and improved tools for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, **35**, 1550–1552.

Xia, X., Xie, Z., Salemi, M., Chen, L. & Wang, Y. (2003) An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, **26**, 1–7.

Yong, S. (2017) A remarkable addition to the Caribbean phasmid fauna: A new tribe, genus and species of desert stick-insect from Cuba (Phasmida: Phasmatidae: Cladomorphinae). *Boletin de la Sociedad Entomológica Aragonesa (S.E.A.),* **60**, 227–234.

Yuan, M.L., Zhang, Q.L., Zhang, L., Guo, Z.L., Liu, Y.J., Shen, Y.Y. & Shao, R. (2016) High-level phylogeny of the Coleoptera inferred with mitochondrial genome sequences. *Molecular Phylogenetics and Evolution*, **104**, 99-111.

Zhang, J. & Madden, T.L. (1997) PowerBLAST: a new network blast application for interactive or automated sequence analysis and annotation. *Genome Research*, **7**, 649–656.

Zhang, R., Wang, Y.H., Jin, J.J., Stull, G.W., Bruneau, A., Cardoso, D., de Queiroz, L.P., Moore, M.J., Zhang, S.D., Chen, S.Y. & Wang, J. (2020) Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Systematic Biology*, **69**, 613–622.

Zompro, O. (2001a) A generic revision of the insect order Phasmatodea: The New World genera of the stick insect subfamily Diapheromeridae: Diapheromerinae = Heteronemiidae: Heteronemiinae *sensu* Bradley & Galil, 1977. *Revue Suisse de Zoologie*, **108**, 189-255.

Zompro, O. (2001b) A review of Eurycanthinae: Eurycanthini, with a key to genera, notes on the subfamily and designation of type species. *Phasmid Studies*, **10**, 19-23.

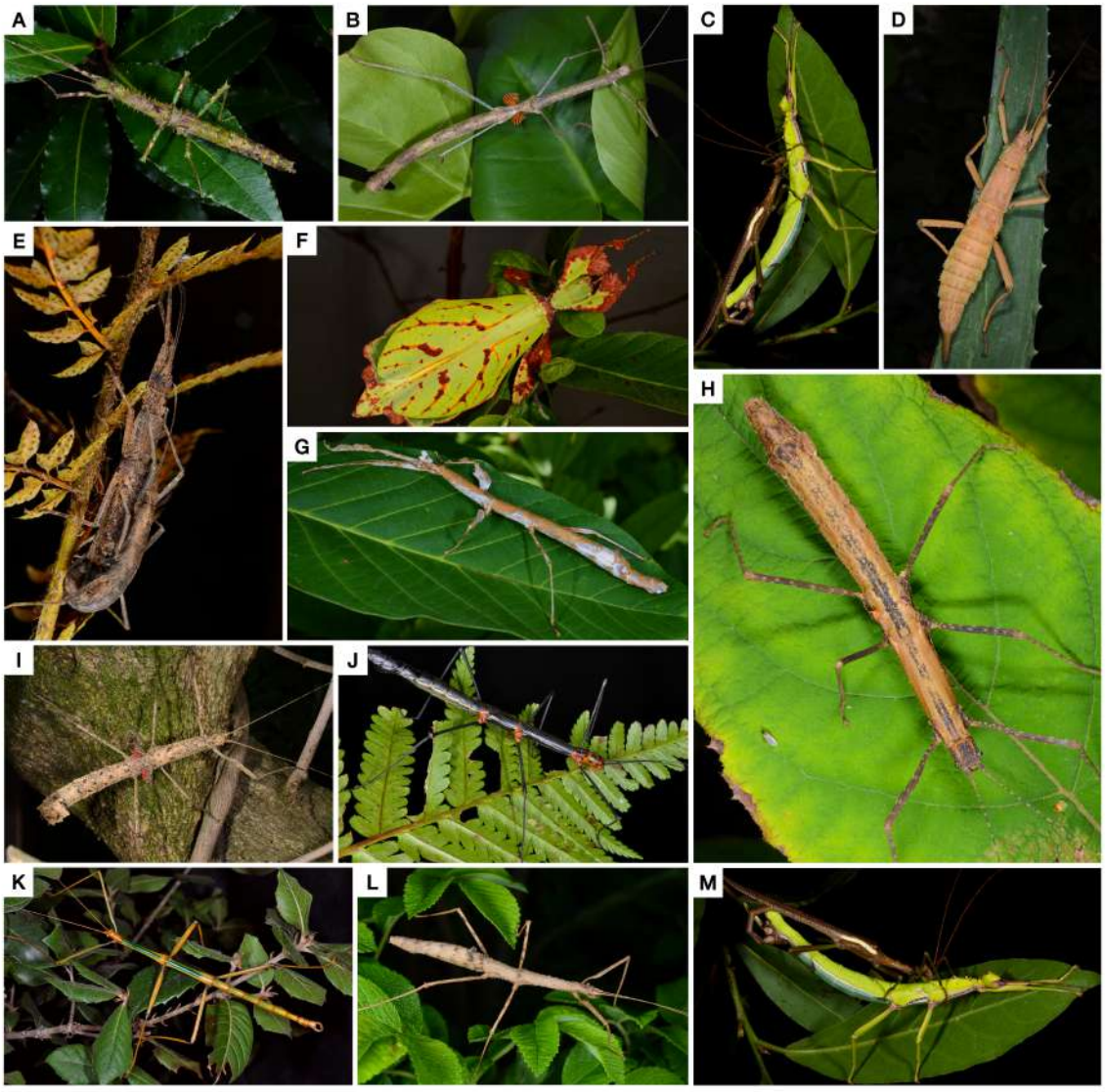


Fig. 1 – Examples of species of Euphasmatodea analyzed in this study: (A) *Centrophasma hadrillum spinocephalum* female; (B) *Phaenopharos struthioneus* female; (C) *Diapherodes martinicensis* couple; (D) *Eurycantha insularis* female; (E) *Neooxyartes sp* couple; (F) *Phyllium mabantai* female; (G) *Phenacephorus latifemur* female; (H) *Neooxyartes zomproi* female; (I) *Oxyartes lamellatus* female; (J) *Oreophoetes topoense* female; (K) *Megaphasma dentricus* male; (L) *Libethra rabdota* female; (M) *Diapherodes martinicensis* couple.

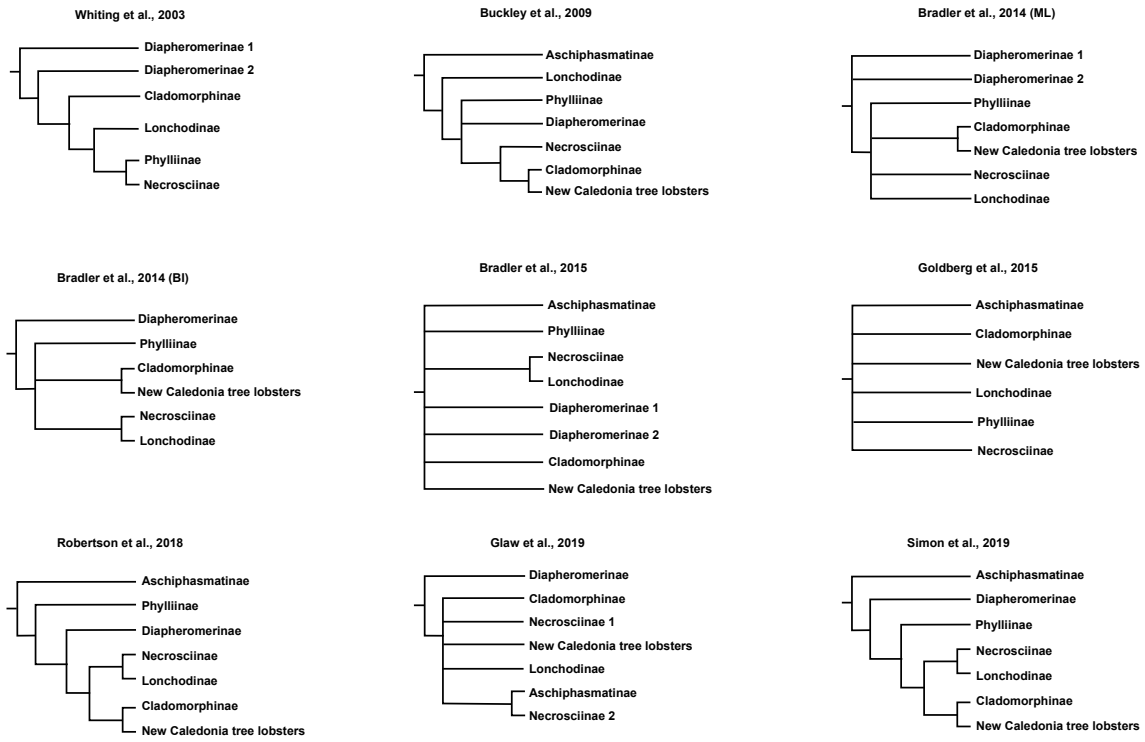


Fig. 2 - Phylogenetic hypotheses from major papers regarding the relationships among Euphasmatodea clades analyzed in this paper.

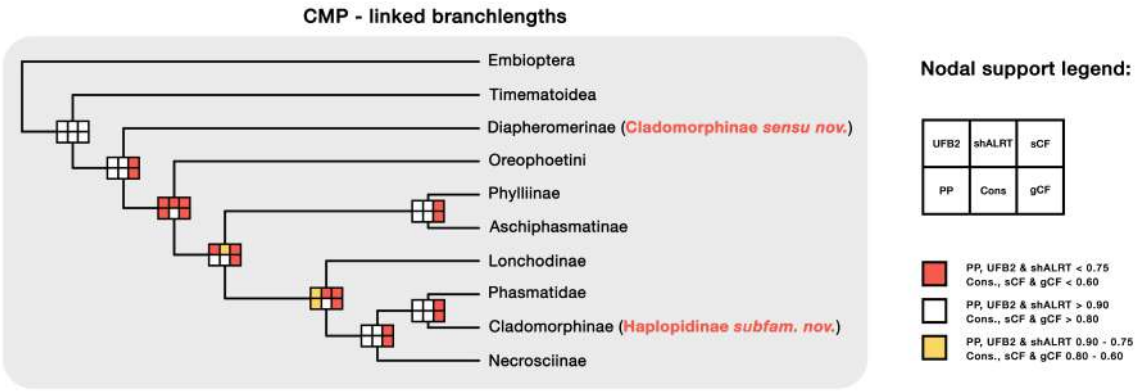


Fig. 3 - Schematic drawing of phylogenetic relationships among the Euphasmatodea clades considered, as inferred in the CMP - linked branchlengts analysis. Names reported in red are after taxonomic revision (see Discussion). Diagrams at nodes represent nodal supports: different metrics are represented in different positions within the diagram, with range of values indicated by different colors, as reported in the legend. (BS = IQ-TREE ultrafast bootstrap; PP = posterior probability; sh-ALRT = Shimodaira-Hasegawa approximate likelihood ratio test; Cons. = consensus among different IQ-TREE runs; sCF and gCF = sites and genes concordance factors, respectively).

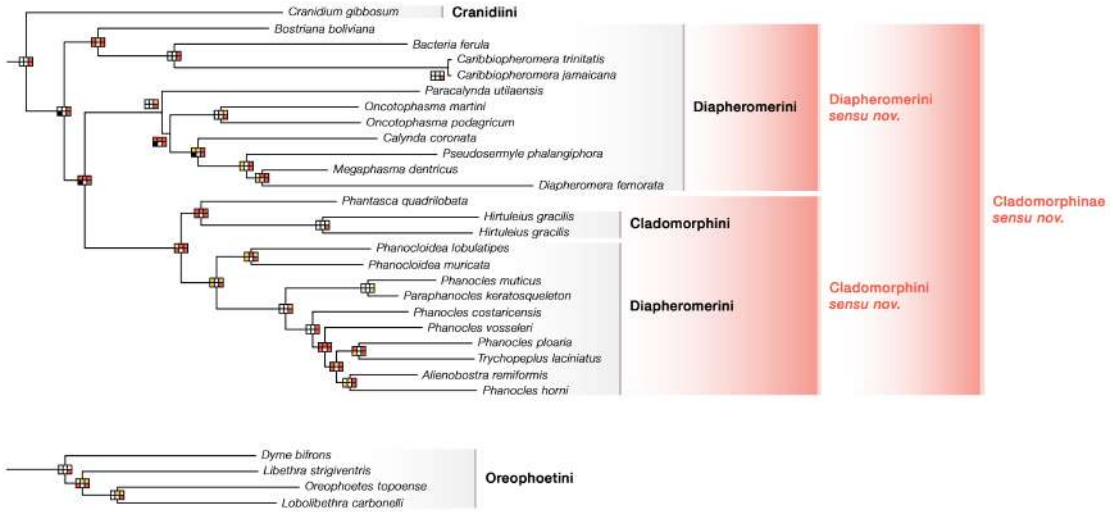


Fig. 4 - Subtree representing the phylogeny of Cladomorphinae *sensu nov*, along with Cranidinii and Oreopoethini. Nodal support diagrams as per legend in Figure 3. In red are reported clade names after taxonomic revision (see Discussion).

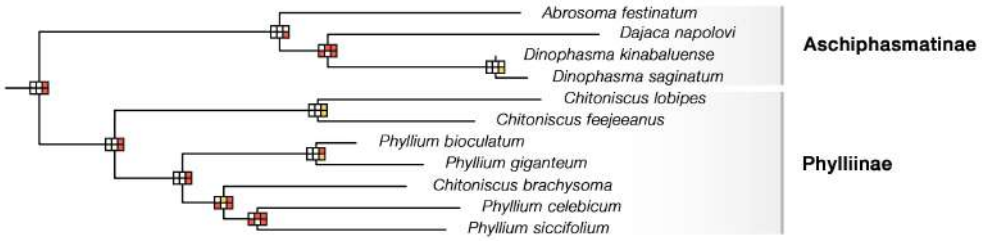


Fig. 5 - Subtree representing the phylogeny of Aschiphasmatinae and Phylliinae. Nodal support diagrams as per legend in Figure 3.

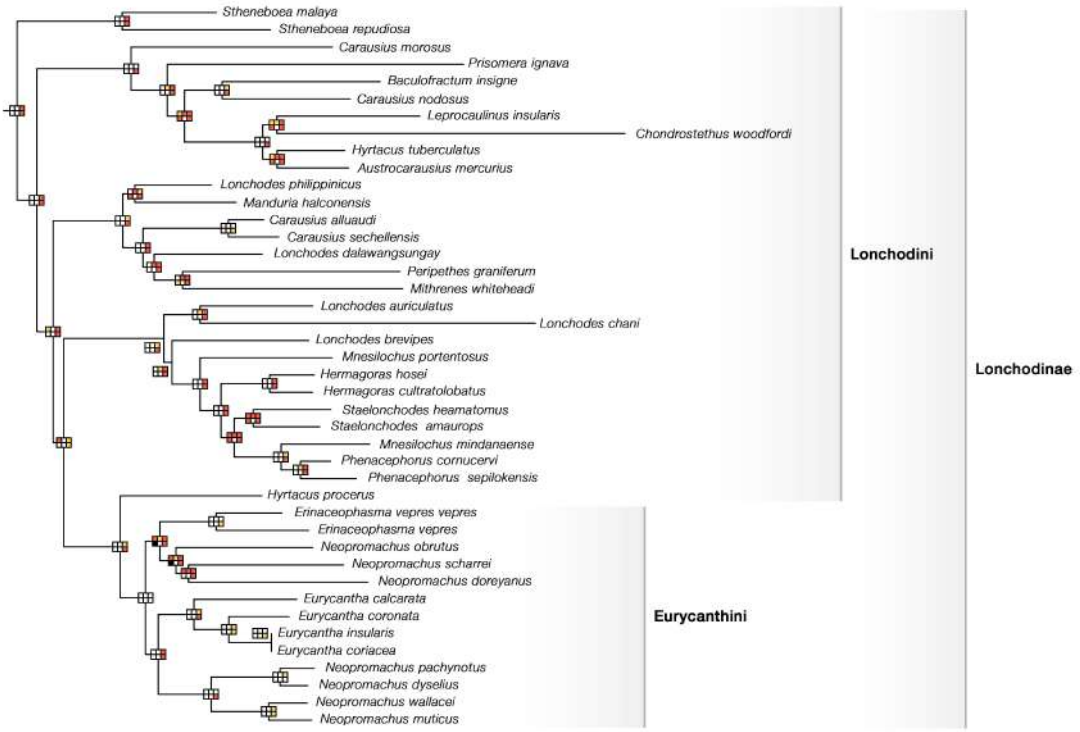


Fig. 6 - Subtree representing the phylogeny of Lonchodinae. Nodal support diagrams as per legend in Figure 3.

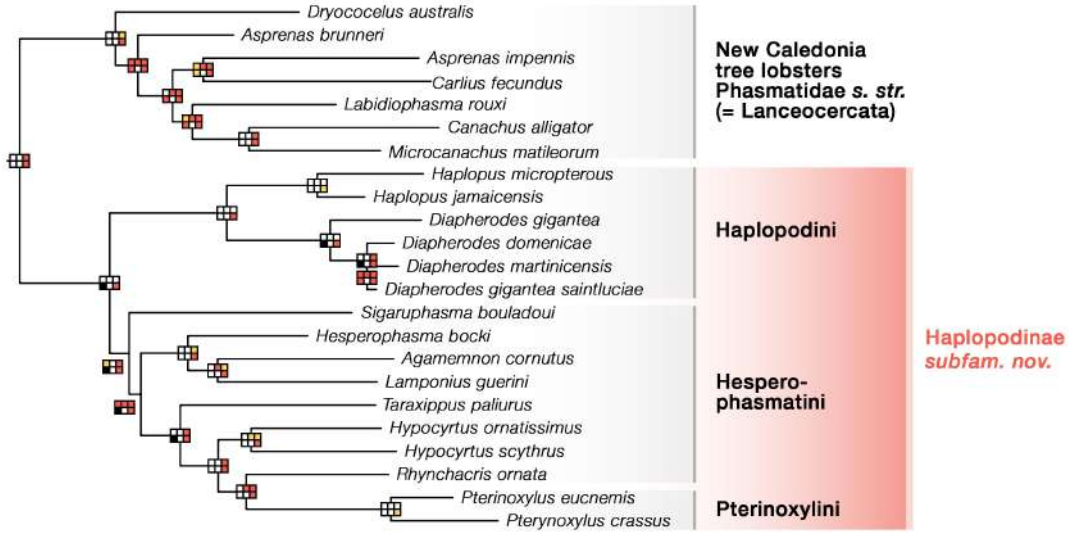


Fig. 7 - Subtree representing the phylogeny of Haplopodinae *subfam. nov.,* and New Caledonia tree lobsters (Phasmatidae s. str = Lanceocercata). Nodal support diagrams as per legend in Figure 3. In red are reported clade names after taxonomic revision (see Discussion).

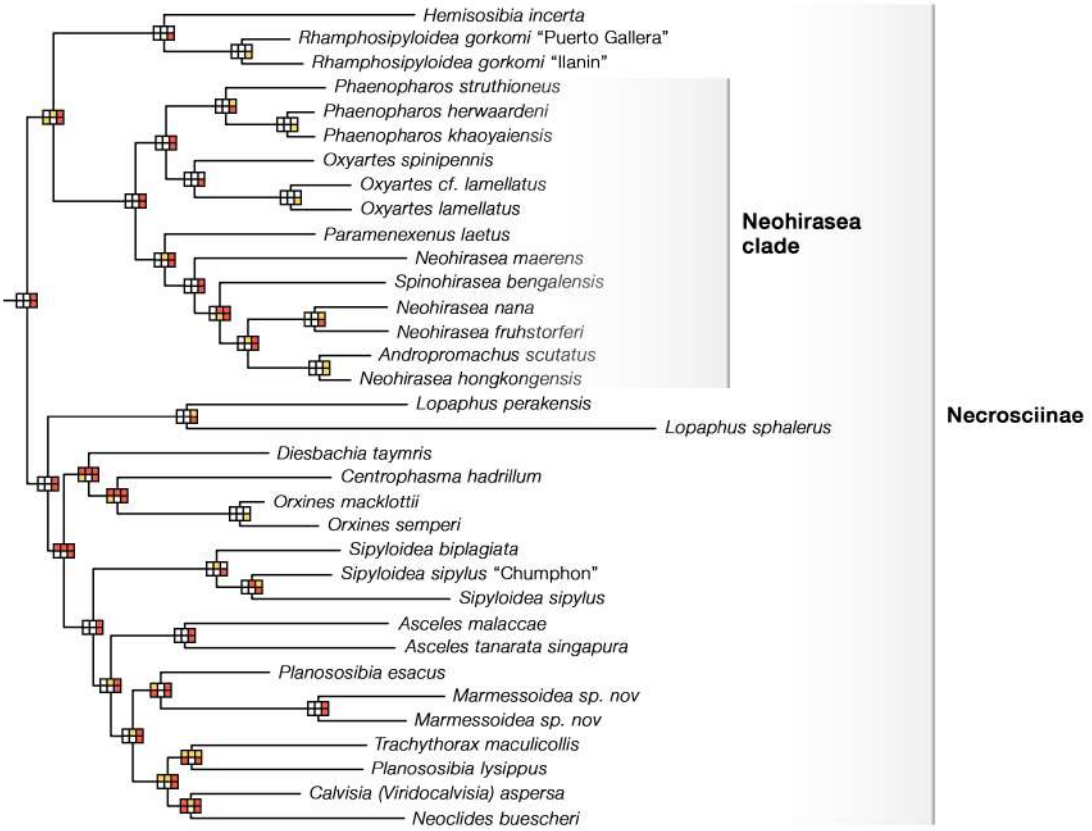


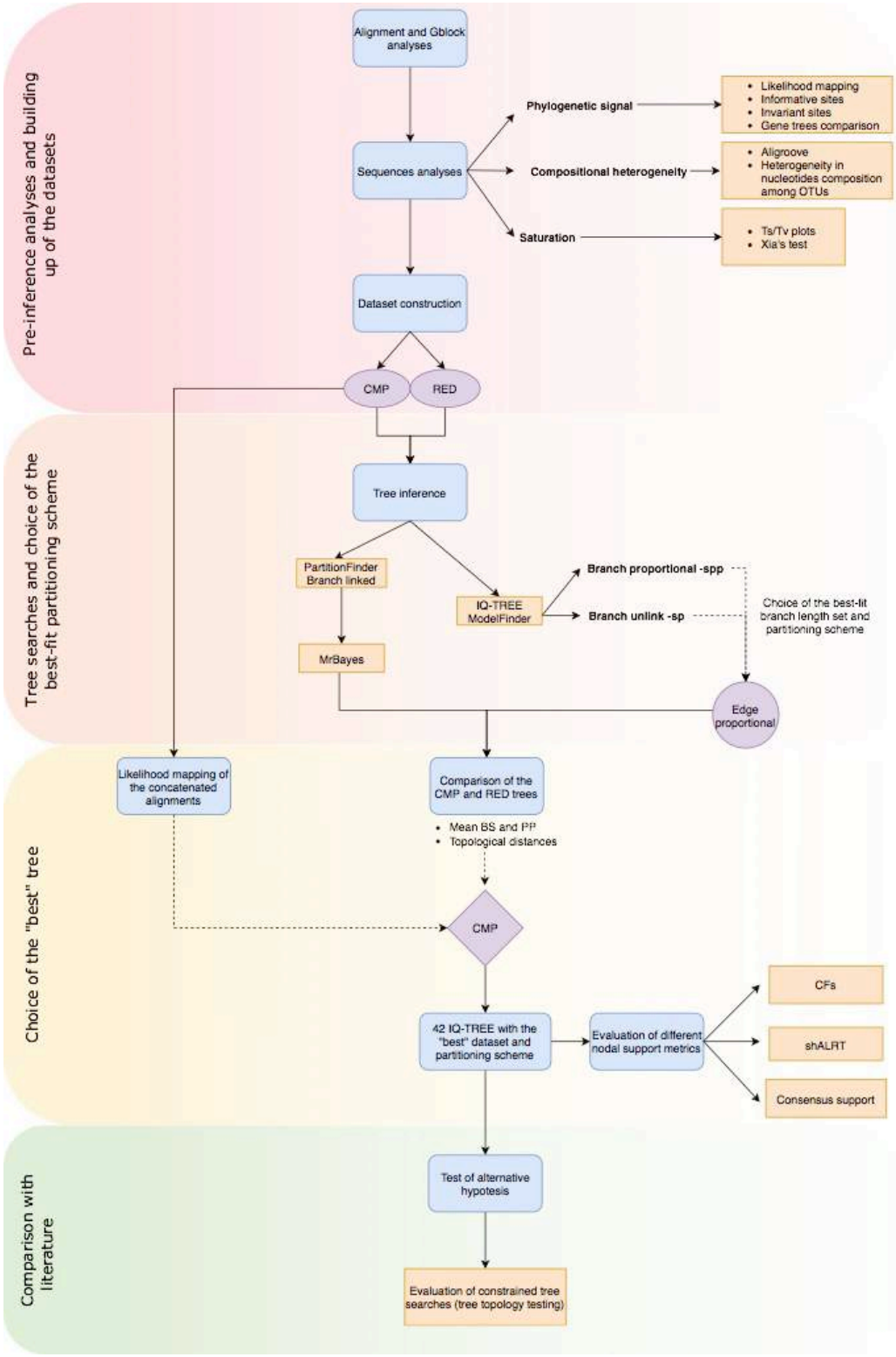
Fig. 8 - Subtree representing the phylogeny of Necrosciinae. Nodal support diagrams as per legend in Figure 3.

TREE	-logL	Free parameters	AIC	AICc	BIC
CMP spp	-93242.4575	391	187690.9682	187762.1086	190215.0806
CMP sp	-92956.3376	906	187724.6753	188157.8550	193573.3859
RED spp	-61229.2200	372	123202.4400	123278.0769	125547.7121
RED sp	-60951.1905	789	123480.3810	123863.7205	128454.6275

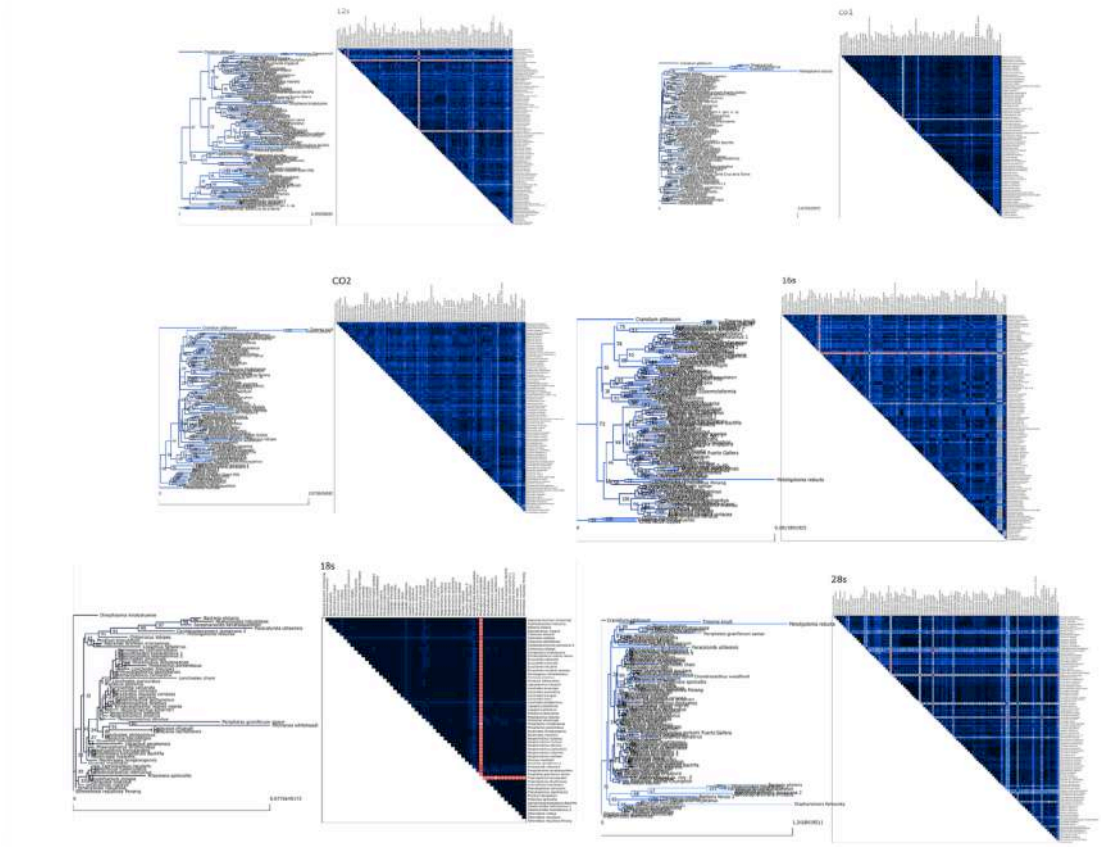
Table 1: log-likelihood values, number of free parameters, AIC, AICc and BIC scores of the IQ-TREE tree searches performed with the two dataset CMP and RED and both an unlink branch set strategy (sp) and a proportional one (spp).

Tree	logL	deltaL	Bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
Lonchodidae	-92350.36497	180.5	0 -	0 -	0 -	0 -	0 -	1.16e-43 -	6.91e-05 -
Robertson et al., (2018)	-92366.88319	197.02	0 -	0 -	0 -	0 -	0 -	8.51e-50 -	3.56e-07 -
Dipheromerinae	-92170.30904	0.44112	0.486 +	0.496 +	0.742 +	0.496 +	0.739 +	0.486 +	0.494 +
FULLnuc spp	-92169..86792	0	0.514 +	0.504 +	1 +	0.504 +	0.749 +	0.514 +	0.506 +

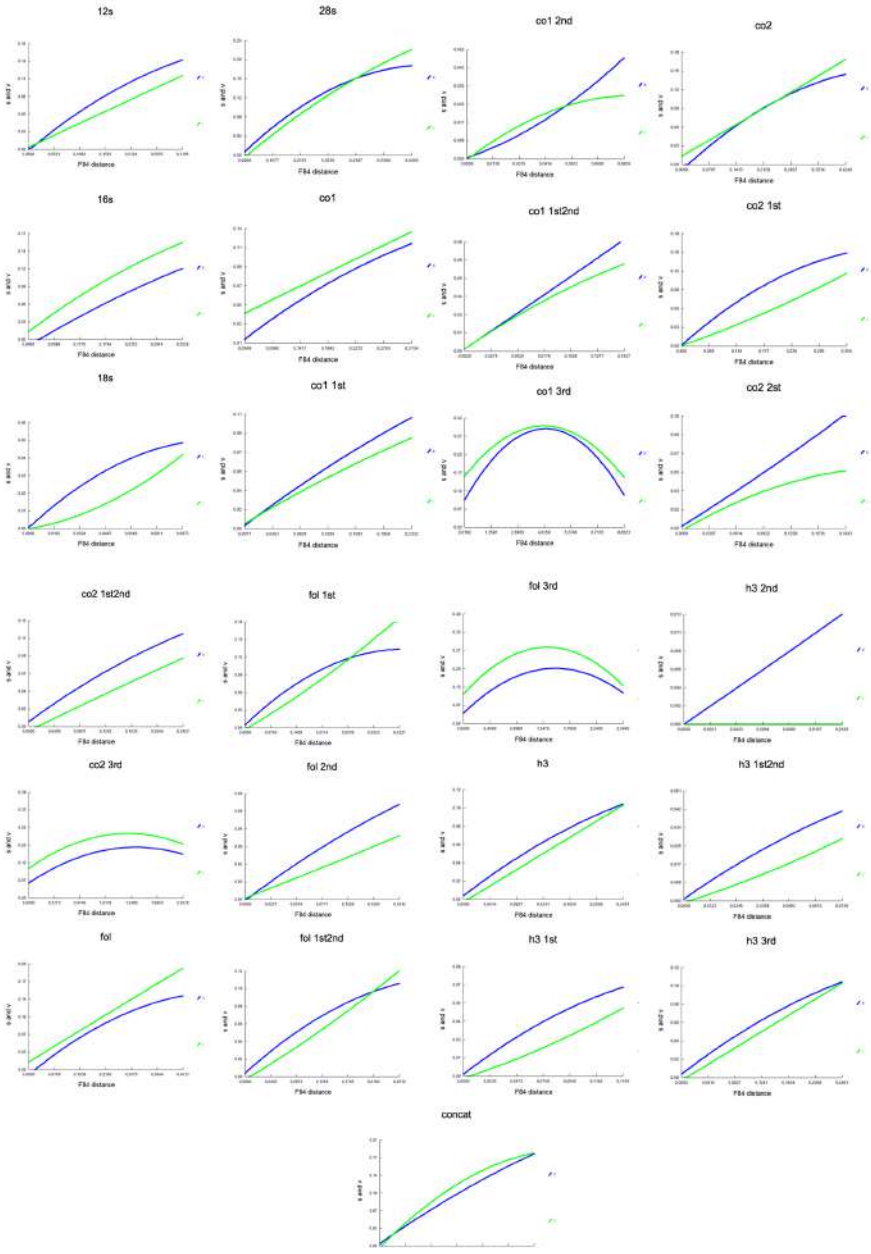
Table 2: Topology test performed with IQ-TREE 2 between the best ML tree and the results of the constrained tree searches.



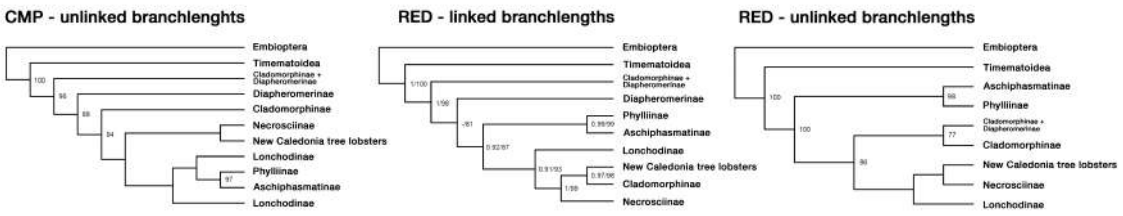
Sup. Fig. 1 – Analytical pipeline describing data processing, model selection, phylogenetic inferences, calculation of support metrics and tree topology testing.



Sup. Fig. 2 – Heterogeneity of sequence divergence analysis using AliGROOVE v 1.06 similarity, including score distance matrices and associated Maximum Likelihood topologies.



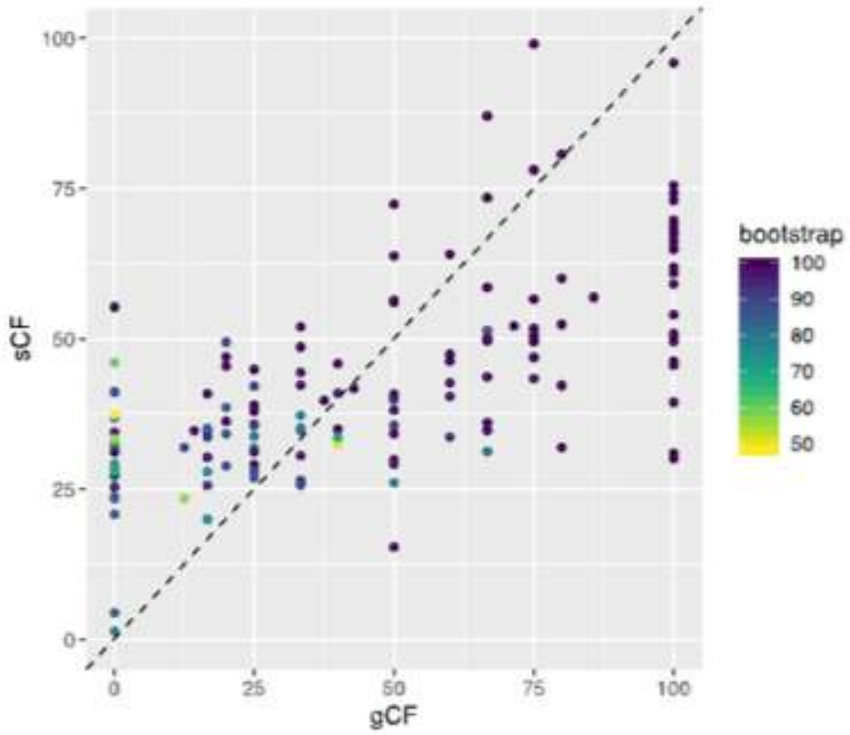
Sup. Fig. 3 – Plots of the transitions and transversions against the F84 distances calculated by DAMBE v. 7.0.28.



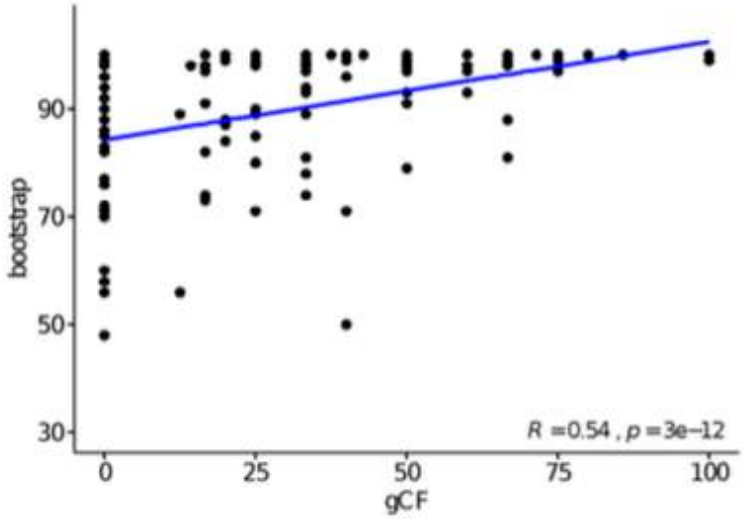
Sup. Fig. 4. Additional topologies obtained using different data processing (CMP and RED dataset) and model selections (unlinking branch lengths or linking them proportionally) approaches.

Sup. Fig. 5

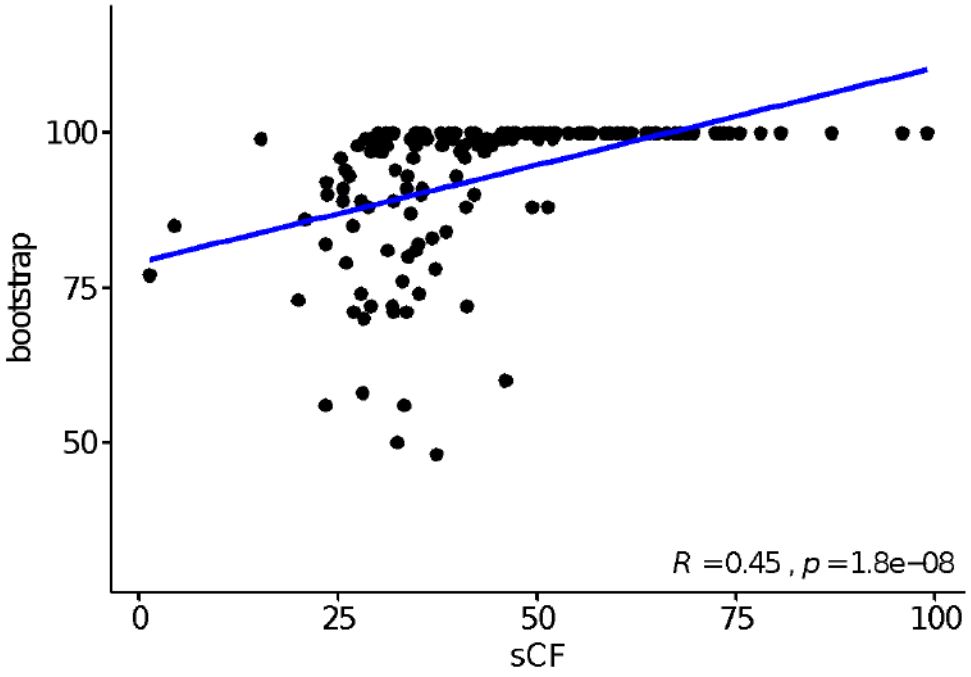
a) Correlation analysis of site concordance factors (sCF) *versus* gene concordance factor (gCF). The colour gradient represents the respective bootstrap value.



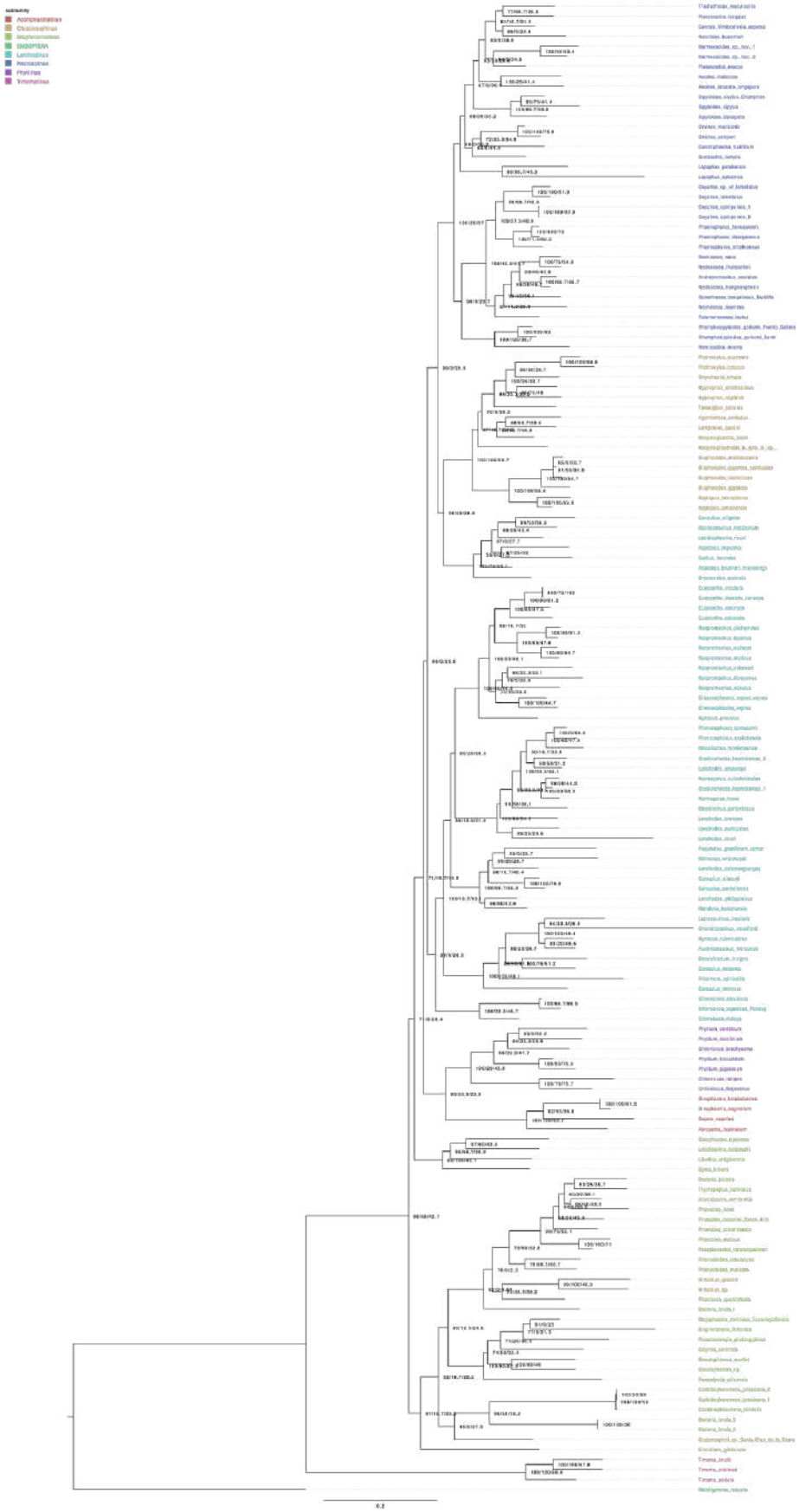
b) Correlation analysis of gCFs *versus* bootstrap values. The blue line represents the regression line. R = Pearson correlation coefficient. p= p-value.



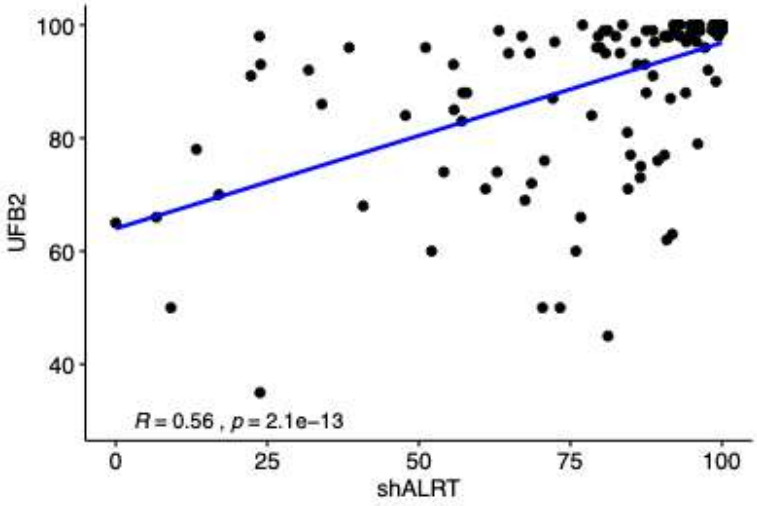
c) sCFs against bootstrap values. The blue line represents the regression line. R = Pearson correlation coefficient. *p*= p-value.



d) FULLnuc spp tree where node labels are annotated respectively with bootstrap values, gCF and sCF. Tips are coloured following the current systematics as reported in Phasmida Species File.



e) shALRT against bootstrap values. The blue line represents the regression line. R = Pearson correlation coefficient. *p*= p-value.



Sup. Tab. 1 - Space name, taxonomy and NCBI sequences used in the present study.

ID	Species	Subfamily	18S	28S	12S	16S	CO1	CO2	H3
0002	Diapherodes gigantea	Cladomorphinae							
0010	Phantasca quadrilobata	Diapheromerinae							
0016	Hermagoras cultratolobatus	Lonchodinae							
0029	Libethra strigiventris	Diapheromerinae							
0038	Phaenopharos struthioneus	Necrosciinae							
0040	Calynda coronata	Diapheromerinae							
0047	Oncotophasma sp.	Diapheromerinae							
0048	Oncotophasma martini	Diapheromerinae							
0049	Trychopeplus laciniatus	Diapheromerinae							
0051	Rhynchacris ornata	Cladomorphinae							
0070	Lamponius guerini	Cladomorphinae							
0099	Taraxippus paliurus	Cladomorphinae							
0105	Diapherodes martinicensis	Cladomorphinae							
0128	Chitoniscus lobipes	Phylliinae							
0142	Sipyloidea biplagiata	Necrosciinae							
0150	Hesperophasma bocki	Cladomorphinae							
0152	Carausius alluaudi	Lonchodinae							
0153	Carausius sechellensis	Lonchodinae							
0176	Prisomera ignava	Lonchodinae							
0181	Hypocyrtus ornatissimus	Cladomorphinae							
0188	Pterinoxylus crassus	Cladomorphinae							
0206	Pterinoxylus eucnemis	Cladomorphinae							
0211	Hirtuleius sp.	Cladomorphinae							
0234	Spinohirasea bengalensis 'Bach'Ma'	Necrosciinae							

0300	Andropromachus scutatus	Necrosciinae
0325	Agamemnon cornutus	Cladomorphinae
0337	Sipyloidea sipylus 'Chumphon'	Necrosciinae
0411	Eurycantha insularis coriacea	Lonchodinae
0442	Haplopus micropterus	Cladomorphinae
0449	Centrophasma hadrillum	Necrosciinae
0493	Phenacephorus cornucervi	Lonchodinae
0527	Lonchodes brevipes	Lonchodinae
0529	Planososibia lysippus	Necrosciinae
0530	Carausius nodosus	Lonchodinae
0540	Mnesilochus portentosus	Lonchodinae
0553	Caribbiopheromera trinitatis	Diapheromerinae
0560	Bostriana boliviana	Cladomorphinae
0588	Lonchodes dalawangsungay	Lonchodinae
0597	Eurycantha calcarata	Lonchodinae
0602	Phanocles vosseleri 'Green Hills'	Diapheromerinae
0610	Hemisosibia incerta	Necrosciinae
0612	Calvisia (Viridocalvisia) aspersa	Necrosciinae
0615	Marmessoidea sp. nov.	Necrosciinae
0616	Marmessoidea sp. Nov.	Necrosciinae
0617	Baculofractum insigne	Lonchodinae
0621	Hermagoras hosei	Lonchodinae
0622	Asceles malaccae	Necrosciinae
0624	Asceles tanarata singapura	Necrosciinae
0626	Staelonchos heamatomus	Lonchodinae
0627	Stheneboea malaya	Lonchodinae

0628	Stheneboea repudiosa	Lonchodinae
0639	Planososibia esacus	Necrosciinae
0664	Hirtuleius gracilis	Cladomorphinae
0704	Phanocloidea muricata	Diapheromerinae
0747	Periphetes graniferum (samar)	Lonchodinae
0789	Rhamphosipyloidea gorkomi (Puerto Gallera)	Necrosciinae
0799	Rhamphosipyloidea gorkomi (Ilanin)	Necrosciinae
0887	Lonchodes philippinicus	Lonchodinae
0910	Paracalynda utilaensis	Diapheromerinae
0915	Phanocloidea lobulatipes	Diapheromerinae
0917	Diapherodes dominicae	Cladomorphinae
0924	Diapherodes gigantea saintluciae	Cladomorphinae
0925	Paraphanocles keratosqueleton	Diapheromerinae
F_0004	Sigaruphasma bouladoi	Cladomorphinae
F_0039	Oxyartes sp. (cf lamellatus)	Necrosciinae
F_0050	Neohirasea nana	Necrosciinae
F_0057	Phanocles costaricensis	Diapheromerinae
F_0073	Phanocles muticus	Diapheromerinae
F_0075	Phanocles horni	Diapheromerinae
F_0077	Oreophoetes topoense	Diapheromerinae
F_0095	Neohirasea fruhstorferi	Necrosciinae
F_0105	Phenacephorus sepilokensis	Lonchodinae
F_0111	Mnesilochus mindanaense	Lonchodinae
F_0118	Hypocyrtus scythrus	Cladomorphinae
F_0124	Megaphasma dentricus "Eusermyleformia"	Diapheromerinae
F_0127	Lobolibethra carbonelli	Diapheromerinae

F_0136	Diesbachia tamyris	Necrosciinae							
F_0150	Mithrenes whiteheadi	Lonchodinae							
F_0168	Oxyartes spinipennis	Necrosciinae							
F_0170	Alienobostra remiformis	Diapheromerinae							
F_0173	Cranidium gibbosum	Cladomorphinae							
F_0176	Bacteria ferula	Diapheromerinae							
F_0178	Dyme bifrons	Diapheromerinae							
F_0180	Caribbiopheromera jamaicana	Diapheromerinae							
F_0187	Manduria halconensis	Lonchodinae							
F_0189	Asprenas brunneri 'microwings'	Lonchodinae							
F_0190	Paramenexenus laetus	Necrosciinae							
F_0210	Neoclides buescheri	Necrosciinae							
PB_0078	Austrocarausius mercurius	Lonchodinae							
Abrosoma_festinatum_isolate_ABF1	Abrosoma festinatum	Aschiphasmatinae		FJ474100.1			FJ474257.1	FJ474334.1	FJ474178.1
Asprenas_impennis	Asprenas impennis	Lonchodinae		KF383453.1			FJ474264.1	FJ474340.1	FJ474185.1
Bacteria_ploiaria_isolate_WS045	Phanocles ploiaria	Diapheromerinae	MK291755.1		MK291604.1	MK291678.1		MK297241.1	MK291528.1
Canachus_alligator	Canachus alligator	Lonchodinae		FJ474110.1			FJ474267.1	FJ474343.1	FJ474190.1
Carausius_morosus	Carausius morosus	Lonchodinae		FJ474111.1			FJ474268.1	FJ474344.1	FJ474191.1
Carlius_fecundus	Carlius fecundus	Lonchodinae		GQ927330.1			FJ474265.1	FJ474341.1	FJ474185.1
Chitoniscus_brachysoma	Chitoniscus brachysoma	Phylliinae		FJ474114.1			FJ474271.1		FJ474194.1
Chitoniscus_feejeeanus	Chitoniscus feejeeanus	Phylliinae		FJ474115.1			FJ474272.1	FJ474347.1	FJ474195.1
Chondrostethus_woodfordi	Chondrostethus woodfordi	Lonchodinae		MK291895.1		MK291740.1			MK291588.1
Dajaca_napolovi_isolate_WS316	Dajaca napolovi	Aschiphasmatinae		MK291911.1	MK291673.1	MK291753.1	MK296741.1		
Diapheromera_femorata_isolate_DIF1	Diapheromera femorata	Diapheromerinae		FJ474119.1			FJ474276.1	FJ474351.1	FJ474199.1
Dinophasma_kinabaluense_isolate_WS141	Dinophasma kinabaluense	Aschiphasmatinae	MK291810.1	MK291885.1	MK291655.1	MK291731.1		MK297279.1	MK291579.1
Dinophasma_saginatum_isolate_DIS1	Dinophasma saginatum	Aschiphasmatinae		FJ474122.1			FJ474279.1		

Dryococelus_australis	Dryococelus australis	Lonchodinae		FJ474123.1			FJ474280.1	FJ474354.1	FJ474202.1
Erinaceophasma_vepres	Erinaceophasma vepres	Lonchodinae		FJ474125.1				FJ474356.1	FJ474204.1
Erinaceophasma_vepres_vepres	Erinaceophasma vepres vepres	Lonchodinae	MK291780.1	MK291855.1	MK291627.1			MK297257.1	MK291551.1
Eurycantha_coronata	Eurycantha coronata	Lonchodinae	MK291767.1	MK291842.1	MK291614.1	MK291690.1		MK297248.1	MK291539.1
Eurycantha_insularis	Eurycantha insularis	Lonchodinae	AY121154.1	AY125294.1	KJ024474.1	KJ024425.1			AY125238.1
Haplopus_jamaicensis_isolate_WS165	Haplopus jamaicensis	Cladomorphinae				KJ024423.1		KJ024538.1	KJ024514.1
Hyrtacus_procerus	Hyrtacus procerus	Lonchodinae	MK291770.1	MK291845.1	MK291617.1	MK291693.1		MK297250.1	
Hyrtacus_tuberculatus	Hyrtacus tuberculatus	Lonchodinae	MK291821.1	MK291897.1	MK291664.1	MK291742.1			MK291590.1
Labidiophasma_rouxi	Labidiophasma rouxi	Lonchodinae		FJ474141.1			FJ474297.1	FJ474371.1	FJ474219.1
Leprocaulinus_insularis	Leprocaulinus insularis	Lonchodinae	KJ024405.1	KJ024382.1	KJ024487.1	KJ024441.1		KJ024552.1	KJ024519.1
Lonchodes_amaurops	Staelonchodes amaurops	Lonchodinae	KJ024402.1	KJ024379.1	KJ024483.1	KJ024437.1		KJ024549.1	
Lonchodes_auriculatus	Lonchodes auriculatus	Lonchodinae	MK291799.1		MK291646.1	MK291721.1		MK297271.1	
Lonchodes_chani	Lonchodes chani	Lonchodinae	MK291819.1	MK291894.1	MK291663.1	MK291739.1			
Lopaphus_perakensis	Lopaphus perakensis	Necrosciinae	AY121171.1	AY125311.1	KJ024485.1	KJ024439.1			AY125254.1
Lopaphus_sphalerus	Lopaphus sphalerus	Necrosciinae	AY121182.1	AY125322.1					AY125265.1
Metoligotoma_reducta	Metoligotoma reducta	EMBIOPTERA	JQ907231.1			EU157047.1	EU157064.1		EU157034.1
Microcanachus_matileorum	Microcanachus matileorum	Lonchodinae		FJ474144.1			FJ474300.1	FJ474374.1	FJ474222.1
Neohirasea_hongkongensis	Neohirasea hongkongensis	Necrosciinae	KJ024403.1	AY125309.1	KJ024484.1				KJ024517.1
Neohirasea_maerens	Neohirasea maerens	Necrosciinae	AY121168.1	AY125308.1	KJ024480.1	KJ024432.1		KJ024546.1	AY125251.1
Neopromachus_doreyanus	Neopromachus doreyanus	Lonchodinae		FJ474149.1			FJ474305.1	FJ474379.1	FJ474227.1
Neopromachus_dyselius	Neopromachus dyselius	Lonchodinae	MK291786.1	MK291861.1	MK291633.1	MK291709.1			MK291557.1
Neopromachus_muticus	Neopromachus muticus	Lonchodinae	MK291784.1	MK291859.1	MK291631.1	MK291707.1		MK297260.1	MK291555.1
Neopromachus_obrutus	Neopromachus obrutus	Lonchodinae	KJ024400.1	KJ024377.1	KJ024481.1	KJ024435.1		KJ024548.1	KJ024515.1
Neopromachus_pachynotus	Neopromachus pachynotus	Lonchodinae	MK291772.1	MK291847.1	MK291619.1	MK291695.1			MK291544.1
Neopromachus_scharreri	Neopromachus scharreri	Lonchodinae	MK291785.1	MK291860.1	MK291632.1	MK291708.1		MK297261.1	MK291556.1
Neopromachus_wallacei	Neopromachus wallacei	Lonchodinae	MK291779.1	MK291854.1	MK291626.1	MK291702.1			MK291550.1

Orxines_macklottii	Pseudodiacantha macklottii	Necrosciinae	AY121153.1	AY125293.1		KJ024438.1			AY125237.1
Orxines_semperi	Orxines semperi	Necrosciinae		FJ474150.1			FJ474306.1		FJ474228.1
Oxyartes_lamellatus	Oxyartes lamellatus	Necrosciinae		FJ474151.1			FJ474307.1	FJ474380.1	FJ474229.1
Phaenopharos_herwaardeni	Phaenopharos herwaardeni	Necrosciinae	KJ024414.1	KJ024384.1	KJ024498.1	KJ024451.1			KJ024530.1
Phaenopharos_khaoyaiensis	Phaenopharos khaoyaiensis	Necrosciinae		FJ474157.1			FJ474313.1	FJ474383.1	FJ474235.1
Phyllium_bioculatum	Phyllium bioculatum	Phylliinae	AY121161.1	AY125301.1					AY125245.1
Phyllium_celebicum	Phyllium celebicum	Phylliinae		FJ474155.1			FJ474311.1		FJ474233.1
Phyllium_giganteum	Phyllium giganteum	Phylliinae		FJ474156.1			FJ474312.1		FJ474234.1
Phyllium_siccifolium	Phyllium siccifolium	Phylliinae		FJ474159.1			FJ474315.1	FJ474384.1	FJ474237.1
Pseudosermyle_phalangiphora_isolate_WS162	Pseudosermyle phalangiphora isolate WS162	Diapheromerinae						MK297286.1	
Sipyloidea_sipylus	Sipyloidea sipylus	Necrosciinae		FJ474168.1			FJ474324.1	FJ474393.1	FJ474246.1
Timema_cristinae	Timema cristinae	Timematinae				DQ890283.1	DQ890210.1		
Timema_knulli	Timema knulli	Timematinae		AY125302.1	KJ024471.1	KJ024420.1	JX237530.1	KJ024536.1	MK291560.1
Timema_podura	Timema podura	Timematinae			KF855798.1	KF855873.1	HQ184599.1	KF855918.1	KF855895.1
Trachythorax_maculicollis	Trachythorax maculicollis	Necrosciinae			KJ024495.1	KJ024448.1		KJ024559.1	KJ024527.1

Sup. Tab. 2 - Gene target, primer name, temperature and time of annealing and amplicon size (bp) used in the present study.

Gene target	Primer Name	Primer sequence	Annealing	Amplicon Size
CO1$_{FOL}$	LCO1490	GTCAACAAATCATAAAGATATTGG	55°C 45s	708
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA		
	FG1F	AAACTAWTAACCTTCAAAGTTA	50°C 60s	1049
	FG2F	CCWACWGTRAATATRTGRTGWGC		
CO1	TL2-N-3014	TTGATTTTTTGGTCATCCAGAAGT	61°C 60s	819
	C1-J-2195	TCCAATGCACTAATCTGCCATATTA		
16S	16sa	CGCCTGTTTATCAAAAACAT	56°C 30s	548
	16sb	CTCCGGTTTGAACTCAGATCA		
12S	12S_F_G	CTGCACCTTGACYTGAAATA	56°C 30s	399
	12S_R_G	CAGCATACGCGGTTATACAA		
CO2	TL2-J-3034	CAGATAAGTGCATTGGATTT	50°C 50s	1000
	TK-N-3785	GTTTAAGAGACCAGTACTTG		
18S	18sai	CCTGAGAAACGGCTACCACATC	57°C 30s	950
	18sbi	GAGTCTCGTTCGTTATCGGA		
28S	28S-1009	TCAAGACGGGTCGGGAGA	58°C 30s	564
	28S-356	AGAACTTTGAAGAGAGAGTTCAAGA		
H3	H3F	ATGGCTCGTACCAAGCAGAC	58°C 30s	500
	H3R	ATATCCTTRGGCATRATRGTGAC		

Sup. Table 3a - Results of the model selection performed with ModelFinder2 with the -m TESTNEWMERGE option (CMP with linked branchlengths, RED with linked branchlengths, CMP with unlinked branchlengths and RED with unlinked branchlengths) along with mean bootstrap value (BS), mean posterior probability (PP) and the percentage of unresolved quartets of the CMP and RED datasets with linked branchlengths.

Tree	Best-fit partitioning scheme	Best-fit model of evolution	Mean BS	Mean PP	% unresolved quartets
CMP with linked branchlengths	12S = 1-266;	GTR + F + I + G4	91.2	0.92	30.2%
	16S = 267-822;	GTR + F + I + G4			
	18S, H3 1st = 823-1788, 4381-4701\3;	TIM2e + I + G4			
	28S = 1789-2403;	GTR + F + I + G4			
	CO1 1st; CO2 2nd = 2404-3159\3, 3161-3777\3;	TIM + F + I + G4			
	CO1 2nd , $CO1_{FOL}$ 2nd = 2405-3159\3, 3779-4380\3;	HKY + F + R5			
	CO1 3rd , $CO1_{FOL}$ 3rd , CO2 3rd = 2406-3159\3, 3780-4380\3, 3162-377713;	TVM + F + R7			
	CO2 1st , $CO1_{FOL}$ 1st = 3160-3777\3, 3778-4380\3;	GTR + F + I + G4			
	H3 2nd = 4382-4701\3	K2P			
	H3 3rd = 4383-4701\3;	K2P + G4			
RED with linked branchlengths	12S = 1-266	GTR + F + I + G4	91.7	0.90	36%
	18S, H3 1st = 267-1232, 3722-4042\3;	TIM2e + I + G4			
	16S = 1233-1788;	GTR + F + I + G4			
	28S = 1789-2403	GTR + F + I + G4			
	CO1 1st = 2404-2907\2;	GTR + F + I + G4			
	CO2 1st , $CO1_{FOL}$ 1st = 2908-3319\2, 3320-3721\2	GTR + F + I + G4			
	$CO1_{FOL}$ 2nd , CO1 2nd , CO2 2nd = 3321-3721\2, 2405-2907\2, 2909-3319\2;	TVM + F + I + G4			
	H3 2nd = 3723-4042\3	K2P			
	H3 3rd = 3724–4042\3	K2P + G4			
CMP with	12S, 16S, H3 1st, CO1 1st, CO2 2nd, CO1 2nd, $CO1_{FOL}$ 2nd, CO2 1st,$CO1_{FOL}$	TVM+F+I+G4	91.3	/	/

unlinked branchlengths	1st, H3 2nd, H3 3rd = 1-266 267-822 4381-4701\3 2404-3159\3 3161-3777\3 2405-3159\3 3779-4380\3 3160-3777\3 3778-4380\3 4382-4701\3 4383-4701\3;				
	18S = 823-1788;	TIM3e+R3			
	28S = 1789-2403;	GTR+F+I+G4			
	CO1 3rd, $CO1_{FOL}$ 3rd, CO2 3rd = 2406-3159\3 3780-4380\3 3162-3777\3;	GTR+F+R10			
RED with unlinked branchlengths	12S, 16S = 1-266 1233-1788;	GTR+F+R5	86.7	/	/
	18S, H3 1st, CO1 1st, CO2 1st, $CO1_{FOL}$ 1st, $CO1_{FOL}$ 2nd, CO1 2nd, CO2 2nd, H3 2nd, H3 3rd = 267-1232 3722-4042\3 2404-2907\2 2908-3319\2 3320-3721\2 3321-3721\2 2405-2907\2 2909-3319\2 3723-4042\3 3724-4042\3;	GTR+F+R10			
	28S = 1789-2403;	GTR+F+I+G4			

Sup. Table 3b - Results of the model selection carried on PartitionFinder2 with a *greedy* strategy and used for the Bayesian Inference with mrBayes.

Dataset	**Best-fit partitioning scheme**	**Best-fit evolutionary model**
RED	12S	GTR+I+G
	18S, H3 1st	SYM+I+G
	16S	GTR+I+G
	28S	GTR+I+G
	CO1 1st	GTR+I+G
	CO2 2nd, CO1 2nd, $CO1_{FOL}$ 2nd	GTR+I+G
	CO2 1st, $CO1_{FOL}$ 1st	GTR+I+G
	H3 2nd	JC+I
	H3 3rd	K80+G
CMP	12S	GTR+I+G

	16S	GTR+I+G
	18S, H3 1st	SYM+I+G
	28S	GTR+I+G
	CO1 1st	GTR+I+G
	CO2 2nd, CO1 2nd, $CO1_{FOL}$ 2nd	GTR+I+G
	CO2 3rd, FOL 3rd, CO1 3rd	GTR+G
	CO2 1st, $CO1_{FOL}$ 1st	GTR+I+G
	H3 2nd	JC+I
	H3 3rd	K80+G

Sup. Table 4 - -logL values for each run performed with the CMP dataset and the best-fit partitioning scheme on IQ-TREE 2. The bold row highlights the tree with the higher value.

Tree	-logL
run 01	-93289.0381
run 02	-93368.9133
run 03	-93367.3830
run 04	-93390.8547
run 05	-93379.4626
run 06	-93290.9280
run 07	-93398.8280
run 08	-93389.2264
run 09	-93382.5260
run 10	**-93242.4575**
run 11	-93388.3312
run 12	-93373.5192
run 13	-93377.4030
run 14	-93364.3166
run 15	-93389.8751
run 16	-93372.3130
run 17	-93381.6843
run 18	-93396.6093
run 19	-93391.2559
run 20	-93371.5534
run 21	-93383.0038
run 22	-93355.9989
run 23	-93374.1535
run 24	-93377.9723
run 25	-93382.1360
run 26	-93381.2844
run 27	-93385.3481
run 28	-93386.5614
run 29	-93278.5060
run 30	-93378.0952
run 31	-93373.3746
run 32	-93387.8558
run 33	-93295.8129
run 34	-93383.8612
run 35	-93360.1160
run 36	-93394.2369
run 37	-93369.2530
run 38	-93384.2511
run 39	-93384.2661
run 40	-93387.2129
run 41	-93389.1834
run 42	-93399.9426

Sup. Table 5 - Compositional bias analysis performed using DAMBE v. 7.0.28.

Dataset	Statistic	Value	Degrees of freedom	Probability
$CO1_{FOL}$	Chi-square	142.04	165	0.9016
	Likelihood ratio chi-square	150.24	165	0.7884
	Cramer's V	0.0378		
$CO1_{FOL}$ 1st	Chi-square	40.07	165	1.0000
	Likelihood ratio chi-square	41.76	165	1.0000
	Cramer's V	0.0348		
$CO1_{FOL}$ 2nd	Chi-square			
	Likelihood ratio chi-square			
	Cramer's V			
$CO1_{FOL}$3rd	Chi-square	381.24	165	0.0000
	Likelihood ratio chi-square	401.57	165	0.0000
	Cramer's V	0.1073		
$CO1_{FOL}$1st+2nd	Chi-square	27.04	165	1.0000
	Likelihood ratio chi-square	17.06	165	1.0000
	Cramer's V	0.0202		
CO1	Chi-square	204.26	240	0.9544
	Likelihood ratio chi-square	194.94	240	0.9850
	Cramer's V	0.0335		
CO1 1st	Chi-square	43.55	240	1.0000
	Likelihood ratio chi-square	39.48	240	1.0000
	Cramer's V	0.0268		
CO1 2nd	Chi-square	12.46	240	1.0000
	Likelihood ratio chi-square	16.43	240	1.0000
	Cramer's V	0.0143		
CO1 3rd	Chi-square	654.93	240	0.0000
	Likelihood ratio chi-square	649.85	240	0.0000
	Cramer's V	0.1038		
CO1 1st+2nd	Chi-square	30.63	240	1.0000
	Likelihood ratio chi-square	42.64	240	1.0000
	Cramer's V	0.0159		
CO2	Chi-square	237.95	294	0.9928
	Likelihood ratio chi-square	251.33	294	0.9660
	Cramer's V	0.0364		

CO2 1st	Chi-square	60.36	294	1.0000
	Likelihood ratio chi-square	76.30	294	1.0000
	Cramer's V	0.0318		
CO2 2nd	Chi-square	25.17	294	1.0000
	Likelihood ratio chi-square	21.23	294	1.0000
	Cramer's V	0.0205		
CO2 3rd	Chi-square	610.85	294	0.0000
	Likelihood ratio chi-square	610.20	294	0.0000
	Cramer's V	0.1011		
CO2 1st+2nd	Chi-square	52.04	294	1.0000
	Likelihood ratio chi-square	39.97	294	1.0000
	Cramer's V	0.0209		
12S	Chi-square	81.56	309	1.0000
	Likelihood ratio chi-square	86.04	309	1.0000
	Cramer's V	0.0339		
16S	Chi-square	126.93	378	1.0000
	Likelihood ratio chi-square	138.86	378	1.0000
	Cramer's V	0.0270		
18S	Chi-square	18.33	207	1.0000
	Likelihood ratio chi-square	32.37	207	1.0000
	Cramer's V	0.0096		
28S	Chi-square			
	Likelihood ratio chi-square			
	Cramer's V			
H3	Chi-square	107.75	402	1.0000
	Likelihood ratio chi-square	111.50	402	1.0000
	Cramer's V	0.0290		
H3 1st	Chi-square			
	Likelihood ratio chi-square			
	Cramer's V			
H3 2nd	Chi-square	4.18	402	1.0000
	Likelihood ratio chi-square	16.16	402	1.0000
	Cramer's V	0.0099		
H3 3rd	Chi-square	299.52	402	1.0000
	Likelihood ratio chi-square	294.58	402	1.0000
	Cramer's V	0.0836		

H3 1st+2nd	Chi-square	6.87	402	1.0000
	Likelihood ratio chi-square	150.84	402	1.0000
	Cramer's V	0.0090		
concatenated	Chi-square	6240.12	513	1.0000
	Likelihood ratio chi-square	6277.35	513	1.0000
	Cramer's V	0.0717		

Sup. Table 6 - Saturation analyses performed using DAMBE v. 7.0.28.

Dataset	I_{ss}	$I_{ss.c}$ Sym	T	DF	P	$I_{ss.c}$ Asym	T	DF	P
$CO1_{FOL}$	0.338	0.713	15.953	523	0.0000	0.385	1.997	523	0.0463
$CO1_{FOL}$ 1st	0.255	0.691	12.663	174	0.0000	0.376	3.519	174	0.0006
$CO1_{FOL}$ 2nd	0.146	0.691	20.197	150	0.0000	0.376	8.523	150	0.0000
$CO1_{FOL}$ 3rd	0.708	0.691	0.517	199	0.6058	0.376	10.027	199	0.0000
$CO1_{FOL}$ 1st+2nd	0.199	0.691	21.486	322	0.0000	0.362	7.125	322	0.0000
CO1	0.285	0.728	21.957	666	0.0000	0.407	6.038	666	0.0000
CO1 1st	0.191	0.683	16.611	211	0.0000	0.357	5.630	211	0.0000
CO1 2nd	0.115	0.683	22.305	182	0.0000	0.357	9.533	182	0.0000
CO1 3rd	0.650	0.683	1.155	250	0.2490	0.357	10.336	250	0.0000
CO1 1st+2nd	0.153	0.703	27.500	391	0.0000	0.378	11.242	391	0.0000
CO2	0.377	0.714	11.171	566	0.0000	0.387	0.331	566	0.7408
CO2 1st	0.300	0.689	7.556	190	0.0000	0.373	1.420	190	0.1574
CO2 2nd	0.213	0.689	9.304	172	0.0000	0.373	3.128	172	0.0021
CO2 3rd	0.729	0.689	0.909	204	0.3644	0.373	8.207	204	0.0000
CO2 1st+2nd	0.245	0.692	12.414	368	0.0000	0.364	3.291	368	0.0011
12S	0.717	0.682	0.339	244	0.7351	0.355	3.511	244	0.0005
16S	0.800	0.708	1.198	513	0.2313	0.380	5.474	513	0.0000
18S	0.164	0.745	21.743	680	0.0000	0.438	10.244	680	0.0000
28S	0.937	0.714	2.894	614	0.0039	0.386	7.145	614	0.0000
H3	0.239	0.684	13.326	288	0.0000	0.354	4.414	288	0.0006
H3 1st	0.126	0.785	19.025	72	0.0000	0.552	12.301	72	0.0000
H3 2nd	0.036	0.785	46.614	105	0.0000	0.552	32.117	105	0.0000
H3 3rd	0.555	0.785	4.283	106	0.0000	0.552	0.056	106	0.9553
H3 1st+2nd	0.091	0.687	26.975	143	0.0000	0.369	12.570	143	0.0000
concat.	1.658	0.811	61.668	4700	0.0000	0.563	79.662	4700	0.0000

Sup. Table 7 - Results of the likelihood mapping analysis, reported as the number of unresolved quartets, performed using IQ-TREE 1.6.12 with the construction of 15.000 quartets and number of parsimony informative sites.

Dataset	N. of unresolved quartets	Parsimony informative sites
$CO1_{FOL}$	34.9%	270/603
CO1	31.0%	367/756
CO2	29.6%	377/618
12S	33.40%	137/266
16S	25.70%	287/556
18S	20.30%	94/966
28S	32.30%	335/615
H3	30.05%	123/321

Sup. Table 8a - Tree distances between gene trees obtained on TreeCmp. Quartet = Quartet distances; RF = Robinson-Foulds distances; Matching Split = Matching Split distances. Each metrics has been normalized indicating whether two trees are more dissimilar than two random trees with the same number of tips, according to a given metric. Values of 1 indicate that the two trees are no more similar than two random trees, values of 0 indicate the same topology.

Tree1	Tree2	Common taxa	Quartet	Quartet to Yule Avg	RF(0.5)	RF(0.5) to Yule Avg	Matching Split	Matching Split to Yule Avg
12S	**16S**	89	729331.0000	0.4483	61.0000	0.711	306.0000	0.4297
12S	**18S**	41	36529.0000	0.5404	33.0000	0.874	162.0000	0.7229
12S	**28S**	63	115951.0000	0.2921	49.0000	0.8199	189.0000	0.4381
12S	**CO1**	42	40098.0000	0.5371	26.0000	0.6702	122.0000	0.5247
12S	**CO2**	59	100824.0000	0.3324	41.0000	0.7356	154.0000	0.3935
12S	**$CO1_{FOL}$**	42	40624.0000	0.5441	29.0000	0.7475	128.0000	0.5506
12S	**H3**	78	246716.0000	0.2592	49.0000	0.6552	241.0000	0.409
16S	**18S**	52	103992.0000	0.5763	37.0000	0.7585	211.0000	0.6532
16S	**28S**	74	367052.0000	0.4785	50.0000	0.7063	244.0000	0.447
16S	**CO1**	51	84975.0000	0.51	29.0000	0.6075	152.0000	0.4837
16S	**CO2**	66	247185.0000	0.5139	43.0000	0.6845	232.0000	0.5022
16S	**$CO1_{FOL}$**	51	73872.0000	0.4434	35.0000	0.7332	155.0000	0.4933
16S	**H3**	92	927028.0000	0.4976	59.0000	0.6646	336.0000	0.4501
18S	**28S**	54	106740.0000	0.5063	37.0000	0.7288	206.0000	0.6023
18S	**CO1**	15	406.0000	0.4479	9.0000	0.7682	21.0000	0.5246
18S	**CO2**	24	3101.0000	0.4379	12.0000	0.5778	48.0000	0.512
18S	**$CO1_{FOL}$**	13	274.0000	0.5782	8.0000	0.8238	16.0000	0.526
18S	**H3**	46	44915.0000	0.4133	31.0000	0.7244	165.0000	0.6173
28S	**CO1**	49	73265.0000	0.5186	25.0000	0.5461	154.0000	0.5237
28S	**CO2**	57	130025.0000	0.4936	38.0000	0.7066	168.0000	0.452
28S	**$CO1_{FOL}$**	25	7165.0000	0.8489	190.000	0.8732	87.0000	0.8666
28S	**H3**	83	432317.0000	0.3526	56.0000	0.7019	247.0000	0.3825
CO1	**CO2**	47	72326.0000	0.6075	31.0000	0.7084	147.0000	0.5299
CO1	**$CO1_{FOL}$**	33	14161.0000	0.5178	24.0000	0.8056	86.0000	0.5398
CO1	**H3**	64	266061.0000	0.6283	44.0000	0.7239	253.0000	0.5737
CO2	**$CO1_{FOL}$**	36	17338.0000	0.4417	20.0000	0.6104	79.0000	0.4323

CO2	**H3**	73	434395.0000	0.5985	51.0000	0.7309	282.0000	0.5258
$CO1_{FOL}$	**H3**	47	728.390.000	0.6118	33.0000	0.6075	172.0000	0.6200
MEAN				0.4991		0.7144		0.5259

Sup. Table 8b) Tree distances between FULLnuc and REDnuc datasets Maximum Likelihood (with linked branch length model) and Bayesian trees.

Tree1	Tree2	Quartet	Quartet to Yule Avg	RF(0.5)	RF(0.5) to Yule Avg	Matching Split	MatchingSplit to Yule Avg
FULLnuc ML	**REDnuc ML**	56206.0000	0.0043	15.0000	0.1029	69.0000	0.0471
REDnuc BI	**FULLnuc BI**	150500.0000	0.0115	14.5.0000	0.0995	89.0000	0.0607

Sup. Table 8c) Tree distances between the trees obtained with the same dataset (FULLnuc or REDnuc) but different type of inference (Maximum likelihood vs Bayesian).

Tree1	Tree2	Quartet	Quartet to Yule Avg	RF(0.5)	RF(0.5) to Yule Avg	Matching Split	Matching Split to YuleAvg
REDnuc BI	**REDnuc ML**	218892.0000	0.0162	11.0000	0.0749	102.0000	0.0688
FULLnuc BI	**FULLnuc ML**	133756.0000	0.0099	15.5000	0.1056	109.0000	0.0735

Sup. Table 8d) Tree distances between FULLnuc spp tree and the three constrained tree searches.

Tree1	Tree2	Quartet	Quartet to Yule Avg	RF(0.5)	RF(0.5) to Yule Avg	Matching Split	Matching Split to Yule Avg
FULLnuc	Diapheromerinae constrain	620580.000	0.0047	8.0000	0.0549	45.0000	0.0307
FULLnuc	Lonchodidae constrain	2,038590.0000	0.1551	13.0000	0.0892	165.0000	0.1126
FULLnuc	Robertson et al., (2018) constrain	2,368134.0000	0.1802	19.0000	0.1303	149.0000	0.1017

3. Trait Evolution in Phasmids.

Macroevolutionary analyses provide new evidences of phasmids wings evolution as a reversible process.

Giobbe Forni*, Jacopo Martelossi*, Pablo Valero, Frank H. Hennemann, Oskar Conle, Andrea Luchetti and Barbara Mantovani.

* These authors contributed equally to this work.

Abstract - The concept that complex ancestral traits can never be re-acquired after their loss has grown popular since its initial formulation and it's often referred to as Dollo's law. Nonetheless, several macroevolutionary evidences - along with molecular ones - suggest instances where complex phenotypes could have been lost throughout a clade evolutionary history and subsequently reverted to their former state in derived lineages. One of the first and most notable rejection of Dollo's law is represented by wing evolution in phasmids: this polyneopteran order of insects - which comprises stick and leaf insects - has played a central role in initiating a long-standing debate on the topic. In this study, a new and comprehensive molecular phylogeny of over 300 Phasmatodea species is used as a framework for investigating wing's evolutionary patterns in the clade, taking into consideration several sources of uncertainty and all the methodological recommendations which have been proposed to test Dollo's law rejection. Macroevolutionary analyses support a dynamic and reversible evolution of wings, with multiple transitions to ancestral states taking place after their loss. Our findings suggest that neither wings or flight have acted as drivers of Phasmatodea species diversification and that brachyptery is an unstable state, when not co-opted for non-aerodynamic adaptations. We also explored the impact on our results of different assumptions relative to the probability of reversals and losses: we found that until reversals are assumed over 30 times more unlikely than losses, they are consistently inferred despite uncertainty in tree and model parameters. Our findings demonstrate that wings evolution can be a reversible and dynamic process in phasmids and contribute to shape our understanding of complex phenotypes evolution.

Keywords: macroevolution, phylogenetic comparative methods, Phasmatodea, wings, Dollo's law.

Introduction:

Traits are commonly lost during evolution and - although some changes can be easily reverted in the short time (Teotónio and Rose, 2000; Rebolleda and Travisano, 2019) - it can be argued that the loss of complex ones is irreversible over long time spans. This concept is often referred to as Dollo's law (Dollo, 1893), despite in its original formulation reversals of complex traits are considered possible as secondary convergence events (Gould, 1970). Nonetheless, the concept that complex structures lost in the line of evolution cannot revert back to their former state later in the lineage has become popular, due to its intuitiveness and the frequent examples supporting it (Collin and Miglietta, 2008). Although this evolutionary principle is still commonly accepted, a number of cases where it is apparently violated have been proposed.

Most challenges to Dollo's law comes from macroevolutionary approaches and include a large number of examples, such as shell coiling in limpets (Collin and Cipriani, 2003), compound eyes in ostracods (Syme and Oakley, 2012), sex and parasitism in mites (Klimov and OConnor, 2013; Domes et al. 2007), mandibular teeth in frogs (Wiens, 2011), limb evolution (Kohlsdorf and Wagner, 2006), eggshells and oviparity (Lynch and Wagner, 2010; Recknagel et al. 2018; Esquerré et al. 2020) in squamata. At the same time, molecular approaches are contributing to unravel possible mechanisms underlying reversible evolution of complex traits: examples of compensatory mutations which have been able to rescue the once-lost functionality of genes can be found (Esfeld et al. 2018) and it has been observed that reversal to a complex ancestral state could happen through changes in relatively few genes (Seher et al. 2012). Moreover, some observations challenge the concept that the genes associated to a lost trait should be no more under the effect of selection and decay; it has been observed how in some instances the molecular blueprint of a trait can persist despite its phenotypical absence, most likely due to pleiotropic constrains (Marshal et al. 1994; Carlini et al. 2013; Lammers et al. 2019).

Among the many challenges to Dollo's law, the evolution of wings in phasmids stands as one of the first and more iconic examples (Whiting et al. 2003). It also played a central role in rethinking Dollo's law and in initiating a long-standing debate on the topic (Collin and Miglietta, 2008). The polyneopteran order

Phasmida includes 454 genera and 3342 described species which are often mimics of twigs and leaves. Phasmids wings present a high level of anatomical disparity (**Fig. 1**): 40% of these insects are macropterous while the remaining are either brachypterous or apterous; differences can be found at all taxonomic levels, including between congeneric species (Zeng et al. 2020). Macropterous phasmid species present long wings which are able to sustain flight with different efficiency, depending on the wing to body size ratio; however, most of them are considered weak flyers, using wings mainly for controlling free-fall descents from tree canopies (Maginnis, 2006). Some lineages have also evolved non-aerodynamic functions for wings, such as aposematic coloration or stridulation capacity, typical of brachypterous forms.

Wings emerged early in the diversification of insects with frequent losses occurring during their evolutionary history (Wipfler et al. 2019). In 2003, Whiting et al. proposed that an event of wing loss took place in the lineage that led to extant phasmids with subsequent reversals restoring wings throughout their evolutionary history. Shortly after this claim several comments followed (Stone and French, 2003; Trueman et al. 2004) and nowadays - even if there is a growing consensus that lost traits can be re-acquired (Porter and Crandall, 2003; Wiens, 2011) - several methodological flaws have been exposed and many approaches have been proposed to avoid incorrect rejection of the Dollo's law. Root prior probability (Sauquet et al. 2017; Goldberg and Igić, 2008), trait-dependent diversification rates (Goldberg and Igić, 2008; Holland et al. 2020) and tree uncertainty (Rangel et al.; 2015, Bollback, 2006) have all to be considered when testing for the irreversible evolution of a trait.

Phylogenetic tests of trait irreversible evolution are frequently misled by inappropriate assignment of the character state distribution at the root, which should be avoided unless unequivocal evidences are available. Despite phylogenetic analyses support extinct Phasmatodea clades - such as Gallophasmatidae, Pterophasmatidae and Susumanioidea - as stem groups of modern stick and leaf insects, they don't allow to make conclusive considerations of the wing status of the MRCA of the extant species (Yang et al. 2019; Yang et al. 2020). Nonetheless, a stark difference can be observed between the wings of extinct stem phasmids and extant species: the latter have either reduced or absent tegmina while stem fossils specimens present both pairs of wings full-sized (Fig. 1). In insects, wings confer a

plethora of potential advantages, such as evading predators, dispersal and mate-finding (Goldsworthy, 2018), but partial reduction or complete loss can have adaptive value as well: wings loss has been related to increased female fecundity (Roff, 1994), cryptic capacity (Whiting et al. 2003) and also to tradeoffs in resource allocation between different anatomical structures (Maginnis 2006). Wings can be associated with sexual recognition and selection which can drive diversification dynamics (Arnqvist et al. 2000; Masta and Maddison, 2002; Singh and Singh, 2014) and different flight capabilities are associated to differential diversification rates across clades, with both dispersal reduction and increment acting as possible drivers of diversification (Ikeda et al. 2012; Waters et al. 2020; Misof et al. 2014). When the state of a character can affect the diversification rates of a lineage, it can drive the trait distribution along the phylogeny tips; this phenomenon can lead to strong biases in transition rates and ancestral state reconstruction when common MK models - which assume neutral character evolution - are used (Goldberg and Igić, 2008; Holland et al. 2020). Trait subject to Dollo's law are expected to become less frequent throughout the evolution of a lineage, but they can be instead rather frequent if they can drive diversification dynamics of a clade . Thus, the interplay between wings and phasmids diversification rates needs to be carefully considered when testing for the trait irreversible evolution. An additional obstacle in elucidating wings evolutionary patterns comes from the uncertainty concerning systematic relationships among the main clades of the order. Important progresses have been made in the last two decades thanks to numerous molecular phylogenies (Bradler et al. 2014; Robertson et al. 2018; Simon et al. 2019), yet monophyly is questioned at different taxonomic levels while the different phylogenetic hypotheses are conflicting and lack reliable support. This situation most likely reflects the ancient and rapid evolutionary radiation of the suborder Euphasmatodea, comprising all Phasmatodea with the exception of the Timema genus (Forni et al. 2020).

In this study we reconstructed the most comprehensive time-tree so far for phasmids and leveraged it to explore the evolutionary patterns of wings in the clade. We tested the hypothesis that reversals could restore the trait to its former state subsequently to its loss by taking advantage of multiple approaches, including different trait-coding strategies along with global (model-fitting) and local (ancestral state reconstruction and stochastic character mapping) approaches. We implemented all possible

recommendations proposed for testing Dollo's law rejection and took into account common biases in phylogenetic comparative analyses. Furthermore, we tested the impact on our findings of different assumptions about the probability of wings losses and reversals. By addressing these questions relatively to phasmid's wings, we used an iconic case to test to which extent the evolution of a complex trait can be a reversible process and to widen the knowledge of how phenotypes evolve.

Material and Methods:

Phylogenetics dataset:

Sample were collected, morphologically identified and preserved dry or in ethanol until molecular analyses were carried out. Genomic DNA was extracted from leg tissues of 111 individuals (**Sup. Table 1**), using the protocol of "Smarter Nucleic Acid Preparation" (Stratec). Eight molecular markers were amplified, consisting in: two mitochondrial PCGs (two fragments of the Cytochrome Oxidase subunit I and one of the Cytochrome Oxidase subunit II), two mitochondrial rRNAs (12S and 16S), two nuclear rRNAs (28Sand 18S) and one nuclear PCG (Histone subunit 3). PCR reactions were carried out according to standard protocols; primers sequences and annealing conditions can be found in **Sup. Table 2.** PCR products were screened through 1% agarose gel electrophoresis and purified using ExoSAP PCR Product Cleanup Reagent (Thermofisher). Amplicons were Sanger sequenced by Macrogen Europe Lab. Chromatograms were inspected with SeqTrace 0.9.0 (Stucky 2012) and the resulting sequences were visualized using Aliview v1.26 (Larsson 2014). BLASTn (Zhang 1997) on NCBI Genbank database was used to search for potential contaminants and all the sequences produced have been submitted to GenBank under the accession number XXXX-YYYY. Complementary sequences were obtained from unpublished inhouse projects on Euphasmatodea systematics under the accession number MN449491-MN449962 and MT077516-MT077845. Complementary sequences were obtained relative to all specimens for which a species level identification was available from the following papers: Whiting et al. (2003), Buckley et al. (2009), Buckley et al. (2010), Bradler et al. (2014), Bradler et al. (2015), Goldberg et al. (2015), Robertson et al. (2018) and Glaw et al. (2019) plus outgroup sequences belonging

to Notoptera (Grylloblattodea + Mantophasmatodea; Damgaard et al. 2008; Jarvis and Whiting 2006) (**Sup. Table 1**). Taxonomical annotation - following Phasmida species file nomenclature (http://phasmida. speciesfile.org/) - are provided in **Sup. Table 1.** We did not consider Embioptera - which would represent the true sister group of Phasmatodea - as previous analyses showed that their inclusion introduces long-branch attraction artifacts (Song et al. 2016).

Phylogenetic hypothesis and divergence time estimate:

All sequences were aligned using Mafft 7.402 (Katoh and Standley, 2013): PCGs were translated to amino-acids using AliView v1.26, aligned using the "L-INS-i" algorithm and subsequently retro-translated into nucleotides. rRNAs were aligned using the "--X-INS-i" algorithm to take into account their secondary structure. Gblock v. 0.91b (Castresana, 2000) was used to exclude possible misaligned positions, selecting the codon flag for PCGs and the nucleotide one for rRNAs. Others parameters were set as follow: minimum number of sequences for a conserved position, 50% (PCGs and rRNA) of sequences included in the alignment; minimum number of sequences for a flanking position, 70% for PCGs and 60% for rRNAs; maximum number of contiguous no conserved position, 8 for PCGs and 10 for rRNAs; minimum length of a block, 10 for PCGs and 5 for rRNAs; allowed gap position, *all* for PCGs and *with half* for rRNAs. Each MSAs has been then concatenated using Phyutility v. 2.2 (Smith and Dunn 2008). All subsequent phylogenetic inferences were performed on the Cipres Science Gateway (http://www.hylo.org/portal2), using XSEDE (Miller et al 2010).

All Maximum Likelihood (ML) inferences were performed using IQ-TREE 1.6.12 (Nguyen et al. 2015), the concatenated alignment was partitioned *a priori* by gene and codon position. The best-fit partitioning scheme and evolutionary model were chosen using ModelFinder (Kalyaanamoorthy et al. 2017) according to the BIC score, allowing partition merging and linking branch length proportionally (**Sup. Table 3**). For each tree search the number of unsuccessful iterations to stop was set up to 500 (-nstop); branch support was estimated with 1000 UltraFast parametric bootstrap replicates (Minh et al. 2013) with 5000 maximum number of iterations. Preliminaries ML inferences have been conducted to identify rouge taxa with RogueNaRok (Aberer et al. 2013) until no rogue taxa were found, and 240 ML tree

searches were performed using the same parameters indicated above and a semi-random number between 0.1 and 1 as perturbation strength for each inference. Finally, the best ML tree was chosen comparing the *-lnL* values of each run.

Divergence times estimation, tree inference and model averaging were jointly performed with Beast2 (Bouckaert et al. 2014) and the bModelTest package (Bouckaert and Drummond, 2017), using a fully Bayesian framework. In this case the concatenated alignment was partitioned *a priori* by gene with unlinked site models and linked tree and clock models; a relaxed clock with a lognormal distribution and a Birth-Death model were used as clock and tree priors, respectively. We used 3 fossils with an unambiguous placement as calibration points (**Sup. Table 4**). The stem group of Phylliidae was calibrated without including any sister group (*use-originate* option in BEAST) as this subfamily position is still debated, and we did not want to force any specific phylogenetic hypothesis. For all calibration points an exponential distribution was chosen as prior distribution with a minimum hard bound set up to the age of the fossil and a soft maximum bound of 410 Mya. The ML tree was linearized with the *chronopl* function of the R package ape v.5 (Paradis and Schliep, 2019) and used as the starting tree for the Bayesian Inference (BI). Two independent chains were run for 150 million generations and sampled every 7000 states. After convergence and adequate ESS were assessed (> 200) with Tracer v.1.7.1 (Rambaut, et al. 2018), log files and tree files were combined with LogCombiner v. 2.6.2 and a 25% conservative burn-in was removed. 21,432 trees were summarized in a maximum clade credibility tree (MCC) keeping the median node heights in TreeAnnotator v. 2.4.2.

The topologies of ML and BI inferred trees were compared through the TreeCmp web tool (Bogdanowicz et al. 2012) using four different topological distance metrics for rooted trees: (1) Triples (Critchlow et al. 1996), (2) Robinson-Foulds based on clusters (Robinson and Foulds, 1981), (3) Matching Pair (Bogdanowicz and Giaro, 2017) and (4) Matching Cluster (Bogdanowicz and Giaro, 2013). Each metric was normalized to indicate whether two trees are more dissimilar than two random trees with the same number of tips: values of 1 indicate that the two trees are no more similar than two random trees, values of 0 indicate the same topology.

Comparative analyses:

The morphological datasets of wings states were compiled from all available data, including images on Phasmida Species File portal (http://phasmida.speciesfile.org; last accessed in July 2020), and from collected specimens. We considered wings as a species-level trait and in the case of sexual dimorphism we considered the more "complex" structure between the two sexes (*e.g.* a species with apterous females and brachypterous males was considered as brachypterous). In a first dataset, wings morphology was coded as a 2-states trait, with presence as 1 (including brachyptery) and absence as 0. Brachyptery was taken into consideration with an additional 3-states dataset, where 0: apterous, 1: brachypterous and 2: macropterous species. Both datasets are provided in **Sup. Table 1.**

For both the Maximum Clade Credibility (MCC) tree and 1000 trees randomly sampled from the tree distribution provided by BEAST, we pruned the outgroup species and a single ingroup species, *Dimorphodes prostasis*, for which no conclusive morphological information was available.

Comparative analyses considering two states (winged, apterous):

To test the reversible evolution of wings (*i.e.* the rejection of Dollo's law) on the 2-states dataset, we used a model comparison approach, evaluating three common MK models through the *fitDiscrete* function of the R package geiger v2.0 (Pennell et al. 2014) and using the MCC tree. These were: (1) Equal Rates (**ER**), (2) All Rates Different (**ARD**) and (3) Loss Only (**LO**). As an additional null hypothesis to the LO model, we constrained the root as winged in a Loss Only model (**WR-LO**) through the function *rayDISC* of the corHMM package (Beaulieu et al. 2013). It has been suggested to evaluate the latest model when testing Dollo's law rejection, as a root fixed to the "complex" character is a logical requirement when evolution is truly irreversible (Lewis, 2001; Goldberg and Igić, 2008). All the models were then compared looking at the resulting AICc score.

To take into account tree uncertainty in model selection and parameters estimations we used the *influ_discrete* function from the R package sensiPhy v0.8.5 (Paterno et al. 2018), fitting all previously evaluated models to 1000 randomly sampled trees from BEAST posterior distribution.

Ancestral state reconstructions (ASRs) were performed with the best-fit model of evolution and the other models, excluding WR-LO which resulted equal to LO. The *rayDISC* function and a prior root probability weighted accordingly to the method of Maddison et al. (2007) and FitzJohn et al. (2009) were used, so that root states probability is weighted according to their conditional probability given the data. As this assumption can greatly influence ASR and parameters estimation (Sauquet et al. 2017; Goldberg and Igić, 2008) we performed two additional ASRs with ER and ARD models and a flat prior, so that root state is equally likely to be in state 0 and 1.

To take into account parameter and tree uncertainty in ASR we performed stochastic character mapping (SCM) using the *make.simmap* function, implemented in Phytools (Revell, 2012), on the MCC tree under the best-fit model with 100 simulation (here referenced as 100Sim) and across 100 trees with one iteration (100Trees).

As previous approaches leveraged no information on transitions rates between different states other than the trait distribution on the tips of the tree and the tree itself, we explored the impact of different assumptions of the probability of wings reversals on our analyses. We assumed a diminishing probability of reversals compared to loss (from 1:1 to 1:500 with an interval of 10, corresponding to an ER model and an approximation of a LO one) in conjunction with loss rates between 0.001 and 0.015, with an interval of 0.001. These rates have been chosen to reflect the optimized values found in previous analyses (0.00375 for the ER model and 0.00905 for the LO; see Results). We carried out SCM using Phytools function jointly on 10 simulation and 10 random trees sampled from the BI posterior distribution for each combination of reversals relative probability and model rate (100 analyses for each combination, 75000 analyses in total).

Comparative analyses considering three states (macropterous, brachypterous, apterous):

When considering brachyptery, our comparative analyses followed the same approach used for the 2-states dataset, but taking into account a wider range of evolutionary models. We fitted the three unordered MK models previously used: (1) **ER**, (2) **ARD**, (3) **LO**, adding a (4) symmetrical model (**SYM**). We also took into consideration some additional biologically meaningful user-supplied MK models, in which one or more transitions were fixed to 0: (5) Ordered Loss-Only (**O-LO**) in which only losses

between proximal states were allowed; (6) Partial Reversal (**PR-BW**) where all losses and transitions from brachypterous to macropterous forms were allowed; (7) Partial Reversal (**PR-AB**) where all losses and transitions from apterous to brachypterous were allowed; (8) Ordered Complete Gain (**O-CG**), with all losses allowed plus transitions from apterous to brachypterous and from brachypterous to macropterous; (9) Ordered Full (**O-F**), where only transitions between proximal states were allowed.

As before, sensitivity analyses were run to quantify the impact of tree uncertainty in model selection and parameter estimation. For simplicity and due to computational limits, we excluded from these analyses the WR-LO, O-LO and PR-AB models (see Results for justification). As sensiPhy can handle only binary characters, we used a custom script to run the *fitDiscrete* function on 1000 trees randomly sampled from the BEAST posterior distribution.

As the outcome of model-fitting analyses was quite conclusive (see Results), we conducted ASRs using *rayDISC* with the best-fit model only and a "maddfitz" or a "flat" prior. Since we noticed that - differently from the analyses carried on with the 2-states dataset - the transition rates inferred by *rayDISC* and *fitDiscrete* were different, we also carried out an additional ASR to mirror previous model selection analyses with *fitDiscrete* (forcing transition rates calculated by *fitDiscrete* and a "maddfitz" root prior).

As for the binary dataset, tree and parameter uncertainty in ASR was taken into consideration carrying out two SCM analyses respectively with 100 simulations and 100 trees.

Diversification analyses:

To model the evolution of wings or flight capabilities along with their impact on phasmids diversification and to assess if trait-dependent diversification rates could bias our previous analyses, we performed trait-dependent diversification analyses using the Hidden State Speciation and Extinction framework (HiSSE) in R (Beaulieu and O'Meara, 2016). We rescored the 2-states dataset and produced two binary matrices: one with brachyptery coded as presence, as for the previous binary analyses, and a modified version with brachyptery coded as absence. Brachypterous species are indeed not able to sustain flight, therefore if loss and reversal of flight are correlated to variation in speciation rates, we will expect that brachypterous species follow the same pattern of apterous lineages. We tested a wide range of possible

macro-evolutionary scenarios generating several state-dependent diversification models in a HiSSe/BiSSe framework: (1) **Bisse**: a standard Bisse model where diversification rates depend on the wings state; (2) **Full Hisse:** diversification rates depend on one explicit character (wings) and a hidden one, differently for winged and apterous forms (*i.e.* a common full Hisse model with the hidden states A and B); (3) **Hisse**: diversification rates change between apterous and winged species only for one hidden state (in our case B, but is arbitrary since they are unknown), while for the other (A in our case) they result equal; (4) e (5) **Half Hisse 1** and **Half Hisse 2**: diversification rates depend on the explicit character and on a hidden state but only for respectively winged and apterous species. These models were tested against three null models: (6) **Bisse Null**: diversification rates constant through the tree and no hidden states; (7) **CID-2**: diversification rates change only as a function of two hidden state; (8) **CID-4:** a null model in which diversification rates are independent from the explicit trait but depend only on four hidden states, accounting for the same number of diversification parameters of a full hisse model (*i.e.* eight). The CID-2 and CID-4 model has been proposed to be two good null hypotheses in order to avoid common Type I errors (Beaulieu and O'Meara, 2016). As for previous analyses, all the models were compared through AICc scores and two additional ASRs were carried out.

Results:

Molecular and Morphological datasets:

The concatenated alignment resulted in 4112 positions and included 345 taxa, 322 of which were phasmids. Out of the 322 phasmid taxa, 92 represent new specimens and 230 were obtained from NCBI. The taxon sampling reflects the naturally observed species richness of Phasmatodea families: Lonchodidae were the most represented (28.8%), followed by Phasmatidae (25.5%) and Diapheromeridae (14.5%), while Agathemeridae and Damasippoididae accounted respectively for 0.6% and 0.3% of the species, beeing the smallest among the Phasmatodea families (**Sup. Fig**. **1**). We were able to obtain information about wing state for all Phasmatodea species with the exception of *Dimorphodes prostasis* Westwood, 1859, which has been excluded from comparative analyses. Overall,

180 species were identified as apterous (56%), 33 as brachypterous (10%) and 108 as macropterous (34%) (**Sup. Table 1**).

Phylogenetic hypothesis and divergence time estimate:

The best ML tree (-*lnL* = 172822) showed a general support of nodes by bootstrap values (BS), although some of the deepest nodes showed weak support (**Sup. Fig. 2**). The Bayesian MCC tree is in general agreement with the ML one, as pointed out by the tree distances (Normalized Triples = 0.0840; normalized Robinson-Fould cluster = 0.2652; normalized Matching pair = 0.1143 and normalized Matching cluster = 0.1226), with the exception of a few, unsupported nodes (**Fig. 2; Sup. Fig. 2)**. The major phasmids clades recovered by both tree inferences are entirely congruent, except for the position of the species *Macrophasma biroi* (Redtenbacher, 1908) and *Phasmotaenia lanyuhensis* Huang and Brock, 2001 (**Sup. Fig. 2**).

Timema resulted as the sister clade of Euphasmatodea, with maximum support. Within the Euphasmatodea suborder, the Pseudophasmatidae family resulted polyphyletic, being split into two clades: one in sister relationship with all remaining taxa, and another placed closer to Agathemeridae (**Fig. 2; Sup. Fig. 2**). Aschiphasmatidae and Phylliidae also resulted sister clades, with high nodal support (BS = 97; PP = 100); though, this group shows variable position and nodal support among ML and BI inferences (**Sup. Fig. 2**). The family Phasmatidae clearly resulted polyphyletic, the subfamily Clitumninae being distantly related from Lanceocercata and Cladomorphinae subfamilies (**Fig. 2; Sup. Fig. 2**). The subfamily Diapheromerinae results polyphyletic in the MCC tree as well, with one cluster in an unsupported sister relationship with the Prisopodidae, and the other as the sister of the latter two (**Fig. 2; Sup. Fig. 2)**. Instead, in the ML tree the subfamily is recovered as monophyletic with high nodal support value (BS = 94) and in sister relationship with the Prisopodidae (BS = 97; **Sup. Fig. 2**).

The divergence between the two suborders Timematodea and Euphasmatodea is recovered at 180 million years ago (Mya) (95% HPD = 122-249 Mya), with the Euphasmatodea radiation beginning 108 Mya (95% HPD = 75-148 Mya) and the diversification of major lineages happening in the subsequent 30 million years (**Fig. 2; Sup. Fig. 2**).

Comparative analyses considering two states (winged, apterous):

For the 2-states dataset, the best-fit on the Bayesian MCC tree resulted to be the ER model (AICc = 280.13, **Sup. Table 5A**) with a transition rate between the two states equal to 0.00375 (**Sup. Fig. 3A**), followed by the ARD (AICc = 282.14), with transition rates close to those of the ER model. LO and WR-LO models were equally not supported with a ΔAICc = 20.63**.**

Tree sensitivity analyses highlighted a strong preference for the ER model: it was recovered as the best-fit for 804 trees out of the 1000 taken from the BEAST posterior distribution, and its AICc distribution resulted significantly lower than all the others (Kruskal-Wallis test, $p < 0.001$; post-hoc pairwise Wilcoxon test with Bonferroni correction, $p < 0.001$). For the remaining 196 trees, the ARD model was found the best-fitting (**Fig. 3A**).

Under the best-fit model, the root of the MCC tree was reconstructed as apterous with the maximum posterior probability (PP). Eight ancient (*i.e.* along internal branches) and 13 recent (*i.e.* along terminal branches) reversals of wings were inferred, while wing losses were inferred respectively seven and eight times (**Fig. 2B**). Similar results were obtained when the ASR was carried out under the ARD model, due to the similar inferred transition rates (**Sup. Fig. 4**). Instead, in a scenario coherent with Dollo's law, 55 independent losses of wings were necessary to explain the distribution of the trait throughout the tree (**Sup. Fig. 4**). Even using a flat prior, the root was recovered as apterous for both ER and ARD models ($PP > 75$) with transition rates, AICc values and the overall ASRs approximately equal to those obtained in previous analyses (**Sup. Table 6A, Sup. Fig. 4**).

Stochastic character mapping highlights a weak impact of parameter, topology and branch lengths (**Fig. 4**): both 100Sim and 100Trees analyses recovered a high preference for a apterous root (**Fig. 4A**) and a higher number of reversals than losses (**Fig. 4B**). An average of 50.61 and 52.71 changes between states were inferred respectively for the 100Sim and 100Tree, with an average of 28.4 and 30.38 being reversals to winged forms.

When different assumptions on the relative probability of reversals (ratio of wing reversal to loss rates) are tested in a SCM framework, it can be observed that up to a ratio of 1:30 reversals are consistently

inferred in stochastic character mapping (at least one reversal is inferred in each of the 100 SCM analysis) independently from the absolute values of the evolutionary model parameters (**Fig. 5**). The latter strongly affect the consistency of reversals inference when the ratio of reversal to loss is assumed lower than 1:30: at low model rates, reversals are inferred even when assumed to be five hundred times less likely than losses. Remarkably, no evolutionary model considered in this analysis consistently rules out the possibility of a reversal, which are always inferred under certain tree and parameter conditions.

Comparative analyses considering three states (macropterous, brachypterous, apterous):

When implementing brachyptery as a third state, ARD resulted the best-fitting model (AICc=280.1303) on the Bayesian MCC tree, followed by SYM (ΔAICc = 6.05) and ER (ΔAICc = 9.26) (**Sup. Table 5B**). As for previous analyses, forcing the root as macropterous and providing a loss-only model (WR-LO) did not cause changes in the AICc values with respect to the loss only model (LO), indicating the scarce impact of this assumption on the analysis. Again, LO and WR-LO models were rejected with a ΔAICc = 33.12. The two models O-LO and PR-AB resulted as the less-supported, with a ΔAICc of 33.671 and 43.408, respectively; therefore, we discarded them from subsequent analyses. Transition rates of the ARD model are shown in **Sup. Fig. 3B**. The highest transition rates are those which describe the shifts from brachyptery to either macropterous (0.0121) or apterous (0.0154) forms, while the rates from apterous to fully macropterous forms and *vice-versa* resulted to be closely matching (0.00181 and 0.00199).

Sensitivity analyses highlight a strong preference of the ARD model which was recovered as the best-fit for 987 of the 1000 trees randomly sampled from the BI posterior distribution, with a AICc distribution significantly lower with respect to all others (Kruskal-Wallis test, $p < 0.001$; post-hoc pairwise Wilcoxon test with Bonferroni correction, $p < 0.001$), while the SYM model was preferred for the remaining 18 trees (**Fig. 3B**). Rates of transitions departing from bachyptery still resulted to be the highest ones (**Fig. 3C**).

The ASRs carried out on the Bayesian MCC tree under the ARD model resulted in different outcomes depending on the parameters used (**Sup. Fig. 5**), but in all reconstruction reversal to macropterous or

brachypterous forms from apterous ancestors are inferred. Using a flat prior, similar transition rates were recovered by *rayDISC* and *fitDiscrete*; in both analyses the root is recovered as brachypterous with a PP of 0.77 and 0.61, respectively (**Sup. Table 6B, Sup. Fig. 5**). In these two ASRs wings reversal from apterous to brachypterous or macropterous forms are present, but only as recent events on terminal branches. Though, a large number of transitions from brachyptery to both macropterous and apterous are recovered, reflecting the low rates between apterous and brachypterous forms and the high rates that move away from partially developed wings. Using a maddfitz prior with transition rates inferred directly by *rayDISC*, the root is recovered as apterous with a PP of 0.97 and reversals are inferred also throughout internal branches. These results reflect the different transition rates inferred with respect to previous analyses: even though the highest rates are still those which depart from brachyptery, they are lower than the previous ones (**Sup. Table 6B**). In this way the model favored a apterous state for deepest nodes and more transitions to brachyptery (**Sup. Fig. 5**). However, the two ASRs which inferred the root as brachypterous are similarly supported with a slight preference for the one obtained with fixed parameters (ΔAICc = 0.46) (**Sup. Table 6B**).

Regarding SCMs, a brachypterous root was preferred in both analyses (**Fig. 4C**) and an average of 90.04 and 90.84 changes between states were inferred for the 100Sim and the 100Trees analyses, respectively. In both of them, the highest average number of transitions were recovered to be from brachypterous to apterous forms (31.11 for the 100Sim and 31.79 for the 100Trees), followed by transition from brachypterous to fully macropterous (respectively 24.55 and 24.28) (**Fig. 4D**). The same results are reflected by the mean total time spent in each state, with brachyptery being the less represented (15% and 16% for the 100sim and the 100 trees, respectively; **Fig. 4E**).

Diversification analyses:

All our diversification analysis preferred a Hisse model - where diversification rates change between apterous and macropterous species only when in the hidden state B - whereas the Bisse model - which underlines that speciation rates are mainly driven by wings - was always rejected with a minimum ΔAICc of 56 (**Sup. Table 7 A-B**). Under the best-fit model, all extinction fractions, even if allowed to vary freely,

resulted equal among different states (**Sup. Table 7C**). The net diversification rate in the hidden state B were similar when considering either macropterous or apterous forms; moreover, the same can be observed when considering brachyptery as absence of wings (in term of flight ability) (**Sup. Table 7C**). Results of the two ASRs are concordant with those previously obtained with the binary datasets and MK models, as expected by the low impact of wings and flight capability in diversification dynamics. Both of them recovered a apterous root and several reversals along the phylogeny (**Sup. Fig. 6**).

Discussion:

This study represents the most comprehensive overview on Phasmatodea phylogeny until now and highlights once more the discordances of morphological taxonomy and molecular phylogenetics, with paraphyly and polyphyly widespread among the majority of families and subfamilies. In agreement with previous studies (Robertson et al.; 2018; Bradler et al.; 2015; Buckley at al.; 2009), our analysis recovered the polyphyly of the Phasmatidae family - with Clitumninae not belonging to the clade formed by Cladomorphinae and Lanceocercata. At variance with recent studies (Robertson et al. 2018; Bradler et al. 2015), Pseudophasmatidae are also found as polyphyletic, an hypothesis which deserves future consideration. Another contentious point is which clades diverged earliest from all other Euphasmatodea: previous analyses proposed a sister relationship of either Aschiphasmatinae or Diapheromerinae with the rest of Euphasmatodea (Robertson et al. 2018; Simon et al.; 2019; Buckley et al.; 2009; Bradler et al.; 2015), while in our phylogenetic hypothesis one of the two Pseudophasmatinae clades is the earliest to diverge and Aschiphasmatidae are in a sister relationship with Phylliidae. Even if it is generally recognized that phasmids taxonomy should be revised to better reflect molecular phylogenetics, in the light of the high uncertainty that pervades their systematics, we suggest caution in proposing new high-level classifications. While an increase in the number of loci considered might represent an important step forward in providing a solid phylogenetic framework for the order, leveraging sensitivity analyses and stochastic character mappings let us consider a wide range of different phylogenetic hypotheses, so that our results shouldn’t be biased by any specific one.

The divergence times we recovered are significantly older than those previously proposed in studies focusing on Phasmatodea. Previous estimates of the divergence between Euphasmatodea and *Timema* ranged between ~95 Mya and ~122 Mya (Buckley et al. 2009; Bradler et al. 2015; Simon et al. 2019, Robertson et al.; 2018), while our estimate is ~174.5 Mya (95% HPD: 1122.63-249.1 Mya). The same holds for the Euphasmatodea crown node, which is here retrieved in the Mid-Cretaceous at 105.05 Mya (95% HPD: 75-148 Mya), while in previous analyses ranged from ~50 to ~80 Mya. These inconsistencies can be explained by the usage of a calibrations scheme which considered also the outgroup, while previous analyses used only calibrations internal to Phasmatodea and potentially underestimated divergence times (Buckley et al. 2010; Bradler et al. 2015). The differences can also be due to prior specifications for node calibration, as we preferred to use wider upper boundaries of prior distribution compared to previous analyses, in order not to force any hypothesis. Paleontological evidence (Yang et al. 2019) are coherent with our results, indicating the Mid-Cretaceous as an important period of phasmids diversification, possibly related to the contemporary rapid diversification of angiosperms (Peris et al. 2017). Our results also confirm a rapid radiation of Euphasmatodea, which can explain the lack of support and consistency of the previous phylogenetic hypotheses.

A first hint of the reversible evolution of phasmids wings is simply provided by the distribution of the trait states across the phylogeny (**Fig. 2**). Although many apterous, brachypterous and macropterous species form well supported clades, other include a mixture of different states: for example, several macropterous species, like *Bacteria ploiaria* (Westwood, 1859), *Cranidium gibbosum* (Burmeister, 1838) and *Lobofemora bidoupensis* Bresseel and Constant, 2015, are found in clades mainly including apterous taxa. Nonetheless, we carefully excluded rogue taxa from our analyses and their placement appears to be coherent with morphological taxonomy.

Model selection strongly supported models in which reversals occurred (*i.e.* from apterous to winged forms in the 2-states analyses; from apterous to bachypterous and macropterous forms and/or from brachypterous to macropterous ones in the 3-states analyses; **Fig. 3)**. These outcomes were consistent despite uncertainty in the tree topology and branch length, while models consistent with Dollo's law - *i.e.* LO and WR-LO - were never supported for any of the 1000 trees sampled from the BI posterior

distribution. Moreover, brachyptery resulted to be the most unstable state, showing the highest rates of transition departing from it compared to other states. Similar outcomes are provided by SCM analyses: brachyptery is recovered as the state with the greatest number of transitions moving away from it and with the less time spent in, while transitions from apterous and brachypterous forms to macropterous ones are consistently inferred (**Fig. 4**). Altogether, these evidences may reflect the results obtained by Zeng et al. (2020), who proposed a fitness valley of intermediate size wings between two adaptive peaks represented by apterous and macropterous taxa (Stroud and Losos, 2016). Complete development or complete loss of wings convey direct benefits by themselves - such as dispersal and defensive capability, increased mimicry capacity or fecundity in females (Roff, 1994; Whiting et al.; 2003; Zeng et al. 2020) - while brachypterous wings may be positively selected and maintained only when co-opted for non-aerodynamic purposes. For all the brachypterous species we considered, wings are associated to functions such as aposematism or stridulation, with very few exceptions such as *Hypocyrtus ornatissimus* (Brunner von Wattenwyl, 1907).

Despite the known limitations of ASR in inferring precise number and position of transitions (Goldberg and Igić, 2008; Duchêne and Lanfear, 2015), our analyses aim to a global evaluation of wings evolutionary patterns in the clade: reversals to brachypterous and macropterous forms were inferred in all analyses and we never recovered a macropterous root. Using the 2-state character coding under the best-fit model of evolution, extant Phasmatodea MRCA was always reconstructed as apterous even when considering different root's prior probabilities, variation in diversification rates and uncertainty in evolutionary model parameters and tree. ASRs using the 3-states coding strategy recovered either a apterous or brachypterous MRCA of Phasmatodea, depending on the combination of transition rates and root assumptions, yet always rejecting a macropterous MRCA. The two ASRs where a brachypterous MRCA is found and Euphasmatodea have diversified as brachypterous-like are preferred by the AICc and also SCMs point to a brachypterous root. Yet, reversals are inferred only on terminal branches, an outcome that does not have a clear biological sense as it implies an unidentified selective pressure towards wings re-evolution only in recent time.

An extant phasmid MRCA which lacked fully developed wings is coherent with paleontological evidence: extinct species representing the stem-group of all extant Phasmatodea (*e.g.* Susumanioidea, Archipseudophasma and Pterophasmatidae) present wings with tegmina longer than all other extinct and extant Euphasmatodea taxa (Yang et al. 2019; Yang et al. 2020; **Fig. 1**). It is therefore possible to hypothesize that the ancestor lineage of extant phasmids presented two fully developed wing pairs and experienced either a reduction or a loss of wings. Then, subsequent reversals happened in the lineages leading to extant forms, which restored the structure in a partially different form; such differences between the derived state of a trait and its ancestral form may infact represent an evidence of its reversal (Cronk, 2009; Recknagel et al. 2018). Phenotypes are rarely derived from single or few genes, often the resulting from a large number of them: modifications in derived traits with respect to their ancestral form could be explained by the possible co-option of novel genes and the decay of others, along with the preservation of pleiotropic ones due to selection for other traits. In insects, wings and legs are tightly linked in a developmental perspective: their imaginal discs generated from the same group of cells and the genetic pathway that guide both structure development is largely shared (Kim et al. 1996; Cohen et al. 1993). Moreover, it has been observed that leg regeneration during phasmids development leads to smaller wings and weaker flight capability (Maginnis, 2006) and that the neural structures and their functional connectivity necessary to sustain flight are conserved also in apterous forms, showing how loss of flight is not correlated to loss of associated muscles and innervations (Kutsch and Kittmann, 1991). Thus, extant phasmids wings - as other reversals - could blur the boundaries between reversion and novelty, presenting a trait which is only partially built on the same genetic blueprint which produced the once lost structure.

Genomic and transcriptomic studies are contributing to elucidate the outcomes of trait loss and possible mechanisms associated to reversal (Seher et al. 2012; Carlini et al. 2013; Esfeld et al. 2018; Lammers et al. 2019) which can be - theoretically - used as informative priors to be applied in the framework of comparative methods. Standard approaches do not consider any prior assumptions on the mechanism of evolution, leveraging transitions rates estimated on the basis of trait distribution at the tips of the phylogeny and the tree itself. When there are no particular expectations on the relative probability of

transitions between states, this seems a valid approach; however, an equal probability of losing and reverting back to a complex structure - *i.e.* made by several integrated parts - represents a strong deviation from common expectations and assumptions. While the majority of our analyses found comparable rates of losses and reversals, as can be observed from global and local approaches on the 2-states character coding strategy, we also tested the impact of different assumption on the relative probability of reversal compared to loss (**Fig. 5**). As no study has ever explored possible mechanisms of wings reversal in phasmids, we arbitrarily tested different ratios of wings reversal to loss (from 1:10 up to 1:500, in conjunction with absolute values of the parameters consistent with their optimization in the ER and LO models. Despite a big effect of the model parameter absolute values can be observed, until a relative probability of 1:30 is assumed, reversals are consistently inferred. When reversals are assumed as more unlikely events, they are no more inferred consistently; however, even when reversals are considered highly unlikely, a scenario of irreversible evolution is never consistently supported. Previous debates on the possible reversible evolution of phasmids wings found very contrasting results, with the absolute values of the evolutionary model parameters playing a major role in defining a probability threshold above which the reversible evolution scenario is no longer supported (Whiting et al. 2003: 1:1500; Trueman et al. 2004: 1:13). Despite there is no approach to determine such a threshold (Stone and French, 2003), we showed that our findings are solid to the plausible expectation of reversals being more unlikely events than losses.

Common MK models assume neutral character evolution and systematic biases can arise when a trait has the potential to influence probabilities of lineage speciation and extinction. *SSE methods can outperform common MK models in state-dependent diversifications scenarios, allowing parameters to depend on the state of the character (Goldberg and Igić, 2008; Holland et al. 2020). Our analyses show the lack of strong trait-dependent diversification trends and the ASRs under the HISSE model are substantially concordant with the one obtained using common MK models (**Sup. Fig. 6**), allowing us to have more confidence in the outcome of the ASR with the 2-states dataset. Our analyses challenge the hypothesis that loss of wings in stick insects is correlated to increasing in speciation rates (Zeng et al.

2020), suggesting that both wing and flight capabilities don't represent main drivers of diversification in the clade.

Conclusions:

Altogether, our findings support a dynamic and reversible evolution of phasmids wings: our analyses inferred either the absence or an extreme reduction of wings in the MRCA of extant Phasmatodea, with multiple reversals subsequent restoring the once lost structures. Neither wings or flights are found to significantly impact lineage diversification in the clade and brachyptery is recovered as an unstable state. Despite having leveraged multiple complementary approaches, our evidences are limited to a macroevolutionary framework and complementary evo-devo, genomic and transcriptomic approaches should follow.

In our opinion, rejecting Dollo's law in this scenario is a matter of how to consider the homology of the derived and ancestral trait states. Yet, independently from the different perspectives which can be adopted on the topic, phasmids wings represent an extraordinary example of how dynamic the evolution of complex traits can be.

Aknowledgments:

The authors would like to thank Mattia Ragazzini for the help in revising the carachter-coding tables. We also wish to thank Taiping Gao and Hongru Yang for permitting the usage of the photo of *Aclistophasma echinulatum* fossil.

References:

Aberer A.J., Krompass D., Stamatakis A. 2013. Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice. Syst. Biol. 62:162-166.

Arnqvist G., Edvardsson M., Friberg U., Nilsson T. 2000. Sexual conflict promotes speciation in insects. Proc. Natl. Acad. Sci. USA 97:10460-10464.

Beaulieu J.M., O'Meara B.C., Donoghue M.J. 2013. Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. Syst. Biol. 62:725-737.

Beaulieu J.M., O'Meara B.C. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Syst. Biol. 65:583-601.

Bogdanowicz D., Giaro K., Wróbel B. 2012. TreeCmp: Comparison of Trees in Polynomial Time. Evol. Bioinform. Online 8:EBO-S9657.

Bogdanowicz D., Giaro K. 2013. On a matching distance between rooted phylogenetic trees. Int. J. Appl. Math. Comput. Sci. 23:669-684.

Bogdanowicz D., Giaro K. 2017. Comparing Phylogenetic Trees by Matching Nodes Using the Transfer Distance Between Partitions. J. Comput. Biol. 24:422-435.

Bollback J.P. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7:88.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10(4):e1003537.

Bouckaert R.R., Drummond A.J. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 17(1):42.

Bradler S., Robertson J.A., Whiting M.F. 2014. A molecular phylogeny of P hasmatodea with emphasis on Necrosciinae, the most species-rich subfamily of stick insects. Syst. Entomol. 39:205-222.

Bradler S., Cliquennois N., Buckley T.R. 2015. Single origin of the Mascarene stick insects: ancient radiation on sunken islands? BMC Evol. Biol. 15:196.

Buckley T.R., Attanayake D., Bradler S. 2009. Extreme convergence in stick insect evolution: phylogenetic placement of the Lord Howe Island tree lobster. Proc. R. Soc. B Biol. Sci. 276:1055-1062.

Buckley T.R., Attanayake D., Nylander J.A., Bradler, S. 2010. The phylogenetic placement and biogeographical origins of the New Zealand stick insects (Phasmatodea). Syst. Entomol. 35(2):207-225.

Carlini D.B., Satish S., Fong D.W. 2013. Parallel reduction in expression, but no loss of functional constraint, in two opsin paralogs within cave populations of Gammarus minus (Crustacea: Amphipoda). BMC Evol. Biol. 13(1):89.

Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. 17:540-552.

Cohen B., Simcox A.A., Cohen S.M., 1993. Allocation of the thoracic imaginal primordia in the Drosophila embryo. Development 117:597-608.

Collin R. and Cipriani R. 2003. Dollo's law and the re–evolution of shell coiling. Proc. R. Soc. B Biol. Sci. 270(1533):2551-2555.

Collin R., Miglietta M.P. 2008. Reversing opinions on Dollo's Law. Trends Ecol. Evol. 23:602-609.

Critchlow D.E., Pearl D.K., Qian C. 1996. The Triples Distance for Rooted Bifurcating Phylogenetic Trees. Syst. Biol. 45:323-334.

Cronk Q.C.B. 2009. Evolution in Reverse Gear: The Molecular Basis of Loss and Reversal. Cold Spring Harb. Symp. Quant. Biol. 74:259-266.

Damgaard J., Klass K.D., Picker M.D., Buder G. 2008. Phylogeny of the Heelwalkers (Insecta: Mantophasmatodea) based on mtDNA sequences, with evidence for additional taxa in South Africa. Mol. Phylogenetics Evol. 47:443-462.

Dollo, L. 1893. The laws of evolution. Bull. Soc. Bel. Geol. Paleontol, 7:164-166.

Domes K., Norton R.A., Maraun M., Scheu S. 2007. Reevolution of sexuality breaks Dollo's law. Proc. Natl. Acad. Sci. USA 104(17):7139-7144.

Duchêne, S., Lanfear, R. 2015. Phylogenetic uncertainty can bias the number of evolutionary transitions estimated from ancestral state reconstruction methods. J. Exp. Zool. B 324:517-524.

Esfeld K., Berardi A.E., Moser M., Bossolini E., Freitas L., Kuhlemeier C. 2018. Pseudogenization and resurrection of a speciation gene. Curr. Biol. 28(23):3776-3786.

Esquerré D., Brennan I.G., Catullo R.A., Torres-Pérez F., Keogh J.S. 2019. How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America's most species-rich lizard radiation (Squamata: Liolaemidae). Evolution *73*(2):214-230.

FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595-611.

Glaw F., Hawlitschek O., Dunz A., Goldberg J., Bradler S., 2019. When Giant Stick Insects Play With Colors: Molecular Phylogeny of the Achriopterini and Description of Two New Splendid Species (Phasmatodea: Achrioptera) From Madagascar. Front. Ecol. Evol. 7:105.

Goldberg E.E., Igić B. 2008. On phylogenetic tests of irreversible evolution. Evolution 62:2727-2741.

Goldberg J., Bresseel J., Constant J., Kneubühler B., Leubner F., Michalik P., Bradler S. 2015. Extreme convergence in egg-laying strategy across insect orders. Sci. Rep. 5:1–7.

Goldsworthy G.J. 2018. Insect flight. CRC Press.

Gould S.J. 1970. Dollo on Dollo's law: irreversibility and the status of evolutionary laws. J. Hist. Biol. 3(2):189-212.

Holland B.R., Ketelaar-Jones S., O'Mara A.R. Woodhams M.D., Jordan G.J. 2020. Accuracy of ancestral state reconstruction for non-neutral traits. Sci. Rep. 10:7644.

Ikeda H., Nishikawa M., Sota T. 2012. Loss of flight promotes beetle diversification. Nat. Commun. 3(1):1-8.

Jarvis K.J., Whiting M.F. 2006. Phylogeny and biogeography of ice crawlers (Insecta: Grylloblattodea) based on six molecular loci: Designating conservation status for Grylloblattodea species. Mol. Phylogenetics Evol. 41:222–237.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin, L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587-589.

Katoh K., Standley D.M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30:772-780.

Kim J., Sebring A., Esch J.J., Kraus M.E., Vorwerk K., Magee J., Carroll S.B. 1996. Integration of positional signals and regulation of wing formation and identity by Drosophila vestigial gene. Nature 382:133–138.

Klimov P.B., OConnor B., 2013. Is permanent parasitism reversible? Critical evidence from early evolution of house dust mites. Syst. Biol. 62(3):411-423.

Kohlsdorf T., Wagner G.P. 2006. Evidence for the reversibility of digit loss: a phylogenetic study of limb evolution in Bachia (Gymnophthalmidae: Squamata). Evolution, 60(9):1896-1912.

Kutsch W., Kittmann R. 1991. Flight motor pattern in flying and non-flying Phasmida. J. Comp. Physiol. A 168(4):483-490.

Lammers M., Kraaijeveld K., Mariën J., Ellers J. 2019. Gene expression changes associated with the evolutionary loss of a metabolic trait: lack of lipogenesis in parasitoids. BMC genomics 20(1):1-14.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278.

Lewis P.O. 2001. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. Syst. Biol. 50:913-925.

Lynch V.J., Wagner G.P. 2010. Did egg-laying boas break Dollo's law? Phylogenetic evidence for reversal to oviparity in sand boas (ERYX : BOIDAE). Evolution 64:207-216.

Maddison W.P., Midford P.E., S.P. Otto 2007. Estimating a binary characters effect on speciation and extinction. Syst. Biol. 56:701-710.

Maginnis T.L. 2006. Leg regeneration stunts wing growth and hinders flight performance in a stick insect (Sipyloidea sipylus). Proc. R Soc. B. Biol. Sci. 273:1811-1814.

Marshall C.R., Raff E.C., Raff R.A. 1994. Dollo's law and the death and resurrection of genes. Proc. Natl. Acad. Sci. USA 91(25):12283-12287.

Masta S.E., Maddison W.P. 2002. Sexual selection driving diversification in jumping spiders. Proc. Natl. Acad. Sci. USA 99:4442-4447.

Miller M.A., Pfeiffer W., Schwartz T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). Presented at the 2010 Gateway Computing Environments Workshop (GCE), IEEE, New Orleans, LA, USA, pp. 1-8.

Minh B.Q., Nguyen M.A.T., von Haeseler A. 2013. Ultrafast Approximation for Phylogenetic Bootstrap. Mol. Biol. Evol. 30:1188-1195.

Misof B., Liu S., Meusemann K., Peters R.S., Donath A., Mayer C., Frandsen P.B., Ware J., Flouri T., Beutel R.G. Niehuis O. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346(6210):763-767.

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. Mol. Biol. Evol. 32:268-274.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-528.

Paterno G.B., Penone C., Werner G.D.A. 2018. sensiPhy: An r-package for sensitivity analysis in phylogenetic comparative methods. Methods Ecol. Evol. 9:1461-1467.

Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon L.J. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216-2218.

Peris D., Pérez-de la Fuente R., Peñalver E., Delclòs X., Barrón E., Labandeira C.C. 2017. False Blister Beetles and the Expansion of Gymnosperm-Insect Pollination Modes before Angiosperm Dominance. Curr. Biol. 27:897-904.

Porter M.L., Crandall K.A. 2003. Lost along the way: the significance of evolution in reverse. Trends Ecol. Evol. 18(10):541-547.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67(5):901.

Rangel T.F., Colwell R.K., Graves G.R., Fučíková K., Rahbek C., Diniz-Filho J.A.F. 2015. Phylogenetic uncertainty revisited: Implications for ecological analyses. Evolution 69(5):1301-1312.

Rebolleda-Gómez M., Travisano M. 2019. Adaptation, chance, and history in experimental evolution reversals to unicellularity. Evolution 73(1):73-83.

Recknagel H., Kamenos N.A., Elmer K.R. 2018. Common lizards break Dollo's law of irreversibility: Genome-wide phylogenomics support a single origin of viviparity and re-evolution of oviparity. Mol. Phylogenetics Evol. 127:579-588.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217-223.

Robertson J.A., Bradler S., Whiting M.F. 2018. Evolution of Oviposition Techniques in Stick and Leaf Insects (Phasmatodea). Front. Ecol. Evol. 6:216.

Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131-147.

Roff D.A. 1994. The evolution of flightlessness: Is history important? Evol. Ecol. 8:639-657.

Sauquet H., von Balthazar M., Magallón S., Doyle J.A., Endress P.K., Bailes E.J., Barroso de Morais E., Bull-Hereñu K., Carrive L., Chartier M., Chomicki G., Coiro M., Cornette R., El Ottra J.H.L., Epicoco C., Foster C.S.P., Jabbour F., Haevermans A., Haevermans T., Hernández R., Little S.A., Löfstrand S., Luna J.A., Massoni J., Nadot S., Pamperl S., Prieu C., Reyes E., dos Santos P., Schoonderwoerd K.M., Sontag S., Soulebeau A., Staedler Y., Tschan G.F., Wing-Sze Leung A., Schönenberger J. 2017. The ancestral flower of angiosperms and its early diversification. Nat. Commun. 8(1)1:10.

Seher T.D., Ng C.S., Signor S.A., Podlaha O., Barmina O., Kopp A. 2012. Genetic basis of a violation of Dollo's law: re-evolution of rotating sex combs in *Drosophila bipectinata*. Genetics, 192(4):1465-1475.

Simon S., Letsch H., Bank S., Buckley T.R., Donath A., Liu S., Machida R., Meusemann K., Misof B., Podsiadlowski L., Zhou X., Wipfler B., Bradler S. 2019. Old World and New World Phasmatodea: Phylogenomics Resolve the Evolutionary History of Stick and Leaf Insects. Front. Ecol. Evol. 7:345.

Singh A., Singh B.N. 2014. Role of sexual selection in speciation in *Drosophila*. Genetica 142:23-41.

Smith S.A., Dunn C.W. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24:715-716.

Song N., Li H., Song F., Cai W. 2016. Molecular phylogeny of Polyneoptera (Insecta) inferred from expanded mitogenomic data. Sci. Rep. 6:1-10.

Stone G., French V. 2003. Evolution: have wings come, gone and come again? Curr. Biol. 13(11):R436-R438.

Stroud J.T., Losos J.B. 2016. Ecological Opportunity and Adaptive Radiation. Annu. Rev. Ecol. Evol. Syst. 47:507-532.

Stucky B.J. 2012. SeqTrace: A Graphical Tool for Rapidly Processing DNA Sequencing Chromatograms. J. Biomol. Tech. 23:90-93.

Syme A.E., Oakley T.H., 2012. Dispersal between shallow and abyssal seas and evolutionary loss and regain of compound eyes in cylindroleberidid ostracods: conflicting conclusions from different comparative methods. Syst. Biol. 61(2):314.

Teotónio H., Rose M.R. 2000. Variation in the reversibility of evolution. Nature, 408(6811):463-466.

Trueman J.W.H., Pfeil B.E., Kelchner S.A.,Yeates D.K. 2004. Did stick insects really regain their wings? Syst. Entomol. 29(2):138-139.

Waters J.M., Emerson B.C., Arribas P. and McCulloch G.A. 2020. Dispersal reduction: causes, genomic mechanisms, and evolutionary consequences. Trends Ecol. Evol.

Whiting M.F., Bradler S., Maxwell T. 2003. Loss and recovery of wings in stick insects. Nature, 421(6920):264-267.

Wiens J.J. 2011. Re-evolution of lost mandibular teeth in frogs after more than 200 million years, and re-evaluating Dollo's law. Evolution 65(5):1283-1296.

Wipfler B., Letsch H., Frandsen P.B., Kapli P., Mayer C., Bartel D., Buckley T.R., Donath A., Edgerly-Rooks J.S., Fujita M., Liu S., Machida R., Mashimo Y., Misof B., Niehuis O., Peters R.S., Petersen M., Podsiadlowski L., Schütte K., Shimizu S., Uchifune T., Wilbrandt J., Yan E., Zhou X., Simon S. 2019. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. Proc. Natl. Acad. Sci. USA 116(8):3024-3029.

Yang H., Shi C., Engel M.S., Zhao Z., Ren D., Gao, T. 2020. Early specializations for mimicry and defense in a Jurassic stick insect. Natl. Sci. Rev.

Yang H., Yin X., Lin X., Wang C., Shih C., Zhang W., Ren D., Gao T. 2019. Cretaceous winged stick insects clarify the early evolution of Phasmatodea. Proc. R. Soc. B Biol. Sci. 286(1909):20191085.

Zeng Y., O'Malley C., Singhal S., Rahim F., Park S., Chen X., Dudley R. 2020. A Tale of Winglets: Evolution of Flight Morphology in Stick Insects. Front. Ecol. Evol. 8:121.

Zhang J., Madden T.L. 1997. PowerBLAST: A New Network BLAST Application for Interactive or Automated Sequence Analysis and Annotation. Genome Res. 7:649–656.

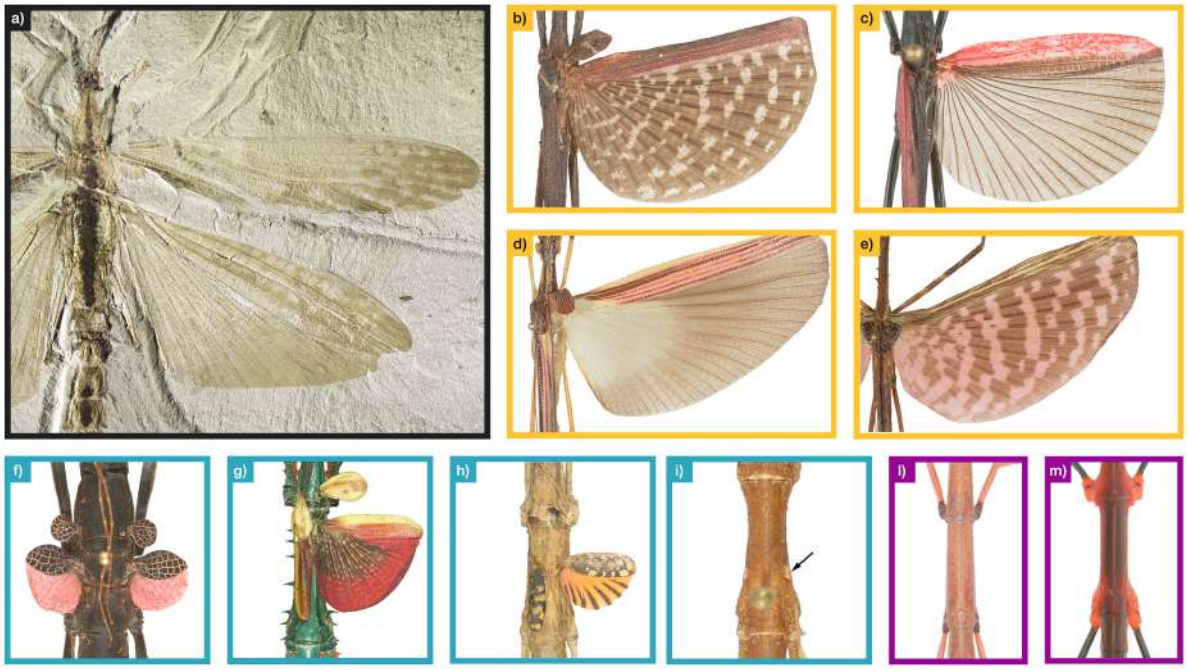


Figure 1 - Wings disparity among extinct and extant Phasmatodea. (a) *Aclistophasma echinulatum* Yang, Shi, Engel, Zhao, Ren and Gao, 2020; (b) *Pterinoxylus crassus* Kirby, 1889; (c) *Orthomeria kangi* Vallotto, Bresseel, Heitzmann and Gottardo, 2016; (d) *Parastratocles fuscomarginatus* Conle, Hennemann, Bellanger, Lelong, Jourdan and Valero, 2020; (e) *Diesbachia tamyris* (Westwood, 1859); (f) *Peruphasma schultei* Conle and Hennemann, 2005; (g) *Achrioptera manga* Glaw, Hawlitschek, Dunz, Goldberg and Bradler, 2019; (h) *Phaenopharos struthioneus* (Westwood, 1859); (i) *Hypocyrtus ornatissimus* (Brunner von Wattenwyl, 1907); (l) *Bacillus atticus* Brunner von Wattenwyl, 1882; (m) *Oreopoethes peruana* (Saussure, 1868). Boxes are colored correspondingly to trait state: gold = Macroptery, blue = Brachyptery, violet = Aptery. All photographs have been taken by Pablo Valero, with the exception of *A. echinulatum* which has been provided by Hongru Yang.

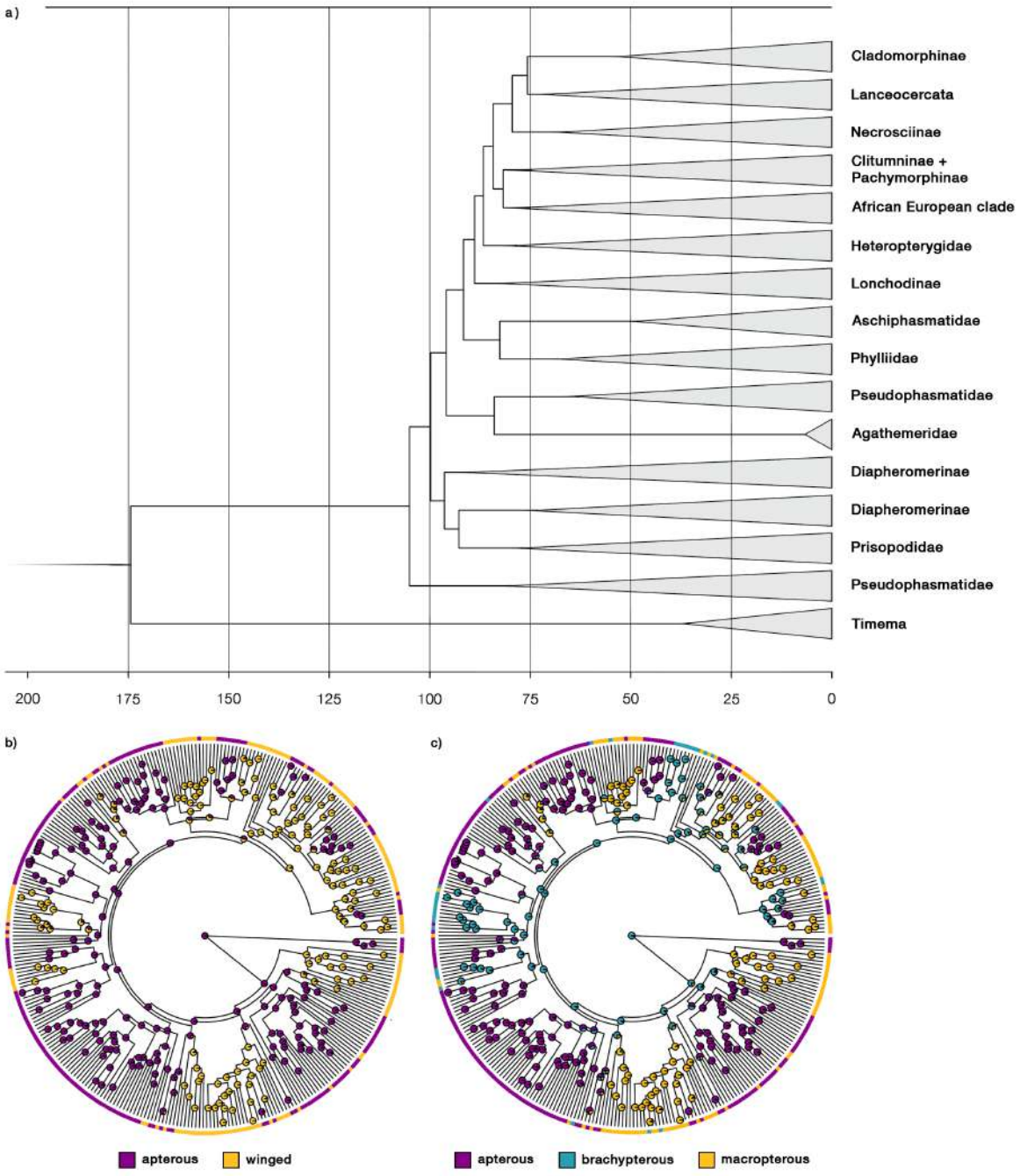


Figure 2 - Timetree of Phasmatodea and ancestral state reconstruction. (a) Phylogenetic and divergence time hypothesis for Phasmatodea major clades; (b) and (c) are ancestral state reconstruction on the Bayesian Inference Maximum Clade Credibility (MCC) tree for the 2-states and 3-states coding schemes under the best-fit model of trait evolution, respectively Equal Rates (ER) and All Rates Different (ARD).

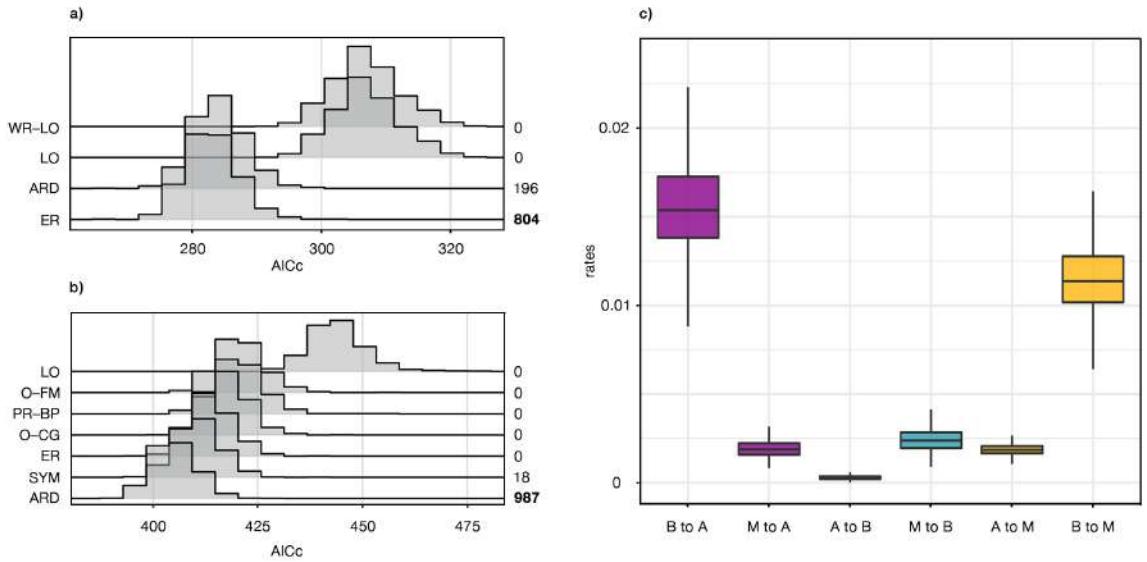


Figure 3 - Model-fitting considering uncertainty in tree topology and branch lengths. Model selection on the 2-states (a) and 3-states (b) coding scheme for 1000 trees randomly sampled from the posterior distribution of the Bayesian Inference. On the right side the number of trees for which each model resulted to be the best-fit is reported, with the more frequent ones (respectively ER and ARD) higlighted in bold. (c) Distribution of inferred rates among states for the 3-states analysis: A = Aptery, B = Brachyptery, M = Macroptery. Boxplots are colored correspondingly to transitions end state: gold = Macroptery, blue = Brachyptery, violet = Aptery.

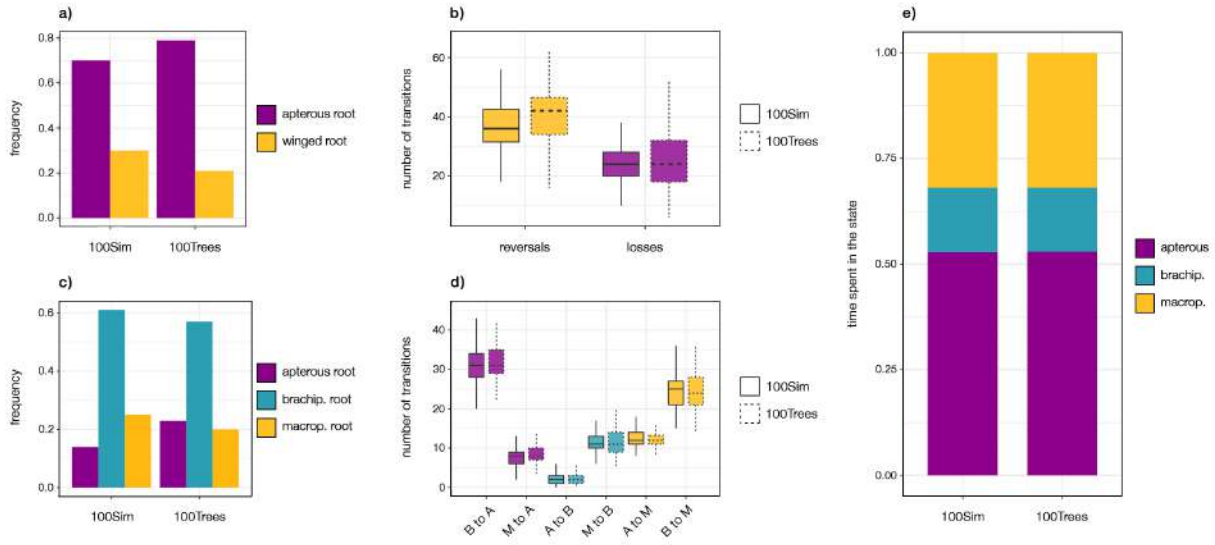


Figure 4 - Stochastic character mapping leveraging 2-states and 3-states coding schemes: the frequency of root state is represented in (a) and (c). (b) and (d) describe the average number of inferred transitions. (e) represents the time spent in each state for the 3-states character coding. Similarly to figure 3, in (b) and (d) boxplots are colored correspondingly to transitions end state. In (b), (d) and (e) 100Sim refers to Stochastic character mapping with 100 simulations, while 100Trees refers to the same analyses carried on 100 trees randomly sampled by Bayesian Inference posterior distribution.

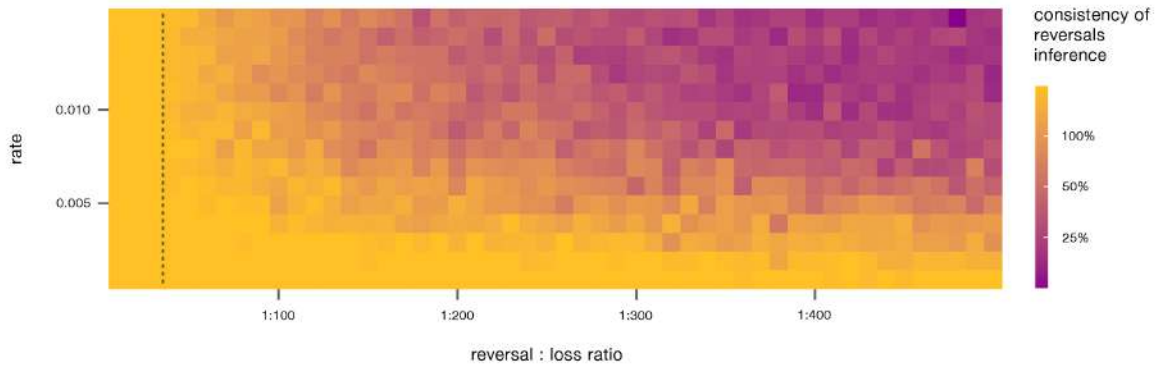
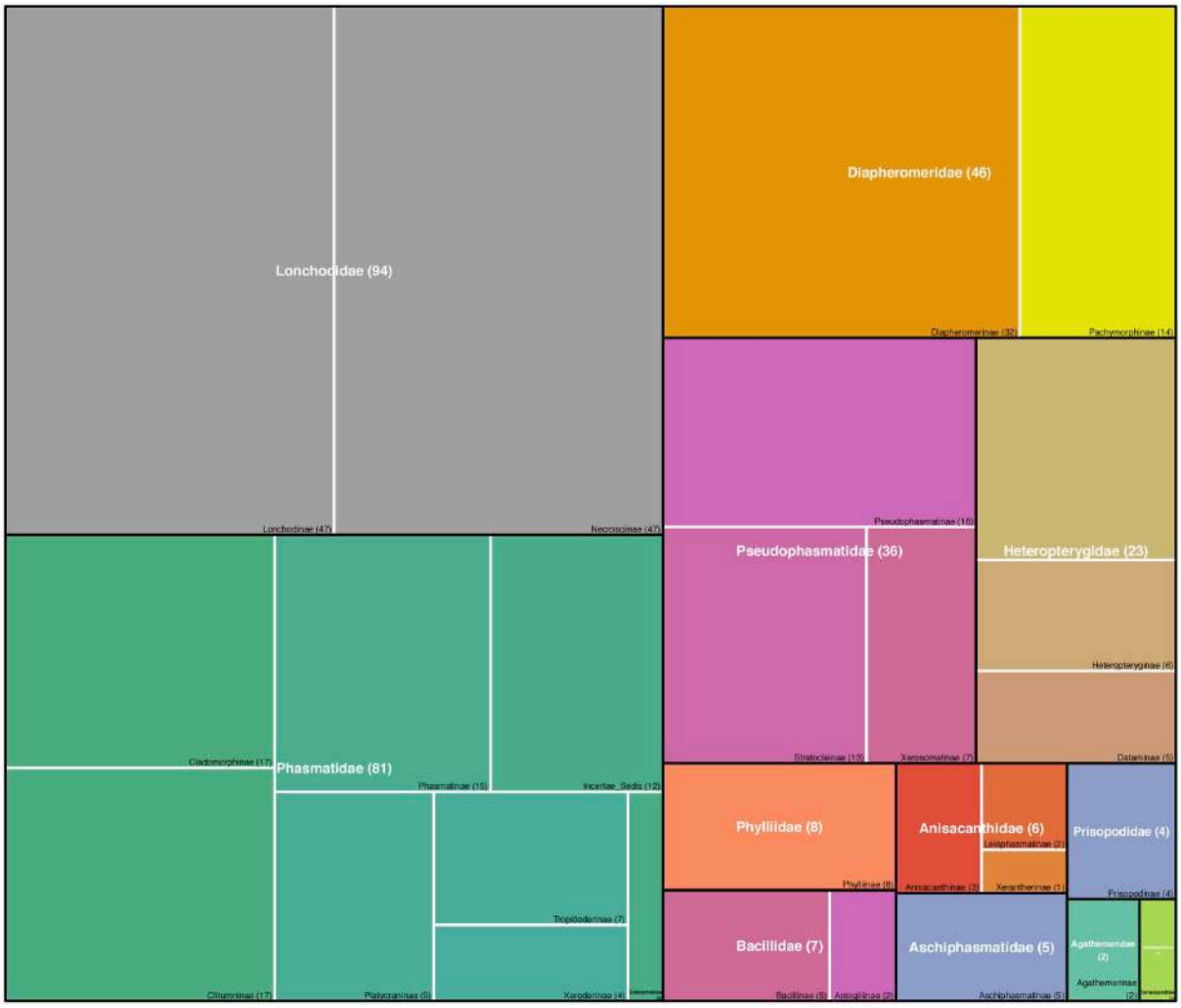
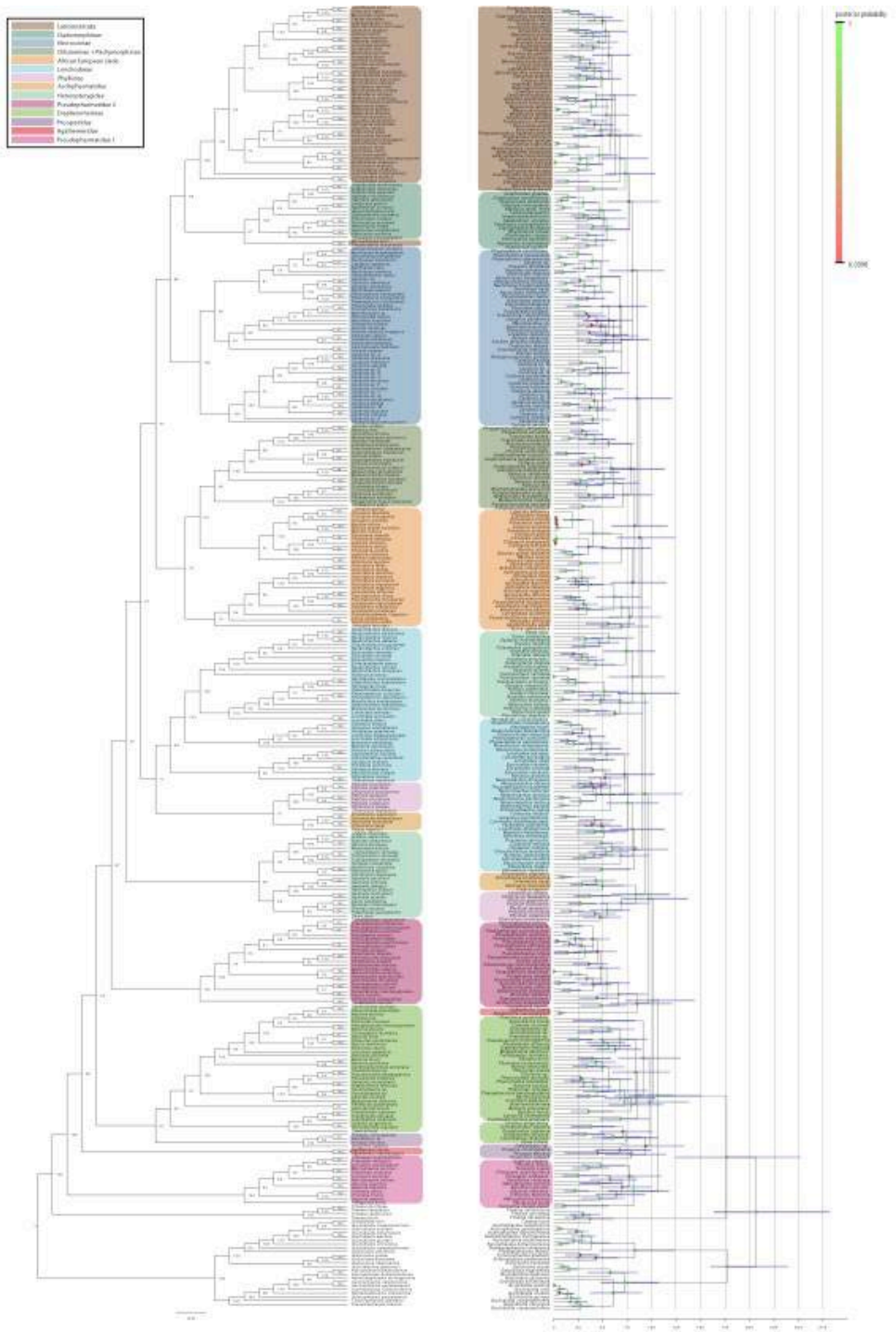


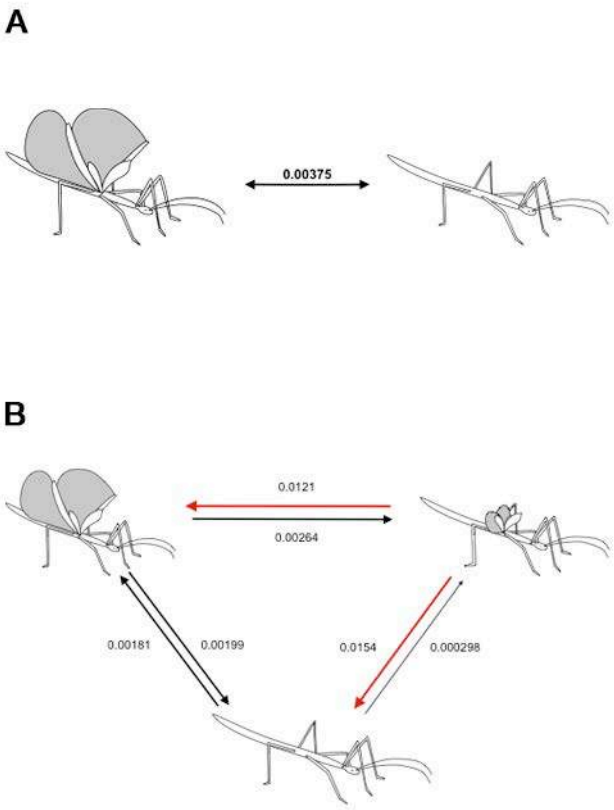
Figure 5 - Different assumption on the probability of wings reversal. The heatmap represents the consistency of reversals inference when different model rates and relative probability of reversal are assumed. When consistency equals 100%, at least one reversal is inferred in each of the 100 SCM analyses for the relative model parameters combination; if consistency <100% and >0% reversals are inferred just on a fraction of the considered trees and character histories. The black dotted line represents the relative reversal probability of 1:30; above this line reversals are no more consistently inferred.



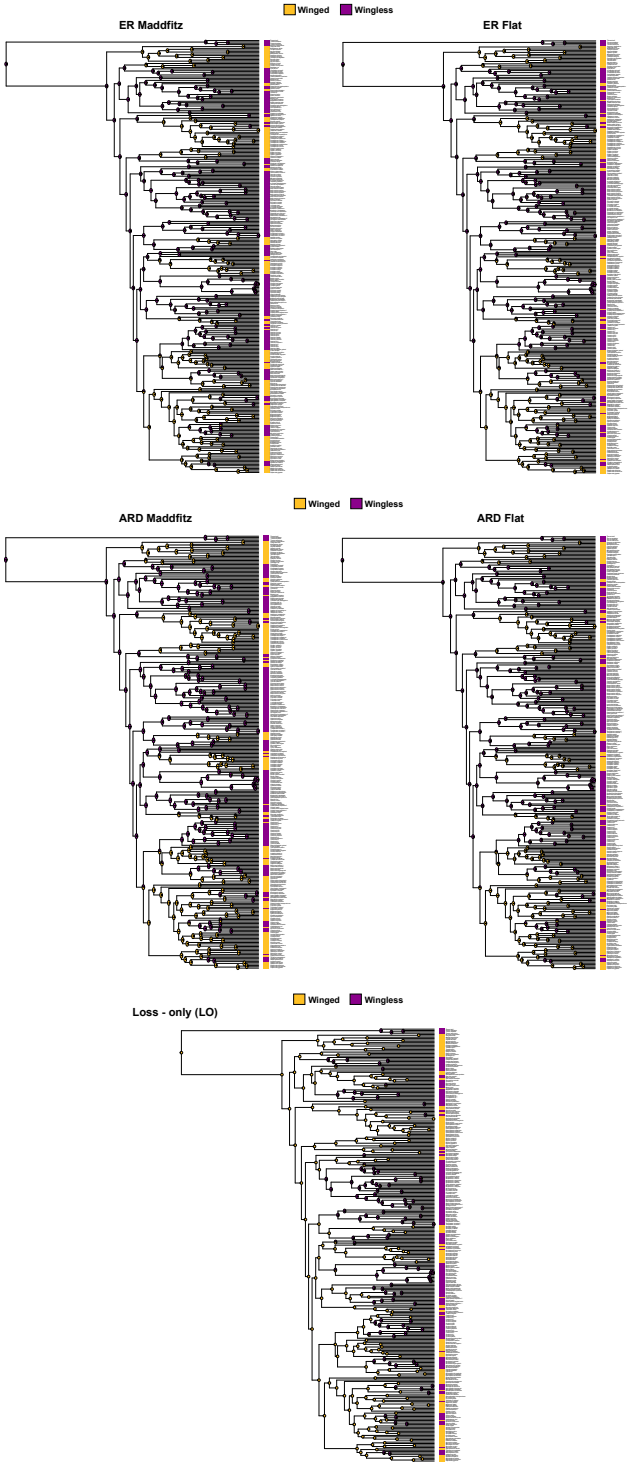
Supplementary Figure 1 - Taxonomic distribution of species at the family and sub-family level; size of rectangles is proportional to the number of specimens.



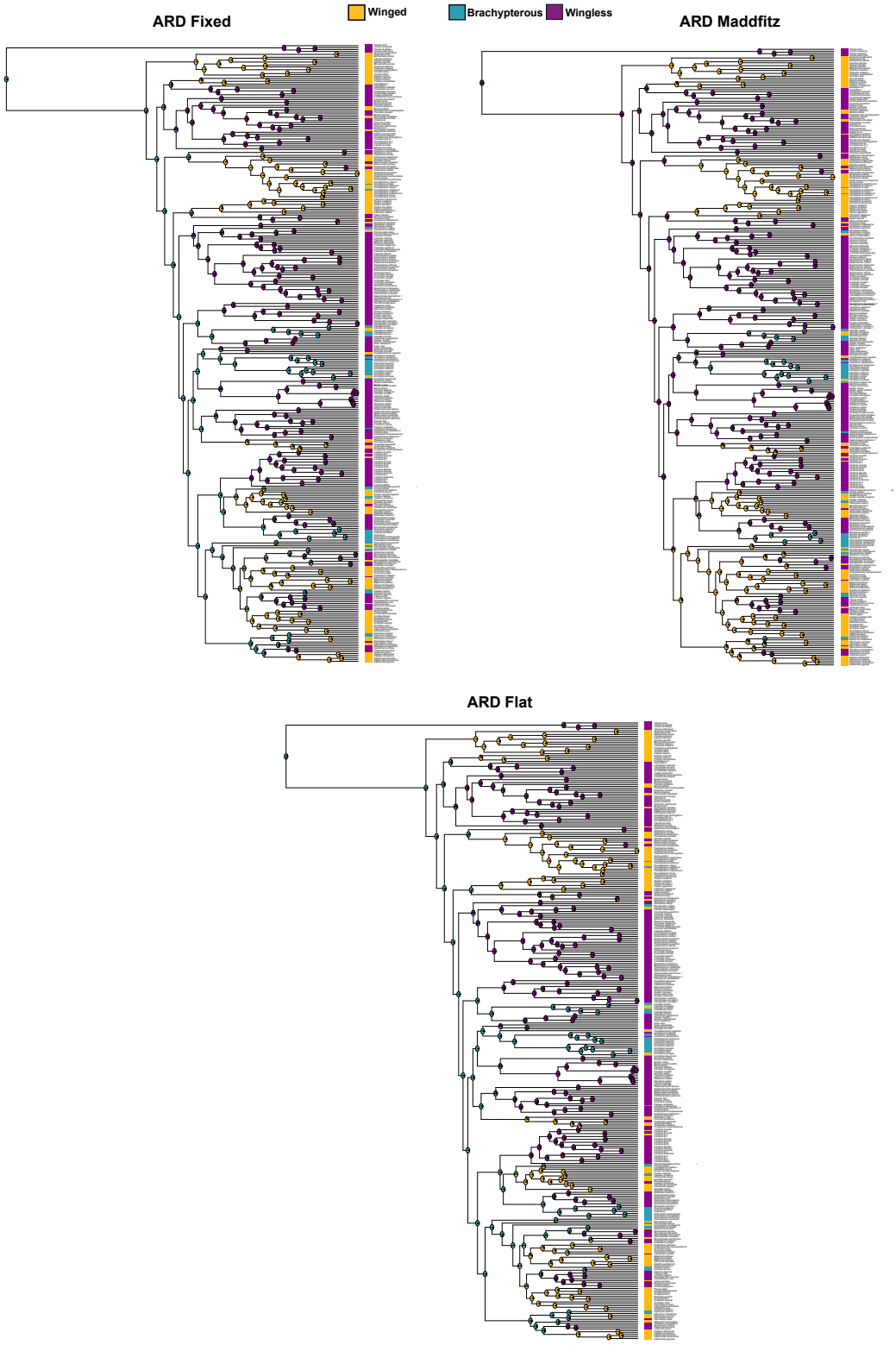
Supplementary Figure 2 - Maximum Likelihood (ML) and Maximum clade credibility (MCC) trees. In the ML tree (left) numbers at nodes represent bootstrap proportions; In the MCC tree (right) the posterior probabilities and the 95% HPD of the dating are reported at each node. Numbers 1, 2 and 3 represent the calibration points described in Supplementary Table 4.



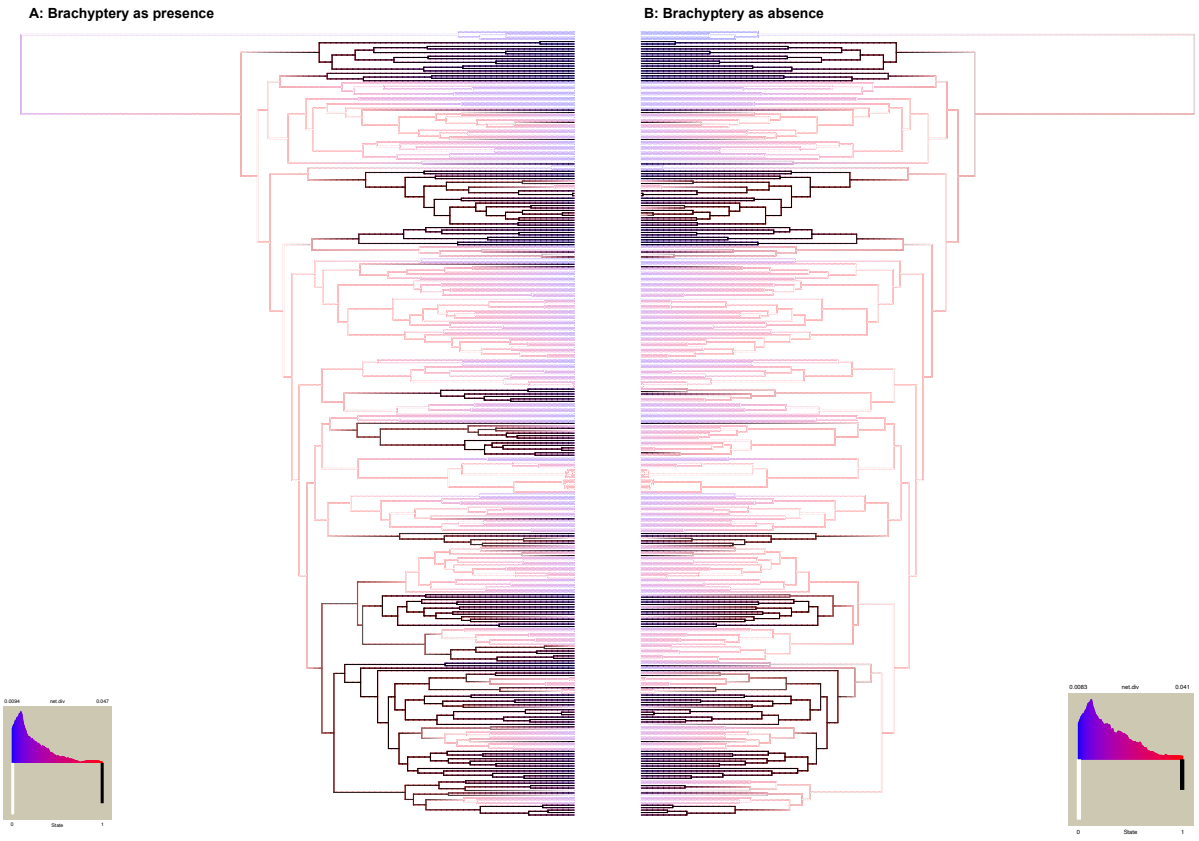
Supplementary Figure 3 - Transition rates inferred with the corresponding best-fit evolutionary models on the MCC tree, respectively Equal Rates (ER) and All Rates Different (ARD). (a) Wing as a 2-states character; (b) Wing as a 3-states character, red color highlights the transition rates which move away from brachyptery.



Supplementary Figure 4 - Ancestral State Resconstruction (ASRs) on the 2-states character coding scheme with different models of evolution and root priors.



Supplementary Figure 5 - ASRs on the 3-states character coding scheme with different root priors.



Supplementary Figure 6 - ASR under the HISSE model. On the left brachyptery was coded as presence, while on the right as absence. White branches represent apterous lineages, while black winged ones. Intermediate colorations are proportional to the posterior probability of the ASR. Outline color of branches is proportional to the net diversification rate, with red branches with higher values and blue smaller, as represent in legends. Width of the white and black rectangles provides the frequency of the observed state.

Supplementary Tables:

Reference	Species	Subfamily	Family	Binary	Multi-state
Newly sequenced	Diapherodes gigantea	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Phantasca quadrilobata	Diapheromerinae	Diapheromeridae	1	2
Newly sequenced	Hermagoras cultratolobatus	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Dares ulula	Dataminae	Heteropterygidae	0	0
Newly sequenced	Libethra strigiventris	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Phaenopharos struthioneus	Necrosciinae	Lonchodidae	1	1
Newly sequenced	Pseudophasma unicolor	Pseudophasmatinae	Pseudophasmatidae	1	2
Newly sequenced	Calynda coronata	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Damasippus sp.	Prisopodinae	Prisopodidae	1	2
Newly sequenced	Oncotophasma sp.	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Oncotophasma sp.	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Trychopeplus laciniatus	Diapheromerinae	Diapheromeridae	1	2
Newly sequenced	Rhynchacris ornata	Cladomorphinae	Phasmatidae	0	0
Newly sequenced	Isagoras asperus	Xerosomatinae	Pseudophasmatidae	1	2
Newly sequenced	Pseudosermyle sp.	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Anthericonia anketeschke	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Eucles unicolor	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Phryganistria heusii yentuensis	Clitumninae	Phasmatidae	0	0
Newly sequenced	Dinophasma saginatum	Aschiphasmatinae	Aschiphasmatidae	1	2
Newly sequenced	Lamponius guerini	Cladomorphinae	Phasmatidae	0	0
Newly sequenced	Taraxippus perezgelaberti	Cladomorphinae	Phasmatidae	0	0
Newly sequenced	Diapherodes martinicensis	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Chitoniscus lobipes	Phylliinae	Phylliidae	1	2
Newly sequenced	Brizoides amabilis	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Sipyloidea biplagiata	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Clonaria conformans	Pachymorphinae	Diapheromeridae	0	0
Newly sequenced	Parapachymorpha zomproi	Clitumninae	Phasmatidae	0	0
Newly sequenced	Tirachoidea biceps	Clitumninae	Phasmatidae	1	2
Newly sequenced	Hesperophasma bocki	Cladomorphinae	Phasmatidae	0	0
Newly sequenced	Carausius alluaudi	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Carausius seychellensis	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Medauroidea extradentata	Clitumninae	Phasmatidae	0	0
Newly sequenced	Prexaspes quadriguttatus	Xerosomatinae	Pseudophasmatidae	1	2
Newly sequenced	Prexaspes ambiguus	Xerosomatinae	Pseudophasmatidae	1	2
Newly sequenced	Paragrostia brulei	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Prisomera spinicollis	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Hypocyrtus ornatissimus	Cladomorphinae	Phasmatidae	1	1

Newly sequenced	Pterinoxylus crassus	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Pterinoxylus eucnemis	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Prisopus conocephalus	Prisopodinae	Prisopodidae	1	2
Newly sequenced	Candovia evoneobertii	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Extatosoma tiaratum	Extatosomatinae	Phasmatidae	1	2
Newly sequenced	Spinohirasea bengalensis	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Orthomeria kangi	Aschiphasmatinae	Aschiphasmatidae	1	2
Newly sequenced	Phyllium giganteum	Phylliinae	Phylliidae	1	2
Newly sequenced	Pseudophasma scabriusculum	Pseudophasmatinae	Pseudophasmatidae	1	2
Newly sequenced	Sceptrophasma langkawicense	Pachymorphinae	Diapheromeridae	0	0
Newly sequenced	Andropromachus scutatus	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Agamemnon cornutus	Cladomorphinae	Phasmatidae	0	0
Newly sequenced	Sipyloidea sipylus	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Trachyaretaon carmelae	Obriminae	Heteropterygidae	0	0
Newly sequenced	Sungaya inexpectata	Obriminae	Heteropterygidae	0	0
Newly sequenced	Clonistria sp.	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Parapachymorpha spinigera	Clitumninae	Phasmatidae	0	0
Newly sequenced	Phyllium bioculatum	Phylliinae	Phylliidae	1	2
Newly sequenced	Graeffea crouanii	Platycraninae	Phasmatidae	1	1
Newly sequenced	Ophicrania palinurus	Platycraninae	Phasmatidae	1	2
Newly sequenced	Aretaon asperrimus	Obriminae	Heteropterygidae	0	0
Newly sequenced	Eurycantha insularis	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Pseudophasma fulvum	Pseudophasmatinae	Pseudophasmatidae	1	2
Newly sequenced	Pseudophasma acanthonotus	Pseudophasmatinae	Pseudophasmatidae	1	2
Newly sequenced	Haplopus micropterus	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Centrophasma hadrillum	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Phenacephorus cornucervi	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Lonchodes brevipes	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Planososibia lysippus	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Carausius nodosus	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Phobaeticus kirbyi	Clitumninae	Phasmatidae	1	2
Newly sequenced	Aretaon muscosus	Obriminae	Heteropterygidae	0	0
Newly sequenced	Mnesilochus portentosus	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Haaniella echinata	Heteropteryginae	Heteropterygidae	1	1
Newly sequenced	Lobofemora bidoupensis	Clitumninae	Phasmatidae	1	1
Newly sequenced	Bostriana boliviana	Cladomorphinae	Phasmatidae	0	0
Newly sequenced	Lonchodes dalawangsungay	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Eurycantha calcarata	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Phanocles vosseleri	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Marmessoidea sp.	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Asceles malaccae	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Asceles tanarata singapura	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Haaniella macroptera	Heteropteryginae	Heteropterygidae	1	2

Newly sequenced	Staelonchodes haematomus	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Stheneboea malaya	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Stheneboea repudiosa	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Hermagoras hosei	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Acanthoxyla geisovii	Phasmatinae	Phasmatidae	0	0
Newly sequenced	Clitarchus tepaki	Phasmatinae	Phasmatidae	0	0
Newly sequenced	Spinotectarchus acornutus	Pachymorphinae	Diapheromeridae	0	0
Newly sequenced	Planososibia esacus	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Paraphasma lateralis	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Dinelytron cahureli	Prisopodinae	Prisopodidae	1	2
Newly sequenced	Cesaphasma servillei	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Bacteria baculus	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Periphloea olivaceus	Pseudophasmatinae	Pseudophasmatidae	1	2
Newly sequenced	Tenerella affinis	Pseudophasmatinae	Pseudophasmatidae	1	2
Newly sequenced	Phanocloidea muricata	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Pseudophasma flavipes	Pseudophasmatinae	Pseudophasmatidae	1	2
Newly sequenced	Acrophylla wuelfingi	Phasmatinae	Phasmatidae	1	2
Newly sequenced	Mearnsiana bullosa	Obriminae	Heteropterygidae	0	0
Newly sequenced	Periphetes graniferum	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Rhaphiderus spiniger	Tropidoderinae	Phasmatidae	0	0
Newly sequenced	Ramulus nematodes	Clitumninae	Phasmatidae	0	0
Newly sequenced	Rhamphosipyloidea gorkomi	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Phyllium jacobsoni	Phylliinae	Phylliidae	1	2
Newly sequenced	Pseudophasma subapterum	Pseudophasmatinae	Pseudophasmatidae	1	1
Newly sequenced	Pseudophasma velutinum	Pseudophasmatinae	Pseudophasmatidae	1	2
Newly sequenced	Tenerella cneius	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Parastratocles tessulatus	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Parastratocles fuscomarginatus	Stratocleinae	Pseudophasmatidae	1	2
Newly sequenced	Lonchodes philippinicus	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Paracalynda utilaensis	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Phanocloidea lobulatipes	Diapheromerinae	Diapheromeridae	1	2
Newly sequenced	Diapherodes dominicae	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Pharaphanocles keratosqueleton	Diapheromerinae	Diapheromeridae	0	0
Robertson et al., 2018	Abrosoma festinatum	Aschiphasmatinae	Aschiphasmatidae	0	0
Glaw et al., 2019	Achrioptera cliquennoisi	Incertae Sedis	Phasmatidae	1	1
Goldberg et al., 2015	Achrioptera fallax	Incertae Sedis	Phasmatidae	1	1
Glaw et al., 2019	Achrioptera impennis	Incertae Sedis	Phasmatidae	1	1
Glaw et al., 2019	Achrioptera magnifica	Incertae Sedis	Phasmatidae	1	1
Glaw et al., 2019	Achrioptera manga	Incertae Sedis	Phasmatidae	1	1
Glaw et al., 2019	Achrioptera maroloko	Incertae Sedis	Phasmatidae	1	1
Robertson et al., 2018	Achrioptera punctipes	Incertae Sedis	Phasmatidae	1	2
Buckley et al., 2009	Acrophylla thoon	Phasmatinae	Phasmatidae	1	2
Buckley et al., 2009	Acrophylla titan	Phasmatinae	Phasmatidae	1	2

Whiting et al., 2003	Agathemera crassa	Agathemerinae	Agathemeridae	0	0
Whiting et al., 2003	Agathemera maculafulgens	Agathemerinae	Agathemeridae	0	0
Robertson et al., 2018	Alienobostra brocki	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Leptynia attenuata	Pachymorphinae	Diapheromeridae	0	0
Goldberg et al., 2015	Anisacantha difformis	Anisacanthinae	Anisacanthidae	1	1
Buckley et al., 2009	Anisomorpha buprestoides	Pseudophasmatinae	Pseudophasmatidae	0	0
Whiting et al., 2003	Anisomorpha ferruginea	Pseudophasmatinae	Pseudophasmatidae	0	0
Bradler et al., 2015	Antongilia muricata	Antongiliinae	Bacillidae	0	0
bradler et al., 2015	Apterograeffea marshallae	Platycraninae	Phasmatidae	0	0
Bradler et al., 2015	Apterograeffea reunionensis	Platycraninae	Phasmatidae	0	0
Buckley et al., 2009	Asprenas impennis	Lonchodinae	Lonchodidae	0	0
newly sequenced	Bacillus atticus	Bacillinae	Bacillidae	0	0
Damgaard et al., 2008	Austrophasma caledonense	NA	NA	NA	NA
Damgaard et al., 2008	Austrophasma gansbaaiense	NA	NA	NA	NA
Damgaard et al., 2008	Austrophasma rawsonvillense	NA	NA	NA	NA
newly sequenced	Agrostia rugicollis	Stratocleinae	Pseudophasmatidae	1	2
newly sequenced	Isagoras taeniatus	Xerosomatinae	Pseudophasmatidae	1	2
newly sequenced	Xerosoma canaliculatum	Xerosomatinae	Pseudophasmatidae	1	2
newly sequenced	Metriophasma pericles	Stratocleinae	Pseudophasmatidae	1	2
Robertson et al., 2018	Bacteria ploiaria	Diapheromerinae	Diapheromeridae	1	2
Bradlet et al., 2014	Baculofractum insigne	Lonchodinae	Lonchodidae	1	1
newly sequenced	Candovia pallida	Necrosciinae	Lonchodidae	0	0
newly sequenced	Candovia sp. L	Necrosciinae	Lonchodidae	0	0
newly sequenced	Candovia sp. M	Necrosciinae	Lonchodidae	0	0
Robertson et al., 2018	Brasidas samarensis	Obriminae	Heteropterygidae	0	0
Buckley et al., 2010	Canachus alligator	Lonchodinae	Lonchodidae	0	0
Buckley et al., 2009	Carausius morosus	Lonchodinae	Lonchodidae	0	0
Whiting et al., 2003	Caribbiopheromera jamaicana	Diapheromerinae	Diapheromeridae	0	0
Buckley et al., 2009	Carlius fecundus	Lonchodinae	Lonchodidae	1	2
Newly sequenced	Clonopsis gallica	Bacillinae	Bacillidae	0	0
Newly sequenced	Clonopsis soumiae	Bacillinae	Bacillidae	0	0
Buckley et al., 2009	Chitoniscus brachysoma	Phylliinae	Phylliidae	1	2
Buckley et al., 2009	Chitoniscus feejeeanus	Phylliinae	Phylliidae	1	2
Buckley et al., 2009	Chondrostethus woodfordi	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Clonaria natalis	Pachymorphinae	Diapheromeridae	0	0
Buckley et al., 2010	Cnipsus rachis	Xeroderinae	Phasmatidae	0	0
Robertson et al., 2018	Creoxylus spinosus	Xerosomatinae	Pseudophasmatidae	1	2
Buckley et al., 2009	Ctenomorpha marginipennis	Phasmatinae	Phasmatidae	1	2
Robertson et al., 2018	Cuniculina cunicula	Clitumninae	Phasmatidae	0	0
Robertson et al., 2018	Dajaca napolovi	Aschiphasmatinae	Aschiphasmatidae	0	0
Robertson et al., 2018	Dares validispinus	Dataminae	Heteropterygidae	0	0
Buckley et al., 2009	Diapheromera femorata	Diapheromerinae	Diapheromeridae	0	0
Buckley et al., 2009	Dimorphodes mancus	Xeroderinae	Phasmatidae	1	1

Whiting et al., 2003	Dimorphodes prostasis	Xeroderinae	Phasmatidae	NA	NA
Robertson et al., 2018	Dinophasma kinabaluense	Aschiphasmatinae	Aschiphasmatidae	0	0
Buckley et al., 2009	Dryococelus australis	Lonchodinae	Lonchodidae	0	0
Bradler et al., 2015	Epicharmus marchali	Xeroderinae	Phasmatidae	1	2
Robertson et al., 2018	Epidares nolimetangere	Dataminae	Heteropterygidae	0	0
Robertson et al., 2018	Erinaceophasma vepres	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Eurycantha coronata	Lonchodinae	Lonchodidae	0	0
Buckley et al., 2010	Eurycnema goliath	Phasmatinae	Phasmatidae	1	2
Robertson et al., 2018	Eurycnema osiris	Phasmatinae	Phasmatidae	1	2
Robertson et al., 2018	Extatosoma popa	Extatosomatinae	Phasmatidae	1	2
Newly sequenced	Sigaruphasma bouladoui	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Oxyartes sp.	Necrosciinae	Lonchodidae	1	1
Newly sequenced	Medauromorpha foedata	Clitumninae	Phasmatidae	0	0
Newly sequenced	Bacteria horni	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Neohirasea nana	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Phanocles costaricensis	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Phanocles mutica	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Oreophoetes topoense	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Neohirasea fruhstorferi	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Oreophoetes peruana	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Phenacephorus sepilokensis	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Mnesilochus mindanaense	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Megaphasma denticrus	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Lobolibethra carbonelli	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Diesbachia tamyris	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Obrimus bicolanus	Obriminae	Heteropterygidae	0	0
Newly sequenced	Mithrenes whiteheadi	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Ramulus artemis	Clitumninae	Phasmatidae	0	0
Newly sequenced	Hypocyrtus scythrus	Cladomorphinae	Phasmatidae	1	1
Newly sequenced	Oxyartes spinipennis	Necrosciinae	Lonchodidae	1	1
Newly sequenced	Bostra remiformis	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Cranidium gibbosum	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Bacteria yersiniana	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Bacteria ferula	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Dyme bifrons	Diapheromerinae	Diapheromeridae	0	0
Newly sequenced	Manduria halconensis	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Asprenas brunneri	Lonchodinae	Lonchodidae	1	1
Newly sequenced	Paramenexenus laetus	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Haaniella gorochovi	Heteropteryginae	Heteropterygidae	1	1
Newly sequenced	Neoclides buescheri	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Trachyaretaon echinatus	Obriminae	Heteropterygidae	0	0
Schoville et al., 2013	Galloisiana kiyosawai	NA	NA	NA	NA
Jarvis and Whiting, 2006	Galloisiana nipponensis	NA	NA	NA	NA

Schoville et al., 2013	Galloisiana yezoensis	NA	NA	NA	NA
Jarvis and Whiting, 2006	Galloisiana yuasai	NA	NA	NA	NA
Bradler et al., 2015	Graeffea seychellensis	Platycraninae	Phasmatidae	1	2
Newly sequenced	Bacillus grandi marettimi	Bacillinae	Bacillidae	0	0
Whinting et al., 2003	Sceptrophasma hispidulum	Pachymorphinae	Diapheromeridae	0	0
Schoville and Roderick, 2010	Grylloblatta bifratrilecta	NA	NA	NA	NA
Jarvis and Whiting, 2006	Grylloblatta campodeiformis	NA	NA	NA	NA
Whinting et al., 2003	Grylloblatta campodeiformis	NA	NA	NA	NA
Jarvis and Whiting, 2006	Grylloblatta chirurgica	NA	NA	NA	NA
Jarvis and Whiting, 2006	Grylloblatta gurneyi	NA	NA	NA	NA
Jarvis and Whiting, 2006	Grylloblatta rothi	NA	NA	NA	NA
Jarvis and Whiting, 2006	Grylloblatta sculleni	NA	NA	NA	NA
Schoville et al., 2013	Grylloblatta washoa	NA	NA	NA	NA
Jarvis and Whiting, 2006	Grylloblattina djakonovi	NA	NA	NA	NA
Goldberg et al., 2015	Haaniella dehaanii	Heteropteryginae	Heteropterygidae	1	1
Robertson et al., 2018	Haaniella muelleri	Heteropteryginae	Heteropterygidae	1	1
Newly sequenced	Pijnackeria lelongi	Pachymorphinae	Diapheromeridae	0	0
Bradlet et al., 2014	Haplopus jamaicensis	Cladomorphinae	Phasmatidae	1	2
Newly sequenced	Pijnackeria masettii	Pachymorphinae	Diapheromeridae	0	0
Damgaard et al., 2008	Hemilobophasma montaguense	NA	NA	NA	NA
Newly sequenced	Pijnackeria barbarae	Pachymorphinae	Diapheromeridae	0	0
Newly sequenced	Leptynia montana	Pachymorphinae	Diapheromeridae	0	0
Buckley et al., 2009	Heteropteryx dilatata	Heteropteryginae	Heteropterygidae	1	2
Robertson et al., 2018	Hoploclonia cuspidata	Obriminae	Heteropterygidae	0	0
Glaw et al., 2019	Hoploclonia gecko	Obriminae	Heteropterygidae	0	0
Newly sequenced	Pijnackeria originis	Pachymorphinae	Diapheromeridae	0	0
Robertson et al., 2018	Hyrtacus procerus	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Hyrtacus tuberculatus	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Anchiale briareus	Phasmatinae	Phasmatidae	1	2
Newly sequenced	Parapodacanthus hasenpuschorum	Tropidoderinae	Phasmatidae	1	2
Newly sequenced	Candovia sp. A	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Cigarrophasma tesselatum	Phasmatinae	Phasmatidae	1	2
Newly sequenced	Candovia sp. C	Necrosciinae	Lonchodidae	1	2
Djernæs et al., 2015	Karoophasma biedouwense	NA	NA	NA	NA
Damgaard et al., 2008	Karoophasma botterkloofense	NA	NA	NA	NA
Buckley et al., 2009	Labidiophasma rouxi	Lonchodinae	Lonchodidae	0	0
Goldberg et al., 2015	Leiophasma lucubense	Leiophasmatinae	Anisacanthidae	0	0
Glaw et al., 2019	Leiophasma nigrolineatum	Leiophasmatinae	Anisacanthidae	0	0
Bradlet et al., 2014	Leprocaulinus insularis	Lonchodinae	Lonchodidae	1	2
Whiting et al.,2013	Ocnophiloidea regularis	Diapheromerinae	Diapheromeridae	0	0
Bradlet et al., 2014	Staelonchodes amaurops	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Lonchodes auriculatus	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Lonchodes chani	Lonchodinae	Lonchodidae	0	0

Whiting et al.,2003	Lopaphus perakensis	Necrosciinae	Lonchodidae	0	0
Whiting et al.,2003	Lopaphus sphalerus	Necrosciinae	Lonchodidae	1	1
Bradlet et al., 2014	Macrophasma biroi	Incertae Sedis	Phasmatidae	1	2
Robertson et al., 2018	Malacomorpha cyllarus	Pseudophasmatinae	Pseudophasmatidae	1	2
Robertson et al., 2018	Malacomorpha jamaicana	Pseudophasmatinae	Pseudophasmatidae	0	0
Buckley et al., 2009	Malandania pulchra	Tropidoderinae	Phasmatidae	1	2
Bradler et al., 2015	Mauritiophasma motalai	Phasmatinae	Phasmatidae	1	1
Buckley et al., 2009	Megacrania aplheus	Platycraninae	Phasmatidae	1	2
Bradlet et al., 2014	Megacrania batesii	Platycraninae	Phasmatidae	1	2
Buckley et al., 2009	Megacrania phelaus	Platycraninae	Phasmatidae	1	2
Robertson et al., 2018	Metriophasma diocles	Xerosomatinae	Pseudophasmatidae	1	2
Buckley et al., 2010	Micrarchus hystriculeus	Pachymorphinae	Diapheromeridae	0	0
Buckley et al., 2009	Monandroptera acanthomera	Tropidoderinae	Phasmatidae	1	2
Bradler et al., 2015	Monoiognosis bipunctata	Incertae Sedis	Phasmatidae	0	0
Bradler et al., 2015	Monoiognosis spinosa	Incertae Sedis	Phasmatidae	0	0
Damgaard et al., 2008	Namaquaphasma ookiepense	NA	NA	NA	NA
Bradlet et al., 2014	Neohirasea hongkongensis	Necrosciinae	Lonchodidae	0	0
Bradlet et al., 2014	Neohirasea maerens	Necrosciinae	Lonchodidae	0	0
Whiting et al.,2003	Neohirasea maerens	Necrosciinae	Lonchodidae	0	0
Buckley et al., 2009	Neopromachus doreyanus	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Neopromachus dyselius	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Neopromachus muticus	Lonchodinae	Lonchodidae	0	0
Bradlet et al., 2014	Neopromachus obrutus	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Neopromachus pachynotus	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Neopromachus scharreri	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Neopromachus wallacei	Lonchodinae	Lonchodidae	0	0
Bradler et al., 2015	Onchestus rentzi	Phasmatinae	Phasmatidae	1	2
Robertson et al., 2018	Ophicrania bifasciatus	Platycraninae	Phasmatidae	1	2
Robertson et al., 2018	Orestes mouhotii	Dataminae	Heteropterygidae	0	0
Buckley et al., 2009	Orxines semperi	Necrosciinae	Lonchodidae	1	1
Buckley et al., 2009	Oxyartes lamellatus	Necrosciinae	Lonchodidae	1	1
Robertson et al., 2018	Parapachymorpha spinosa	Clitumninae	Phasmatidae	0	0
Robertson et al., 2018	Paraphasma amabile	Stratocleinae	Pseudophasmatidae	1	2
Whiting et al.,2003	Pseudophasma rufipes	Pseudophasmatinae	Pseudophasmatidae	1	2
Glaw et al., 2019	Parectatosoma minus	Anisacanthinae	Anisacanthidae	1	1
Bradler et al., 2015	Parectatosoma mocquerysi	Anisacanthinae	Anisacanthidae	1	1
Bradler et al., 2015	Paronchestus charon	Phasmatinae	Phasmatidae	1	2
Newly sequenced	Candovia sp. B	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Tropidoderus childrenii	Tropidoderinae	Phasmatidae	1	2
Newly sequenced	Anchiale austrotessulata	Phasmatinae	Phasmatidae	1	2
Newly sequenced	Candovia robinsoni	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia spurcata	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia aberrata	Necrosciinae	Lonchodidae	0	0

Newly sequenced	Candovia coenosa	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia peridromes	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia sp. E	Necrosciinae	Lonchodidae	1	2
Newly sequenced	Candovia sp. B	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia annulata	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia strumosa	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia granulosa	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia sp. F	Necrosciinae	Lonchodidae	0	0
Newly sequenced	Candovia sp. G	Necrosciinae	Lonchodidae	0	0
Buckley et al., 2009	Peruphasma schultei	Pseudophasmatinae	Pseudophasmatidae	1	1
Bradlet et al., 2014	Phaenopharos herwaardeni	Necrosciinae	Lonchodidae	1	1
Buckley et al., 2009	Phaenopharos khaoyaiensis	Necrosciinae	Lonchodidae	1	1
Bradler et al., 2015	Phalces tuberculatus	Bacillinae	Bacillidae	0	0
Buckley et al., 2009	Pharnacia ponderosa	Clitumninae	Phasmatidae	1	2
Robertson et al., 2018	Phasma gigas	Phasmatinae	Phasmatidae	1	2
Bradler et al., 2015	Phasmotaenia lanyuhensis	Incertae Sedis	Phasmatidae	1	1
Robertson et al., 2018	Phobaeticus serratipes	Clitumninae	Phasmatidae	1	2
Buckley et al., 2009	Phyllium celebicum	Phylliinae	Phylliidae	1	2
Buckley et al., 2009	Phyllium siccifolium	Phylliinae	Phylliidae	1	2
Buckley et al., 2010	Podacanthus wilkinsoni	Tropidoderinae	Phasmatidae	1	2
Damgaard et al., 2008	Praedatophasma maraisi	NA	NA	NA	NA
Robertson et al., 2018	Prisopus berosus	Prisopodinae	Prisopodidae	1	2
Bradler et al., 2015	Pseudoleosthenes irregularis	Damasippoididae	Damasippoidinae	1	2
Robertson et al., 2018	Pseudophasma phthisicum	Pseudophasmatinae	Pseudophasmatidae	1	2
Bradlet et al., 2014	Pseudophasma rufipes	Pseudophasmatinae	Pseudophasmatidae	1	2
Buckley et al., 2009	Pseudosermyle phalangiphora	Diapheromerinae	Diapheromeridae	0	0
Buckley et al., 2009	Pterobrimus depressus	Obriminae	Heteropterygidae	0	0
Goldberg et al., 2015	Pylaemenes guangxiensis	Dataminae	Heteropterygidae	0	0
Buckley et al., 2009	Ramulus thaii	Clitumninae	Phasmatidae	0	0
Robertson et al., 2018	Rhamphophasma spinicorne	Clitumninae	Phasmatidae	0	0
Robertson et al., 2018	Rhaphiderus scabrosus	Tropidoderinae	Phasmatidae	0	0
Newly sequenced	Bacillus rossius	Bacillinae	Bacillidae	0	0
Buckley et al., 2009	Sceptrophasma hispidulum	Pachymorphinae	Diapheromeridae	0	0
Jarvis and Whiting, 2006	Sclerophasma paresisensis	NA	NA	NA	NA
Bradler et al., 2015	Spathomorpha adefa	Incertae Sedis	Phasmatidae	0	0
Robertson et al., 2018	Staelonchodes haematomus	Lonchodinae	Lonchodidae	0	0
Newly sequenced	Clonopsis androgenes	Bacillinae	Bacillidae	0	0
Newly sequenced	Clonopsis felicitatis	Bacillinae	Bacillidae	0	0
Robertson et al., 2018	Thaumatobactron guentheri	Lonchodinae	Lonchodidae	0	0
Robertson et al., 2018	Timema californicum	Timematinae	Timematidae	0	0
Robertson et al., 2018	Timema dorotheae	Timematinae	Timematidae	0	0
Whiting et al., 2003	Timema knulli	Timematinae	Timematidae	0	0
Robertson et al., 2018	Timema nevadense	Timematinae	Timematidae	0	0

Robertson et al., 2018	Tirachoidea westwoodii	Clitumninae	Phasmatidae	0	0
Buckley et al., 2009	Trachyaretaon carmelae	Obriminae	Heteropterygidae	0	0
Bradlet et al., 2014	Trachythorax maculicollis	Necrosciinae	Lonchodidae	1	2
Jarvis and Whiting, 2006	Tyrannophasma gladiator	NA	NA	NA	NA
Glaw et al., 2019	Xerantherix nossibianus	Xerantherinae	Anisacanthidae	1	1
Robertson et al., 2018	Xylica oedematosa	Antongiliinae	Bacillidae	0	0
Buckley et al., 2009	Zehntneria mystica	Pachymorphinae	Diapheromeridae	0	0

Supplementary Table 1 - Species included in the analyses along with the references of the species downloaded from NCBI and the 2- and 3-states character coding scheme. In the binary dataset 0 represents wingless species while 1 winged. In the multi-state dataset 0 represents wingless species, 1 brachipterous and 2 macropterous.

Gene target	Primer Name	Primer sequence	Annealing	Amplicon Size
CO1	LCO1490	GTCAACAAATCATAAAGATATTGG	55°C 45s	708
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA		
	FG1F	AAACTAWTAACCTTCAAAGTTA	50°C 60s	1049
	FG2F	CCWACWGTRAATATRTGRTGWGC		
CO1	TL2-N-3014	TTGATTTTTTGGTCATCCAGAAGT	61°C 60s	819
	C1-J-2195	TCCAATGCACTAATCTGCCATATTA		
16S	16sa	CGCCTGTTTATCAAAAACAT	56°C 30s	548
	16sb	CTCCGGTTTGAACTCAGATCA		
12S	12S_F_G	CTGCACCTTGACYTGAAATA	56°C 30s	399
	12S_R_G	CAGCATACGCGGTTATACAA		
CO2	TL2-J-3034	CAGATAAGTGCATTGGATTT	50°C 50s	1000
	TK-N-3785	GTTTAAGAGACCAGTACTTG		
18S	18sai	CCTGAGAAACGGCTACCACATC	57°C 30s	950
	18sbi	GAGTCTCGTTCGTTATCGGA		
28S	28S-1009	TCAAGACGGGTCGGGAGA	58°C 30s	564
	28S-356	AGAACTTTGAAGAGAGAGTTCAAGA		
H3	H3F	ATGGCTCGTACCAAGCAGAC	58°C 30s	500
	H3R	ATATCCTTRGGCATRATRGTGAC		

Supplementary Table 2 - Primers sequence and PCR annealing condition.

Partition	Model of Evolution	N. of sites
Charset 1: 12s	GTR+F+I+G4	195
Charset 2: 16s	GTR+F+R6	397
Charset 3: 18s; COI 2- ; COI_RL 2-	TNe+R7	1337
Charset 4: 28s	TIM2e+G4	256
Charset 5: COI 1-	GTR+F+I+G4	253
Charset 6: COI 3-	TIM2+F+ASC+R10	253
Charset 7: COII 1-	GTR+F+I+G4	217
Charset 8: COII 2-	TPM2u+F+ASC+R6	217
Charset 9: COII 3-	GTR+F+R5	217
Charset 10: COI_RL 1-	TIM2+F+G4	224
Charset 11: COI_RL 3-	TIM2+F+R5	224
Charset 12: H3 1-	GTR+F+I+G4	107
Charset 13: H3 2-	K3P	107
Charset 14: H3 3-	K2P+I+G4	107

Supplementary Table 3 - Maximum Likelihood model selection performed by ModelFinder with the -MERGE option and in according to the BIC score.

Fossil	Age	Geological formation	Description	Reference	Calibration node
Juramantophasma sinica	158.1 Ma	Daohugou, China, Callovian to Callovian (164.7 – 155.7 Ma).	Crown Notoptera	Huang et al., 2008	1
Echinosomiscus primoticus	98.8 Ma	Myanmar; Early/lower Cenomanin (99.7 – 94.3 Ma).	Crown Euphasmatodea	Engel et al., 2016	2
Eophyllium messelensis	47 Ma	Messel formation, Germany (48.6 – 40.4 Ma).	Crown Phyllinae	Wedmann et al., 2007	3

Supplementary Table 4 - Fossil calibrations used for the divergence times analysis. The calibration nodes are reported in Sup. Fig. 2.

Sup. Tab. 5a)

Models	AICc	-lnL	ΔAICc
WR-LO	302.7572	157.40	22.6269
LO	302.7572	157.40	22.6269
ARD	282.1424	146.96	2.0121
ER	280.1303	145.34	/

Sup. Tab. 5b)

Model:	AICc	-lnL	ΔAICc
O-LO	448.2000	-222081142	43.408
PG-AB	438.4630	-215168225	33.671
WR-LO	437.9144	-2159194	33.122
LO	437.9144	-215919352	33.122
O-F	419.6980	-205785746	14.906
PG-BW	418.3981	-205135779	13.606
O-CG	416.9163	-203362907	12.124
ER	414.0534	-206020426	92.61
SYM	410.8401	-202382223	60.48
ARD	404.7921	-196262311	/

Supplementary Table 5 - AICc, -lnL, and ΔAICc values of the evolutionary models fitted on the Maximum Clade Credibility tree with the binary (a) and the multi-state (b) datasets. Description of the evolutionary models can be found in the body of the text.

Sup. Tab. 6a)

Model	root.p	q01	q10	root.state	AICc
ARD	maddfitz	0.00373	0.00379	wingless - 1 PP	282.1424
ER	maddfitz	0.00375	0.00375	wingless - 1 PP	280.1303
ARD	flat	0.00371	0.00395	wingless - 0.77 PP	283.0994
ER	flat	0.00381	0.00381	wingless - 0.79 PP	281.0992
LO	fixed	0	0.00905	winged - fixed	302.4318

Sup. Tab. 6b)

Model	root.p	q01	q02	q12	q10	q20	q21	root.state	AICc
ARD_fixed	maddfitz	0.00029796	0.0018135	0.012060802	0.0154381	0.0019847	0.00264779	brachipetorus - 0.95	404.79
ARD	maddfitz	0.00069608	0.0030986	0.006150559	0.0062354	0.0030782	0.00315926	macropterous - 0.97 PP	410.57
ARD	flat	0.00032446	0.0018244	0.012103716	0.0155935	0.0020871	0.00284703	brachipterous - 0.63 PP	405.24

Supplementary Table 6 - Ancestral State Reconstructions (ASRs) results for the 2-states (a) and 3-states (b) character coding scheme. Root.p = root prior probability, root.state = probability of ASRs at the root were PP = posterior probability. In (a) q01 = tranistion rates between wingless and winged; q10 = transition rate between winged and wingless. In (b) q01 = tranistion rates between wingless and brachypterous; q02 = transition rate between wingless and macropterous; q12 = transition rates between brachypterous and macropterous; q10 = transition rate between brachypterous and wingless; q20 = transition rate between macropterous and wingless; q21 = transition rate between macropterous and brachypterous

Sup. Tab. 7a)

Model	AICc	-lnL	ΔAICc
Bisse	3296	1642	64
Bisse null	3292	1642	60
Half hisse 1	3267	1623	35
Half hisse 2	3262	1620	30
CID-4	3261	1621	29
CID-2	3241	1608	9
Full Hisse	3237	1601	5
Hisse	3232	1601	/

Sup. Tab. 7b)

Model	AICc	-lnL	ΔAICc
CID-2	3305	1640	59
Bisse	3302	1645	56
Bisse null	3299	1645	53
CID-4	3266	1624	20
Half hisse 1	3264	1622	18
Half hisse 2	3263	1621	17
Full Hisse	3248	1607	2
Hisse	3246	1609	/

Sup. Tab.7 c)

Dataset	0A *r*	0A *a*	1A *r*	1A *a*	0B *r*	0B *a*	1B *r*	1B *a*
1	0.047	0.002061	0.047	0.002061	0.008624	0.002061	0.009576	0.002061
2	0.041307	0.002061	0.041307	0.002061	0.008310	0.002061	0.008339	0.002061

Supplementary Table 7 - (a) and (b) model selection results using *SSE models on respectively the original binary dataset and on the modified version with brachiptery coded as absence. Details on the evolutionary models can be found in the body of the text. In (c) are reported the diversification parameters of the best-fit model for each dataset: 1 =original binary dataset, 2 = modified version; 0 –1 = explicit character states (respectively wingless and winged); A-B = hidden states; r = net diversification rates (= λ – μ); a = extinction fraction (= μ/λ).

Evolution of automixis through different cytological mechanisms is associated with convergent transcriptional reprogramming in *Bacillus* insects.

Giobbe Forni, Alexander Sasha Mikheyev, Andrea Luchetti and Barbara Mantovani

Abstract - Understanding the relative contribution of novel and pre-existing genes and the degree of convergence in the processes underlying the establishment of novelties represent central themes in evolutionary biology. Automixis - *i.e.* meiotic parthenogenesis - requires several modifications to overcome the obstacles of reproducing without male contribution, including ploidy restoration after meiosis and activation of oocytes. Here we characterize the gonads transcriptomic profiles of two *Bacillus* automictic stick-insects - which restore ploidy through different cytological mechanisms and represent independent shifts to parthenogenesis - and compare them with a bisexual congeneric species. Changes in expression of hundreds of genes were observed in parthenogens' gonads when compared to the bisexual species and their transcriptional program is found to be largely assembled from genes that were already present before the establishment of automixis. Despite the different underlying mechanisms, we identify a large shared core of genes with gonads-biased patterns of expression in parthenogens which are either male gonads-biased in the sexual species or are not differentially expressed there. While convergence seems to have played a prominent role at the expression level, it appears to have had a marginal one in sequence evolution: just five genes were identified as candidates for convergent modifications at the sequence level. This work is the first to explore the molecular blueprint of automixis in a comparative framework and highlights how similar outcomes can still be associated with convergent evolutionary processes even when attained through different mechanisms.

Keywords: parthenogenesis; automixis; RNA-seq; dNdS; phylostratigraphy; convergence.

Introduction

Understanding the mechanisms underlying complex evolutionary innovations is a central theme in biology. One key aspect is the relative contribution of the evolution of novel genes and the modification of pre-existing ones (Khalturin et al, 2009; Shubin et al, 2009; Tautz et al, 2011; Rosenblum et al, 2014; McLysaght and Guerzoni 2015). Both mechanisms can concur to the evolution of novelties, with several traits being associated at the same time with *de-novo* gene origination and the modification of pre-existing ones (Albertin et al, 2015; Aguilera et al, 2017; Hilgers et al, 2018). Nonetheless, the process through which novel traits are established can have different signatures, with some apparently linked to the appearance of novel genes (Milde et al, 2009; Babonis et al, 2016; Sim et al, 2016; Santos et al, 2017) and others more strongly associated with changes in pre-existing ones (Jasper et al, 2015; Fisher et al, 2020; Almundi et al, 2020). In the latter case, changes can result from modifications in *cis*-regulatory elements affecting pattern of gene expression (McGirr et al, 2020) and/or in protein coding sequence parts of genes (Casewell et al, 2019; Jebb et al, 2020). Another key aspect is the degree to which convergent evolution of traits involves modifications in the same sets of genes (Stern, 2013; Sackton and Clark, 2019). While the establishment of some traits seems to have happened through similar trajectories (Foote et al, 2018; Warner et al, 2019; Burskaia et al, 2020; Yuan et al, 2020; van Kruistum et al, 2021) in other cases such evidences are lacking, implying that similar outcomes are achieved through different paths (Zou and Zhang, 2015; Corbett-Detig et al, 2020). These results show that how any single process is responsible for the establishment of novel traits but, instead, their origin occurs through multiple non-mutually exclusive routes (Wagner and Lynch, 2010). Yet, relatively few studies investigate to what extent similar phenotypic endpoints show convergence when they are attained by different mechanisms. To answer this question, we focused on the repeated evolution of automictic parthenogenesis in the polyneopteran insect genus *Bacillus*, trying to elucidate the processes underlying its establishment and the extent of their convergence.

Parthenogenesis provides several advantages, including the transmission of beneficial allele combination to future generations, improved colonization ability and reproductive success under environmental conditions with scarce mating opportunities (Lehtonen et al, 2012; Tilquin et al, 2016; Schwander et al, 2010). Despite being often associated with loss of heterozygosity due to the lack of outcrossing (Svendsen et al, 2013; Engelstädter, 2017), its importance as an adaptive trait is highlighted by the many independent times it has evolved across the tree of life (Mogie, 1986). In parthenogenetically reproducing females, ploidy maintenance can take place either through apomixis - with a mitotic process retaining the ploidy - or through automixis - in which meiosis is maintained but a

different set of mechanisms allows ploidy restoration (Normark, 2013). The latter process can be achieved either by post-meiotic doubling of chromosome sets or by the fusion of meiotic products, which can happen through central fusion - *i.e.* restitutional meiosis at anaphase I or the fusion of its products - or through terminal fusion - *i.e.* restitutional meiosis at anaphase II or the fusion of its products (Mirzaghaderi and Hörandl, 2016). Independently from the mechanism through which egg ploidy is either retained or restored, the establishment of parthenogenesis incurs in additional and shared barriers, such as the activation of the oocyte without sperm fertilization (Pál and Papp, 2017).

Several clues to the causes and consequences of apomixis are available in animals (Neiman et al, 2014; Parker et al 2018; Bast et al, 2018; Parker et al, 2019; Tvedte et al, 2019; Jaron et al, 2020), but studies on automixis mainly focus on the consequences of this reproductive strategy (Oxley et al, 2014; Kraaijeveld et al, 2016; Matsuura, 2017; Tvedte et al, 2017) and an understanding of its molecular underpinnings is restricted to few taxa (Wallberg et al, 2016; Yagound et al, 2020). Parthenogenesis is a well known phenomenon in stick- and leaf-insects of the order Phasmida and it has been particularly well characterized in the Mediterranean *Bacillus* species complex (Scali et al., 2003). In this work we analyzed three *Bacillus* species: the strictly bisexual *B. grandii,* the strictly parthenogen *B. atticus*, and the facultative parthenogen *B. rossius*. The *Bacillus* clade - which also includes hybrid lineages and instances of androgenetic reproduction (Scali et al, 2003) - has started its diversification over 20 million years ago (Mantovani et al, 2001). Earlier research has provided evidence that in the obligate parthenogen *B. atticus* - after a regular first meiotic division - nuclei at prophase II fuse in a diploid egg nucleus; this is followed by a second division leading to a degenerating polocyte and a quickly dividing, unreduced nucleus originating the parthenogenetic embryo (Marescalchi et al, 1993). On the contrary, the egg of *Bacillus rossius* parthenogenetic populations are produced through a canonical meiosis, but during the embryonic development some cells of the haploid blastula achieve diploidization *via* anaphase restitution (Pijnacker et al, 1969).

The *Bacillus* species complex represents an exceptional system to understand the evolutionary processes underlying the establishment of different automictic mechanisms and to test the extent of their convergence, due to the independent shifts in reproductive strategies of *B. atticus* and *B. rossius* and their different means to restore ploidy. Here, we leveraged new transcriptomics resources of these two species and of the bisexual congeneric *B. grandii* to carry out comparative analyses of gene expression and molecular evolution trying to elucidate: (1) the relative contribution to the evolution of automixis of the trascriptional reprogramming and sequence modification of pre-exhisting genes *versus* the emergence of novel ones and (2) the degree of convergence in the molecular changes associated with automixis. Our analyses represent the first investigation to explore the modifications

associated with different automictic mechanisms and the results from this work establish the *Bacillus* species complex as a new testing ground to explore causes and consequences of these reproductive strategies.

Materials and Methods

Study Design, RNA Extraction and Sequencing: *B. grandii* (Marettimo), *B. atticus* (Necropoli Camerina) and *B. rossius* (Massa San Nicola; Bonandin et al., 2014) specimens were raised at 23°C and fed with lentisk and bramble. *B. grandii* males were sacrificed 1 day following their final moult and females from all the three species within 24h from their first egg laying, in order to ensure individuals were reproductively active and that reproductive structures were completely developed. For non-reproductive tissues samples, three legs were used from each individual (one foreleg, one mid-leg and one hind-leg) while reproductive tracts (ovaries in females and testes in males) were dissected. The same individuals were used for leg and reproductive tract samples. Tissues were conserved in Triazol (Life Technologies) and RNA was extracted following the manufacturer's protocol. RNA quantity and quality were measured using NanoDrop (Thermo Scientific) and Bioanalyzer (Agilent). Library preparation and paired-end sequencing (150bp, HiSeqX) were performed by Macrogen Korea. All the sequenced libraries were uploaded to the SRA under the accession PRJNA578804.

Transcriptome Assembly, Orthology Inference and Phylostratigraphy: all custom scripts associated with this project are deposited at (https://github.com/for-giobbe/bacillus_spp_rna). All available reads for each species have been pooled and used for de-novo assembly using Trinity (Grabherr et al, 2011; Haas et al, 2013) with default parameters. We used TransDecoder (v5.5.0) on the raw assemblies to detect coding regions and UTRs in assembled transcripts, also integrating homology information inferred using BLASTp (Camacho et al, 2009) against the SwissProt protein database and HMMER (Finn et al, 2016) against Pfam domain database; when multiple ORFs were predicted for a transcript we chose the single best one. To filter out of our assemblies all the transcript which could be of contaminant origin, we carried out a BLASTp search of each species proteome against the nt database using an e-val cutoff of 1e-3 and subsequently assigned each hit to a specific lineage using taxonkit (Shen & Xiong, 2019). We retained only those transcripts which had all the twenty best hits as Metazoans and at least half of them as Panarthropoda; we excluded all the transcripts which did not have any hit over the e-value cutoff. To provide a comparative framework across species, we carried out an inference to identify homologous gene clusters - including ortholog and paralog - using Orthofinder2 (Emms & Kelly, 2019). We included the traduced proteomes of the three

Bacillus species along with the *Phyllium pilippinicum* one (TSA accession: GCPM00000000; Wipfler et al, 2019) as an outgroup; the latter has been chosen since it has been found in a sister relationship with *Bacillus rossius* in a recent phylogenomic publication, where the divergence between the two clades is estimated to have happened around 50 mya (Simon et al, 2019). A species tree was inferred according to the STAG method (Emms & Kelly, 2018).

Subsequently, we leveraged the orthology inference to estimate the evolutionary age of genes using phylostratigraphy (Domazet-Lošo et al, 2007; Domazet-Lošo and Tautz, 2010); if a species shared a homolog with any of the other species, we assumed that the last common ancestor of the two already possessed a copy of this gene. Using a custom script, we defined for each species: (1) species-specific genes, as those which were found in orthogroups consisting of sequences only of the species considered (also known as Taxon Restricted Genes or orphan genes); (2) genes shared among *Bacillus*, as those which were found in orthogroups with at least one CDS of another *Bacillus* species, but lacking any *Phyllium* sequence; (3) genes shared with outgroup, as those which were found in orthogroups with a *Phyllium* sequence.

Cross-Species Differential Expression: for each species, reads from each sample and biological replicate were mapped separately on the reference transcriptome using Bowtie2 (Langmead & Salzberg 2012) with default parameters and RSEM was used to estimate expression values (Li and Dewie, 2011). Using the gene level abundance estimates for each of our samples we constructed a matrix of normalized TPM expression values containing the expression values that are cross-sample normalized using the TMM method to adjust for differences in sample composition (Robinson & Oshlack, 2010). We then run the differential gene expression analyses between reproductive (gonads) and non-reproductive tissues (legs) - separately for each species and sex using Deseq2 with Benjamini and Hochberg's method for multiple tests correction (Love et al, 2014). Subsequently, we gathered the FDR values and LogFC for each transcript based on the orthology inference results, using a custom script. Gonads-biased genes were defined: 1) in parthenogens as those that showed an FDR<0.05 and a LogFC>1 between reproductive and non-reproductive tissues; 2) in *B. grandii* as those that showed an FDR<0.05 and a LogFC>1 between reproductive and non-reproductive tissues, in both sexes. For *B. grandii* we also defined male gonads-biased and female gonads-biased genes as those which had an FDR<0.05 and a LogFC>1 in one sex and were either not differentially expressed or downregulated in the other sex. Shifts in expression patterns between the two parthenogens and the sexual species were visualized using Sankey plots implemented in ggplot2 (Wickham, 2016).

Positive Selection Inference: we screened the 2840 protein-coding genes inferred as one-to-one orthologues across the four species for signatures of positive selection on the branches leading to *Bacillus atticus* and *Bacillus rossius*. Initially we aligned all orthogroups as aminoacids using two local alignment algorithms - l-INS-i and g-INS-i - implemented in MAFFT (Katoh and Standley, 2013). The amino acid alignments were then retro-translated to nucleotides using pal2nal (Suyama et al, 2006) and Gblocks was used to exclude spurious signals coming from misalignments (Talavera and Castresana, 2007). We then inferred instances of positive selection using the branch-site model implemented in codeml (Yang, 2007; Yang and Nielsen 2002; Zhang et al, 2005); all codeml analyses have been carried out using BASE (Forni et al, 2020). Separate analyses have been carried out for each terminal branch of our phylogeny, comparing two models, (a) one which had model = 2 & NSsites = 2 with omega fixed at one and (b) the same model, with omega free to vary. All codeml analyses have been carried out using the fixed species tree and with branch lengths inferred using RAxML with a codon-aware partitioning scheme and a GTR model (Stamatakis, 2014). We compared the likelihoods of the two models with a Likelihood Ratio Test in R (R Core Team, 2017), to assess whether the model including positive selection was a better fit than the one which was not considering it. The analyses were repeated for the two datasets aligned with l-INS.i and g-INS-i strategies: only the genes with a consistent signal of positive selection - independently from the alignment methods - were considered for subsequent analysis. To gain further lines of evidence for a possible role of these genes in the parthenogenetic reproductive process, we only considered genes which had an average TMM-normalized TPM >10 in gonads. We extracted per-gene and per-branch number of sites which were found to be under positive selection by Bayes Empirical Bayes test ($p<0.95$) and retrieved (1) genes undergoing species-specific positive selection (*i.e.* which had at least one site under positive selection in a single parthenogen terminal branch and none in the other species) and/or (2) genes undergoing convergent positive selection (*i.e.* which had at least one site under positive selection in both parthenogen terminal branch and none in the other species). To further confirm that our results were not driven by the alignment strategy, the candidate genes undergoing convergent positive selection were aligned using the homology extension approach implemented in PSI-Coffee (Chang et al, 2012). Codeml analyses were repeated as before and additionally we cross-checked the results obtained by codeml using the aBSREL (adaptive Branch-Site Random Effects Likelihood; Smith et al, 2015) approach implemented in HyPHY (Pond et al, 2005).

Genes Overlaps and GO Enrichment: Fisher's exact test implemented in the R package GeneOverlap (Shen et al, 2020) has been used to determine *p* and odds ratio between the two parthenogens' gene groups described above, in comparison to a genomic background which consisted in all the orthogroups which shared at least two *Bacillus*

species (n=15972). Functional annotation of genes has been carried out separately for each species through blatstp searches against Uniref database (eval<1e-3), combined with hmmer searches against the pfam database; subsequently we generated GO-terms using Argot 2.5 with a TotalScore > 200 (Lavezzo et al, 2016). Subsequently we gathered all GO-terms associated with each OG across the three *Bacillus* species, collapsing multiple entries of the same term. Enrichment analyses were performed with the TopGO package in Bioconductor, using Fisher exact test and both elim and weight algorithms - which took into account GO hierarchy - and a node size of 2 (Alexa and Rahnenführer, 2009). GO-terms were considered to be significantly enriched when elim $p<0.05$; genes associated with enriched terms of interest were retrieved and further characterized using BLASTP and HMMER online implementations (Finn et al, 2011; Potter et al, 2018).

Results

Orthology Inference and Phylostratigraphy of Gonads-biased Genes: definitive assemblies had a comparable number of coding sequences (CDSs): 13,666 for *B. grandii*, 13,703 for *B. atticus* and 14,162 for *B. rossius*. The orthology inference yielded 2,840 orthogroups consisting of single copy genes shared by all four species (*i.e.,* including the outgroup) and 5,283 single copy genes shared between the three *Bacillus* species. Single-copy genes which were found across each parthenogen comparison with the sexual species were respectively 6,329 for *B. atticus* and 6,201 for *B. rossius*. The species tree - inferred according to the STAG (Species Tree inference from All Genes) algorithm implemented in Orthofinder2 - is coherent with previous hypotheses on the clade systematic relationships (Figure 1a; Scali et al, 2003).

The orthogroups clustering was used to estimate a measure of gene origin of each species gene using phylostratigraphy, to understand to which extent the establishment of parthenogenesis is coupled with the evolution of novel genes (Figure 1b). Gonads-biased genes in parthenogens appeared composed by a smaller proportion of species-specific genes compared to what could be observed in the complete transcriptome assemblies, both for *B. atticus* (respectively 2.2% and 3.8%) and *B. rossius* (3.6% and 5.4%, respectively).

Convergent and Species-specific Patterns of Gonads Gene Expression Changes across Automictic Species: to investigate the convergent gene expression changes in parthenogen gonads compared to the bisexual species, we carried out cross-species comparison combining the results of the DE analyses and the orthology inference (Figure 2). Genes with gonads-biased expression in parthenogens appeared to be mainly composed of either gonads-biased (*B. atticus* = 293; *B. rossius* = 311) or female gonads-biased (*B. atticus* = 229; *B. rossius* = 243) in the

bisexual species *B. grandii*. Nonetheless, we also retrieved genes which are either male gonads-biased (*B. atticus* = 93; *B. rossius* = 147) or not gonads-biased in the bisexual species (*B. atticus* = 141; *B. rossius* = 271). Moreover, some gonads-biased (*B. atticus* = 110; *B. rossius* = 85) and female gonads-biased (*B. atticus* = 126; *B. rossius* = 110) genes in the bisexual species were not found to have gonads-biased patterns of expression in parthenogens. As species-specific changes in gene expression can be the result not only of lineage-specific adaptations but also of stochastic changes, we tested whether the overlap across the two interspecific comparisons *B. atticus vs B. grandii* and *B. rossius vs B. grandii* (Figure 3) were significant. All genes subsets considered resulted to be significatively similar across the two parthenogens, a pattern which is consistent with convergent expression changes in parthenogens and similarly constrained processes. As expected, we retrieved a substantial similarity for genes which showed gonads-biased (*p*: 0; odds ratio: 505.9) and non gonads-biased (*p*: 0; odd ratio: 50.5) expression profiles in the bisexual and parthenogenetic species (Supplementary Figure 1). Genes showing a female gonads-biased pattern of expression in the sexual species and in parthenogens largely overlap across parthenogens (pval: 4e-281; odd ratio: 548.7); interestingly, convergent changes were found in parthenogens gonads-biased genes which are male gonads-biased in the bisexual species (*p*: 2e-50; odd ratio: 89.6). Genes which showed a gonads-biased expression in parthenogens but were not gonads-biased in bisexual species showed a significant overlap between the two parthenogens (*p*: 6e-47; odds ratio: 33.5), as those which exhibited a gonads-biased expression pattern in gonads of either bisexual females or across both sexes and were not gonads-biased in parthenogens (respectively *p*: 3.6e-55; odds ratio: 146.5 and *p*: 1.3e-83; odds ratio: 189.3).

Guided by the previous knowledge on the different mechanisms of automixis, we used the results of the GO-terms enrichment as a mean to explore their possible physiological function. Among parthenogens, gonads-biased genes which are male gonads-biased in the bisexual species, many appeared to be related to centrosome formation and males' reproductive mechanisms (Supplementary Table 1): these include the homologs of *centrosomal protein of 152 kDa* (CEP152 in OG0009808; Chang et al, 2010) and *centromere protein J* (CENPJ/CPAP in OG0007643; Kodani et al, 2015). We also retrieved a homolog to *trimethyllysine dioxygenase* (TMLHE in OG0008327), an enzyme which catalyzes the first step in the carnitine biosynthesis pathway; this molecule is found in male seminal fluid of mammals and appears related to sperm count and motility (Blackman et al, 2004; Poiani, 2006). Genes which showed a gonads-biased expression in parthenogens, but which were not gonads-biased in the sexual species (Supplementary Table 2) included a homolog to *spermidine synthase* (SRM in OG0002943); spermidine is a polyamine, a class of molecules which are essential to male and female reproductive processes, embryo development, mating and fertilization efficiency (Bauer et al, 2013; Lefèvre et al, 2011). We also found a *folate*

transporter 1 homolog (SLC19A1 in OG0003254); the latter has been experimentally proven that folate elimination can impede fertility and egg viability in insects (Akov 1967; King 1959). Non gonads-biased genes in parthenogens which are either gonads-biased or female gonads-biased in the bisexual species included genes related to glycans potentially involved in oogenesis and egg coating, such as a homologs to the *mannosyl-oligosaccharide 1,2-alpha-mannosidase* and *galactose-1-phosphate uridylyltransferase* (found in MAN1A1 in OG0006579 and GALT in OG0004972; Cornwall et al, 1991; Akintayo et al, 2019; Nishimura et al, 2019; Avilès et al, 2000). A gene homolog to *mitotic checkpoint protein* (BUB3, found in OG0005535) was also found among them: its product regulates chromosome segregation during oocyte meiosis, with the dual function of spindle-assembly checkpoint signaling and establishment of correct kinetochore-microtubule attachments (Kalitsis et al 2003).

Species-specific changes in gene transcription can reflect a wide range of phenomena at play or can just derive from stochasticity, but they can also be informative of the different mechanisms through which diploidy is restored in the two automictic species. In *B. atticus*, a gonads-biased gene - which doesn't show the same pattern in the gonads of the other species considered - presents a substantial homology with *Anoctamin 6 (*TMEM16F in OG0002369), a phospholipid scramblase involved in endocytosis and which mediates cell-cell fusion in human trofoblast (Bricogne et al, 2019; Zhang et al, 2020).

Convergent and Species-Specific Sequence Modifications across Automictic Species: using our species phylogeny, we explored instances of positive selection within a protein using branch-site codon models. Our screen identified 207 and 166 genes, respectively in *B. rossius* and *B. atticus*, which have at least one site under positive selection ($p<0.95$ in the Bayes Empirical Bayes test) across the replicate analyses carried out using different local alignment strategies (g-INS-i and l-INS-i). To explore the amount of sequence modifications convergently affecting the same genes across the two automictic species we used stringent criteria: (1) we considered the impact of the alignment strategy on our results and we filtered out possible misaligned region; (2) we cross-checked with two different approaches (aBSREL and codeml); (3) we focused on signals shared exclusively among the parthenogenetic species and (4) considered gonads gene expression. Six genes were found to have at least one site under positive selection in both parthenogens and not in the bisexual species within the codeml analyses, but one was not found to have undergone positive selection when cross-checked with aBSREL (Figure 4; Supplementary Table 5).

As for modification of genes expression patterns, we leverage the GO enrichment analyses to characterize genes inferred to have undergone species-specific sequence modifications (Supplementary Table 6). In *B. atticus* we

retrieved two genes homolog to *exocyst complex component 5* and *vacuolar protein sorting 39*, whose products are known to mediate vesicle trafficking and cytokinesis (respectively: EXOC5 in OG0002865 and VPS39 in OG0005132; Wang et al, 2002; Richardson et al, 2004). In *B. rossius* we found two genes involved in spindle orientation: *dynactin subunit one* plays a role in metaphase spindle orientation and is required for microtubule anchoring at the mother centriole (DCTN1 in OG0003368; Kiyomitsu et al 2012; Kodani et al, 2013); *lethal(2)giant larvae protein* is involved in oocyte axis specification and in the regulation of mitotic spindle orientation through microtubule cytoskeleton organization (l(2)gl in G0005517; Bilder et al, 2000; Albertson et al, 2003; Li et al, 2008; Carvalho et al, 2015). Among the five genes which underwent convergent sequence modifications, the products of *pleiotropic regulator 1*, *protein E3 ubiquitin-protein ligase NEDD4* and *OTU deubiquitinase 6B* homologs (PLRG1 in OG0006021; NEDD4 in OG0005237; OTUD6B in OG0004336) are all regulators involved in cell proliferation and differentiation in embryonic and reproductive tissues (Kleinridders et al, 2009; Zhou et al, 2015; Sakata et al, 2004; Sabol et al, 2017). Homolog to *histidine protein methyltransferase 1* (METTL18 in OG0004353) has been identified in silico as a maternal factor potentially interacting with sperm factors (Ntostis et al, 2017). *Mitochondrial amidoxime reducing component 1* (MTARC1 in OG0003097) gene is a signal-anchored protein of the outer mitochondrial membrane in humans (Klein et al, 2012).

Discussion

Our results are consistent with previous findings that a combination of novel genes and modifications of pre-existing ones are associated with novel traits (Albertin et al, 2015; Aguilera et al, 2017; Hilgers et al, 2018; Barua and Mikheyev, 2020). We found that the convergent expression reprogramming of hundreds of genes is coupled with the establishment of automixis. As the sensitivity of our analyses can be affected by the strength of expression changes considered, taking into account more subtle expression modifications could potentially reveal additional instances of convergent changes, thus making our estimate a conservative one. Additionally, the transcriptional program associated with parthenogen gonads is found to be assembled largely from genes that were already present before the establishment of the novel reproductive strategies. Since they are based on trascription data, our phylostratigraphic analyses may have not included some species lowly expressed genes; this potential bias is expected to impact similarly reproductive and non-reproductive tissues and in both cases the number of ancient genes could potentially be even bigger. As such, the large majority of gonads-biased genes in parthenogens originated before the establishment of automixis, coherently with patterns observed for other traits whose

appearance is not clearly coupled with the *de-novo* origination of genes (Jasper et al, 2015; Fisher et al, 2020; Almudi et al 2020). The extent of convergence in gene expression changes across the two independent parthenogens suggests that their transcriptional program results from a constrained process, as observed in other phasmids apomictic species (Parker et al, 2019). While convergent changes seem to have played a prominent role at the transcriptional level, they marginally affected sequence evolution. Only five genes out of the thousands analyzed were found having undergone positive selection across both parthenogens, an outcome which is in stark contrast to instances where evolutionary novelties are associated with pervasive convergent sequence modifications (Wang and Yang, 2021).

The physiological significance of many observed modifications can be associated with the restoration of ploidy without male genetic contribution, such as for the expression reprogramming of homologs to *mitotic checkpoint protein*, *centromere protein J* and *centrosomal protein of 152 kDa*. *Mitotic checkpoint protein* product regulates chromosome segregation during oocyte meiosis, with the dual function of spindle-assembly checkpoint signaling and establishment of correct kinetochore-microtubule attachments (Kalitsis et al 2003). The depletion of *centromere protein J* is essential for centrosome integrity and spindle bipolarity and its overexpression causes the generation of multinucleated cells and abnormalities in spindle formation, positioning and orientation (Cho et al, 2006; Kohlmaier et al, 2009; Lee et al, 2014). C*entrosomal protein of 152 kDa* (known as *asterless* in *Drosophila*) is associated with the failure of centrosome duplication, while its overexpression led to *de novo* centriole formation and duplication of centrosomes in embryos (Dzhindzhev et al, 2010). Aforementioned changes could underlie a partially shared background for further species-specific modifications associated with the different mechanisms through which ploidy is restored in the two species. The diploidy of *B. atticus* is the result of the fusion of the products of the first meiotic division. Homologs to *exocyst complex component 5* and *vacuolar protein sorting 39* were found to have undergone species-specific modifications in *B. atticus*: they are known to mediate vesicle trafficking/membrane fusion and could underlie the fusion of the nuclei produced by the first meiotic division (Wang et al, 2002; Richardson et al, 2004; Bricogne et al, 2019; Zhang et al, 2020). Furthermore, genes which could be involved in the anaphase restitution responsible for *B. rossius* blastula diploidization are the *dynactin subunit one* and *lethal(2)giant larvae* homologs, which are involved in spindle orientation and oocyte axis specification; their modification could alter karyokinesis and cytokinesis in the blastula, impeding the separation of chromatids and resulting in restitutional nuclei.

Other than ploidy, an additional obstacle to the establishment of reproduction in the absence of males is represented by egg development activation: in bisexual species the contribution of centrioles and the removal of the

meiotic block are usually accomplished by sperm and seminal fluids. Yet, several arthropods appear to have overcome these constrains long before the shift to single-sex reproduction as the initiation of oocytes embryonic development is typically triggered by events independent from males (Went, 1982). Moreover, in *Bacillus*, centrosomes are assembled from maternal components without any paternal contribution (Marescalchi et al, 2002). Nonetheless, insect ovideposition appears to be tightly linked to mating, so that modifications could be necessary to maintain a high rate in the absence of mating-related clues. The latter represents a strong physiological constrain, as highlighted by instances where males are maintained without any other clear purpose (Miyakawa and Mikheyev, 2015). Some of the genes which underwent convergent transcriptional reprogramming in the parthenogen gonads are associated with the production of molecules - such as carnitine or spermidine - which are present in male seminal fluids and have a role in fertilization or even act as pheromones in some animals (Lefèvre et al, 2011; Scott et al, 2019). Regarding this aspect, an interesting gene identified by convergent selection analyses is the homolog to *histidine methyltransferase*; this was found associated with developmental competence in oocytes and it has been identified *in silico* as a maternal factor potentially interacting with the sperm factor Hdac11 and thus with a potential role in processes downstream the sperm-egg fusion (Ntostis et al, 2017; Biase et al, 2018). While convergent changes in parthenogens could be explained by a plethora of non-exlcusive phenomena, they could potentially compensate for the lack of products which are provided by sperm in bisexual systems and are necessary to sustain female fertility (Parker et al, 2019).

Not all of the observed modifications are expected to concur to the establishment of automixis; many could also result from the removal of selective pressures on mating-associated mechanisms and their subsequent decay (Kraaijeveld et al, 2016; Parker et al, 2019). Interestingly, some gene expression changes seem related the production of glycans involved in oogenesis and egg coating. Their gonads-biased pattern of expression is lost in parthenogens, possibly because the absence of sperm-egg interaction makes their maintenance unnecessary. As for other systems, the convergent changes in ancient and more recent parthenogens let us hypothesize that selection is the driving force behind these processes (Schwander, et al. 2013).

Our analyses didn't find changes associated with a large number of genes with similar functions, but this is consistent with the observed association of reproductive strategies with few - or even single - genes (Wallberg et al, 2016; Parker et al, 2018; Yagound et al, 2020; Ma et al, 2021). Genes reported above have known functions in model species which are somehow compatible with a role in the establishment of automixis; yet, the phenotypic effects of their sequence and expression modification is unknown in the context of the species considered and the changes could also be the result of unrelated phenomena or stochastic drift. Even without considering the

physiological role of the modifications, the repeated evolution of automixis in *Bacillus* appears to be coupled with the convergent transcriptional reprogramming of a core set of genes. Our interpretation of the differences between the sexual and automictic species is sustained by the concept that the latter are the result of two independent shifts in reproductive strategy, with the ancestral state of the clade being sexual. While reversions from parthenogenesis to sexuality have been proposed (Domes et al, 2017), the fact that the species carry out two entirely different automictic mechanisms of reproduction make the possibility of a parthenogenetic ancestor rather improbable. As such, the strong signature of convergent transcriptional reprogramming associated with the independent establishment of automixis highlights how similar novelties can still be associated with convergent evolutionary paths even when attained by different mechanisms.

References

Aguilera, F., McDougall, C. and Degnan, B.M., 2017. Co-option and de novo gene evolution underlie molluscan shell diversity. Molecular biology and evolution, 34(4), pp.779-792.

Akintayo, A.A. and Stanley, P., 2019. Roles for Golgi glycans in oogenesis and spermatogenesis. Frontiers in Cell and Developmental Biology, 7, p.98.

Akov, S., 1967. Effect of folic acid antagonists on larval development and egg production in *Aedes aegypti*. Journal of insect physiology, 13(6), pp.9 13-923.

Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W. and Rokhsar, D.S., 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature, 524(7564), pp.220-224.

Albertson, R. and Doe, C.Q., 2003. Dlg, Scrib and Lgl regulate neuroblast cell size and mitotic spindle asymmetry. Nature cell biology, 5(2), pp.166-170.

Alexa, A., & Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. Bioconductor Improv, 27.

Almudi, I., Vizueta, J., Wyatt, C.D., de Mendoza, A., Marlétaz, F., Firbas, P.N., Feuda, R., Masiero, G., Medina, P., Alcaina-Caro, A. and Cruz, F., 2020. Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings. Nature communications, 11(1), pp.1-11

Avilés, M., Okinaga, T., Shur, B.D. and Ballesta, J., 2000. Differential expression of glycoside residues in the mammalian zona pellucida. Molecular Reproduction and Development: Incorporating Gamete Research, 57(3), pp.296-308.

Babonis, L.S., Martindale, M.Q. and Ryan, J.F., 2016. Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone Nematostella vectensis. BMC evolutionary biology, 16(1), pp.1-22.

Barua, Agneesh, and Alexander S. Mikheyev. An ancient, conserved gene regulatory network led to the rise of oral venom systems. bioRxiv (2020).

Bast, J., Parker, D.J., Dumas, Z., Jalvingh, K.M., Tran Van, P., Jaron, K.S., Figuet, E., Brandt, A., Galtier, N. and Schwander, T., 2018. Consequences of asexuality in natural populations: insights from stick insects. Molecular biology and evolution, 35(7), pp.1668-1677.

Bauer, M.A., Carmona-Gutiérrez, D., Ruckenstuhl, C., Reisenbichler, A., Megalou, E.V., Eisenberg, T., Magnes, C., Jungwirth, H., Sinner, F.M., Pieber, T.R. and Fröhlich, K.U., 2013. Spermidine promotes mating and fertilization efficiency in model organisms. Cell Cycle, 12(2), pp.346-352.

Biase, F.H. and Kimble, K.M., 2018. Functional signaling and gene regulatory networks between the oocyte and the surrounding cumulus cells. BMC genomics, 19(1), pp.1-13.

Bilder, D., Li, M. and Perrimon, N., 2000. Cooperative regulation of cell polarity and growth by Drosophila tumor suppressors. Science, 289(5476), pp.113-116.

Bonandin, L., Scavariello, C., Luchetti, A. and Mantovani, B. (2014). Evolutionary dynamics of R2 retroelement and insertion inheritance in the genome of bisexual and parthenogenetic *Bacillus rossius* populations (Insecta Phasmida). Insect Molecular Biology, 23(6): 808-820

Bricogne, C., Fine, M., Pereira, P.M., Sung, J., Tijani, M., Wang, Y., Henriques, R., Collins, M.K. and Hilgemann, D.W., 2019. TMEM16F activation by Ca 2+ triggers plasma membrane expansion and directs PD-1 trafficking. Scientific reports, 9(1), pp.1-13.

Burskaia, V., Naumenko, S., Schelkunov, M., Bedulina, D., Neretina, T., Kondrashov, A., Yampolsky, L. and Bazykin, G.A., 2020. Excessive parallelism in protein evolution of Lake Baikal amphipod species flock. Genome biology and evolution, 12(9), pp.1493-1503.

Carvalho, C.A., Moreira, S., Ventura, G., Sunkel, C.E. and Morais-de-Sá, E., 2015. Aurora A triggers Lgl cortical release during symmetric division to control planar spindle orientation. Current Biology, 25(1), pp.53-60.

Casewell, N.R., Petras, D., Card, D.C., Suranse, V., Mychajliw, A.M., Richards, D., Koludarov, I., Albulescu, L.O., Slagboom, J., Hempel, B.F. and Ngum, N.M., 2019. Solenodon genome reveals convergent evolution of venom in eulipotyphlan mammals. Proceedings of the National Academy of Sciences, 116(51), pp.25745-25755.

Chang, J., Cizmecioglu, O., Hoffmann, I. and Rhee, K., 2010. PLK2 phosphorylation is critical for CPAP function in procentriole formation during the centrosome cycle. The EMBO journal, 29(14), pp.2395-2406.

Chang, J.M., Di Tommaso, P., Taly, J.F. and Notredame, C., 2012. Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. BMC bioinformatics, 13(S4), p.S1.

Cho, Jun-Hung, Chih-Jui Chang, Chiung-Ya Chen, and Tang K. Tang. "Depletion of CPAP by RNAi disrupts centrosome integrity and induces multipolar spindles." Biochemical and biophysical research communications 339, no. 3 (2006): 742-747.

Corbett-Detig, R.B., Russell, S.L., Nielsen, R. and Losos, J., 2020. Phenotypic convergence is not mirrored at the protein level in a lizard adaptive radiation. Molecular Biology and Evolution, 37(6), pp.1604-1614.

Cornwall, G.A., Tulsiani, D.R.P. and Orgebin-Crist, M.C., 1991. Inhibition of the mouse sperm surface α-D-mannosidase inhibits sperm-egg binding in vitro. Biology of Reproduction, 44(5), pp.913-921.

Dalbies-Tran, R., Cadoret, V., Desmarchais, A., Elis, S., Maillard, V., Monget, P., Monniaux, D., Reynaud, K., Saint-Dizier, M. and Uzbekova, S., 2020. A Comparative Analysis of Oocyte Development in Mammals. Cells, 9(4), p.1002.

Domazet-Lošo, T., Brajković, J. and Tautz, D., 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. Trends in Genetics, 23(11), pp.533-539.

Domazet-Lošo, T. and Tautz, D., 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature, 468(7325), pp.815-818.

Domes, K., Norton, R.A., Maraun, M. and Scheu, S., 2007. Reevolution of sexuality breaks Dollo's law. Proceedings of the National Academy of Sciences, 104(17), pp.7139-7144.

Dzhindzhev, N.S., Quan, D.Y., Weiskopf, K., Tzolovsky, G., Cunha-Ferreira, I., Riparbelli, M., Rodrigues-Martins, A., Bettencourt-Dias, M., Callaini, G. and Glover, D.M., 2010. Asterless is a scaffold for the onset of centriole assembly. Nature, 467(7316), pp.714-718.

Engelstädter, J., 2017. Asexual but not clonal: evolutionary processes in automictic populations. Genetics, 206(2), pp.993-1009.

Finn, R.D., Clements, J. and Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic acids research, 39(suppl_2), pp.W29-W37.

Fisher, C.R., Wegrzyn, J.L. and Jockusch, E.L., 2020. Co-option of wing-patterning genes underlies the evolution of the treehopper helmet. Nature Ecology & Evolution, 4(2), pp.250-260.

Foote, A.D., Liu, Y., Thomas, G.W., Vinař, T., Alföldi, J., Deng, J., Dugan, S., van Elk, C.E., Hunter, M.E., Joshi, V. and Khan, Z., 2015. Convergent evolution of the genomes of marine mammals. Nature genetics, 47(3), pp.272-275.

Forni, G., Ruggeri, A.A., Piccinini, G. and Luchetti, A., 2020. BASE: a novel workflow to integrate non-ubiquitous genes in genomics analyses for selection. bioRxiv.

Hilgers, L., Hartmann, S., Hofreiter, M. and von Rintelen, T., 2018. Novel genes, ancient genes, and gene co-option contributed to the genetic basis of the radula, a molluscan innovation. Molecular biology and evolution, 35(7), pp.1638-1652.

Jaron, K.S., Bast, J., Nowell, R.W., Ranallo-Benavidez, T.R., Robinson-Rechavi, M. and Schwander, T., 2020. Genomic features of parthenogenetic animals. bioRxiv, p.497495.

Jasper, W.C., Linksvayer, T.A., Atallah, J., Friedman, D., Chiu, J.C. and Johnson, B.R., 2015. Large-scale coding sequence change underlies the evolution of post developmental novelty in honeybees. Molecular biology and evolution, 32(2), pp.334-346.

Jebb, D., Huang, Z., Pippel, M., Hughes, G.M., Lavrichenko, K., Devanna, P., Winkler, S., Jermiin, L.S., Skirmuntt, E.C., Katzourakis, A. and Burkitt-Gray, L., 2020. Six reference-quality genomes reveal evolution of bat adaptations. Nature, 583(7817), pp.578-584.

Kalitsis, P., Earle, E., Fowler, K.J. and Choo, K.A., 2000. Bub3 gene disruption in mice reveals essential mitotic spindle checkpoint function during early embryogenesis. Genes & development, 14(18), pp.2277-2282.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), pp.772-780.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. and Bosch, T.C., 2009. More than just orphans: are taxonomically-restricted genes important in evolution? Trends in Genetics, 25(9), pp.404-413.

King, R.C. and Sang, J.H., 1959. Oogenesis in adult *Drosophila melanogaster*. 8. The role of folic acid in oogenesis. Growth, 23, pp.37-53.

Kiyomitsu, T. and Cheeseman, I.M., 2012. Chromosome-and spindle-pole-derived signals generate an intrinsic code for spindle position and orientation. Nature cell biology, 14(3), pp.311-317.

Kleinridders, A., Pogoda, H.M., Irlenbusch, S., Smyth, N., Koncz, C., Hammerschmidt, M. and Brüning, J.C., 2009. PLRG1 is an essential regulator of cell proliferation and apoptosis during vertebrate development and tissue homeostasis. Molecular and cellular biology, 29(11), pp.3173-3185.

Kodani, A., Salomé Sirerol-Piquer, M., Seol, A., Manuel Garcia-Verdugo, J. and Reiter, J.F., 2013. Kif3a interacts with Dynactin subunit p150Glued to organize centriole subdistal appendages. The EMBO journal, 32(4), pp.597-607.

Kodani, A., Timothy, W.Y., Johnson, J.R., Jayaraman, D., Johnson, T.L., Al-Gazali, L., Sztriha, L., Partlow, J.N., Kim, H., Krup, A.L. and Dammermann, A., 2015. Centriolar satellites assemble centrosomal microcephaly proteins to recruit CDK2 and promote centriole duplication. Elife, 4, p.e07519.

Kohlmaier, G., Lončarek, J., Meng, X., McEwen, B.F., Mogensen, M.M., Spektor, A., Dynlacht, B.D., Khodjakov, A. and Gönczy, P., 2009. Overly long centrioles and defective cell division upon excess of the SAS-4-related protein CPAP. Current Biology, 19(12), pp.1012-1018.

Kraaijeveld, K., Anvar, S.Y., Frank, J., Schmitz, A., Bast, J., Wilbrandt, J., Petersen, M., Ziesmann, T., Niehuis, O., De Knijff, P. and Den Dunnen, J.T., 2016. Decay of sexual trait genes in an asexual parasitoid wasp. Genome biology and evolution, 8(12), pp.3685-3695.

Lavezzo, E., Falda, M., Fontana, P., Bianco, L. and Toppo, S., 2016. Enhancing protein function prediction with taxonomic constraints–The Argot2. 5 web server. Methods, 93, pp.15-23.

Lee, Miseon, Jaerak Chang, Sunghoe Chang, Kyung S. Lee, and Kunsoo Rhee. "Asymmetric spindle pole formation in CPAP-depleted mitotic cells." Biochemical and Biophysical Research Communications 444, no. 4 (2014): 644-650.

Lefèvre, P.L., Palin, M.F. and Murphy, B.D., 2011. Polyamines on the reproductive landscape. Endocrine reviews, 32(5), pp.694-712.

Lehtonen, J., Jennions, M.D. and Kokko, H., 2012. The many costs of sex. Trends in ecology & evolution, 27(3), pp.172-178.

Ma, W.-J., Pannebakker, B.A., Li, X., Geuverink, E., Anvar, S.Y., Veltsos, P., Schwander, T., van de Zande, L. and Beukeboom, L.W., 2021. A single QTL with large effect is associated with female functional virginity in an asexual parasitoid wasp. Molecular Ecology. Accepted Author Manuscript.

Mantovani, B., Passamonti, M. and Scali, V., 2001. The mitochondrial cytochrome oxidase II gene in *Bacillus* stick insects: ancestry of hybrids, androgenesis, and phylogenetic relationships. Molecular phylogenetics and evolution, 19(1), pp.157-163.

Marescalchi, O., Pijnacker, L.P. and Scali, V., 1993. Automictic parthenogenesis and its genetic consequence in *Bacillus atticus atticus* (Insecta Phasmatodea). Invertebrate reproduction & development, 24(1), pp.7-12.

Marescalchi, O., Zauli, C. and Scali, V., 2002. Centrosome dynamics and inheritance in related sexual and parthenogenetic *Bacillus* (Insecta Phasmatodea). Molecular Reproduction and Development: Incorporating Gamete Research, 63(1), pp.89-95.

Matsuura, K. 2017. Evolution of the asexual queen succession system and its underlying mechanisms in termites. Journal of Experimental Biology 220, 63-72

McGirr, J.A. and Martin, C.H., 2020. Few fixed variants between trophic specialist pupfish species reveal candidate cis-regulatory alleles underlying rapid craniofacial divergence. Molecular Biology and Evolution.

McLysaght, A. and Guerzoni, D., 2015. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1678), p.20140332.

Milde, S., Hemmrich, G., Anton-Erxleben, F., Khalturin, K., Wittlieb, J. and Bosch, T.C., 2009. Characterization of taxonomically restricted genes in a phylum-restricted cell type. Genome biology, 10(1), pp.1-16.

Mirzaghaderi, G. and Hörandl, E., 2016. The evolution of meiotic sex and its alternatives. Proceedings of the Royal Society B: Biological Sciences, 283(1838), p.20161221.

Miyakawa, M.O. and Mikheyev, A.S., 2015. Males are here to stay: fertilization enhances viable egg production by clonal queens of the little fire ant (*Wasmannia auropunctata*). The Science of Nature, 102(3-4), p.15.

Mogie, M., 1986. Automixis: its distribution and status. Biological Journal of the Linnean Society, 28(3), pp.321-329.

Neiman, M., Sharbel, T.F. and Schwander, T., 2014. Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. Journal of evolutionary biology, 27(7), pp.1346-1359.

NG, C.M., Blackman, M.R., Wang, C. and Swerdloff, R.S., 2004. The role of carnitine in the male reproductive system. Annals of the New York Academy of Sciences, 1033(1), pp.177-188.

Nishimura, K., Dioguardi, E., Nishio, S., Villa, A., Han, L., Matsuda, T. and Jovine, L., 2019. Molecular basis of egg coat cross-linking sheds light on ZP1-associated female infertility. Nature communications, 10(1), pp.1-15.

Normark, 2013 : Brenner's Encyclopedia of Genetics, 2nd edition, Volume 5 doi:10.1016/B978-0-12-374984-0.01123-2 233

Ntostis, P., Carter, D., Iles, D., Huntriss, J., Tzetis, M. and Miller, D., 2017. Potential sperm contributions to the murine zygote predicted by in silico analysis. Reproduction, 154(6), pp.777-788.

Oxley, P.R., Ji, L., Fetter-Pruneda, I., McKenzie, S.K., Li, C., Hu, H., Zhang, G. and Kronauer, D.J., 2014. The genome of the clonal raider ant *Cerapachys biroi*. Current Biology, 24(4), pp.451-458.

Pál, C. and Papp, B., 2017. Evolution of complex adaptations in molecular systems. Nature ecology & evolution, 1(8), pp.1084-1092.

Parker, D.J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M. and Schwander, T., 2019. Repeated evolution of asexuality involves convergent gene expression changes. Molecular biology and evolution, 36(2), pp.350-364.

Parker, D.J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M. and Schwander, T., 2019. Sex-biased gene expression is repeatedly masculinized in asexual females. Nature communications, 10(1), pp.1-11.

Pijnacker, L.P., 1969. Automictic parthenogenesis in the stick insect *Bacillus rossius* Rossi (Cheleutoptera, phasmidae). Genetica, 40(1), pp.393-399.

Poiani, A. (2006). Complexity of seminal fluid: a review. Behavioral Ecology and Sociobiology, 60(3), 289-310.

Pond, S.L.K. and Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. In Statistical methods in molecular evolution (pp. 125-181). Springer, New York, NY.

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R. and Finn, R.D., 2018. HMMER web server: 2018 update. Nucleic acids research, 46(W1), pp.W200-W204.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richardson, S.C., Winistorfer, S.C., Poupon, V., Luzio, J.P. and Piper, R.C., 2004. Mammalian late vacuole protein sorting orthologues participate in early endosomal fusion and interact with the cytoskeleton. Molecular biology of the cell, 15(3), pp.1197-1210.

Rosenblum, E.B., Parent, C.E. and Brandt, E.E., 2014. The molecular basis of phenotypic convergence. Annual Review of Ecology, Evolution, and Systematics, 45, pp.203-226.

Sackton TB and Clark N, 2019. Convergent evolution in the genomics era: new insights and directions. Phil. Trans. R. Soc. B37420190102

Sakata, T., Sakaguchi, H., Tsuda, L., Higashitani, A., Aigaki, T., Matsuno, K. and Hayashi, S., 2004. *Drosophila Nedd4* regulates endocytosis of notch and suppresses its ligand-independent activation. Current biology, 14(24), pp.2228-2236.

Santos, M.E., Le Bouquin, A., Crumière, A.J. and Khila, A., 2017. Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. Science, 358(6361), pp.386-390.

Scali, V., Passamonti, M., Marescalchi, O. and Mantovani, B., 2003. Linkage between sexual and asexual lineages: genome evolution in *Bacillus* stick insects. Biological Journal of the Linnean Society, 79(1), pp.137-150.

Schwander, T., Vuilleumier, S., Dubman, J. and Crespi, B.J., 2010. Positive feedback in the transition from sexual reproduction to parthenogenesis. Proceedings of the Royal Society B: Biological Sciences, 277(1686), pp.1435-1442.

Scott, A.M., Zhang, Z., Jia, L., Li, K., Zhang, Q., Dexheimer, T., Ellsworth, E., Ren, J., Chung-Davidson, Y.W., Zu, Y. and Neubig, R.R., 2019. Spermine in semen of male sea lamprey acts as a sex pheromone. PLoS biology, 17(7), p.e3000332.

Shen L, Sinai ISoMaM (2020). GeneOverlap: Test and visualize gene overlaps. R package version 1.26.0, http://shenlab-sinai.github.io/shenlab-sinai/.

Shubin, N., Tabin, C. and Carroll, S., 2009. Deep homology and the origins of evolutionary novelty. Nature, 457(7231), pp.818-823.

Sim, A.D. and Wheeler, D., 2016. The venom gland transcriptome of the parasitoid wasp *Nasonia vitripennis* highlights the importance of novel genes in venom function. BMC genomics, 17(1), p.571.

Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K. and Kosakovsky Pond, S.L., 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Molecular biology and evolution, 32(5), pp.1342-1353.

Sobol, A., Askonas, C., Alani, S., Weber, M.J., Ananthanarayanan, V., Osipo, C. and Bocchetta, M., 2017. Deubiquitinase OTUD6B isoforms are important regulators of growth and proliferation. Molecular Cancer Research, 15(2), pp.117-127.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), pp.1312-1313.

Stern, D.L., 2013. The genetic causes of convergent evolution. Nature Reviews Genetics, 14(11), pp.751-764.

Suomalainen, E., Saura, A. and Lokki, J., 1987. Cytology and evolution in parthenogenesis. CRC Press.

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic acids research, 34(suppl_2), W609-W612.

Svendsen, N., Reisser, C.M., Dukić, M., Thuillier, V., Ségard, A., Liautard-Haag, C., Fasel, D., Hürlimann, E., Lenormand, T., Galimov, Y. and Haag, C.R., 2015. Uncovering cryptic asexuality in *Daphnia magna* by RAD sequencing. Genetics, 201(3), pp.1143-1155.

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

Tautz, D. and Domazet-Lošo, T., 2011. The evolutionary origin of orphan genes. Nature Reviews Genetics, 12(10), pp.692-702.

Tilquin, A. and Kokko, H., 2016. What does the geography of parthenogenesis teach us about sex? Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1706), p.20150538.

Tvedte, E.S., Forbes, A.A. and Logsdon Jr, J.M., 2017. Retention of core meiotic genes across diverse Hymenoptera. Journal of Heredity, 108(7), pp.791-806.

Tvedte, E.S., Logsdon Jr, J.M. and Forbes, A.A., 2019. Sex loss in insects: causes of asexuality and consequences for genomes. Current opinion in insect science, 31, pp.77-83.

van Kruistum, H., Nijland, R., Reznick, D.N., Groenen, M.A., Megens, H.J. and Pollux, B.J., 2021. Parallel genomic changes drive repeated evolution of placentas in live-bearing fish. Molecular Biology and Evolution.

Wagner, G.P. and Lynch, V.J., 2010. Evolutionary novelties. Current Biology, 20(2), pp.R48-R52.

Wallberg, A., Pirk, C.W., Allsopp, M.H. and Webster, M.T., 2016. Identification of multiple loci associated with social parasitism in honeybees. PLoS genetics, 12(6), p.e1006097.

Wang, Y. and Yang, L., 2021. Genomic evidence for convergent molecular adaptation in electric fishes. Genome Biology and Evolution.

Wang, H., Tang, X., Liu, J., Trautmann, S., Balasundaram, D., McCollum, D. and Balasubramanian, M.K., 2002. The multiprotein exocyst complex is essential for cell separation in *Schizosaccharomyces pombe*. Molecular biology of the cell, 13(2), pp.515-529.

Warner, M.R., Qiu, L., Holmes, M.J., Mikheyev, A.S. and Linksvayer, T.A., 2019. Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. Nature communications, 10(1), pp.1-11.

Went, D.F., 1982. Egg activation and parthenogenetic reproduction in insects. Biological Reviews, 57(2), pp.319-344.

Wipfler, B., Letsch, H., Frandsen, P.B., Kapli, P., Mayer, C., Bartel, D., Buckley, T.R., Donath, A., Edgerly-Rooks, J.S., Fujita, M. and Liu, S., 2019. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. Proceedings of the National Academy of Sciences, 116(8), pp.3024-3029.

Yagound, B., Dogantzis, K.A., Zayed, A., Lim, J., Broekhuyse, P., Remnant, E.J., Beekman, M., Allsopp, M.H., Aamidor, S.E., Dim, O. and Buchmann, G., 2020. A Single Gene Causes Thelytokous Parthenogenesis, the Defining Feature of the Cape Honeybee *Apis mellifera capensis*. Current Biology.

Yang, Z. and Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular biology and evolution, 19(6), pp.908-917.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution, 24(8), pp.1586-1591.

Yuan, J., Zhang, X., Gao, Y., Zhang, X., Liu, C., Xiang, J. and Li, F., 2020. Adaptation and molecular evidence for convergence in decapod crustaceans from deep-sea hydrothermal vent environments. Molecular Ecology, 29(20), pp.3954-3969.

Zhang, J., Nielsen, R. and Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular biology and evolution, 22(12), pp.2472-2479.

Zhang, Y., Le, T., Grabau, R., Mohseni, Z., Kim, H., Natale, D.R., Feng, L., Pan, H. and Yang, H., 2020. TMEM16F phospholipid scramblase mediates trophoblast fusion and placental development. Science Advances, 6(19), p.eaba0310.

Zhou, Z., Kawabe, H., Suzuki, A., Shinmyozu, K. and Saga, Y., 2017. NEDD4 controls spermatogonial stem cell homeostasis and stress response by regulating messenger ribonucleoprotein complexes. Nature communications, 8(1), pp.1-13.

Zou, Z. and Zhang, J., 2015. No genome-wide protein sequence convergence for echolocation. Molecular Biology and Evolution, 32(5), pp.1237-1241.

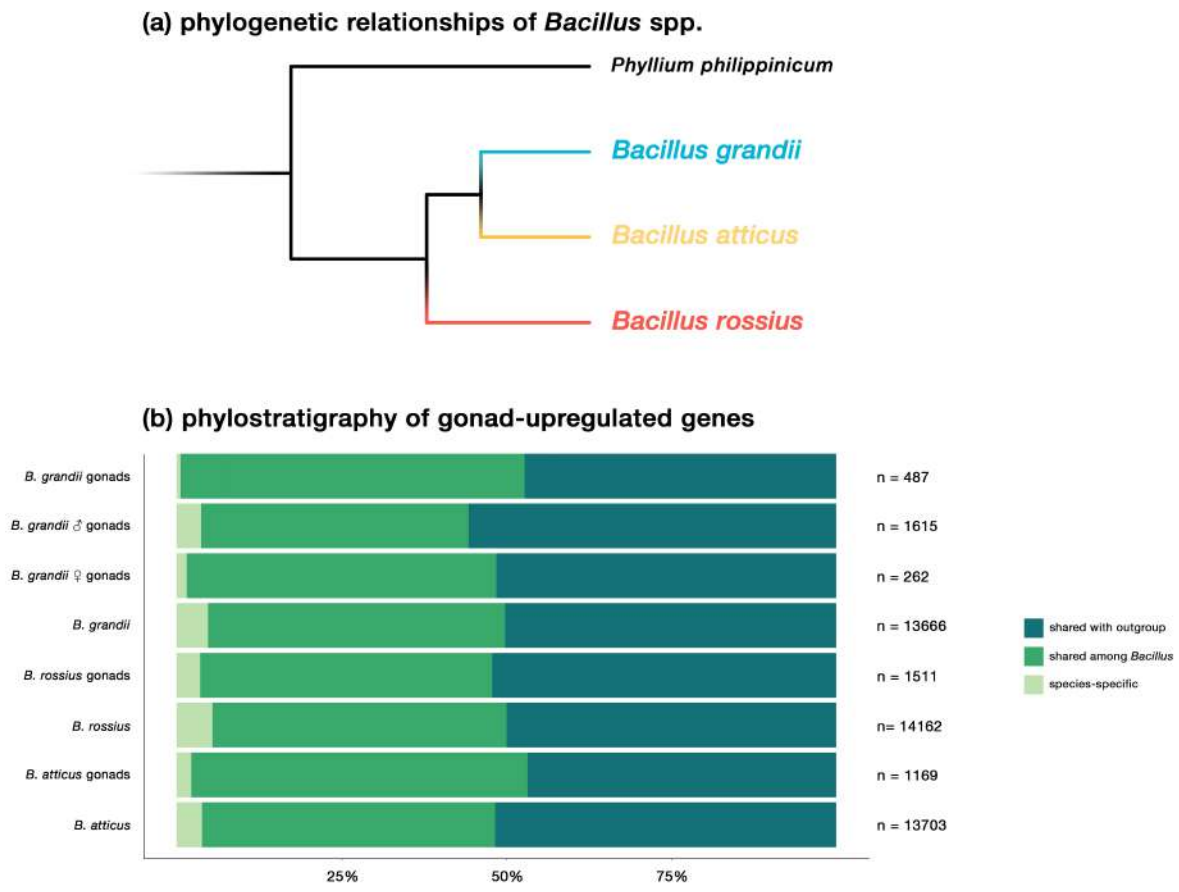


Figure 1 - Phylogenetic relationships of *Bacillus* species and phylostratigraphy of gonad upregulated genes. (a) Species tree inferred using the STAG method; asterisks correspond to the shift from bisexual reproduction to automixis. (b) Phylostratigraphy of gonad upregulated genes; numbers on the right represent the number of transcripts present in each subset / assembly.

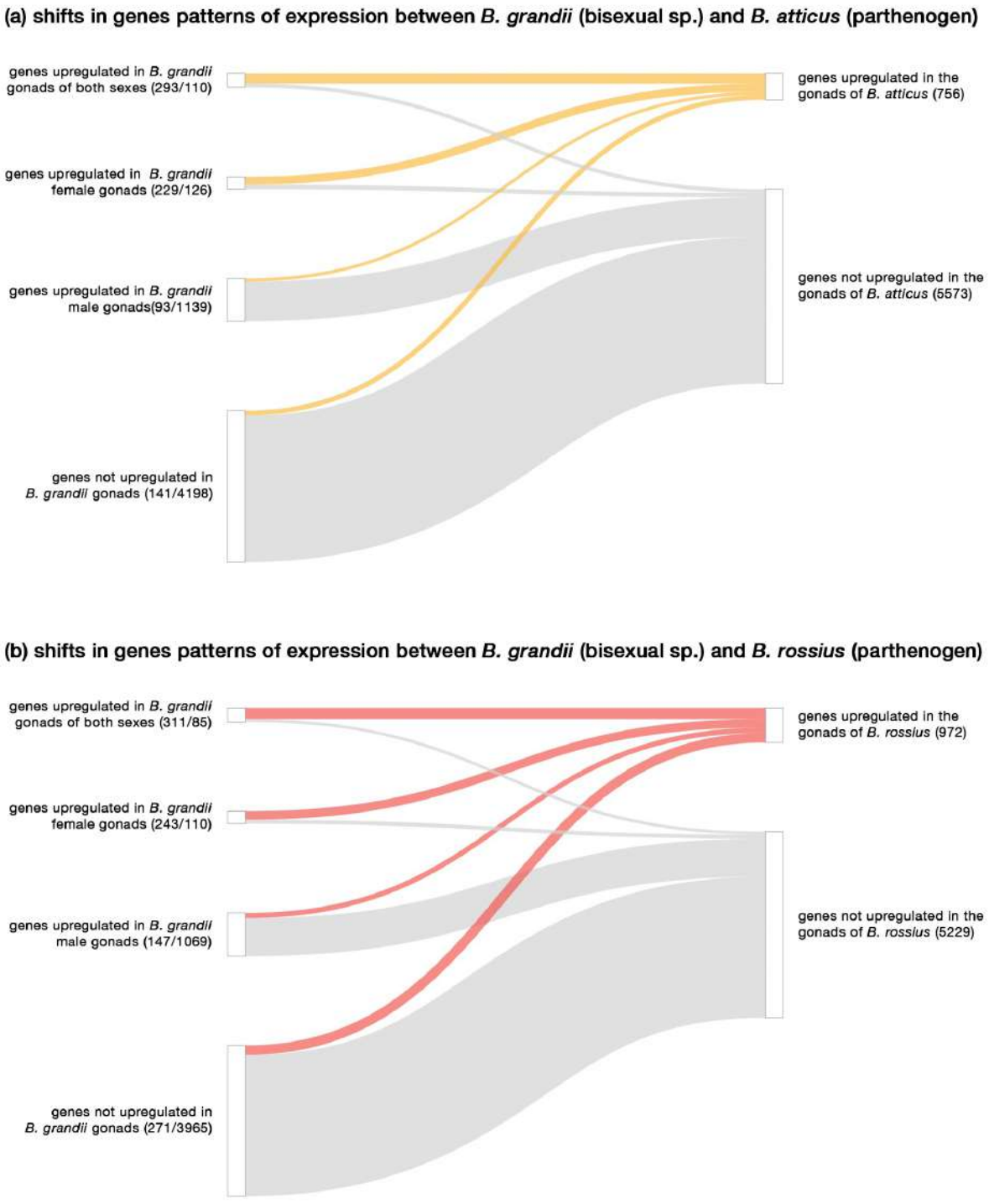


Figure 2 - Shifts in composition of genes with gonad-specific expression: (a) comparison between the automictic *Bacillus rossius* and the bisexual *Bacillus grandii*; (b) comparison between the automictic *Bacillus atticus* and the bisexual *Bacillus grandii*. Numbers in parentheses are the genes with gonad-specific or not-specific expression regulation, respectively.

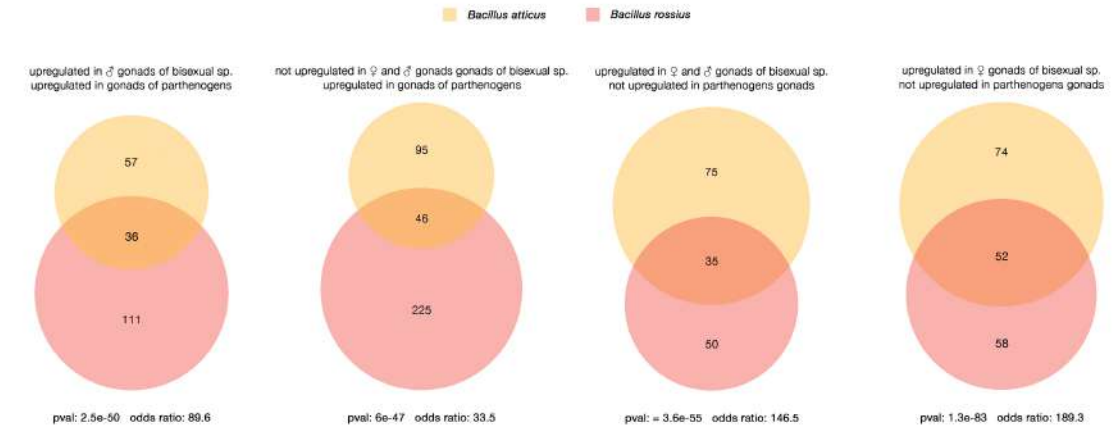


Figure 3 - Convergent transcriptional modification across the two parthenogen species: Venn-diagrams represent the genes with changes in expression between parthenogens and the bisexual species. Overlaps represent shared changes across the two parthenogens for which Fisher's exact test has been used to determine *p* and odds ratio, in comparison to the genomic background (n=15972).

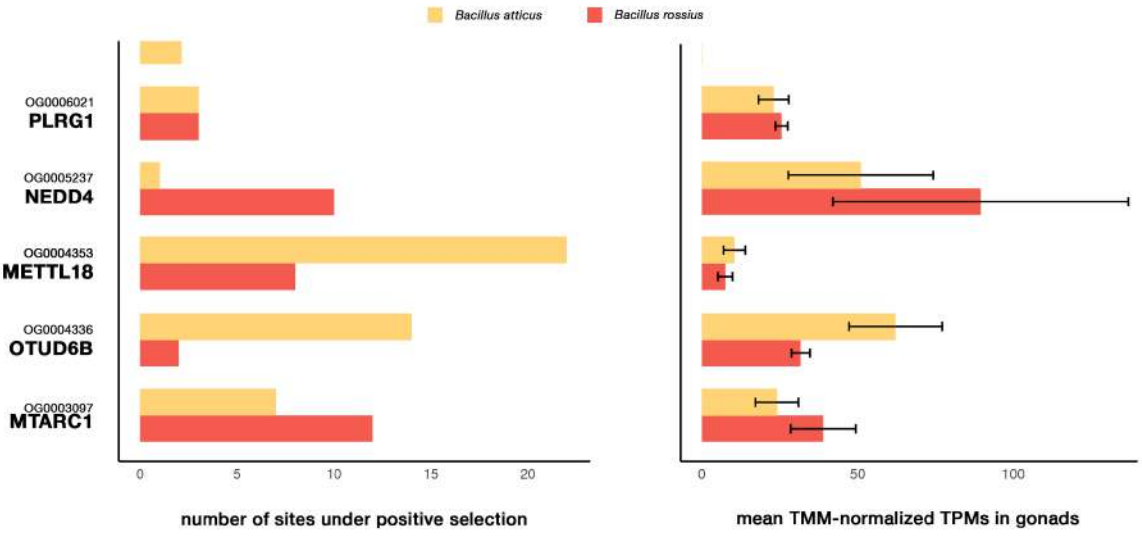


Figure 4 - Convergent instances of positive selection: on the right, number of sites inferred to have undergone positive selection (BEB>95) in the two parthenogens; on the left, TMM-normalized expression values for the same genes. Genes with <10 TMMs in parthenogen gonads were not considered. On the y axis the orthogroups in which the genes are found and the putative homologs (in bold) are given.

Supplementary Figure 1 - Convergent transcriptional modification across the two parthenogen species: Venn-diagrams represent the genes with changes in expression between parthenogens and the bisexual species. Overlaps represent shared changes across the two parthenogens for which Fisher's exact test has been used to determine *p* and odds ratio, in comparison to the genomic background (n=15972).

Sup. Tab. 1 a - genes upregulated in male gonads of *B. grandii* and upregulated in gonads of *B. atticus* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0006434	seryl-tRNA aminoacylation	1	1	0.01	1	0.0057	0.0057
GO:0019370	leukotriene biosynthetic process	1	1	0.01	2	0.0057	0.0057
GO:1902975	mitotic DNA replication initiation	1	1	0.01	3	0.0057	0.0057
GO:0000467	exonucleolytic trimming to generate matu...	1	1	0.01	4	0.0057	0.0057
GO:0097056	selenocysteinyl-tRNA(Sec) biosynthetic p...	1	1	0.01	5	0.0057	0.0057
GO:0031087	deadenylation-independent decapping of n...	1	1	0.01	6	0.0057	0.0057
GO:0071340	skeletal muscle acetylcholine-gated chan...	1	1	0.01	7	0.0057	0.0057
GO:0019303	D-ribose catabolic process	1	1	0.01	8	0.0057	0.0057
GO:0007288	sperm axoneme assembly	1	1	0.01	9	0.0057	0.0057
GO:0060628	regulation of ER to Golgi vesicle-mediat...	2	1	0.01	10	0.0114	0.0114
GO:0016125	sterol metabolic process	4	1	0.02	11	0.0227	0.0227
GO:0006888	endoplasmic reticulum to Golgi vesicle-m...	5	1	0.03	15	0.0282	10.000

Sup. Tab. 1 b - genes upregulated in male gonads of *B. grandii* and upregulated in gonads of *B. atticus* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0042277	peptide binding	15	2	0.11	36	0.0048	10.000
GO:0004463	leukotriene-A4 hydrolase activity	1	1	0.01	1	0.0072	0.0072
GO:0004972	NMDA glutamate receptor activity	1	1	0.01	2	0.0072	0.0072
GO:0005546	phosphatidylinositol-4 5-bisphosphate bi…	1	1	0.01	3	0.0072	0.0072
GO:0071207	histone pre-mRNA stem-loop binding	1	1	0.01	4	0.0072	0.0072
GO:0016430	tRNA (adenine-N6-)-methyltransferase act...	1	1	0.01	5	0.0072	0.0072
GO:0004828	serine-tRNA ligase activity	1	1	0.01	6	0.0072	0.0072
GO:0008176	tRNA (guanine-N7-)-methyltransferase act...	1	1	0.01	7	0.0072	0.0072
GO:0050614	delta24-sterol reductase activity	1	1	0.01	8	0.0072	0.0072
GO:0004747	ribokinase activity	1	1	0.01	9	0.0072	0.0072
GO:0004301	epoxide hydrolase activity	1	1	0.01	10	0.0072	0.0072
GO:0000246	delta24(24-1) sterol reductase activity	1	1	0.01	11	0.0072	0.0072
GO:0047485	protein N-terminus binding	2	1	0.01	14	0.0144	0.0144
GO:0008270	zinc ion binding	3	1	0.02	15	0.0216	0.0216
GO:0017137	Rab GTPase binding	4	1	0.03	16	0.0286	0.0286

Sup. Tab. 1 c - genes upregulated in male gonads of *B. grandii* and upregulated in gonads of *B. rossius* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0051103	DNA ligation involved in DNA repair	2	2	0.03	1	0.00021	0.00021
GO:0000097	sulfur amino acid biosynthetic process	4	2	0.06	29	0.00124	0.02837
GO:0051606	detection of stimulus	10	2	0.15	51	0.00880	100.000
GO:0031047	gene silencing by RNA	12	2	0.18	39	0.01268	0.06964
GO:0070570	regulation of neuron projection regenera...	1	1	0.01	52	0.01475	100.000
GO:0070571	negative regulation of neuron projection...	1	1	0.01	53	0.01475	100.000
GO:0015886	heme transport	1	1	0.01	54	0.01475	100.000
GO:0046439	L-cysteine metabolic process	1	1	0.01	55	0.01475	100.000
GO:0071267	L-methionine salvage	1	1	0.01	56	0.01475	100.000
GO:0080182	histone H3-K4 trimethylation	1	1	0.01	2	0.01475	0.01475
GO:1901678	iron coordination entity transport	1	1	0.01	57	0.01475	100.000
GO:1903564	regulation of protein localization to ci...	1	1	0.01	58	0.01475	100.000
GO:1903565	negative regulation of protein localizat...	1	1	0.01	59	0.01475	100.000
GO:1903567	regulation of protein localization to ci...	1	1	0.01	60	0.01475	100.000
GO:1903568	negative regulation of protein localizat...	1	1	0.01	3	0.01475	0.01475
GO:0060251	regulation of glial cell proliferation	1	1	0.01	61	0.01475	100.000
GO:0060253	negative regulation of glial cell prolif...	1	1	0.01	62	0.01475	100.000
GO:0007620	copulation	1	1	0.01	4	0.01475	0.01475
GO:0014010	Schwann cell proliferation	1	1	0.01	63	0.01475	100.000

GO:0014011	Schwann cell proliferation involved in a...	1	1	0.01	64	0.01475	100.000
GO:0033089	positive regulation of T cell differenti...	1	1	0.01	5	0.01475	0.01475
GO:0014012	peripheral nervous system axon regenerat...	1	1	0.01	65	0.01475	100.000
GO:0042420	dopamine catabolic process	1	1	0.01	6	0.01475	0.01475
GO:0014014	negative regulation of gliogenesis	1	1	0.01	66	0.01475	100.000
GO:0042424	catecholamine catabolic process	1	1	0.01	67	0.01475	100.000
GO:0042412	taurine biosynthetic process	1	1	0.01	7	0.01475	0.01475
GO:0043102	amino acid salvage	1	1	0.01	68	0.01475	100.000
GO:0019336	phenol-containing compound catabolic pro...	1	1	0.01	69	0.01475	100.000
GO:0097037	heme export	1	1	0.01	8	0.01475	0.01475
GO:0051571	positive regulation of histone H3-K4 met...	1	1	0.01	9	0.01475	0.01475
GO:0016048	detection of temperature stimulus	1	1	0.01	70	0.01475	100.000
GO:0060245	detection of cell density	1	1	0.01	71	0.01475	100.000
GO:0035281	pre-miRNA export from nucleus	1	1	0.01	10	0.01475	0.01475
GO:0019530	taurine metabolic process	1	1	0.01	72	0.01475	100.000
GO:0014009	glial cell proliferation	1	1	0.01	73	0.01475	100.000
GO:0031102	neuron projection regeneration	1	1	0.01	74	0.01475	100.000
GO:0031103	axon regeneration	1	1	0.01	75	0.01475	100.000
GO:1904376	negative regulation of protein localizat...	1	1	0.01	76	0.01475	100.000
GO:0009595	detection of biotic stimulus	1	1	0.01	77	0.01475	100.000
GO:0019509	L-methionine salvage from methylthioaden...	1	1	0.01	11	0.01475	0.01475
GO:0000967	rRNA 5'-end processing	1	1	0.01	78	0.01475	100.000
GO:0050961	detection of temperature stimulus involv...	1	1	0.01	79	0.01475	100.000
GO:0009372	quorum sensing	1	1	0.01	12	0.01475	0.01475
GO:0046513	ceramide biosynthetic process	1	1	0.01	13	0.01475	0.01475
GO:0050965	detection of temperature stimulus involv...	1	1	0.01	14	0.01475	0.01475
GO:0042795	snRNA transcription by RNA polymerase II	1	1	0.01	15	0.01475	0.01475
GO:0019450	L-cysteine catabolic process to pyruvate	1	1	0.01	80	0.01475	100.000
GO:0019694	alkanesulfonate metabolic process	1	1	0.01	81	0.01475	100.000
GO:0019451	L-cysteine catabolic process to pyruvate...	1	1	0.01	16	0.01475	0.01475
GO:0050951	sensory perception of temperature stimul...	1	1	0.01	82	0.01475	100.000
GO:0046305	alkanesulfonate biosynthetic process	1	1	0.01	83	0.01475	100.000
GO:0048681	negative regulation of axon regeneration	1	1	0.01	84	0.01475	100.000
GO:1900368	regulation of RNA interference	1	1	0.01	85	0.01475	100.000
GO:0043249	erythrocyte maturation	1	1	0.01	17	0.01475	0.01475
GO:0042769	DNA damage response detection of DNA da...	1	1	0.01	18	0.01475	0.01475
GO:0048679	regulation of axon regeneration	1	1	0.01	86	0.01475	100.000
GO:1903035	negative regulation of response to wound...	1	1	0.01	87	0.01475	100.000
GO:0010624	regulation of Schwann cell proliferation	1	1	0.01	88	0.01475	100.000
GO:0042755	eating behavior	1	1	0.01	89	0.01475	100.000
GO:0010626	negative regulation of Schwann cell prol...	1	1	0.01	90	0.01475	100.000
GO:1905044	regulation of Schwann cell proliferation...	1	1	0.01	91	0.01475	100.000
GO:1905045	negative regulation of Schwann cell prol...	1	1	0.01	19	0.01475	0.01475
GO:0006288	base-excision repair DNA ligation	1	1	0.01	92	0.01475	100.000
GO:1900370	positive regulation of RNA interference	1	1	0.01	20	0.01475	0.01475
GO:0001659	temperature homeostasis	1	1	0.01	21	0.01475	0.01475
GO:0019448	L-cysteine catabolic process	1	1	0.01	93	0.01475	100.000
GO:0006659	phosphatidylserine biosynthetic process	1	1	0.01	22	0.01475	0.01475
GO:0006685	sphingomyelin catabolic process	1	1	0.01	23	0.01475	0.01475
GO:0009093	cysteine catabolic process	1	1	0.01	94	0.01475	100.000
GO:0019614	catechol-containing compound catabolic p...	1	1	0.01	95	0.01475	100.000
GO:1904894	positive regulation of receptor signalin...	1	1	0.01	96	0.01475	100.000
GO:0062149	detection of stimulus involved in sensor...	1	1	0.01	97	0.01475	100.000
GO:0000480	endonucleolytic cleavage in 5'-ETS of tr...	1	1	0.01	24	0.01475	0.01475

GO:0046427	positive regulation of receptor signalin...	1	1	0.01	25	0.01475	0.01475
GO:0000256	allantoin catabolic process	1	1	0.01	26	0.01475	0.01475
GO:0002021	response to dietary excess	1	1	0.01	98	0.01475	100.000
GO:0000472	endonucleolytic cleavage to generate mat...	1	1	0.01	27	0.01475	0.01475
GO:0002023	reduction of food intake in response to ...	1	1	0.01	28	0.01475	0.01475
GO:0042737	drug catabolic process	15	2	0.22	99	0.01962	100.000
GO:0051704	multi-organism process	45	3	0.66	100	0.02713	100.000
GO:0016458	gene silencing	18	2	0.27	101	0.02781	100.000
GO:0007260	tyrosine phosphorylation of STAT protein	2	1	0.03	102	0.02929	100.000
GO:0009086	methionine biosynthetic process	2	1	0.03	103	0.02929	100.000
GO:0009071	serine family amino acid catabolic proce...	2	1	0.03	104	0.02929	100.000
GO:0071466	cellular response to xenobiotic stimulus	2	1	0.03	105	0.02929	100.000
GO:0006805	xenobiotic metabolic process	2	1	0.03	106	0.02929	100.000
GO:0042509	regulation of tyrosine phosphorylation o...	2	1	0.03	30	0.02929	0.02929
GO:0050870	positive regulation of T cell activation	2	1	0.03	107	0.02929	100.000
GO:0071265	L-methionine biosynthetic process	2	1	0.03	108	0.02929	100.000
GO:0030218	erythrocyte differentiation	2	1	0.03	109	0.02929	100.000
GO:1902107	positive regulation of leukocyte differe...	2	1	0.03	110	0.02929	100.000
GO:0040020	regulation of meiotic nuclear division	2	1	0.03	31	0.02929	0.02929
GO:0045582	positive regulation of T cell differenti...	2	1	0.03	111	0.02929	100.000
GO:0051569	regulation of histone H3-K4 methylation	2	1	0.03	112	0.02929	100.000
GO:0014013	regulation of gliogenesis	2	1	0.03	113	0.02929	100.000
GO:0046606	negative regulation of centrosome cycle	2	1	0.03	114	0.02929	100.000
GO:0042417	dopamine metabolic process	2	1	0.03	115	0.02929	100.000
GO:0007631	feeding behavior	2	1	0.03	116	0.02929	100.000
GO:0016444	somatic cell DNA recombination	2	1	0.03	32	0.02929	0.02929
GO:0034101	erythrocyte homeostasis	2	1	0.03	117	0.02929	100.000
GO:1903708	positive regulation of hemopoiesis	2	1	0.03	118	0.02929	100.000
GO:1904375	regulation of protein localization to ce...	2	1	0.03	119	0.02929	100.000
GO:0071356	cellular response to tumor necrosis fact...	2	1	0.03	33	0.02929	0.02929
GO:0061515	myeloid cell development	2	1	0.03	120	0.02929	100.000
GO:0034471	ncRNA 5'-end processing	2	1	0.03	121	0.02929	100.000
GO:0048821	erythrocyte development	2	1	0.03	122	0.02929	100.000
GO:0010826	negative regulation of centrosome duplic...	2	1	0.03	34	0.02929	0.02929
GO:0051251	positive regulation of lymphocyte activa...	2	1	0.03	123	0.02929	100.000
GO:0048678	response to axon injury	2	1	0.03	124	0.02929	100.000
GO:1905476	negative regulation of protein localizat...	2	1	0.03	125	0.02929	100.000
GO:1903039	positive regulation of leukocyte cell-ce...	2	1	0.03	126	0.02929	100.000
GO:0045621	positive regulation of lymphocyte differ...	2	1	0.03	127	0.02929	100.000
GO:0006684	sphingomyelin metabolic process	2	1	0.03	128	0.02929	100.000
GO:0051445	regulation of meiotic cell cycle	2	1	0.03	129	0.02929	100.000
GO:0060148	positive regulation of posttranscription...	2	1	0.03	130	0.02929	100.000
GO:0006826	iron ion transport	2	1	0.03	131	0.02929	100.000
GO:0019233	sensory perception of pain	2	1	0.03	132	0.02929	100.000
GO:0043605	cellular amide catabolic process	2	1	0.03	133	0.02929	100.000
GO:0000255	allantoin metabolic process	2	1	0.03	134	0.02929	100.000
GO:0002262	myeloid cell homeostasis	2	1	0.03	135	0.02929	100.000
GO:0010468	regulation of gene expression	123	5	1.81	136	0.03151	100.000
GO:1903827	regulation of cellular protein localizat...	20	2	0.29	137	0.03391	100.000
GO:0030097	hemopoiesis	21	2	0.31	138	0.03714	100.000
GO:0048534	hematopoietic or lymphoid organ developm...	22	2	0.32	139	0.04049	100.000
GO:0046825	regulation of protein export from nucleu...	3	1	0.04	35	0.04361	0.04361
GO:2000241	regulation of reproductive process	3	1	0.04	140	0.04361	100.000
GO:0033081	regulation of T cell differentiation in ...	3	1	0.04	141	0.04361	100.000

GO:0051568	histone H3-K4 methylation	3	1	0.04	142	0.04361	100.000
GO:0006555	methionine metabolic process	3	1	0.04	143	0.04361	100.000
GO:0044764	multi-organism cellular process	3	1	0.04	144	0.04361	100.000
GO:0006534	cysteine metabolic process	3	1	0.04	145	0.04361	100.000
GO:0007159	leukocyte cell-cell adhesion	3	1	0.04	146	0.04361	100.000
GO:0000966	RNA 5'-end processing	3	1	0.04	147	0.04361	100.000
GO:0048872	homeostasis of number of cells	3	1	0.04	148	0.04361	100.000
GO:1903034	regulation of response to wounding	3	1	0.04	149	0.04361	100.000
GO:1905475	regulation of protein localization to me...	3	1	0.04	150	0.04361	100.000
GO:1903037	regulation of leukocyte cell-cell adhesi...	3	1	0.04	151	0.04361	100.000
GO:0000098	sulfur amino acid catabolic process	3	1	0.04	152	0.04361	100.000
GO:1903441	protein localization to ciliary membrane	3	1	0.04	153	0.04361	100.000
GO:0006658	phosphatidylserine metabolic process	3	1	0.04	154	0.04361	100.000
GO:0009301	snRNA transcription	3	1	0.04	155	0.04361	100.000
GO:0034612	response to tumor necrosis factor	3	1	0.04	156	0.04361	100.000
GO:0002696	positive regulation of leukocyte activat...	3	1	0.04	157	0.04361	100.000
GO:0050867	positive regulation of cell activation	3	1	0.04	158	0.04361	100.000
GO:0007610	behavior	23	2	0.34	48	0.04394	0.27240

Sup. Tab. 1 d - genes upregulated in male gonads of *B. grandii* and upregulated in gonads of *B. rossius* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0140142	nucleocytoplasmic carrier activity	1	1	0.02	32	0.015	1.000
GO:0070883	pre-miRNA binding	1	1	0.02	1	0.015	0.015
GO:0000049	tRNA binding	1	1	0.02	2	0.015	0.015
GO:0004333	fumarate hydratase activity	1	1	0.02	3	0.015	0.015
GO:0017172	cysteine dioxygenase activity	1	1	0.02	4	0.015	0.015
GO:0019841	retinol binding	1	1	0.02	5	0.015	0.015
GO:0004037	allantoicase activity	1	1	0.02	6	0.015	0.015
GO:0003882	CDP-diacylglycerol-serine O-phosphatidyl...	1	1	0.02	7	0.015	0.015
GO:0016907	G protein-coupled acetylcholine receptor...	1	1	0.02	8	0.015	0.015
GO:0046570	methylthioribulose 1-phosphate dehydrata...	1	1	0.02	9	0.015	0.015
GO:0050290	sphingomyelin phosphodiesterase D activi...	1	1	0.02	10	0.015	0.015
GO:0050291	sphingosine N-acyltransferase activity	1	1	0.02	11	0.015	0.015
GO:0010484	H3 histone acetyltransferase activity	1	1	0.02	12	0.015	0.015
GO:0005049	nuclear export signal receptor activity	1	1	0.02	13	0.015	0.015
GO:0035614	snRNA stem-loop binding	1	1	0.02	14	0.015	0.015
GO:0016429	tRNA (adenine-N1-)-methyltransferase act...	1	1	0.02	15	0.015	0.015
GO:0015226	carnitine transmembrane transporter acti...	1	1	0.02	16	0.015	0.015
GO:1990226	histone methyltransferase binding	1	1	0.02	17	0.015	0.015
GO:0016836	hydro-lyase activity	17	2	0.26	33	0.026	1.000
GO:0016410	N-acyltransferase activity	18	2	0.27	34	0.029	1.000
GO:0008227	G protein-coupled amine receptor activit...	2	1	0.03	35	0.030	1.000
GO:0005501	retinoid binding	2	1	0.03	36	0.030	1.000
GO:0015232	heme transporter activity	2	1	0.03	18	0.030	0.030
GO:0044325	ion channel binding	2	1	0.03	19	0.030	0.030
GO:0019840	isoprenoid binding	2	1	0.03	37	0.030	1.000
GO:0008381	mechanosensitive ion channel activity	2	1	0.03	20	0.030	0.030
GO:0016426	tRNA (adenine) methyltransferase activit...	2	1	0.03	38	0.030	1.000
GO:0035613	RNA stem-loop binding	2	1	0.03	39	0.030	1.000

GO:0004767	sphingomyelin phosphodiesterase activity	2	1	0.03	21	0.030	0.030
GO:0015651	quaternary ammonium group transmembrane ...	2	1	0.03	40	0.030	1.000
GO:0015464	acetylcholine receptor activity	2	1	0.03	41	0.030	1.000
GO:0016835	carbon-oxygen lyase activity	19	2	0.29	42	0.032	1.000
GO:0016747	transferase activity transferring acyl …	49	3	0.74	43	0.035	1.000
GO:0016407	acetyltransferase activity	21	2	0.32	28	0.038	0.198
GO:0005504	fatty acid binding	3	1	0.05	22	0.045	0.045
GO:0098960	postsynaptic neurotransmitter receptor a...	3	1	0.05	44	0.045	1.000
GO:0070008	serine-type exopeptidase activity	3	1	0.05	23	0.045	0.045
GO:0016813	hydrolase activity acting on carbon-nit…	3	1	0.05	45	0.045	1.000
GO:0099528	G protein-coupled neurotransmitter recep...	3	1	0.05	46	0.045	1.000
GO:0005215	transporter activity	134	5	2.02	47	0.045	1.000

Sup. Tab. 1 e - genes upregulated in male gonads of *B. grandii* and upregulated in gonads of both parthenogens (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0098535	de novo centriole assembly involved in m...	1	1	0	1	0.0043	0.0043
GO:0035999	tetrahydrofolate interconversion	1	1	0	2	0.0043	0.0043
GO:0045329	carnitine biosynthetic process	1	1	0	3	0.0043	0.0043
GO:0006348	chromatin silencing at telomere	1	1	0	4	0.0043	0.0043
GO:0016539	intein-mediated protein splicing	1	1	0	5	0.0043	0.0043
GO:0006336	DNA replication-independent nucleosome a...	1	1	0	6	0.0043	0.0043
GO:0031936	negative regulation of chromatin silenci...	1	1	0	7	0.0043	0.0043
GO:0042426	choline catabolic process	1	1	0	8	0.0043	0.0043
GO:0046599	regulation of centriole replication	2	1	0.01	9	0.0085	0.0085
GO:0006545	glycine biosynthetic process	2	1	0.01	10	0.0085	0.0085
GO:1901096	regulation of autophagosome maturation	2	1	0.01	11	0.0085	0.0085
GO:0010038	response to metal ion	9	1	0.04	12	0.0379	0.0379

Sup. Tab. 1 f - genes upregulated in male gonads of *B. grandii* and upregulated in gonads of both parthenogens (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0008480	sarcosine dehydrogenase activity	1	1	0.01	1	0.006	0.006
GO:0004561	alpha-N-acetylglucosaminidase activity	1	1	0.01	2	0.006	0.006
GO:0035174	histone serine kinase activity	1	1	0.01	3	0.006	0.006
GO:0047865	dimethylglycine dehydrogenase activity	1	1	0.01	4	0.006	0.006
GO:0003721	telomerase RNA reverse transcriptase act...	1	1	0.01	5	0.006	0.006
GO:0004809	tRNA (guanine-N2-)-methyltransferase act...	1	1	0.01	6	0.006	0.006
GO:0050353	trimethyllysine dioxygenase activity	1	1	0.01	7	0.006	0.006
GO:0004066	asparagine synthase (glutamine-hydrolyzi...	2	1	0.01	8	0.012	0.012
GO:0005096	GTPase activator activity	2	1	0.01	9	0.012	0.012
GO:0005484	SNAP receptor activity	2	1	0.01	10	0.012	0.012
GO:0010485	H4 histone acetyltransferase activity	2	1	0.01	11	0.012	0.012
GO:0061733	peptide-lysine-N-acetyltransferase activ...	5	1	0.03	12	0.030	1.000
GO:0004402	histone acetyltransferase activity	5	1	0.03	13	0.030	1.000
GO:0016884	carbon-nitrogen ligase activity with gl…	6	1	0.04	14	0.036	1.000
GO:0030695	GTPase regulator activity	6	1	0.04	15	0.036	1.000
GO:0034212	peptide N-acetyltransferase activity	7	1	0.04	16	0.042	1.000

Sup. Tab. 2 a - genes not upregulated in gonads of *B. grandii* and upregulated in gonads of *B. atticus* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0007188	adenylate cyclase-modulating G protein-c...	3	2	0.03	21	0.00037	0.022
GO:0098656	anion transmembrane transport	20	3	0.23	32	0.00131	0.173
GO:0015698	inorganic anion transport	6	2	0.07	27	0.00182	0.054
GO:0015748	organophosphate ester transport	11	2	0.13	35	0.00646	1.000
GO:0015711	organic anion transport	37	3	0.42	36	0.00788	1.000
GO:1901264	carbohydrate derivative transport	13	2	0.15	37	0.00903	1.000
GO:0010496	intercellular transport	1	1	0.01	1	0.01142	0.011
GO:0002566	somatic diversification of immune recept...	1	1	0.01	38	0.01142	1.000
GO:0032410	negative regulation of transporter activ...	1	1	0.01	2	0.01142	0.011
GO:1902476	chloride transmembrane transport	1	1	0.01	3	0.01142	0.011
GO:1901028	regulation of mitochondrial outer membra...	1	1	0.01	39	0.01142	1.000
GO:1901029	negative regulation of mitochondrial out...	1	1	0.01	4	0.01142	0.011
GO:0046398	UDP-glucuronate metabolic process	1	1	0.01	40	0.01142	1.000
GO:0019520	aldonic acid metabolic process	1	1	0.01	41	0.01142	1.000
GO:0019521	D-gluconate metabolic process	1	1	0.01	5	0.01142	0.011
GO:0015701	bicarbonate transport	1	1	0.01	6	0.01142	0.011
GO:0046963	3'-phosphoadenosine 5'-phosphosulfate tr...	1	1	0.01	42	0.01142	1.000
GO:0030206	chondroitin sulfate biosynthetic process	1	1	0.01	7	0.01142	0.011
GO:0008293	torso signaling pathway	1	1	0.01	8	0.01142	0.011
GO:1902559	3'-phospho-5'-adenylyl sulfate transmemb...	1	1	0.01	9	0.01142	0.011
GO:0006065	UDP-glucuronate biosynthetic process	1	1	0.01	10	0.01142	0.011
GO:0071880	adenylate cyclase-activating adrenergic ...	1	1	0.01	11	0.01142	0.011
GO:0007362	terminal region determination	1	1	0.01	12	0.01142	0.011
GO:0071875	adrenergic receptor signaling pathway	1	1	0.01	43	0.01142	1.000
GO:0007474	imaginal disc-derived wing vein specific...	1	1	0.01	13	0.01142	0.011
GO:0040018	positive regulation of multicellular org...	1	1	0.01	14	0.01142	0.011
GO:0016446	somatic hypermutation of immunoglobulin ...	1	1	0.01	15	0.01142	0.011
GO:0043615	astrocyte cell migration	1	1	0.01	16	0.01142	0.011
GO:0006751	glutathione catabolic process	1	1	0.01	17	0.01142	0.011
GO:0036085	GDP-fucose import into Golgi lumen	1	1	0.01	18	0.01142	0.011
GO:0006821	chloride transport	1	1	0.01	44	0.01142	1.000
GO:0030259	lipid glycosylation	1	1	0.01	19	0.01142	0.011
GO:0051051	negative regulation of transport	15	2	0.17	45	0.01198	1.000
GO:0016051	carbohydrate biosynthetic process	15	2	0.17	20	0.01198	0.012
GO:0005975	carbohydrate metabolic process	48	3	0.55	46	0.01615	1.000
GO:0003002	regionalization	20	2	0.23	47	0.02095	1.000
GO:0050650	chondroitin sulfate proteoglycan biosynt...	2	1	0.02	48	0.02271	1.000
GO:0050654	chondroitin sulfate proteoglycan metabol...	2	1	0.02	49	0.02271	1.000
GO:1905710	positive regulation of membrane permeabi...	2	1	0.02	50	0.02271	1.000
GO:1902110	positive regulation of mitochondrial mem...	2	1	0.02	51	0.02271	1.000
GO:0072530	purine-containing compound transmembrane...	2	1	0.02	52	0.02271	1.000

GO:1901679	nucleotide transmembrane transport	2	1	0.02	53	0.02271	1.000
GO:0030178	negative regulation of Wnt signaling pat...	2	1	0.02	22	0.02271	0.023
GO:0008347	glial cell migration	2	1	0.02	54	0.02271	1.000
GO:1902108	regulation of mitochondrial membrane per...	2	1	0.02	55	0.02271	1.000
GO:0030204	chondroitin sulfate metabolic process	2	1	0.02	56	0.02271	1.000
GO:1902686	mitochondrial outer membrane permeabiliz...	2	1	0.02	57	0.02271	1.000
GO:0007189	adenylate cyclase-activating G protein-c...	2	1	0.02	58	0.02271	1.000
GO:0035794	positive regulation of mitochondrial mem...	2	1	0.02	59	0.02271	1.000
GO:0040014	regulation of multicellular organism gro...	2	1	0.02	60	0.02271	1.000
GO:0016445	somatic diversification of immunoglobuli...	2	1	0.02	61	0.02271	1.000
GO:0097345	mitochondrial outer membrane permeabiliz...	2	1	0.02	62	0.02271	1.000
GO:0002377	immunoglobulin production	2	1	0.02	63	0.02271	1.000
GO:0090480	purine nucleotide-sugar transmembrane tr...	2	1	0.02	64	0.02271	1.000
GO:0015783	GDP-fucose transmembrane transport	2	1	0.02	65	0.02271	1.000
GO:0002440	production of molecular mediator of immu...	2	1	0.02	66	0.02271	1.000
GO:0007389	pattern specification process	24	2	0.27	67	0.02959	1.000
GO:0010823	negative regulation of mitochondrion org...	3	1	0.03	68	0.03388	1.000
GO:0019933	cAMP-mediated signaling	3	1	0.03	69	0.03388	1.000
GO:0006749	glutathione metabolic process	3	1	0.03	70	0.03388	1.000
GO:0098661	inorganic anion transmembrane transport	3	1	0.03	71	0.03388	1.000
GO:0043171	peptide catabolic process	3	1	0.03	72	0.03388	1.000
GO:0034763	negative regulation of transmembrane tra...	3	1	0.03	23	0.03388	0.034
GO:0007354	zygotic determination of anterior/poster...	3	1	0.03	73	0.03388	1.000
GO:0006024	glycosaminoglycan biosynthetic process	3	1	0.03	74	0.03388	1.000
GO:0035264	multicellular organism growth	3	1	0.03	75	0.03388	1.000
GO:0048639	positive regulation of developmental gro...	3	1	0.03	76	0.03388	1.000
GO:0017121	plasma membrane phospholipid scrambling	3	1	0.03	24	0.03388	0.034
GO:0002200	somatic diversification of immune recept...	3	1	0.03	77	0.03388	1.000
GO:0015931	nucleobase-containing compound transport	26	2	0.3	78	0.03437	1.000
GO:0048856	anatomical structure development	297	7	3.39	33	0.04238	0.502
GO:0051187	cofactor catabolic process	4	1	0.05	79	0.04493	1.000
GO:0051503	adenine nucleotide transport	4	1	0.05	80	0.04493	1.000
GO:0019935	cyclic-nucleotide-mediated signaling	4	1	0.05	81	0.04493	1.000
GO:0045927	positive regulation of growth	4	1	0.05	82	0.04493	1.000
GO:0090559	regulation of membrane permeability	4	1	0.05	83	0.04493	1.000
GO:0045332	phospholipid translocation	4	1	0.05	84	0.04493	1.000
GO:0044273	sulfur compound catabolic process	4	1	0.05	85	0.04493	1.000
GO:0015865	purine nucleotide transport	4	1	0.05	86	0.04493	1.000
GO:0015868	purine ribonucleotide transport	4	1	0.05	87	0.04493	1.000
GO:0006023	aminoglycan biosynthetic process	4	1	0.05	88	0.04493	1.000
GO:0050806	positive regulation of synaptic transmis...	4	1	0.05	25	0.04493	0.045
GO:0042219	cellular modified amino acid catabolic p...	4	1	0.05	89	0.04493	1.000
GO:0048813	dendrite morphogenesis	4	1	0.05	26	0.04493	0.045

GO:0046902	regulation of mitochondrial membrane per...	4	1	0.05	90	0.04493	1.000

Sup. Tab. 2 b - genes not upregulated in gonads of *B. grandii* and upregulated in gonads of *B. atticus* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0003979	UDP-glucose 6-dehydrogenase activity	1	1	0.01	1	0.009	0.009
GO:0008510	sodium:bicarbonate symporter activity	1	1	0.01	2	0.009	0.009
GO:0047837	D-xylose 1-dehydrogenase (NADP+) activit...	1	1	0.01	3	0.009	0.009
GO:0004994	somatostatin receptor activity	1	1	0.01	4	0.009	0.009
GO:0005229	intracellular calcium activated chloride...	1	1	0.01	5	0.009	0.009
GO:0030165	PDZ domain binding	1	1	0.01	6	0.009	0.009
GO:0042923	neuropeptide binding	1	1	0.01	7	0.009	0.009
GO:0004581	dolichyl-phosphate beta-glucosyltransfer...	1	1	0.01	8	0.009	0.009
GO:0004582	dolichyl-phosphate beta-D-mannosyltransf...	1	1	0.01	9	0.009	0.009
GO:0003839	gamma-glutamylcyclotransferase activity	1	1	0.01	10	0.009	0.009
GO:0005452	inorganic anion exchanger activity	1	1	0.01	11	0.009	0.009
GO:0004615	phosphomannomutase activity	1	1	0.01	12	0.009	0.009
GO:0001537	N-acetylgalactosamine 4-O-sulfotransfera...	1	1	0.01	13	0.009	0.009
GO:0080132	fatty acid alpha-hydroxylase activity	1	1	0.01	14	0.009	0.009
GO:0017128	phospholipid scramblase activity	1	1	0.01	15	0.009	0.009
GO:0046964	3'-phosphoadenosine 5'-phosphosulfate tr...	1	1	0.01	16	0.009	0.009
GO:0004743	pyruvate kinase activity	1	1	0.01	17	0.009	0.009
GO:0009881	photoreceptor activity	3	1	0.03	18	0.027	0.027
GO:0005227	calcium activated cation channel activit...	3	1	0.03	19	0.027	0.027
GO:0140323	solute:anion antiporter activity	5	1	0.05	22	0.044	1.000
GO:0015301	anion:anion antiporter activity	5	1	0.05	20	0.044	0.044

Sup. Tab. 2 c - genes not upregulated in gonads of *B. grandii* and upregulated in gonads of *B. rossius* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:1901566	organonitrogen compound biosynthetic pro...	193	18	6.34	96	0.0010	10.000
GO:0046415	urate metabolic process	2	2	0.07	2	0.0011	0.0011
GO:0006560	proline metabolic process	3	2	0.1	97	0.0031	10.000
GO:0048814	regulation of dendrite morphogenesis	3	2	0.1	3	0.0031	0.0031
GO:0006906	vesicle fusion	10	3	0.33	98	0.0034	10.000
GO:0043604	amide biosynthetic process	74	9	2.43	99	0.0055	10.000
GO:0071482	cellular response to light stimulus	4	2	0.13	100	0.0061	10.000
GO:0046394	carboxylic acid biosynthetic process	39	5	1.28	101	0.0080	10.000
GO:0043101	purine-containing compound salvage	5	2	0.16	102	0.0100	10.000
GO:0043558	regulation of translational initiation i...	5	2	0.16	103	0.0100	10.000
GO:0043648	dicarboxylic acid metabolic process	16	3	0.53	104	0.0140	10.000
GO:0000209	protein polyubiquitination	6	2	0.2	81	0.0146	0.1518
GO:0034754	cellular hormone metabolic process	6	2	0.2	4	0.0146	0.0146
GO:0042398	cellular modified amino acid biosyntheti...	7	2	0.23	5	0.0200	0.0200
GO:0006418	tRNA aminoacylation for protein translat...	20	3	0.66	7	0.0259	0.0259
GO:0009065	glutamine family amino acid catabolic pr...	8	2	0.26	105	0.0262	10.000

GO:0043038	amino acid activation	21	3	0.69	106	0.0295	10.000
GO:0043039	tRNA aminoacylation	21	3	0.69	107	0.0295	10.000
GO:0090174	organelle membrane fusion	11	4	0.36	1	0.0315	0.0003
GO:0019428	allantoin biosynthetic process	1	1	0.03	8	0.0328	0.0328
GO:0035210	prepupal development	1	1	0.03	108	0.0328	10.000
GO:0046874	quinolinate metabolic process	1	1	0.03	109	0.0328	10.000
GO:0035526	retrograde transport plasma membrane to…	1	1	0.03	9	0.0328	0.0328
GO:0006642	triglyceride mobilization	1	1	0.03	10	0.0328	0.0328
GO:0032469	endoplasmic reticulum calcium ion homeos...	1	1	0.03	11	0.0328	0.0328
GO:0018377	protein myristoylation	1	1	0.03	110	0.0328	10.000
GO:0032793	positive regulation of CREB transcriptio...	1	1	0.03	12	0.0328	0.0328
GO:0046086	adenosine biosynthetic process	1	1	0.03	111	0.0328	10.000
GO:0010121	arginine catabolic process to proline vi...	1	1	0.03	13	0.0328	0.0328
GO:0051664	nuclear pore localization	1	1	0.03	112	0.0328	10.000
GO:0060415	muscle tissue morphogenesis	1	1	0.03	14	0.0328	0.0328
GO:1902730	positive regulation of proteoglycan bios...	1	1	0.03	113	0.0328	10.000
GO:0006169	adenosine salvage	1	1	0.03	15	0.0328	0.0328
GO:0019627	urea metabolic process	1	1	0.03	114	0.0328	10.000
GO:0019628	urate catabolic process	1	1	0.03	115	0.0328	10.000
GO:0010908	regulation of heparan sulfate proteoglyc...	1	1	0.03	116	0.0328	10.000
GO:0010909	positive regulation of heparan sulfate p...	1	1	0.03	16	0.0328	0.0328
GO:0018008	N-terminal peptidyl-glycine N-myristoyla...	1	1	0.03	17	0.0328	0.0328
GO:0044209	AMP salvage	1	1	0.03	18	0.0328	0.0328
GO:0010133	proline catabolic process to glutamate	1	1	0.03	19	0.0328	0.0328
GO:0000050	urea cycle	1	1	0.03	20	0.0328	0.0328
GO:0060997	dendritic spine morphogenesis	1	1	0.03	21	0.0328	0.0328
GO:1904591	positive regulation of protein import	1	1	0.03	117	0.0328	10.000
GO:1902914	regulation of protein polyubiquitination	1	1	0.03	118	0.0328	10.000
GO:1902915	negative regulation of protein polyubiqu...	1	1	0.03	119	0.0328	10.000
GO:0019896	axonal transport of mitochondrion	1	1	0.03	22	0.0328	0.0328
GO:0006148	inosine catabolic process	1	1	0.03	23	0.0328	0.0328
GO:0000290	deadenylation-dependent decapping of nuc...	1	1	0.03	24	0.0328	0.0328
GO:0033183	negative regulation of histone ubiquitin...	1	1	0.03	120	0.0328	10.000
GO:1903830	magnesium ion transmembrane transport	1	1	0.03	25	0.0328	0.0328
GO:0000019	regulation of mitotic recombination	1	1	0.03	26	0.0328	0.0328
GO:1900044	regulation of protein K63-linked ubiquit...	1	1	0.03	121	0.0328	10.000
GO:1900045	negative regulation of protein K63-linke...	1	1	0.03	122	0.0328	10.000
GO:0036151	phosphatidylcholine acyl-chain remodelin...	1	1	0.03	27	0.0328	0.0328
GO:0007344	pronuclear fusion	1	1	0.03	28	0.0328	0.0328
GO:0018201	peptidyl-glycine modification	1	1	0.03	123	0.0328	10.000
GO:0016056	rhodopsin mediated signaling pathway	1	1	0.03	124	0.0328	10.000
GO:0016059	deactivation of rhodopsin mediated signa...	1	1	0.03	29	0.0328	0.0328
GO:0051865	protein autoubiquitination	1	1	0.03	30	0.0328	0.0328

GO:0009052	pentose-phosphate shunt non-oxidative b...	1	1	0.03	31	0.0328	0.0328
GO:0046532	regulation of photoreceptor cell differe...	1	1	0.03	125	0.0328	10.000
GO:0035970	peptidyl-threonine dephosphorylation	1	1	0.03	32	0.0328	0.0328
GO:0042554	superoxide anion generation	1	1	0.03	33	0.0328	0.0328
GO:0046469	platelet activating factor metabolic pro...	1	1	0.03	126	0.0328	10.000
GO:0042307	positive regulation of protein import in...	1	1	0.03	34	0.0328	0.0328
GO:0042573	retinoic acid metabolic process	1	1	0.03	127	0.0328	10.000
GO:0070535	histone H2A K63-linked ubiquitination	1	1	0.03	128	0.0328	10.000
GO:0071110	histone biotinylation	1	1	0.03	35	0.0328	0.0328
GO:0019805	quinolinate biosynthetic process	1	1	0.03	36	0.0328	0.0328
GO:1902358	sulfate transmembrane transport	1	1	0.03	37	0.0328	0.0328
GO:0006562	proline catabolic process	1	1	0.03	129	0.0328	10.000
GO:0046473	phosphatidic acid metabolic process	1	1	0.03	130	0.0328	10.000
GO:0061472	karyomere membrane fusion	1	1	0.03	131	0.0328	10.000
GO:0034499	late endosome to Golgi transport	1	1	0.03	38	0.0328	0.0328
GO:0050774	negative regulation of dendrite morphoge...	1	1	0.03	132	0.0328	10.000
GO:0050775	positive regulation of dendrite morphoge...	1	1	0.03	133	0.0328	10.000
GO:0032055	negative regulation of translation in re...	1	1	0.03	134	0.0328	10.000
GO:0032056	positive regulation of translation in re...	1	1	0.03	135	0.0328	10.000
GO:0032057	negative regulation of translational ini...	1	1	0.03	39	0.0328	0.0328
GO:0032058	positive regulation of translational ini...	1	1	0.03	136	0.0328	10.000
GO:0006529	asparagine biosynthetic process	1	1	0.03	137	0.0328	10.000
GO:0009305	protein biotinylation	1	1	0.03	138	0.0328	10.000
GO:0006499	N-terminal protein myristoylation	1	1	0.03	139	0.0328	10.000
GO:0042478	regulation of eye photoreceptor cell dev...	1	1	0.03	40	0.0328	0.0328
GO:0006535	cysteine biosynthetic process from serin...	1	1	0.03	41	0.0328	0.0328
GO:0022400	regulation of rhodopsin mediated signali...	1	1	0.03	140	0.0328	10.000
GO:2001160	regulation of histone H3-K79 methylation	1	1	0.03	141	0.0328	10.000
GO:2001162	positive regulation of histone H3-K79 me...	1	1	0.03	42	0.0328	0.0328
GO:0071043	CUT metabolic process	1	1	0.03	142	0.0328	10.000
GO:0034418	urate biosynthetic process	1	1	0.03	143	0.0328	10.000
GO:0035073	pupariation	1	1	0.03	43	0.0328	0.0328
GO:0097359	UDP-glucosylation	1	1	0.03	44	0.0328	0.0328
GO:0071034	CUT catabolic process	1	1	0.03	45	0.0328	0.0328
GO:0000722	telomere maintenance via recombination	1	1	0.03	46	0.0328	0.0328
GO:0019493	arginine catabolic process to proline	1	1	0.03	144	0.0328	10.000
GO:0019544	arginine catabolic process to glutamate	1	1	0.03	47	0.0328	0.0328
GO:0000741	karyogamy	1	1	0.03	145	0.0328	10.000
GO:0048280	vesicle fusion with Golgi apparatus	1	1	0.03	146	0.0328	10.000
GO:0019532	oxalate transport	1	1	0.03	48	0.0328	0.0328
GO:1901314	regulation of histone H2A K63-linked ubi...	1	1	0.03	147	0.0328	10.000

Sup. Tab. 2 d - genes not upregulated in gonads of *B. grandii* and upregulated in gonads of *B. rossius* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0004520	endodeoxyribonuclease activity	6	2	0.2	73	0.015	1.000
GO:0016874	ligase activity	66	6	2.23	74	0.021	1.000
GO:0140030	modification-dependent protein binding	8	2	0.27	75	0.028	1.000
GO:0004812	aminoacyl-tRNA ligase activity	20	3	0.68	1	0.028	0.028
GO:0016875	ligase activity forming carbon-oxygen b…	20	3	0.68	76	0.028	1.000
GO:0004080	biotin-[propionyl-CoA-carboxylase (ATP-h...	1	1	0.03	2	0.034	0.034
GO:1901611	phosphatidylglycerol binding	1	1	0.03	77	0.034	1.000
GO:1901612	cardiolipin binding	1	1	0.03	3	0.034	0.034
GO:0005458	GDP-mannose transmembrane transporter ac...	1	1	0.03	4	0.034	0.034
GO:0050104	L-gulonate 3-dehydrogenase activity	1	1	0.03	5	0.034	0.034
GO:0004329	formate-tetrahydrofolate ligase activity	1	1	0.03	6	0.034	0.034
GO:0035497	cAMP response element binding	1	1	0.03	7	0.034	0.034
GO:0004077	biotin-[acetyl-CoA-carboxylase] ligase a...	1	1	0.03	8	0.034	0.034
GO:0003980	UDP-glucose:glycoprotein glucosyltransfe...	1	1	0.03	9	0.034	0.034
GO:0018271	biotin-protein ligase activity	1	1	0.03	78	0.034	1.000
GO:0016215	acyl-CoA desaturase activity	1	1	0.03	79	0.034	1.000
GO:0004502	kynurenine 3-monooxygenase activity	1	1	0.03	10	0.034	0.034
GO:0015095	magnesium ion transmembrane transporter ...	1	1	0.03	11	0.034	0.034
GO:0004018	N6-(1 2-dicarboxyethyl)AMP AMP-lyase (fu…	1	1	0.03	12	0.034	0.034
GO:0019531	oxalate transmembrane transporter activi...	1	1	0.03	13	0.034	0.034
GO:0019107	myristoyltransferase activity	1	1	0.03	80	0.034	1.000
GO:0008140	cAMP response element binding protein bi...	1	1	0.03	14	0.034	0.034
GO:0070403	NAD+ binding	1	1	0.03	15	0.034	0.034
GO:0034450	ubiquitin-ubiquitin ligase activity	1	1	0.03	16	0.034	0.034
GO:0004824	lysine-tRNA ligase activity	1	1	0.03	81	0.034	1.000
GO:0004825	methionine-tRNA ligase activity	1	1	0.03	82	0.034	1.000
GO:0016717	oxidoreductase activity acting on paire…	1	1	0.03	83	0.034	1.000
GO:0050811	GABA receptor binding	1	1	0.03	17	0.034	0.034
GO:0016744	transferase activity transferring aldeh…	1	1	0.03	18	0.034	0.034
GO:0000014	single-stranded DNA endodeoxyribonucleas...	1	1	0.03	19	0.034	0.034
GO:0016155	formyltetrahydrofolate dehydrogenase act...	1	1	0.03	20	0.034	0.034
GO:0004122	cystathionine beta-synthase activity	1	1	0.03	21	0.034	0.034
GO:0004587	ornithine-oxo-acid transaminase activity	1	1	0.03	22	0.034	0.034
GO:0004846	urate oxidase activity	1	1	0.03	23	0.034	0.034
GO:1990889	H4K20me3 modified histone binding	1	1	0.03	24	0.034	0.034
GO:0032184	SUMO polymer binding	1	1	0.03	25	0.034	0.034
GO:0000293	ferric-chelate reductase activity	1	1	0.03	26	0.034	0.034
GO:0070626	(S)-2-(5-amino-1-(5-phospho-D-ribosyl)im...	1	1	0.03	27	0.034	0.034
GO:0004830	tryptophan-tRNA ligase activity	1	1	0.03	84	0.034	1.000
GO:0017050	D-erythro-sphingosine kinase activity	1	1	0.03	28	0.034	0.034
GO:0070615	nucleosome-dependent ATPase activity	1	1	0.03	29	0.034	0.034
GO:0016175	superoxide-generating NADPH oxidase acti...	1	1	0.03	30	0.034	0.034

GO:0004140	dephospho-CoA kinase activity	1	1	0.03	31	0.034	0.034
GO:0070025	carbon monoxide binding	1	1	0.03	32	0.034	0.034
GO:0008481	sphinganine kinase activity	1	1	0.03	33	0.034	0.034
GO:0004657	proline dehydrogenase activity	1	1	0.03	34	0.034	0.034
GO:0038132	neuregulin binding	1	1	0.03	35	0.034	0.034
GO:0004768	stearoyl-CoA 9-desaturase activity	1	1	0.03	36	0.034	0.034
GO:0003691	double-stranded telomeric DNA binding	1	1	0.03	37	0.034	0.034
GO:0004750	ribulose-phosphate 3-epimerase activity	1	1	0.03	38	0.034	0.034
GO:0004053	arginase activity	1	1	0.03	39	0.034	0.034
GO:0016663	oxidoreductase activity acting on other...	1	1	0.03	85	0.034	1.000
GO:0031994	insulin-like growth factor I binding	1	1	0.03	40	0.034	0.034
GO:0036312	phosphatidylinositol 3-kinase regulatory...	1	1	0.03	41	0.034	0.034
GO:0004530	deoxyribonuclease I activity	1	1	0.03	42	0.034	0.034
GO:0004379	glycylpeptide N-tetradecanoyltransferase...	1	1	0.03	43	0.034	0.034

Sup. Tab. 2 e - genes not upregulated in gonads of *B. grandii* and upregulated in gonads of parthenogens (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0006465	signal peptide processing	1	1	0.01	2	0.0076	0.0076
GO:0042532	negative regulation of tyrosine phosphor...	1	1	0.01	3	0.0076	0.0076
GO:0007455	eye-antennal disc morphogenesis	1	1	0.01	4	0.0076	0.0076
GO:0006428	isoleucyl-tRNA aminoacylation	1	1	0.01	5	0.0076	0.0076
GO:0071934	thiamine transmembrane transport	1	1	0.01	6	0.0076	0.0076
GO:0008295	spermidine biosynthetic process	1	1	0.01	7	0.0076	0.0076
GO:0016024	CDP-diacylglycerol biosynthetic process	1	1	0.01	19	0.0076	10.000
GO:0006777	Mo-molybdopterin cofactor biosynthetic p...	1	1	0.01	8	0.0076	0.0076
GO:0015771	trehalose transport	1	1	0.01	9	0.0076	0.0076
GO:0015884	folic acid transport	1	1	0.01	10	0.0076	0.0076
GO:0045002	double-strand break repair via single-st...	1	1	0.01	11	0.0076	0.0076
GO:0061015	snRNA import into nucleus	1	1	0.01	12	0.0076	0.0076
GO:0008645	hexose transmembrane transport	2	1	0.02	20	0.0152	10.000
GO:1904659	glucose transmembrane transport	2	1	0.02	21	0.0152	10.000
GO:0007478	leg disc morphogenesis	2	1	0.02	13	0.0152	0.0152
GO:0035218	leg disc development	2	1	0.02	22	0.0152	10.000
GO:0046323	glucose import	2	1	0.02	14	0.0152	0.0152
GO:0015749	monosaccharide transmembrane transport	3	1	0.02	23	0.0227	10.000
GO:0007602	phototransduction	3	1	0.02	15	0.0227	0.0227
GO:0034219	carbohydrate transmembrane transport	4	1	0.03	24	0.0301	10.000
GO:0009583	detection of light stimulus	4	1	0.03	25	0.0301	10.000
GO:0003333	amino acid transmembrane transport	5	1	0.04	16	0.0375	0.0375
GO:0010632	regulation of epithelial cell migration	5	1	0.04	17	0.0375	0.0375

Sup. Tab. 2 f - genes not upregulated in gonads of *B. grandii* and upregulated in gonads of parthenogens (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0008518	folate:anion antiporter activity	1	1	0.01	2	0.0096	0.0096

GO:0005131	growth hormone receptor binding	1	1	0.01	3	0.0096	0.0096
GO:0004822	isoleucine-tRNA ligase activity	1	1	0.01	4	0.0096	0.0096
GO:0004366	glycerol-3-phosphate O-acyltransferase a...	1	1	0.01	5	0.0096	0.0096
GO:0008269	JAK pathway signal transduction adaptor ...	1	1	0.01	6	0.0096	0.0096
GO:0015234	thiamine transmembrane transporter activ...	1	1	0.01	7	0.0096	0.0096
GO:0003943	N-acetylgalactosamine-4-sulfatase activi...	1	1	0.01	8	0.0096	0.0096
GO:0004588	orotate phosphoribosyltransferase activi...	1	1	0.01	9	0.0096	0.0096
GO:0005542	folic acid binding	1	1	0.01	22	0.0096	10.000
GO:0004590	orotidine-5'-phosphate decarboxylase act...	1	1	0.01	10	0.0096	0.0096
GO:0045183	translation factor activity non-nucleic…	1	1	0.01	11	0.0096	0.0096
GO:0043422	protein kinase B binding	1	1	0.01	12	0.0096	0.0096
GO:0003680	AT DNA binding	1	1	0.01	13	0.0096	0.0096
GO:0004766	spermidine synthase activity	1	1	0.01	14	0.0096	0.0096
GO:0050220	prostaglandin-E synthase activity	1	1	0.01	15	0.0096	0.0096
GO:0005499	vitamin D binding	1	1	0.01	23	0.0096	10.000
GO:0002161	aminoacyl-tRNA editing activity	2	1	0.02	16	0.0192	0.0192
GO:0005527	macrolide binding	2	1	0.02	24	0.0192	10.000
GO:0005528	FK506 binding	2	1	0.02	17	0.0192	0.0192
GO:0009008	DNA-methyltransferase activity	2	1	0.02	18	0.0192	0.0192
GO:0034596	phosphatidylinositol phosphate 4-phospha...	4	1	0.04	19	0.0381	0.0381

Sup. Tab. 3 a - genes upregulated in gonads of *B. grandii* and not upregulated in gonads of *B. atticus* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0006289	nucleotide-excision repair	5	2	0.05	1	0.0010	0.001
GO:0006259	DNA metabolic process	89	6	0.93	32	0.0071	1.000
GO:0006323	DNA packaging	14	2	0.15	33	0.0088	1.000
GO:0000070	mitotic sister chromatid segregation	15	2	0.16	2	0.0101	0.010
GO:0071578	zinc ion import across plasma membrane	1	1	0.01	3	0.0105	0.010
GO:0017196	N-terminal peptidyl-methionine acetylati...	1	1	0.01	4	0.0105	0.010
GO:0006297	nucleotide-excision repair DNA gap fill…	1	1	0.01	34	0.0105	1.000
GO:0032288	myelin assembly	1	1	0.01	35	0.0105	1.000
GO:0032289	central nervous system myelin formation	1	1	0.01	5	0.0105	0.010
GO:1903753	negative regulation of p38MAPK cascade	1	1	0.01	6	0.0105	0.010
GO:0006409	tRNA export from nucleus	1	1	0.01	7	0.0105	0.010
GO:0061098	positive regulation of protein tyrosine ...	1	1	0.01	8	0.0105	0.010
GO:1900744	regulation of p38MAPK cascade	1	1	0.01	36	0.0105	1.000
GO:0099587	inorganic ion import across plasma membr...	1	1	0.01	37	0.0105	1.000
GO:0000183	rDNA heterochromatin assembly	1	1	0.01	9	0.0105	0.010
GO:0046087	cytidine metabolic process	1	1	0.01	38	0.0105	1.000
GO:0071431	tRNA-containing ribonucleoprotein comple...	1	1	0.01	39	0.0105	1.000
GO:0010032	meiotic chromosome condensation	1	1	0.01	10	0.0105	0.010
GO:0006216	cytidine catabolic process	1	1	0.01	40	0.0105	1.000
GO:0070221	sulfide oxidation using sulfide:quinone…	1	1	0.01	11	0.0105	0.010

GO:0038066	p38MAPK cascade	1	1	0.01	41	0.0105	1.000
GO:0070212	protein poly-ADP-ribosylation	1	1	0.01	12	0.0105	0.010
GO:0019418	sulfide oxidation	1	1	0.01	42	0.0105	1.000
GO:0009972	cytidine deamination	1	1	0.01	13	0.0105	0.010
GO:0045004	DNA replication proofreading	1	1	0.01	14	0.0105	0.010
GO:0000012	single strand break repair	1	1	0.01	15	0.0105	0.010
GO:0007076	mitotic chromosome condensation	1	1	0.01	43	0.0105	1.000
GO:0098659	inorganic cation import across plasma me...	1	1	0.01	44	0.0105	1.000
GO:0006261	DNA-dependent DNA replication	16	2	0.17	45	0.0115	1.000
GO:0000819	sister chromatid segregation	16	2	0.17	46	0.0115	1.000
GO:0051347	positive regulation of transferase activ...	17	2	0.18	26	0.0129	0.105
GO:0006260	DNA replication	19	2	0.2	47	0.0161	1.000
GO:0098813	nuclear chromosome segregation	21	2	0.22	48	0.0195	1.000
GO:0071577	zinc ion transmembrane transport	2	1	0.02	49	0.0208	1.000
GO:0022010	central nervous system myelination	2	1	0.02	50	0.0208	1.000
GO:0051031	tRNA transport	2	1	0.02	51	0.0208	1.000
GO:0050731	positive regulation of peptidyl-tyrosine...	2	1	0.02	52	0.0208	1.000
GO:0018206	peptidyl-methionine modification	2	1	0.02	53	0.0208	1.000
GO:0097064	ncRNA export from nucleus	2	1	0.02	54	0.0208	1.000
GO:0006273	lagging strand elongation	2	1	0.02	16	0.0208	0.021
GO:0070303	negative regulation of stress-activated ...	2	1	0.02	55	0.0208	1.000
GO:0032873	negative regulation of stress-activated ...	2	1	0.02	56	0.0208	1.000
GO:0006911	phagocytosis engulfment	2	1	0.02	17	0.0208	0.021
GO:0030261	chromosome condensation	2	1	0.02	57	0.0208	1.000
GO:0032291	axon ensheathment in central nervous sys...	2	1	0.02	58	0.0208	1.000
GO:0140014	mitotic nuclear division	22	2	0.23	59	0.0213	1.000
GO:0007059	chromosome segregation	22	2	0.23	60	0.0213	1.000
GO:0061097	regulation of protein tyrosine kinase ac...	3	1	0.03	61	0.0311	1.000
GO:0046131	pyrimidine ribonucleoside metabolic proc...	3	1	0.03	62	0.0311	1.000
GO:0046133	pyrimidine ribonucleoside catabolic proc...	3	1	0.03	63	0.0311	1.000
GO:2000279	negative regulation of DNA biosynthetic ...	3	1	0.03	18	0.0311	0.031
GO:0014003	oligodendrocyte development	3	1	0.03	64	0.0311	1.000
GO:0045005	DNA-dependent DNA replication maintenanc...	3	1	0.03	65	0.0311	1.000
GO:0032508	DNA duplex unwinding	3	1	0.03	19	0.0311	0.031
GO:0006471	protein ADP-ribosylation	3	1	0.03	66	0.0311	1.000
GO:0098739	import across plasma membrane	3	1	0.03	67	0.0311	1.000
GO:0048709	oligodendrocyte differentiation	3	1	0.03	68	0.0311	1.000
GO:0042552	myelination	3	1	0.03	69	0.0311	1.000
GO:0006829	zinc ion transport	3	1	0.03	70	0.0311	1.000
GO:0043085	positive regulation of catalytic activit...	28	2	0.29	71	0.0335	1.000
GO:0045934	negative regulation of nucleobase-contai...	28	2	0.29	72	0.0335	1.000
GO:0051172	negative regulation of nitrogen compound...	71	3	0.74	73	0.0359	1.000
GO:2000113	negative regulation of cellular macromol...	31	2	0.32	74	0.0405	1.000

GO:0008366	axon ensheathment	4	1	0.04	75	0.0412	1.000
GO:0046135	pyrimidine nucleoside catabolic process	4	1	0.04	76	0.0412	1.000
GO:0006271	DNA strand elongation involved in DNA re...	4	1	0.04	77	0.0412	1.000
GO:0010324	membrane invagination	4	1	0.04	78	0.0412	1.000
GO:0022616	DNA strand elongation	4	1	0.04	79	0.0412	1.000
GO:0099024	plasma membrane invagination	4	1	0.04	80	0.0412	1.000
GO:0006544	glycine metabolic process	4	1	0.04	20	0.0412	0.041
GO:0000387	spliceosomal snRNP assembly	4	1	0.04	21	0.0412	0.041
GO:0006474	N-terminal protein amino acid acetylatio...	4	1	0.04	81	0.0412	1.000
GO:0042981	regulation of apoptotic process	32	2	0.33	82	0.0429	1.000
GO:0010558	negative regulation of macromolecule bio...	33	2	0.35	83	0.0454	1.000
GO:0000280	nuclear division	34	2	0.36	84	0.0479	1.000
GO:0043067	regulation of programmed cell death	34	2	0.36	85	0.0479	1.000

Sup. Tab. 3 b - genes upregulated in gonads of *B. grandii* and not upregulated in gonads of *B. atticus* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0003677	DNA binding	45	4	0.46	23	0.00087	1.00
GO:0016799	hydrolase activity hydrolyzing N-glycos…	9	2	0.09	24	0.00341	1.00
GO:0140097	catalytic activity acting on DNA	43	3	0.44	25	0.00856	1.00
GO:0070224	sulfide:quinone oxidoreductase activity	1	1	0.01	1	0.01025	0.01
GO:0032810	sterol response element binding	1	1	0.01	2	0.01025	0.01
GO:0071208	histone pre-mRNA DCP binding	1	1	0.01	3	0.01025	0.01
GO:0005212	structural constituent of eye lens	1	1	0.01	4	0.01025	0.01
GO:0008477	purine nucleosidase activity	1	1	0.01	5	0.01025	0.01
GO:0034039	8-oxo-7 8-dihydroguanine DNA N-glycosyla...	1	1	0.01	6	0.01025	0.01
GO:0032356	oxidized DNA binding	1	1	0.01	26	0.01025	1.00
GO:0032357	oxidized purine DNA binding	1	1	0.01	7	0.01025	0.01
GO:0015643	toxic substance binding	1	1	0.01	8	0.01025	0.01
GO:0043139	5'-3' DNA helicase activity	1	1	0.01	9	0.01025	0.01
GO:0008353	RNA polymerase II CTD heptapeptide repea...	1	1	0.01	10	0.01025	0.01
GO:0003910	DNA ligase (ATP) activity	1	1	0.01	11	0.01025	0.01
GO:0004427	inorganic diphosphatase activity	1	1	0.01	12	0.01025	0.01
GO:0045027	DNA end binding	1	1	0.01	13	0.01025	0.01
GO:0003909	DNA ligase activity	1	1	0.01	27	0.01025	1.00
GO:0004126	cytidine deaminase activity	1	1	0.01	14	0.01025	0.01
GO:0051539	4 iron 4 sulfur cluster binding	1	1	0.01	15	0.01025	0.01
GO:1990837	sequence-specific double-stranded DNA bi...	21	2	0.22	19	0.01852	0.12
GO:0042803	protein homodimerization activity	2	1	0.02	16	0.02040	0.02
GO:0004596	peptide alpha-N-acetyltransferase activi...	2	1	0.02	17	0.02040	0.02
GO:0016672	oxidoreductase activity acting on a sul…	2	1	0.02	28	0.02040	1.00
GO:0003684	damaged DNA binding	2	1	0.02	29	0.02040	1.00
GO:0008534	oxidized purine nucleobase lesion DNA N-...	2	1	0.02	30	0.02040	1.00
GO:0043565	sequence-specific DNA binding	26	2	0.27	31	0.02781	1.00

GO:0003690	double-stranded DNA binding	27	2	0.28	32	0.02985	1.00
GO:0016886	ligase activity forming phosphoric este…	3	1	0.03	33	0.03045	1.00
GO:0000702	oxidized base lesion DNA N-glycosylase a...	3	1	0.03	34	0.03045	1.00
GO:0016505	peptidase activator activity involved in...	3	1	0.03	18	0.03045	0.03
GO:0042802	identical protein binding	4	1	0.04	35	0.04040	1.00
GO:0051540	metal cluster binding	4	1	0.04	36	0.04040	1.00
GO:0051536	iron-sulfur cluster binding	4	1	0.04	37	0.04040	1.00

Sup. Tab. 3 c - genes upregulated in gonads of *B. grandii* and not upregulated in gonads of *B. rossius* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0040001	establishment of mitotic spindle localiz...	1	1	0	1	0.0038	0.0038
GO:0090116	C-5 methylation of cytosine	1	1	0	2	0.0038	0.0038
GO:0006610	ribosomal protein import into nucleus	1	1	0	3	0.0038	0.0038
GO:0001188	RNA polymerase I preinitiation complex a...	1	1	0	4	0.0038	0.0038
GO:0007079	mitotic chromosome movement towards spin...	1	1	0	5	0.0038	0.0038
GO:1901838	positive regulation of transcription of ...	1	1	0	6	0.0038	0.0038
GO:0006306	DNA methylation	3	2	0.01	7	0.0067	0.0067
GO:0045898	regulation of RNA polymerase II transcri...	2	1	0.01	9	0.0076	0.0076
GO:0030488	tRNA methylation	3	1	0.01	10	0.0114	0.0114
GO:0006360	transcription by RNA polymerase I	5	2	0.02	8	0.0133	0.0067
GO:0001510	RNA methylation	11	1	0.04	14	0.0412	10.000
GO:0006333	chromatin assembly or disassembly	13	1	0.05	11	0.0485	0.0485

Sup. Tab. 3 d - genes upregulated in gonads of *B. grandii* and not upregulated in gonads of *B. rossius* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0016428	tRNA (cytosine-5-)-methyltransferase act...	1	1	0.01	1	0.0072	0.0072
GO:0003886	DNA (cytosine-5-)-methyltransferase acti...	1	1	0.01	2	0.0072	0.0072
GO:0004315	3-oxoacyl-[acyl-carrier-protein] synthas...	1	1	0.01	3	0.0072	0.0072
GO:0004806	triglyceride lipase activity	1	1	0.01	4	0.0072	0.0072
GO:0001054	RNA polymerase I activity	1	1	0.01	5	0.0072	0.0072
GO:0004573	mannosyl-oligosaccharide glucosidase act...	1	1	0.01	6	0.0072	0.0072
GO:0008139	nuclear localization sequence binding	1	1	0.01	7	0.0072	0.0072
GO:0031995	insulin-like growth factor II binding	1	1	0.01	8	0.0072	0.0072
GO:0036374	glutathione hydrolase activity	1	1	0.01	9	0.0072	0.0072
GO:0005009	insulin-activated receptor activity	1	1	0.01	10	0.0072	0.0072
GO:0043559	insulin binding	1	1	0.01	11	0.0072	0.0072
GO:0001972	retinoic acid binding	1	1	0.01	12	0.0072	0.0072
GO:0098519	nucleotide phosphatase activity acting …	1	1	0.01	13	0.0072	0.0072
GO:0036143	kringle domain binding	1	1	0.01	14	0.0072	0.0072
GO:0051879	Hsp90 protein binding	1	1	0.01	15	0.0072	0.0072
GO:0005200	structural constituent of cytoskeleton	2	1	0.01	16	0.0144	0.0144
GO:0035064	methylated histone binding	3	1	0.02	17	0.0216	0.0216
GO:0140034	methylation-dependent protein binding	3	1	0.02	18	0.0216	10.000
GO:0042393	histone binding	6	1	0.04	19	0.0427	10.000

Sup. Tab. 3 e - genes upregulated in gonads of *B. grandii* and not upregulated in gonads of parthenogens (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0033499	galactose catabolic process via UDP-gala...	1	1	0.01	1	0.0076	0.0076
GO:0051131	chaperone-mediated protein complex assem...	1	1	0.01	2	0.0076	0.0076
GO:0006285	base-excision repair AP site formation	1	1	0.01	3	0.0076	0.0076
GO:0006478	peptidyl-tyrosine sulfation	1	1	0.01	4	0.0076	0.0076
GO:0032781	positive regulation of ATPase activity	1	1	0.01	5	0.0076	0.0076
GO:2000001	regulation of DNA damage checkpoint	1	1	0.01	6	0.0076	0.0076
GO:0006296	nucleotide-excision repair DNA incision…	1	1	0.01	7	0.0076	0.0076
GO:0036066	protein O-linked fucosylation	1	1	0.01	8	0.0076	0.0076
GO:1901642	nucleoside transmembrane transport	1	1	0.01	9	0.0076	0.0076
GO:0035621	ER to Golgi ceramide transport	1	1	0.01	10	0.0076	0.0076
GO:0010259	multicellular organism aging	2	1	0.02	11	0.0152	0.0152
GO:0000244	spliceosomal tri-snRNP complex assembly	2	1	0.02	12	0.0152	0.0152
GO:0071173	spindle assembly checkpoint	3	1	0.02	13	0.0227	0.0227
GO:0006303	double-strand break repair via nonhomolo...	3	1	0.02	14	0.0227	0.0227
GO:0031577	spindle checkpoint	3	1	0.02	20	0.0227	10.000
GO:0007043	cell-cell junction assembly	4	1	0.03	21	0.0301	10.000
GO:0000387	spliceosomal snRNP assembly	4	1	0.03	22	0.0301	10.000
GO:0006505	GPI anchor metabolic process	4	1	0.03	15	0.0301	0.0301
GO:0000726	non-recombinational repair	4	1	0.03	23	0.0301	10.000
GO:0043297	apical junction assembly	4	1	0.03	16	0.0301	0.0301
GO:0007568	aging	5	1	0.04	24	0.0375	10.000

Sup. Tab. 3 f - genes upregulated in gonads of *B. grandii* and not upregulated in gonads of parthenogens (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0042030	ATPase inhibitor activity	1	1	0.01	1	0.0066	0.0066
GO:0046922	peptide-O-fucosyltransferase activity	1	1	0.01	2	0.0066	0.0066
GO:0008108	UDP-glucose:hexose-1-phosphate uridylylt...	1	1	0.01	3	0.0066	0.0066
GO:0001070	RNA-binding transcription regulator acti...	1	1	0.01	4	0.0066	0.0066
GO:0000703	oxidized pyrimidine nucleobase lesion DN...	1	1	0.01	5	0.0066	0.0066
GO:0008476	protein-tyrosine sulfotransferase activi...	1	1	0.01	6	0.0066	0.0066
GO:0005337	nucleoside transmembrane transporter act...	1	1	0.01	7	0.0066	0.0066
GO:0070678	preprotein binding	1	1	0.01	8	0.0066	0.0066
GO:0097001	ceramide binding	2	1	0.01	9	0.0132	0.0132
GO:0046625	sphingolipid binding	3	1	0.02	14	0.0198	10.000
GO:0035258	steroid hormone receptor binding	3	1	0.02	10	0.0198	0.0198
GO:0042162	telomeric DNA binding	3	1	0.02	11	0.0198	0.0198
GO:0035257	nuclear hormone receptor binding	7	1	0.05	15	0.0456	10.000

Sup. Tab. 4 a - genes upregulated in female gonads of *B. grandii* and not upregulated in gonads of *B. atticus* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:1990258	histone glutamine methylation	1	1	0.01	1	0.0095	0.0095

GO:1990592	protein K69-linked ufmylation	1	1	0.01	2	0.0095	0.0095
GO:0048096	chromatin-mediated maintenance of transc...	1	1	0.01	3	0.0095	0.0095
GO:0040035	hermaphrodite genitalia development	1	1	0.01	4	0.0095	0.0095
GO:0000494	box C/D snoRNA 3'-end processing	1	1	0.01	5	0.0095	0.0095
GO:0019919	peptidyl-arginine methylation to asymme…	1	1	0.01	6	0.0095	0.0095
GO:0008587	imaginal disc-derived wing margin morpho...	1	1	0.01	7	0.0095	0.0095
GO:0061504	cyclic threonylcarbamoyladenosine biosyn...	1	1	0.01	8	0.0095	0.0095
GO:0051897	positive regulation of protein kinase B ...	1	1	0.01	9	0.0095	0.0095
GO:0007469	antennal development	1	1	0.01	10	0.0095	0.0095
GO:0061357	positive regulation of Wnt protein secre...	1	1	0.01	11	0.0095	0.0095
GO:0007480	imaginal disc-derived leg morphogenesis	1	1	0.01	12	0.0095	0.0095
GO:0010961	cellular magnesium ion homeostasis	1	1	0.01	13	0.0095	0.0095
GO:2000757	negative regulation of peptidyl-lysine a...	2	1	0.02	30	0.0189	10.000
GO:0046500	S-adenosylmethionine metabolic process	2	1	0.02	14	0.0189	0.0189
GO:0035067	negative regulation of histone acetylati...	2	1	0.02	15	0.0189	0.0189
GO:1901984	negative regulation of protein acetylati...	2	1	0.02	31	0.0189	10.000
GO:0034729	histone H3-K79 methylation	2	1	0.02	16	0.0189	0.0189
GO:2000779	regulation of double-strand break repair	3	1	0.03	17	0.0283	0.0283
GO:0006891	intra-Golgi vesicle-mediated transport	3	1	0.03	18	0.0283	0.0283
GO:0018205	peptidyl-lysine modification	31	2	0.29	32	0.0339	10.000
GO:2000756	regulation of peptidyl-lysine acetylatio...	4	1	0.04	33	0.0375	10.000
GO:0006282	regulation of DNA repair	4	1	0.04	34	0.0375	10.000
GO:0035065	regulation of histone acetylation	4	1	0.04	35	0.0375	10.000
GO:1901983	regulation of protein acetylation	4	1	0.04	36	0.0375	10.000
GO:0006085	acetyl-CoA biosynthetic process	4	1	0.04	19	0.0375	0.0375
GO:0010906	regulation of glucose metabolic process	5	1	0.05	21	0.0467	0.0467
GO:1902749	regulation of cell cycle G2/M phase tran...	5	1	0.05	37	0.0467	10.000
GO:0010389	regulation of G2/M transition of mitotic...	5	1	0.05	22	0.0467	0.0467
GO:0031057	negative regulation of histone modificat...	5	1	0.05	38	0.0467	10.000
GO:0045727	positive regulation of translation	5	1	0.05	23	0.0467	0.0467
GO:0010675	regulation of cellular carbohydrate meta...	5	1	0.05	39	0.0467	10.000
GO:0006732	coenzyme metabolic process	38	2	0.36	40	0.0493	10.000
GO:0006790	sulfur compound metabolic process	38	2	0.36	41	0.0493	10.000

Sup. Tab. 4 b - genes upregulated in female gonads of *B. grandii* and not upregulated in gonads of *B. atticus* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0016595	glutamate binding	1	1	0.01	2	0.006	0.00603
GO:0071566	UFM1 activating enzyme activity	1	1	0.01	3	0.006	0.00603
GO:0061503	tRNA threonylcarbamoyladenosine dehydrat...	1	1	0.01	4	0.006	0.00603
GO:0016596	thienylcyclohexylpiperidine binding	1	1	0.01	5	0.006	0.00603
GO:1990259	histone-glutamine methyltransferase acti...	1	1	0.01	14	0.006	100.000
GO:0031151	histone methyltransferase activity (H3-K...	1	1	0.01	6	0.006	0.00603
GO:0017174	glycine N-methyltransferase activity	1	1	0.01	7	0.006	0.00603

GO:0035242	protein-arginine omega-N asymmetric meth...	1	1	0.01	8	0.006	0.00603
GO:0004035	alkaline phosphatase activity	1	1	0.01	9	0.006	0.00603
GO:0005524	ATP binding	2	1	0.01	10	0.012	0.01202
GO:0008469	histone-arginine N-methyltransferase act...	2	1	0.01	15	0.012	100.000
GO:0070491	repressing transcription factor binding	2	1	0.01	11	0.012	0.01202
GO:0019212	phosphatase inhibitor activity	3	1	0.02	12	0.018	0.01799
GO:0031491	nucleosome binding	3	1	0.02	13	0.018	0.01799
GO:0042054	histone methyltransferase activity	8	3	0.05	1	0.029	0.00054
GO:0035639	purine ribonucleoside triphosphate bindi...	5	1	0.03	16	0.030	100.000
GO:0032559	adenyl ribonucleotide binding	8	1	0.05	17	0.047	100.000
GO:0003682	chromatin binding	8	1	0.05	18	0.047	100.000

Sup. Tab. 4 c - genes upregulated in female gonads of *B. grandii* and not upregulated in gonads of *B. rossius* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0007168	receptor guanylyl cyclase signaling path...	1	1	0.01	1	0.0057	0.0057
GO:0008611	ether lipid biosynthetic process	1	1	0.01	2	0.0057	0.0057
GO:0009103	lipopolysaccharide biosynthetic process	1	1	0.01	3	0.0057	0.0057
GO:0006432	phenylalanyl-tRNA aminoacylation	1	1	0.01	4	0.0057	0.0057
GO:0006182	cGMP biosynthetic process	1	1	0.01	5	0.0057	0.0057
GO:0042351	'de novo' GDP-L-fucose biosynthetic proc...	1	1	0.01	6	0.0057	0.0057
GO:0043605	cellular amide catabolic process	2	1	0.01	7	0.0114	0.0114
GO:0019673	GDP-mannose metabolic process	2	1	0.01	8	0.0114	0.0114
GO:0072583	clathrin-dependent endocytosis	2	1	0.01	9	0.0114	0.0114
GO:0043112	receptor metabolic process	2	1	0.01	10	0.0114	0.0114
GO:0080111	DNA demethylation	2	1	0.01	11	0.0114	0.0114
GO:0072698	protein localization to microtubule cyto...	2	1	0.01	12	0.0114	0.0114
GO:0044380	protein localization to cytoskeleton	2	1	0.01	20	0.0114	10.000
GO:0048488	synaptic vesicle endocytosis	3	1	0.02	13	0.0170	0.0170
GO:0006614	SRP-dependent cotranslational protein ta...	3	1	0.02	14	0.0170	0.0170
GO:0035510	DNA dealkylation	3	1	0.02	21	0.0170	10.000
GO:0140238	presynaptic endocytosis	3	1	0.02	22	0.0170	10.000
GO:0009200	deoxyribonucleoside triphosphate metabol...	4	1	0.02	15	0.0227	0.0227
GO:0006613	cotranslational protein targeting to mem...	4	1	0.02	23	0.0227	10.000
GO:0045047	protein targeting to ER	4	1	0.02	24	0.0227	10.000
GO:0072599	establishment of protein localization to...	4	1	0.02	25	0.0227	10.000
GO:0009263	deoxyribonucleotide biosynthetic process	5	1	0.03	16	0.0282	0.0282
GO:0044728	DNA methylation or demethylation	5	1	0.03	26	0.0282	10.000
GO:0033365	protein localization to organelle	50	2	0.29	27	0.0314	10.000
GO:0006304	DNA modification	6	1	0.03	28	0.0338	10.000
GO:0036465	synaptic vesicle recycling	6	1	0.03	29	0.0338	10.000
GO:0070972	protein localization to endoplasmic reti...	6	1	0.03	30	0.0338	10.000
GO:0006898	receptor-mediated endocytosis	6	1	0.03	31	0.0338	10.000
GO:0009262	deoxyribonucleotide metabolic process	7	1	0.04	32	0.0393	10.000

GO:0006612	protein targeting to membrane	7	1	0.04	33	0.0393	10.000
GO:0048511	rhythmic process	8	1	0.05	17	0.0448	0.0448

Sup. Tab. 4 d - genes upregulated in female gonads of *B. grandii* and not upregulated in gonads of *B. rossius* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0004383	guanylate cyclase activity	1	1	0.01	1	0.0054	0.0054
GO:0070573	metallodipeptidase activity	1	1	0.01	2	0.0054	0.0054
GO:0030942	endoplasmic reticulum signal peptide bin...	1	1	0.01	3	0.0054	0.0054
GO:0004826	phenylalanine-tRNA ligase activity	1	1	0.01	4	0.0054	0.0054
GO:0008312	7S RNA binding	1	1	0.01	5	0.0054	0.0054
GO:0016941	natriuretic peptide receptor activity	1	1	0.01	6	0.0054	0.0054
GO:0052634	C-19 gibberellin 2-beta-dioxygenase acti...	1	1	0.01	7	0.0054	0.0054
GO:0052635	C-20 gibberellin 2-beta-dioxygenase acti...	1	1	0.01	8	0.0054	0.0054
GO:0003924	GTPase activity	1	1	0.01	9	0.0054	0.0054
GO:0004748	ribonucleoside-diphosphate reductase act...	1	1	0.01	10	0.0054	0.0054
GO:0008609	alkylglycerone-phosphate synthase activi...	1	1	0.01	11	0.0054	0.0054
GO:0008446	GDP-mannose 4 6-dehydratase activity	1	1	0.01	12	0.0054	0.0054
GO:0019798	procollagen-proline dioxygenase activity	2	1	0.01	13	0.0108	0.0108
GO:0019003	GDP binding	2	1	0.01	14	0.0108	0.0108
GO:0005525	GTP binding	3	1	0.02	15	0.0162	0.0162
GO:0031543	peptidyl-proline dioxygenase activity	3	1	0.02	19	0.0162	10.000
GO:0001883	purine nucleoside binding	4	1	0.02	20	0.0215	10.000
GO:0032550	purine ribonucleoside binding	4	1	0.02	21	0.0215	10.000
GO:0035639	purine ribonucleoside triphosphate bindi...	5	1	0.03	22	0.0269	10.000
GO:0001882	nucleoside binding	5	1	0.03	23	0.0269	10.000
GO:0015662	ion transmembrane transporter activity …	5	1	0.03	16	0.0269	0.0269
GO:0032549	ribonucleoside binding	5	1	0.03	24	0.0269	10.000
GO:0032561	guanyl ribonucleotide binding	6	1	0.03	25	0.0322	10.000
GO:0019001	guanyl nucleotide binding	7	1	0.04	26	0.0374	10.000
GO:0042562	hormone binding	9	1	0.05	17	0.0479	0.0479

Sup. Tab. 4 e - genes upregulated in female gonads of *B. grandii* and not upregulated in gonads of parthenogens (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0010637	negative regulation of mitochondrial fus...	1	1	0.01	1	0.0057	0.0057
GO:0019254	carnitine metabolic process CoA-linked	1	1	0.01	2	0.0057	0.0057
GO:0051791	medium-chain fatty acid metabolic proces...	1	1	0.01	3	0.0057	0.0057
GO:0070206	protein trimerization	1	1	0.01	4	0.0057	0.0057
GO:0097310	cap2 mRNA methylation	1	1	0.01	5	0.0057	0.0057
GO:0022417	protein maturation by protein folding	1	1	0.01	6	0.0057	0.0057
GO:1901532	regulation of hematopoietic progenitor c...	1	1	0.01	7	0.0057	0.0057
GO:2000648	positive regulation of stem cell prolife...	1	1	0.01	8	0.0057	0.0057
GO:1904381	Golgi apparatus mannose trimming	1	1	0.01	9	0.0057	0.0057
GO:1902033	regulation of hematopoietic stem cell pr...	2	1	0.01	10	0.0114	0.0114

GO:0000463	maturation of LSU-rRNA from tricistronic...	2	1	0.01	11	0.0114	0.0114
GO:0042254	ribosome biogenesis	31	2	0.18	19	0.0127	10.000
GO:0016082	synaptic vesicle priming	3	1	0.02	12	0.0170	0.0170
GO:0071425	hematopoietic stem cell proliferation	3	1	0.02	20	0.0170	10.000
GO:0042058	regulation of epidermal growth factor re...	4	1	0.02	13	0.0227	0.0227
GO:1901184	regulation of ERBB signaling pathway	4	1	0.02	21	0.0227	10.000
GO:0000470	maturation of LSU-rRNA	4	1	0.02	22	0.0227	10.000
GO:0007173	epidermal growth factor receptor signali...	5	1	0.03	23	0.0282	10.000
GO:0038127	ERBB signaling pathway	6	1	0.03	24	0.0338	10.000
GO:0006898	receptor-mediated endocytosis	6	1	0.03	14	0.0338	0.0338
GO:0022613	ribonucleoprotein complex biogenesis	52	2	0.3	25	0.0338	10.000
GO:0016079	synaptic vesicle exocytosis	7	1	0.04	26	0.0393	10.000
GO:0042273	ribosomal large subunit biogenesis	7	1	0.04	27	0.0393	10.000
GO:0099643	signal release from synapse	8	1	0.05	28	0.0448	10.000
GO:0140029	exocytic process	8	1	0.05	29	0.0448	10.000
GO:0007269	neurotransmitter secretion	8	1	0.05	30	0.0448	10.000
GO:0042274	ribosomal small subunit biogenesis	8	1	0.05	15	0.0448	0.0448

Sup. Tab. 4 f - genes upregulated in female gonads of *B. grandii* and not upregulated in gonads of parthenogens (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0031177	phosphopantetheine binding	1	1	0.01	1	0.006	0.006
GO:0070991	medium-chain-acyl-CoA dehydrogenase acti...	1	1	0.01	2	0.006	0.006
GO:0070524	11-beta-hydroxysteroid dehydrogenase (NA...	1	1	0.01	3	0.006	0.006
GO:0034189	very-low-density lipoprotein particle bi...	1	1	0.01	4	0.006	0.006
GO:0047874	dolichyldiphosphatase activity	1	1	0.01	5	0.006	0.006
GO:0005041	low-density lipoprotein particle recepto...	1	1	0.01	6	0.006	0.006
GO:0008650	rRNA (uridine-2'-O-)-methyltransferase a...	1	1	0.01	7	0.006	0.006
GO:0004556	alpha-amylase activity	1	1	0.01	8	0.006	0.006
GO:0070742	C2H2 zinc finger domain binding	1	1	0.01	9	0.006	0.006
GO:0030229	very-low-density lipoprotein particle re...	1	1	0.01	10	0.006	0.006
GO:0038025	reelin receptor activity	1	1	0.01	11	0.006	0.006
GO:0034185	apolipoprotein binding	2	1	0.01	12	0.012	0.012
GO:0008187	poly-pyrimidine tract binding	2	1	0.01	13	0.012	0.012
GO:0030971	receptor tyrosine kinase binding	2	1	0.01	14	0.012	0.012
GO:1990782	protein tyrosine kinase binding	2	1	0.01	17	0.012	1.000
GO:0003964	RNA-directed DNA polymerase activity	3	1	0.02	15	0.018	0.018
GO:0034061	DNA polymerase activity	4	1	0.02	18	0.024	1.000
GO:0003727	single-stranded RNA binding	4	1	0.02	19	0.024	1.000

Sup. Tab. 5 a -GO enrichment of genes with at leas one codon under positive selection (BEB P>0.95) in *B. atticus* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0044264	cellular polysaccharide metabolic proces...	10	3	0.3	1	0.0027	0.0027
GO:0045595	regulation of cell differentiation	41	5	1.23	40	0.0067	0.0845

GO:0002682	regulation of immune system process	30	4	0.9	121	0.0112	10.000
GO:0002683	negative regulation of immune system pro...	6	2	0.18	2	0.0123	0.0123
GO:0033692	cellular polysaccharide biosynthetic pro...	6	2	0.18	122	0.0123	10.000
GO:0000271	polysaccharide biosynthetic process	7	2	0.21	123	0.0169	10.000
GO:1903706	regulation of hemopoiesis	7	2	0.21	124	0.0169	10.000
GO:0097485	neuron projection guidance	8	2	0.24	3	0.0220	0.0220
GO:0006974	cellular response to DNA damage stimulus	55	5	1.65	101	0.0226	0.7199
GO:0034660	ncRNA metabolic process	78	6	2.34	125	0.0271	10.000
GO:2001020	regulation of response to DNA damage sti...	9	2	0.27	71	0.0278	0.1895
GO:0034637	cellular carbohydrate biosynthetic proce...	9	2	0.27	126	0.0278	10.000
GO:0006935	chemotaxis	12	2	0.36	127	0.0481	10.000
GO:0042330	taxis	12	2	0.36	128	0.0481	10.000
GO:0033554	cellular response to stress	113	7	3.39	107	0.0491	0.8646
GO:0016070	RNA metabolic process	214	11	6.41	129	0.0494	10.000

Sup. Tab. 5 b -GO enrichment of genes with at leas one codon under positive selection (BEB P>0.95) in *B. atticus* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0019783	ubiquitin-like protein-specific protease...	6	2	0.18	1	0.013	0.013
GO:0003723	RNA binding	54	5	1.66	66	0.023	0.716
GO:0004857	enzyme inhibitor activity	22	3	0.68	34	0.028	0.091
GO:0016893	endonuclease activity active with eithe…	9	2	0.28	74	0.029	1.000
GO:0008234	cysteine-type peptidase activity	11	2	0.34	75	0.043	1.000
GO:0140098	catalytic activity acting on RNA	86	6	2.64	76	0.045	1.000

Sup. Tab. 5 c -GO enrichment of genes with at leas one codon under positive selection (BEB P>0.95) in *B. rossius* (BP)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0031929	TOR signaling	6	3	0.29	2	0.0021	0.0068
GO:0006879	cellular iron ion homeostasis	2	2	0.1	1	0.0024	0.0024
GO:0031112	positive regulation of microtubule polym...	3	2	0.15	3	0.0069	0.0069
GO:0046113	nucleobase catabolic process	3	2	0.15	4	0.0069	0.0069
GO:0071229	cellular response to acid chemical	4	2	0.2	6	0.0134	0.0134
GO:0006284	base-excision repair	4	2	0.2	7	0.0134	0.0134
GO:0007019	microtubule depolymerization	4	2	0.2	8	0.0134	0.0134
GO:0043244	regulation of protein-containing complex...	11	3	0.54	5	0.0141	0.0132
GO:0044267	cellular protein metabolic process	317	24	15.53	199	0.0156	0.7863
GO:0031329	regulation of cellular catabolic process	32	5	1.57	244	0.0179	10.000
GO:0051223	regulation of protein transport	22	4	1.08	245	0.0201	10.000
GO:0090087	regulation of peptide transport	22	4	1.08	246	0.0201	10.000
GO:0006998	nuclear envelope organization	5	2	0.25	79	0.0216	0.1388
GO:0006144	purine nucleobase metabolic process	5	2	0.25	9	0.0216	0.0216
GO:0032006	regulation of TOR signaling	5	2	0.25	247	0.0216	10.000
GO:0006661	phosphatidylinositol biosynthetic proces...	5	2	0.25	10	0.0216	0.0216
GO:0006641	triglyceride metabolic process	5	2	0.25	11	0.0216	0.0216

GO:0071692	protein localization to extracellular re...	13	3	0.64	248	0.0228	10.000
GO:0035592	establishment of protein localization to...	13	3	0.64	249	0.0228	10.000
GO:0009306	protein secretion	13	3	0.64	250	0.0228	10.000
GO:0002790	peptide secretion	13	3	0.64	130	0.0228	0.2192
GO:0032984	protein-containing complex disassembly	24	4	1.18	251	0.0271	10.000
GO:0070201	regulation of establishment of protein l...	24	4	1.18	252	0.0271	10.000
GO:0019538	protein metabolic process	351	25	17.2	253	0.0279	10.000
GO:0009894	regulation of catabolic process	36	5	1.76	254	0.0288	10.000
GO:0006508	proteolysis	76	8	3.72	169	0.0299	0.3845
GO:0044260	cellular macromolecule metabolic process	499	33	24.45	223	0.0308	0.9822
GO:0009168	purine ribonucleoside monophosphate bios...	6	2	0.29	12	0.0313	0.0313
GO:0009127	purine nucleoside monophosphate biosynth...	6	2	0.29	255	0.0313	10.000
GO:0016241	regulation of macroautophagy	6	2	0.29	13	0.0313	0.0313
GO:0072522	purine-containing compound biosynthetic ...	26	4	1.27	256	0.0354	10.000
GO:0015031	protein transport	65	7	3.19	257	0.0370	10.000
GO:0044237	cellular metabolic process	957	57	46.89	212	0.0394	0.9453
GO:0045017	glycerolipid biosynthetic process	16	3	0.78	174	0.0402	0.4198
GO:0002791	regulation of peptide secretion	7	2	0.34	258	0.0425	10.000
GO:0042558	pteridine-containing compound metabolic ...	7	2	0.34	111	0.0425	0.1808
GO:0006892	post-Golgi vesicle-mediated transport	7	2	0.34	14	0.0425	0.0425
GO:0050708	regulation of protein secretion	7	2	0.34	15	0.0425	0.0425
GO:0015833	peptide transport	67	7	3.28	259	0.0427	10.000
GO:0032880	regulation of protein localization	28	4	1.37	260	0.0450	10.000
GO:0033036	macromolecule localization	145	12	7.11	208	0.0473	0.8581

Sup. Tab. 5 d - GO enrichment of genes with at leas one codon under positive selection (BEB P>0.95) in *B. rossius* (MF)

GO.ID	Term	Annotated	Significant	Expected	Rank in Weight_Fisher	Elim_Fisher	Weight_Fisher
GO:0060590	ATPase regulator activity	3	2	0.16	1	0.0079	0.0079
GO:0016209	antioxidant activity	12	3	0.63	2	0.0217	0.0217
GO:0008092	cytoskeletal protein binding	34	5	1.78	52	0.0295	0.1970
GO:0001067	regulatory region nucleic acid binding	14	3	0.73	113	0.0333	10.000
GO:0000976	transcription regulatory region sequence...	14	3	0.73	3	0.0333	0.0333
GO:0000987	cis-regulatory region sequence-specific ...	7	2	0.37	114	0.0481	10.000
GO:0016684	oxidoreductase activity acting on perox…	7	2	0.37	4	0.0481	0.0481
GO:0008171	O-methyltransferase activity	7	2	0.37	34	0.0481	0.1480
GO:0008378	galactosyltransferase activity	7	2	0.37	5	0.0481	0.0481

Sup. Tab. 6 a - codeml results for convergent positive selection in parthengoens - *Phyllium philippinicum*

OG	branch_model_g	site_model_g	model_g_np	model_g_LnL	branch_model_a	site_model_a	model_a_np	model_a_LnL	p.value	significance	sites BEB > 0.95
OG0003097	2	2	4	-1884,150469	2	2	5	-1884,150469	1	n/s	0
OG0004336	2	2	4	-2459,34765	2	2	5	-2459,34765	1	n/s	0
OG0004353	2	2	4	-2161,777299	2	2	5	-2161,777299	1	n/s	0
OG0005196	2	2	4	-2459,837244	2	2	5	-2459,837244	1	n/s	0

OG0005237	2	2	4	-10380,8868	2	2	5	-10380,8868	1	n/s	0
OG0006021	2	2	4	-1880,032196	2	2	5	-1880,032196	1	n/s	0

Sup. Tab. 6 b - codeml results for convergent positive selection in parthengoens - *B. grandii*

OG	branch_model_g	site_model_g	model_g_np	model_g_LnL	branch_model_a	site_model_a	model_a_np	model_a_LnL	p.value	significance	sites BEB > 0.95
OG0003097	2	2	4	-1884,253161	2	2	5	-1884,252972	0,9845	n/s	0
OG0004336	2	2	4	-2459,347651	2	2	5	-2459,34765	0,9989	n/s	0
OG0004353	2	2	4	-2161,777304	2	2	5	-2161,777332	1	n/s	0
OG0005196	2	2	4	-2459,820609	2	2	5	-2459,793326	0,8153	n/s	0
OG0005237	2	2	4	-10394,63138	2	2	5	-10394,63138	1	n/s	0
OG0006021	2	2	4	-1879,890214	2	2	5	-1879,890214	1	n/s	0

Sup. Tab. 6 c - codeml results for convergent positive selection in parthengoens - *B. atticus*

OG	branch_model_g	site_model_g	model_g_np	model_g_LnL	branch_model_a	site_model_a	model_a_np	model_a_LnL	p.value	significance	sites BEB > 0.95
OG0003097	2	2	4	1880,840439	2	2	5	-1868,754847	0	***	7
OG0004336	2	2	4	2450,720433	2	2	5	-2419,32525	0	***	14
OG0004353	2	2	4	2158,602561	2	2	5	-2113,131414	0	***	22
OG0005196	2	2	4	2459,837244	2	2	5	-2451,521883	0	***	2
OG0005237	2	2	4	10394,63138	2	2	5	-10384,33331	0	***	1
OG0006021	2	2	4	1872,862592	2	2	5	-1867,181731	0,0007	***	3

Sup. Tab. 6 d - codeml results for convergent positive selection in parthengoens - *B. rossius*

OG	branch_model_g	site_model_g	model_g_np	model_g_LnL	branch_model_a	site_model_a	model_a_np	model_a_LnL	p.value	significance	sites BEB > 0.95
OG0003097	2	2	4	-1880,203554	2	2	5	-1848,496384	0	***	12
OG0004336	2	2	4	-2457,80888	2	2	5	-2446,538505	0	***	2
OG0004353	2	2	4	-2158,758771	2	2	5	-2150,072045	0	***	8
OG0005196	2	2	4	-2459,810179	2	2	5	-2450,286185	0	***	2
OG0005237	2	2	4	-10394,63138	2	2	5	-10355,20889	0	***	10
OG0006021	2	2	4	-1875,946689	2	2	5	-1867,42091	0	***	3

Sup. Tab. 6 e - aBSREL results for convergent positive selection in parthengoens

OG	branch	B	LRT	test pval	Uncorrected p-value	ω distribution over sites	model	AICC	log L	Parameters
OG0003097	*B. atticus*	0.0000	505584	0.0000	0.0000	ω1 = 0.00 (88%)	Nucleotide GTR	3532.73	-1908.08	13
						ω2 = 9090 (12%)	Baseline MG94xREV	3532.73	-1817.69	24
	B. rossius	0.0000	672427	0.0000	0.0000	ω1 = 0.00390 (91%)	Full adaptive model	3532.73	-1737.48	28
						ω2 = 77.0 (9.4%)				
OG0004336	*B. atticus*	0.0000	818758	0.0000	0.0000	ω1 = 1.00 (92%)	Nucleotide GTR	4514.05	-2430.48	13
						ω2 = 1560 (7.8%)	Baseline MG94xREV	4514.05	-2286.23	22
	B. rossius	0.0000	153003	0.0002	0.0002	ω1 = 0.314 (97%)	Full adaptive model	4514.05	-2230.42	26
						ω2 = 38.5 (2.8%)				
OG0004353	*B. atticus*	0.0000	1098904	0.0000	0.0000	ω1 = 1.00 (83%)	Nucleotide GTR	3963.27	-2119.48	13
						ω2 = 248 (17%)	Baseline MG94xREV	3963.27	-2036.12	24
	B. rossius	0.0000	479798	0.0000	0.0000	ω1 = 0.675 (79%)	Full adaptive model	3963.27	-1952.73	28

						ω2 = 100000 (21%)				
OG0005196	*B. atticus*	0.0000	0.0000	10000	10000	ω1 = 0.571 (100%)	Nucleotide GTR	4696.41	-2492.04	13
	B. rossius	0.0000	0.0000	10000	10000	ω1 = 0.103 (100%)	Baseline MG94xREV	4696.41	-2375.89	24
							Full adaptive model	4696.41	-2321.68	26
OG0005237	*B. rossius*	0.0000	1240444	0.0000	0.0000	ω1 = 0.463 (96%)	Nucleotide GTR	19909.43	-10321.40	13
						ω2 = 409 (4.5%)	Baseline MG94xREV	19909.43	-10062.53	22
	B. atticus	0.0000	124331	0.0007	0.0007	ω1 = 0.489 (99%)	Full adaptive model	19909.43	-9924.53	30
						ω2 = 73.3 (0.64%)				
OG0006021	*B. atticus*	0.0000	249177	0.0000	0.0000	ω1 = 0.171 (94%)	Nucleotide GTR	3594.21	-1955.71	13
						ω2 = 100000 (5.8%)	Baseline MG94xREV	3594.21	-1806.19	22
	B. rossius	0.0000	221671	0.0000	0.0000	ω1 = 0.214 (96%)	Full adaptive model	3594.21	-1770.37	26

Lost but not forgotten: pleiotropic interactions constrain gene expression and sequence evolution of male gonad genes after the shift to parthenogenesis.

Giobbe Forni, Sasha Alexander Mikheyev, Andrea Luchetti and Barbara Mantovani.

Abstract - Selection - or the lack of it - may lead to the loss of a trait; when this happens, theory predicts that the molecular groundplan responsible for the trait phenotypic expression should decay. Yet, empirical evidence is mixed. As most phenotypes result from large numbers of interacting genes with different levels of pleiotropy, we hypothesize that gene network architecture could constrain the decay and preservation of the different trait-associated genes after trait loss. We tested this hypothesis in the *Bacillus* stick insects, where parthenogenesis evolved twice with different underlying mechanisms. After first retrieving genes associated with male reproductive tissues in a bisexual species, we investigated their modifications in the two parthenogens, focusing on gene regulatory network structure inferred by gene co-expression patterns. We found that the gene co-expression networks associated with male reproductive structures (*i.e.* gonad) were still preserved in parthenogens. Connectivity within the network played a key role, with highly connected (more pleiotropic) genes retaining more gonad-specific transcription patterns. No signature of sequence degradation was observed for male gonad-associated genes in parthenogens. Overall, more connected genes in the bisexual species network have undergone a slower evolutionary pace compared to peripheral ones; in parthenogens this correlation is not found for genes associated to male gonad in the bisexual species, reflecting the different context of their expression. Our findings demonstrate how the genomic blueprint responsible for a trait can be preserved after its loss, with pleiotropic interactions constraining gene expression and sequence evolution.

Keywords: Pleiotropy; Trait Loss; Parthenogenesis; Gene Co-Expression Network; Molecular Evolution.

Introduction:

Selection is a prominent force shaping the phenotypes of species and can be responsible for the establishment of novel traits. But when selective pressures change - as a consequence of environmental shifts or changes in life history - traits may also become useless or disadvantageous, decay and be lost. Trait loss is a common phenomenon across the tree of life and can have adaptive values as strong as the establishment of novel ones (Porter and Crandall, 2003). This process is predicted to have evolutionary consequences for the genes involved in its expression, either via drift or selection (Hall and Colegrave, 2008; van der Kooi & Schwander, 2014). The decay of trait-specific genes after its loss can be due to several causes, including mutations in coding (Meredith et al, 2009; Kraaijeveld et al, 2016) and regulatory parts (Leal and Cohn, 2016; Roscito et al, 2018; Sackton et al, 2019), changes in expression patterns (Ament et al, 2011; Zhang and Reed, 2016) and complete gene losses (Suen et al, 2011; Đaković et al, 2014). Nonetheless, in some instances empirical evidence of decay following trait loss are lacking. For example, parasitic wasps which have lost lipogenesis do not show any gene sequence degradation among those involved in the pathway (Lammers et al., 2019), genes underlying photosynthesis are inferred to be under a strong purifying selection in a parasitic plant (McNeal et al., 2007) and the expression of functional opsins has been observed in cave crustacean with reduced or absent eyes (Carlini et al, 2013; Stern and Crandall, 2018; Perez-Moreno et al, 2018). In *Timema* species which lost males and sexual reproduction different outcomes have been observed for different genes and whether genes expression or sequence evolution is considered. Decay has been associated to expression changes in legs and whole-bodies - but not in reproductive tract - while changes in selection regimes acting on sequence evolution are found only in genes undergoing cross-species convergent expression changes in whole-bodies (Bast et al, 2019). All together, these results highlight the difficulty in making generalizations on the impact of trait loss on its genomic blueprint; this process can have different outcomes on different genes and also vary depending whether expression patterns or sequence changes are considered.

However, these observations on maintenance and decay are not necessarily conflicting: phenotypes are in fact the product of large networks of interactions and different evolutionary trajectory can be expected for the genes underlying a trait after its loss. Genes have variable levels of pleiotropy (He & Zhang, 2006; Paaby & Rockman, 2013; Visscher and Yang, 2016): after trait loss, it can be expected that genes under selection also for traits unrelated to the lost one will be preserved over time, while those specifically associated with the lost trait will decay. Moreover, co-option of pre-existing genes and whole regulatory networks are frequent mechanisms underlying the origin of novelty (Monteiro, 2012; Isabel Almudí and Juan Pascual-Anaya, 2019; Hu et al, 2019; Almudi et al, 2020). In this perspective, few genes could be entirely dispensable after trait loss - either because they are already involved in

other phenotypes or because they can be co-opted for novel purposes - so that those involved in multiple biological processes are less likely to degenerate after the selective constraints on a specific trait are removed (Smith et al, 2015). As pleiotropy is known to modulate selection (Fraïsse et al, 2019) and constrain both gene expression (Papakostas, et al, 2014; Morandin et al, 2017) and sequence evolution (Alvarez-Ponce et al, 2011; Salathé et al, 2015), we hypothesized that the different trajectories of genes underlying a certain trait after its loss could be explained by their position in the associated co-expression network topology. Highly central (*i.e.* highly connected) genes are - in fact - more pleiotropic and more likely to be essential than others (Jeong et al, 2001; MacNeil and Walhoutm, 2011).

Our testing ground has been the Mediterranean *Bacillus* species complex of stick insects: the bisexual *Bacillus grandii*, the obligate parthenogen *Bacillus atticus* and the geographical parthenogen *Bacillus rossius* (Scali et al, 2003). The *Bacillus* clade started its diversification over 20 million years ago (Mantovani et al, 2001): earlier researches provided evidence that the two lineages represent independent shifts to automictic (*i.e.* meiotic) parthenogenesis, as highlighted by the different underlying mechanisms (**Fig. 1**; Scali et al, 2003). Despite the challenges in demonstrating the lack of some level of cryptic sex (Schurko et al, 2009), males are no more present in *B. atticus* and in *B. rossius* parthenogenetic populations as they have been observed - neither in nature or in captivity - during a timespan of over ten years. Parthenogenesis occurrs rather commonly in nature and it has been thoroughly studied at the molecular level: it provides an ideal framework to test our hypotheses on the consequences of trait loss, as several theoretical expectations have been developed (van der Kooi & Schwander, 2014) and many efforts have already been carried out to characterize its causes and consequences at the molecular level (Kraaijeveld et al, 2016; Brandt et al, 2017; Bast et al, 2018; Parker et al, 2019a; Parker et al, 2019b).

Here, we identified gonad-associated genes in the bisexual species, using a weighted gene co-expression network approach; the latter is known to reliably capture interactions among genes and proteins and allowed us to infer the extant of genes pleiotropic interactions (Carlson et al., 2006; MacNeil and Walhoutm, 2011; Allen et al, 2012). Comparative transcriptomic and phylogenetic approaches were then used to investigate the modifications which male gonad-associated genes underwent in the two lineages which shifted to parthenogenetic reproduction and where males are no more present. We specifically tested the hypothesis that genes associated with male reproductive structures in the bisexual species could have undergone different extant of decay and preservation in partenogens, with pleiotropy as of the main forces shaping their expression and sequence changes. This hypothesis predicts that the evolutionary fate of trait-associated genes after its loss will be determined by their ancestral pleiotropic interactions, so that highly pleiotropic genes will undergo fewer changes compared to less connected ones, in both expression and sequence.

Materials and Methods:

Transcriptome Assembly and Annotation, Orthology Inference and Differential Expression Analysis - Data generation, transcriptome assembly, orthology inference and differential expression were all carried out previously (for details see Forni et al, 2020). Non-reproductive (one foreleg, one mid-leg and one hind-leg) and reproductive (ovaries in females and testes in males) tissues samples were collected for *B. grandii* (Marettimo), *B. atticus* (Necropoli Camerina) and *B. rossius* (Massa San Nicola) after they reached sexual maturity. After RNA extraction and sequencing, all the sequenced libraries were uploaded to the SRA under the accession PRJNA578804. We used Trinity to assemble the novel transcriptomes (Haas et al, 2013) and TransDecoder (v5.5.0) to detect coding regions in the raw assemblies. Orthofinder2 (Emms & Kelly, 2019) was used to infer orthogroups among the three *Bacillus* species and the outgroup *Phyllium pilippinicum* (TSA accession: GCPM00000000); the latter is the closest relative to the *Bacillus* clade for which 'omics' data are available, with their divergence taking place *circa* 50 Mya (Simon et al, 2019). Reads were mapped using Bowtie2 (Langmead & Salzberg 2012) and RSEM was used to estimate expression values (Li and Dewie, 2011). Differential expression analyses between reproductive and non-reproductive tissues were performed separately for each species and sex, using Trimmed Mean of M-values normalized Transcript Per Million values (TMM normalized TPMs) in Deseq2 (Love et al., 2014). Subsequently, we gathered the FDR values and LogFC for each transcript based on the orthology inference results.

Gene Co-Expression Network Construction in the Bisexual Species and Gonad-Associated Genes Expression Modifications in Parthenogens - A weighted gene co-expression network was inferred for the bisexual species *B. grandii*. This approach identifies genes that have coordinated expression patterns across samples, using a hierarchical clustering approach; genes are clustered into modules of highly interconnected genes which are then associated with specific traits (Stuart et al, 2003; Hovart and Dong, 2008). TMM normalized TPM values used for the DE analysis were gathered for 24 *B. grandii* samples and for 2842 orthogrups consisting of single-copy genes shared across all four species. We chose to restrict our analysis to this subset of orthogroups in order to make the network analysis results match those of sequence evolution analyses; the latter are carried out in a phylogenetic framework, for which at least four branches are required. The analysis was performed using the R package WGCNA (Langfelder and Horvath 2008). A signed network - which assumes that strongly negatively correlated genes are not interconnected - was inferred based on the criterion of approximate scale-free topology; the latter is known to reflect properties of biological networks, which exhibit few highly connected genes linking the rest of less connected ones (Zhang and Horvath, 2005). A soft thresholding power of 19 - to which co-expression similarity was raised to calculate adjacency - was selected. In addition to constructing relationships between genes, WGCNA computes ‘modules’ of

genes with similar expression patterns; modules were defined to contain at least 30 genes with a dendrogram cut height of 0.2. We then calculated modules signifcance (the average correlation of module's gene expression profiles with external traits) for male and female gonad: we defined as male gonad-associated all genes found in modules which had a positive significance correlation with male gonad while female gonad-associated all genes inside modules which had a postive significance with female gonad; we also considered separately all genes unrelated to male or female gonad. Using the same parameters which we used for network construction, we calculated total and extra-modular connectivity for each gene in the bisexual species gene coexpression network; we used these metric (which measure how correlated the expression of a gene is with others) as an extimation of a gene pleiotropic interactions, respectively within the whole network and outside its module of membership.

After having identified gonad-associated modules in the bisexual *B. grandii*, we tested wether they could be partially preserved in parthenogens, despite male absence. We inferred module preservation statistics in WGCNA for the parthenogens *B. atticus* and *B. rossius*, leveraging gonad expression data only and using 500 permutations (Langfelder et al, 2011). This approach let us quantify how density and connectivity patterns of the bisexual species modules are preserved in the species which shifted to parthenogenesis. We relied on the $Z_{summary}$ metric - which combines statistical summaries of network density and connectivity to get an estimate of whether network characteristics are preserved. Following the thresholds proposed by Langfelder et al, 2011, we considered Zsummary < 2 as no evidence for module preservation, 2 < Zsummary < 10 as weak to moderate evidence and Zsummary > 10 as strong evidence. As Zsummary often depends on the module size (*i.e.* the number of nodes in a module), we also considered the medianRank metric for comparing relative preservation among multiple modules, which appears to be much less dependent on module size.

We then assed wether genes associated to male gonad modules have retained a gonad-biased expression pattern in parthenogens dependending on the extant of their pleiotropic interactions in the bisexual species. Using R (4.0.2; R Core Team, 2017), Spearman correlations were calculated between gene total connectivity in the bisexual *B. grandii* co-expression network and the LogFC found between reproductive and non-reproductive tissues of each *Bacillus* species and sex, for the different subset of genes defined above.

Sequence Evolution of Gonad-Associated Genes Across Bisexual and Parthenogenetic Species - We tested wether we could observe any signature of sequence degradation and/or relaxation of selective pressures for male gonad-associated genes in parthenogens. Protein sequences were aligned using MAFFT (Katoh & Standley, 2013) with default parameters and then retro-translated to a codon-based alignment using pal2nal (Suyama et al., 2006). To infer selection regimes on coding sites we used codeml of the PAML package (v.4.8; Yang, 2007). We calculated dN,

dS and dNdS ratios along the branches of our phylogeny for the 2842 single-copy ubiquitous genes orthogroups; all codeml analyses have been carried out using BASE (Forni et al, 2020). For each alignment, the species tree branch-lengths were optimized in maximum-likelihood framework using a codon-aware GTR+G model of evolution, with RaxML v8.2.12 (Stamatakis, 2014). Then a model with one omega class shared between all the branches of the tree (m0) and one model with a specific omega class for each branch (m1) were inferred and subsequently compared using a Likelihood Ratio Test (LRT) in R (4.0.2; R Core Team 2017). The p values obtained were adjusted for multiple comparisons using Benjamini and Hochberg's correction (Benjamini and Hochberg, 1995) and the best-fit model was selected for subsequent steps. dNdS and substitution rates were retrieved for each terminal branch leading to the three *Bacillus* species and for each orthogroup, considering the best-fit model. We excluded any dNdS values associated to dN>3 (implying substitution saturation) and also any artefactual dNdS values of 999. To estimate the strength of selection on synonymous sites we estimated the bias in codon usage in the three *Bacillus* species: using the alignment used for the dNdS analyses we calculated the Measure Independent of Length and Composition (MILC) metric for each *Bacillus* species, using the R package coRdon (Elek et al, 2019).

Subsequently we tested the hypothesis that gene connectivity (as a proxy for pleiotropic interactions) was driving serquence modifications after trait loss. Using R (4.0.2; R Core Team 2017) we performed Spearman correlations between gene connectivity (total and extra-modular) inferred in the bisexual *B. grandii* co-expression network and dNdS, MILC and substitution rates of each *Bacillus* terminal branch; correlations were calculated for male gonad-associated and female gonad-associated genes along with all remaining genes. After checking for metrics distribution normality using the Shapiro-Wilk test implemented in R (4.0.2; R Core Team 2017), we compared male gonad and female gonad-associated genes metrics - along with all other genes - using the Wilcoxon test implemented in ggpubr (Kassambara and Kassambara, 2020).

GO Term Enrichment Analyses - Gene annotation has been carried out separately for each species leveraging blastp searches against Uniref database (eval<1e-3) and hmmer searches against the pfam database; we then generated GO-terms with Argot 2.5 with a TotalScore > 200 (Lavezzo et al, 2016). Subsequently, we gathered all GO-terms associated with each OG across the three *Bacillus* species and collapsed multiple entries of the same term. Each module enrichment analyses were performed with the TopGO package in Bioconductor, using Fisher exact test and both elim and weight algorithms - which take into account GO hierarchy (Alexa and Rahnenführer, 2009). GO-terms were considered to be significantly enriched when elim $p < 0.05$.

Code Availability - All scripts used for this project are available at https://github.com/for-giobbe/parthenogeneisis-in-the-Bacillus-species-complex.

Results:

Preservation of Male Gonad-associated Genes Expression in Parthenogens - We inferred a weighted gene co-expression network for the bisexual species *B. grandii*, using the expression data of the 2842 single-copy ubiquitous genes also used for sequence evolution analyses. Seven modules of co-expression were identified, three of them (A, D, E) presented a significative positive correlation to male gonad and two (C, F) to female gonad (**Fig. 2A; Sup. Fig. 1**); no module presented a significative positive correlation with male and female gonad at the same time. Each module contained between 909 and 105 genes, with 2.7% of genes left unassigned. Modules of co-expression associated with male and female gonad are enriched for several transcription and meiosis associated functions, such as gene silencing, cell cycle and proliferation, morphogenesis, chromosome organization (**Sup. Tab. 1**).

We then tested to which extent the module of co-expression associated to male and female gonad in the bisexual species are preserved across the two parthenogenetic species *B. atticus* and *B. rossius* (**Fig. 2B, Sup. Fig. 1**). To consider a module as preserved we adopted the Zsummary thresholds of 2 for weak preservation and 10 for strong preservation; as the latter metric is often dependent on module size, we also considered the median rank (Langfelder et al., 2011). Overall the gene network was found to be preserved (Zsummary 16.0 for *B. rossius* and 15.0 for *B. atticus*), while genes which were not part of the co-expression network didn't show any sign of preservation (Zsummary -0.61 for *B. rossius* and -0.17 for *B. atticus*). Female gonad related modules are found to be consistently preserved in parthenogens, while modules associated to male gonad exhibit different degrees of preservation, yet with a consistent pattern across both parthenogens. Yet, some lineage-specific differences in preservation can be found, such as in module C; as it is less preserved in *B. rossius*, the difference can possibly be explained by the latter being the first lineage to diverge among the three. Interestingly, the more preserved module across both parthenogens is module B, which in the bisexual species network is associated with male somatic tissue.

For male gonad-associated genes, a significant correlation between total connectivity in the gene co-expression network and LogFC between reproductive and non-reproductive tissues is found in *B. grandii* males (Spearman correlation $r=0.328$, $p<0.05$; **Fig. 2C**): this is expected, as highly connected genes associated to a tissue should present a strong tissue-specific pattern of expression. A weaker but significative correlation is also found in *B. grandii* females ($r=0.286$, $p<0.05$): highly connected genes associated to male gonad are also expressed in female gonad implying that the former physiology is sustained by genes which have pleiotropic effects in both sexes. Conversely, female gonad associate genes appear to be strongly downregulated in male gonad (**Sup. Fig. 2**). Subsequently we tested whether male gonad-associated genes connectivity in the bisexuals correlated with parthenogens LogFC: the correlation is found to be significative in both *B. atticus* (Spearman correlation: $r=0.187$, $p<0.05$) and *B. rossius*

(Spearman correlation: r=0.192, p<0.05) implying that highly connected male gonad-associated genes have retained gonad-specific expresion patterns in parthenogens.

Preservation of Male Gonad-associated Genes Sequence in Parthenogens - To infer the different branch metrics (dNdS, MILC and substitution rates) we relied on a phylogenetic approach, as our aim was to observe the different changes the genes underwent in the branches leading to parthenogenesis or mantaining bisexuality. In general, parthenogenetic species exhibit signs of a selection strength equal to the bisexual species: there is, in fact, no statistical difference in gene dNdS distribution between *B. rossius* and *B. grandii*, while *B. atticus* seem to have a generally lower dNdS than the other two species. Moreover, no difference in dNdS was retrieved comparing female or male gonad-associated genes across the different species (**Sup. Fig. 3**). When genes associated to male gonad modules of co-expression are compared to either the female gonad-associated ones or to all other network genes, the only statistically supported difference found is that - across all species - female gonad-associated genes have undergone a stronger selection regime compared to both male gonad-associated and gonad-unrelated genes (**Fig. 3A**; **Sup. Fig. 3**). No consistent difference in dNdS has been found comparing separately the genes associated to the different modules, suggesting that the different extant of coexpression patterns preservation is decoupled from selection regimes strength (**Sup. Fig. 4**). Codon usage biase did not appear to have undergone appreciable changes across most of the comparison we tested, possibly due to the low effective population size of *Bacillus* populations and the weak selection associated to them. Overall, these results show that the shift to parthenogenesis didn't cause any form of sequence degradation and relaxation of selection strength for male gonad-associated genes.

For all genes subsets, a significant negative correlation between connectivity (both total and extra-modular) and evolutionary rates (dN) is found (**Sup. Fig. 5**; with the exception of male gonad-associated genes in *B. rossius* - *p*=0.057). This pattern is due to the strong constraint on sequence evolution of genes which affect the expression of many others, with respect to those whose expression affects fewer. When female gonad-associated or gonad-unrelated genes are considered, a significant negative correlation can be observed between substitution rates and connectivity (both total and extra-modular) across all the three *Bacillus* spp, implying that highly connected genes underwent a slower evolutionary pace than less connected ones in both the bisexual species and in parthenogens (**Sup. Fig. 5**). When male gonad-associated genes are considered the correlation is significant only for the bisexual species *B. grandii*, while there is no statistical support for the two parthenogens *B. atticus* and *B. rossius* (**Sup. Fig. 5**). The same pattern is also found between dNdS and extra-modular connectivity (**Fig. 3)**: all gene subsets present a significative negative correlation between the two metrics, with the exception of male gonads-associated genes in

parthenogens. Interestingly, no consisten correlation was found between but dNdS and total connectivity, as for codon usage bias in general (**Sup. Fig. 5**).

Discussion:

Genes act as parts of integrated systems and their co-expression is responsible for complex phenotypes; as such, genes regulatory networks rather than single genes should be considered in respect to trait evolution (Hu et al., 2019; Almundi et al., 2020). When the gene network associated with a lost trait is considered, our results reveal a general preservation of its groundplan in both transcriptional program and sequence evolution. Co-expression modules associated with male gonad in the sexual species show different extent of preservation in parthenogens, some being partially degraded while others strongly preserved; from a sequence evolution perspective, we instead retrieved a consistent preservation among trait-related genes. These findings are coherent with trait loss being coupled mainly with transcriptional modifications (Roscito et al, 2018; Smith et al, 2015) and changes affecting few genes (Zhang and Reed, 2016; Esfeld et al, 2018; Ma et al, 2020; Yagound et al, 2020) - leaving intact a big portion of the trait blueprint (Smith et al, 2015; Stern and Crandall, 2018; Perez-Moreno et al, 2018; Bast et al., 2018; Lammers et al., 2019). In large part this preservation appears to be due to male gonad physiological pathways consisting of genes with pleiotropic effects in both sexes, as previously hypothesized (Kooi and Schwander, 2011; Schwander et al, 2013). Our findings are also compatible with the observed masculinization of sex-biased gene expression following the enstablishment of parthenogenesis (Veltsos et al, 2017; Parker et al, 2019); this shift of gene expression patterns towards male optima in parthenogenetic females has been proposed to stem from selection on female trait reduction and/or the production of male factors necessary to sustain female fertility. As such, the similar transcriptional changes across the two independent shifts to parthenogenesis could result from an adaptive process; yet, our results higlight how prior contrains (*i.e.* gene pleiotropy and essentiality) have also played a major role in the process (Losos et al, 2011).

Genes composing a co-expression network share different properties, with their level of pleiotropy being one of the most striking differences among them (Tyler et al, 2009; Wang et al, 2010). Here we show that the co-expression network topology constrains genes expression after trait loss: male gonad-associated genes whose expression is correlated with many others are most likely to retain gonad-biased transcription patterns in parthenogen, while loosely connected ones had their gonad-specific expression degraded. Highly connected genes are more likely under selection also for functions that are not related to the lost trait, in respect to loosely connected ones; moreover, theoretical and empirical observations predict a strong correlation between connectivity and essentiality, so that

degradation of the genes which are largely responsible for maintaining network connectivity could be detrimental or even lethal (Jeong et al, 2001; Carlson et al., 2006). Network position is also an important determinant of gene sequence evolution: our results remark how highly connected genes have a lower evolutionary rate than those located on the periphery of gene co-expression networks (Stern et al, 2008; Masalia et al, 2017; Mack et al, 2019). In the bisexual species we found a consistent correlation of gene extra-modular connectivity with both gene substitution rate and selection regime strength. For female gonad-associated and gonad-unrelated gene this correlation is present also in the two parthenogenetic lineages: this is expected, as highly connected genes unrelated to male gonads are most likely still interacting with the same targets after the shift to parthenogenesis. Instead, when genes related to the male gonad module of co-expression are considered in the lineages where males are absent, this correlation is not found. This pattern can be explained by their accommodation in a novel context - either due to novel interactions or by the loss of previous ones - and this may lead to subsequent evolutionary changes (Hunt et al, 2011).

A large body of literature focuses on how traits are established (Aguilera et al, 2017; Fisher et al, 2020; Almudi et al, 2020); nonetheless unraveling the processes associated to trait loss can also provide insights on the selective forces that underlie trait evolution. If the groundplan associated to a trait is largely preserved despite the trait beeing phenotypically absent, it could be potentially co-opted for novel purposes or explain reversions to the ancestral state. The latter has been considered an unlikely event (*e.g.* Dollo's law), as the loss was expected to be coupled with the decay of the trait molecular ground plan (Marshall et al, 1994). Yet, the results presented here challenge this concept and numerous examples of such reversals have been proposed, including flowers color (Esfeld et al, 2018), compound eyes in ostracods (Syme and Oakley, 2012), oviparity in lizards (Recknagel et al. 2018; Esquerré et al. 2020), insects flight (Whiting et al, 2003; Forni et al, 2020) and the sex itself (Domes et al, 2017). Functional males can infact occasionally appear in parthenogenetic species, as supported by empirical observations in other insects (Heethoff et al, 2009; Vershinina and Kuznetsova, 2016; Morgan-Richards et al, 2019). Moreover, no signature of a weakened selection is found in parthenogens - consistently with observations on other arthropods (Brandt et al, 2017) - sustaining the idea that parthenogenesis does not represent an evolutionary dead end. As for parthenogenesis, this work higlights how the molecular groundplan of complex traits can persis for long time despite their phenotipical absence and potentially be recruited for novel purposes or re-establish the former trait when specific selective pressures arise.

References:

Aguilera, F., McDougall, C. and Degnan, B.M., 2017. Co-option and de novo gene evolution underlie molluscan shell diversity. Molecular biology and evolution, 34(4), pp.779-792.

Alexa A, Rahnenfuhrer J (2019). topGO: Enrichment Analysis for Gene Ontology. R package version 2.38.1.

Almudi, I. and Pascual-Anaya, J., 2019. How Do Morphological Novelties Evolve? Novel Approaches to Define Novel Morphologies. In Old Questions and Young Approaches to Animal Evolution (pp. 107-132). Springer, Cham.

Almudi, I., Vizueta, J., Wyatt, C.D., de Mendoza, A., Marlétaz, F., Firbas, P.N., Feuda, R., Masiero, G., Medina, P., Alcaina-Caro, A. and Cruz, F., 2020. Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings. Nature communications, 11(1), pp.1-11

Alvarez-Ponce, D., Feyertag, F. and Chakraborty, S., 2017. Position matters: network centrality considerably impacts rates of protein evolution in the human protein–protein interaction network. *Genome biology and evolution*, *9*(6), pp.1742-1756.

Allen, J.D., Xie, Y., Chen, M., Girard, L. and Xiao, G., 2012. Comparing statistical methods for constructing large scale gene networks. *PloS one*, *7*(1), p.e29348.

Ament, S.A., Chan, Q.W., Wheeler, M.M., Nixon, S.E., Johnson, S.P., Rodriguez-Zas, S.L., Foster, L.J. and Robinson, G.E., 2011. Mechanisms of stable lipid loss in a social insect. *Journal of Experimental Biology*, *214*(22), pp.3808-3821.

Bast, J., Parker, D.J., Dumas, Z., Jalvingh, K.M., Tran Van, P., Jaron, K.S., Figuet, E., Brandt, A., Galtier, N. and Schwander, T., 2018. Consequences of asexuality in natural populations: insights from stick insects. Molecular biology and evolution, 35(7), pp.1668-1677.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), pp.289-300.

Brandt, A., Schaefer, I., Glanz, J., Schwander, T., Maraun, M., Scheu, S. and Bast, J., 2017. Effective purifying selection in ancient asexual oribatid mites. *Nature communications*, *8*(1), pp.1-9.

Carlini, D.B., Satish, S. and Fong, D.W., 2013. Parallel reduction in expression, but no loss of functional constraint, in two opsin paralogs within cave populations of Gammarus minus (Crustacea: Amphipoda). *BMC Evolutionary Biology*, *13*(1), p.89.

Carlson, M.R., Zhang, B., Fang, Z., Mischel, P.S., Horvath, S. and Nelson, S.F., 2006. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC genomics*, *7*(1), p.40.

Cayre, M., Strambi, C., Charpin, P., Augier, R. and Renucci, M., 1996. Inhibition of polyamine biosynthesis alters oviposition behavior in female crickets. *Behavioral neuroscience*, *110*(5), p.1117.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J. and Myers, R.M., 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *science*, *327*(5963), pp.302-305.

Đaković, N., Térézol, M., Pitel, F., Maillard, V., Elis, S., Leroux, S., Lagarrigue, S., Gondret, F., Klopp, C., Baeza, E. and Duclos, M.J., 2014. The loss of adipokine genes in the chicken genome and implications for insulin metabolism. *Molecular biology and evolution*, *31*(10), pp.2637-2646.

Domes, K., Norton, R.A., Maraun, M. and Scheu, S., 2007. Reevolution of sexuality breaks Dollo's law. *Proceedings of the National Academy of Sciences*, *104*(17), pp.7139-7144.

Elek, A., Kuzman, M. and Vlahoviček, K., 2019. coRdon: codon usage analysis and prediction of gene expressivity. Bioconductor 3.8.

Emms, D.M. and Kelly, S., 2018. STAG: species tree inference from all genes. *BioRxiv*, p.267914.

Emms, D.M. and Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, *20*(1), pp.1-14.

Erwin, D.H. and Davidson, E.H., 2009. The evolution of hierarchical gene regulatory networks. *Nature Reviews Genetics*, *10*(2), pp.141-148.

Esfeld, K., Berardi, A.E., Moser, M., Bossolini, E., Freitas, L. and Kuhlemeier, C., 2018. Pseudogenization and resurrection of a speciation gene. *Current Biology*, *28*(23), pp.3776-3786.

Esquerré, D., Brennan, I.G., Catullo, R.A., Torres‐Pérez, F. and Keogh, J.S., 2019. How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America's most species‐rich lizard radiation (Squamata: Liolaemidae). *Evolution*, *73*(2), pp.214-230.

Fisher, C.R., Wegrzyn, J.L. and Jockusch, E.L., 2020. Co-option of wing-patterning genes underlies the evolution of the treehopper helmet. Nature Ecology & Evolution, 4(2), pp.250-260.

Forni, G., Ruggeri, A.A., Piccinini, G. and Luchetti, A., 2020. BASE: a novel workflow to integrate non-ubiquitous genes in genomics analyses for selection. *bioRxiv*.

Forni, G., Martelossi, J., Valero, P., Hennemann, F.H., Conle, O., Luchetti, A. and Mantovani, B., 2020. Macroevolutionary Analyses Provide New Evidences of Phasmids Wings Evolution as a Reversible Process. *bioRxiv*.

Fraïsse, C., Puixeu Sala, G. and Vicoso, B., 2019. Pleiotropy modulates the efficacy of selection in Drosophila melanogaster. *Molecular biology and evolution*, *36*(3), pp.500-515.

Gillott, C., 2003. Male accessory gland secretions: modulators of female reproductive physiology and behavior. Annual review of entomology, 48(1), pp.163-184.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. and MacManes, M.D., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, *8*(8), pp.1494-1512.

Hall, A.R. and Colegrave, N., 2008. Decay of unused characters by selection and drift. *Journal of Evolutionary Biology*, *21*(2), pp.610-617.

He, X. and Zhang, J., 2006. Toward a molecular understanding of pleiotropy. *Genetics*, *173*(4), pp.1885-1891.

Heethoff, M., Norton, R.A., Scheu, S. and Maraun, M., 2009. Parthenogenesis in oribatid mites (Acari, Oribatida): evolution without sex. In *Lost Sex* (pp. 241-257). Springer, Dordrecht.

Ho, W.W. and Smith, S.D., 2016. Molecular evolution of anthocyanin pigmentation genes following losses of flower color. *BMC evolutionary biology*, *16*(1), pp.1-10.

Horvath, S. and Dong, J., 2008. Geometric interpretation of gene coexpression network analysis. *PLoS comput biol*, *4*(8), p.e1000117.

Hu, Yonggang, David M. Linz, and Armin P. Moczek. 2019. Beetle Horns Evolved from Wing Serial Homologs. Science 366 (6468): 1004–7.

Hunt, B.G., Ometto, L., Wurm, Y., Shoemaker, D., Soojin, V.Y., Keller, L. and Goodisman, M.A., 2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity. Proceedings of the National Academy of Sciences, 108(38), pp.15936-15941.

Jaron, K.S., Bast, J., Nowell, R.W., Ranallo-Benavidez, T.R., Robinson-Rechavi, M. and Schwander, T., 2019. Genomic features of asexual animals. BioRxiv, p.497495.

Jeong, H., Mason, S., Barabási, A. *et al.* Lethality and centrality in protein networks. *Nature* 411, 41–42 (2001). https://doi.org/10.1038/35075138

Kassambara, A. and Kassambara, M.A., 2020. Package 'ggpubr'.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), pp.772-780.

Kogan, P. H. & Hagedorn, H. H. Polyamines, and effects from reducing their 641 synthesis during egg development in the yellow fever mosquito, Aedes 642 aegypti. J Insect Physiol 46, 1079–1095. (2000).

Kohlsdorf, T. and Wagner, G.P., 2006. Evidence for the reversibility of digit loss: a phylogenetic study of limb evolution in Bachia (Gymnophthalmidae: Squamata). *Evolution*, *60*(9), pp.1896-1912.

Kraaijeveld, K., Anvar, S.Y., Frank, J., Schmitz, A., Bast, J., Wilbrandt, J., Petersen, M., Ziesmann, T., Niehuis, O., De Knijff, P. and Den Dunnen, J.T., 2016. Decay of sexual trait genes in an asexual parasitoid wasp. Genome biology and evolution, 8(12), pp.3685-3695.

Lammers, M., Kraaijeveld, K., Mariën, J. and Ellers, J., 2019. Gene expression changes associated with the evolutionary loss of a metabolic trait: lack of lipogenesis in parasitoids. *BMC genomics*, *20*(1), pp.1-14.

Langfelder, P. and Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, *9*(1), p.559.

Langfelder, P., Luo, R., Oldham, M.C. and Horvath, S., 2011. Is my network module preserved and reproducible?. *PLoS Comput Biol*, *7*(1), p.e1001057.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), p.357.

Lavezzo, E., Falda, M., Fontana, P., Bianco, L. and Toppo, S., 2016. Enhancing protein function prediction with taxonomic constraints–The Argot2. 5 web server. *Methods*, *93*, pp.15-23.

Leal, F. and Cohn, M.J., 2016. Loss and Re-emergence of Legs in Snakes by Modular Evolution of Sonic hedgehog and HOXD Enhancers. *Current Biology*, *26*(21), pp.2966-2973.

Li, B. and Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, *12*(1), pp.1-16.

Losos, J.B., 2011. Convergence, adaptation, and constraint. *Evolution: International Journal of Organic Evolution*, *65*(7), pp.1827-1840.

Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), p.550.

Ma, W.J., Pannebakker, B.A., Li, X., Geuverink, E., Anvar, S.Y., Veltsos, P., Schwander, T., van de Zande, L. and Beukeboom, L.W., 2020. A single QTL with large effect is associated with female functional virginity in an asexual parasitoid wasp. *bioRxiv*.

Mack, K.L., Phifer-Rixey, M., Harr, B. and Nachman, M.W., 2019. Gene expression networks across multiple tissues are associated with rates of molecular evolution in wild house mice. *Genes*, *10*(3), p.225.

MacNeil, L.T. and Walhout, A.J., 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome research*, *21*(5), pp.645-657.

Mantovani, B., Passamonti, M. and Scali, V., 2001. The mitochondrial cytochrome oxidase II gene in Bacillus stick insects: ancestry of hybrids, androgenesis, and phylogenetic relationships. *Molecular phylogenetics and evolution*, *19*(1), pp.157-163.

Marshall, C.R., Raff, E.C. and Raff, R.A., 1994. Dollo's law and the death and resurrection of genes. *Proceedings of the National Academy of Sciences*, *91*(25), pp.12283-12287.

Masalia, R.R., Bewick, A.J. and Burke, J.M., 2017. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. PLoS One, 12(7), p.e0182289.

McNeal, J.R., Kuehl, J.V., Boore, J.L. and de Pamphilis, C.W., 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus Cuscuta. BMC Plant Biology, 7(1), p.57.

Meredith, R.W., Gatesy, J., Murphy, W.J., Ryder, O.A. and Springer, M.S., 2009. Molecular decay of the tooth gene enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genet*, *5*(9), p.e1000634.

Monteiro, A., 2012. Gene regulatory networks reused to build novel traits: Co-option of an eye-related gene regulatory network in eye-like organs and red wing patches on insect wings is suggested by optix expression. BioEssays, 34(3), pp.181-186.

Morandin, C., Mikheyev, A.S., Pedersen, J.S. and Helanterä, H., 2017. Evolutionary constraints shape caste-specific gene expression across 15 ant species. *Evolution*, *71*(5), pp.1273-1284.

Morgan-Richards, M., Trewick, S.A. and Stringer, I.A., 2010. Geographic parthenogenesis and the common tea-tree stick insect of New Zealand. *Molecular ecology*, *19*(6), pp.1227-1238.

Morgan-Richards, M., Langton-Myers, S.S. and Trewick, S.A., 2019. Loss and gain of sexual reproduction in the same stick insect. *Molecular ecology*, *28*(17), pp.3929-3941.

Paaby, A.B. and Rockman, M.V., 2013. The many faces of pleiotropy. *Trends in Genetics*, *29*(2), pp.66-73.

Papakostas, S., Vøllestad, L.A., Bruneaux, M., Aykanat, T., Vanoverbeke, J., Ning, M., Primmer, C.R. and Leder, E.H., 2014. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nature Communications*, *5*(1), pp.1-9.

Parker, D.J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M. and Schwander, T., 2019. Repeated evolution of asexuality involves convergent gene expression changes. *Molecular biology and evolution*, *36*(2), pp.350-364.

Parker, D.J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M. and Schwander, T., 2019. Sex-biased gene expression is repeatedly masculinized in asexual females. *Nature communications*, *10*(1), pp.1-11.

Pérez-Moreno, J.L., Balázs, G. and Bracken-Grissom, H.D., 2018. Transcriptomic insights into the loss of vision in Molnár János Cave's crustaceans. *Integrative and Comparative Biology*, *58*(3), pp.452-464.

Porter, M.L. and Crandall, K.A., 2003. Lost along the way: the significance of evolution in reverse. *Trends in Ecology & Evolution*, *18*(10), pp.541-547.

Protas, M.E., Trontelj, P. and Patel, N.H., 2011. Genetic basis of eye and pigment loss in the cave crustacean, Asellus aquaticus. *Proceedings of the National Academy of Sciences*, *108*(14), pp.5702-5707.

R. Core Team. (R: A Language and Environment for Statistical Computing, 2017).

Recknagel, H., Kamenos, N.A. and Elmer, K.R., 2018. Common lizards break Dollo's law of irreversibility: genome-wide phylogenomics support a single origin of viviparity and re-evolution of oviparity. *Molecular Phylogenetics and Evolution*, *127*, pp.579-588.

Roscito, J.G., Sameith, K., Parra, G., Langer, B.E., Petzold, A., Moebius, C., Bickle, M., Rodrigues, M.T. and Hiller, M., 2018. Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. *Nature communications*, *9*(1), pp.1-15.

Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J.S., Wheeler, N.E., Gardner, P.P., Clarke, J.A., Baker, A.J., Clamp, M. and Edwards, S.V., 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science*, *364*(6435), pp.74-78.

Salathé, M., Ackermann, M. and Bonhoeffer, S., 2005. The effect of multifunctionality on the rate of evolution in yeast. *Molecular biology and evolution*, *23*(4), pp.721-722.

Scali, V., Passamonti, M., Marescalchi, O. and Mantovani, B., 2003. Linkage between sexual and asexual lineages: genome evolution in Bacillus stick insects. *Biological Journal of the Linnean Society*, *79*(1), pp.137-150.

Schurko, A.M., Neiman, M. and Logsdon Jr, J.M., 2009. Signs of sex: what we know and how we know it. *Trends in ecology & evolution*, *24*(4), pp.208-217.

Schwander, T., Crespi, B.J., Gries, R. and Gries, G., 2013. Neutral and selection-driven decay of sexual traits in asexual stick insects. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1764), p.20130823.

Simon, S., Letsch, H., Bank, S., Buckley, T.R., Donath, A., Liu, S., Machida, R., Meusemann, K., Misof, B., Podsiadlowski, L. and Zhou, X., 2019. Old World and New World Phasmatodea: phylogenomics resolve the evolutionary history of stick and leaf insects. *Frontiers in Ecology and Evolution*, *7*, p.345.

Smith, C.R., Helms Cahan, S., Kemena, C., Brady, S.G., Yang, W., Bornberg-Bauer, E., Eriksson, T., Gadau, J., Helmkampf, M., Gotzek, D., Okamoto Miyakawa, M., Suarez, A.V., Mikheyev A., 2015. How do genomes create novel phenotypes? Insights from the loss of the worker caste in ant social parasites. Molecular biology and evolution, 32(11), pp.2919-2931.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), pp.1312-1313.

Stern, D.B. and Crandall, K.A., 2018. The evolution of gene expression underlying vision loss in cave animals. *Molecular biology and evolution*, *35*(8), pp.2005-2014.

Stern, D.L. and Orgogozo, V., 2008. The loci of evolution: how predictable is genetic evolution? *Evolution: International Journal of Organic Evolution*, *62*(9), pp.2155-2177.

Stuart, J.M., Segal, E., Koller, D. and Kim, S.K., 2003. A gene-coexpression network for global discovery of conserved genetic modules. *science*, *302*(5643), pp.249-255.

Suen, G., Teiling, C., Li, L., Holt, C., Abouheif, E., Bornberg-Bauer, E., Bouffard, P., Caldera, E.J., Cash, E., Cavanaugh, A. and Denas, O., 2011. The genome sequence of the leaf-cutter ant Atta cephalotes reveals insights into its obligate symbiotic lifestyle. *PLoS Genet*, *7*(2), p.e1002007.

Suyama, M., Torrents, D. and Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, *34*(suppl_2), pp.W609-W612.

Syme, A.E. and Oakley, T.H., 2012. Dispersal between shallow and abyssal seas and evolutionary loss and regain of compound eyes in cylindroleberidid ostracods: conflicting conclusions from different comparative methods. *Systematic Biology*, *61*(2), p.314.

Tyler, A.L., Asselbergs, F.W., Williams, S.M. and Moore, J.H., 2009. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *Bioessays*, *31*(2), pp.220-227.

van der Kooi, C.J. and Schwander, T., 2014. On the fate of sexual traits under asexuality. *Biological Reviews*, *89*(4), pp.805-819.

Veltsos, P., Fang, Y., Cossins, A.R., Snook, R.R. and Ritchie, M.G., 2017. Mating system manipulation and the evolution of sex-biased gene expression in Drosophila. *Nature communications*, *8*(1), pp.1-11.

Vershinina, A.O. and Kuznetsova, V.G., 2016. Parthenogenesis in Hexapoda: Entognatha and non-holometabolous insects. *Journal of Zoological Systematics and Evolutionary Research*, *54*(4), pp.257-268.

Visscher, P.M. and Yang, J., 2016. A plethora of pleiotropy across complex traits. *Nature genetics*, *48*(7), pp.707-708.

Wang, Z., Liao, B.Y. and Zhang, J., 2010. Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Sciences*, *107*(42), pp.18034-18039.

Went, D.F., 1982. Egg activation and parthenogenetic reproduction in insects. *Biological Reviews*, *57*(2), pp.319-344.

Whiting, M.F., Bradler, S. and Maxwell, T., 2003. Loss and recovery of wings in stick insects. *Nature*, *421*(6920), pp.264-267.

Wyatt, G.R., Rothaus, K., Lawler, D., Herbst, E.J. Ornithine decarboxylase and polyamines in silkmoth pupal tissues: effect of ecdysone and injury. Biochim. Biophys. Acta. 304, 482–494 (1973).

Yagound, B., Dogantzis, K.A., Zayed, A., Lim, J., Broekhuyse, P., Remnant, E.J., Beekman, M., Allsopp, M.H., Aamidor, S.E., Dim, O. and Buchmann, G., 2020. A Single Gene Causes Thelytokous Parthenogenesis, the Defining Feature of the Cape Honeybee Apis mellifera capensis. Current Biology.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, *24*(8), pp.1586-1591.

Zhang, B. and Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, *4*(1).

Zhang, L. and Reed, R.D., 2016. Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns. Nature communications, 7(1), pp.1-7.

Zhao, W., Langfelder, P., Fuller, T., Dong, J., Li, A. and Hovarth, S., 2010. Weighted gene coexpression network analysis: state of the art. *Journal of biopharmaceutical statistics*, *20*(2), pp.281-300.

Figure Captions:

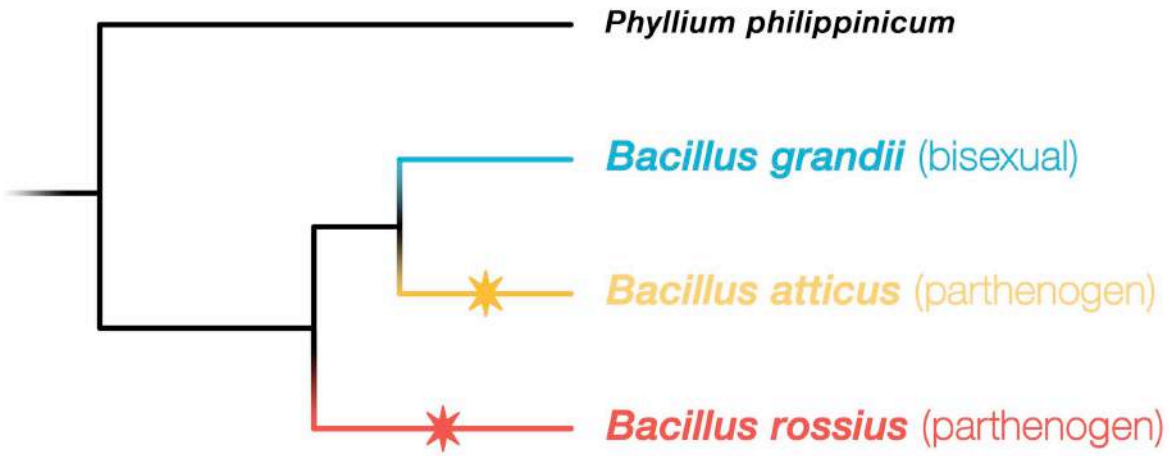


Figure 1 - ***Bacillus*** **phylogeny:** phylogeny of the *Bacillus* species analyzed in this paper. Parthenogens result from two independent shifts in reproductive strategies (highlighted with asterisk) and perform automixis (*i.e.* meiotic parthenogenesis): *Bacillus grandii* is an obligate bisexual lineage, while *Bacillus atticus* is an obligate parthenogen (which carries out central fusion) and *Bacillus rossius* is a geographical parthenogen (which performs post-meiotic doubling). For both parthenogens no male has been reported in over ten years, neither in captive breeding or in the wild.

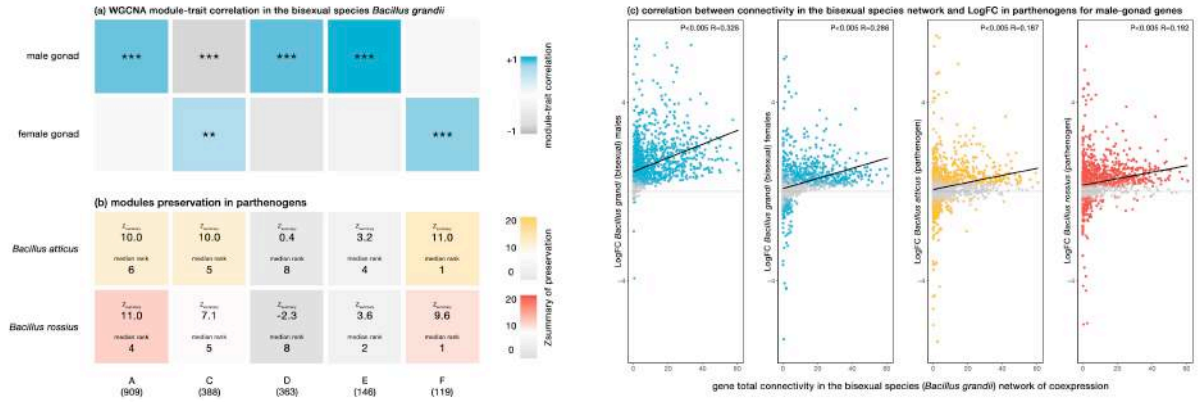


Figure 2 - Expression preservation of male gonad-associated genes in parthenogens: colors used reflect species coloring as in Figure 1; (A) co-expression modules correlation to male and female gonad in the bisexual species *Bacillus grandii*. Each column in the table corresponds to a module while the upper a lower rows correspond to male and female gonad. Tiles colour intensity is promportional to module-trait correlations, with relative *p* values expressed as asterisks (*** p<0.0005; ** p<0.005; * p<0.05; ns > 0.05). (B) gonad-associated modules preservation in the two parthenogens *Bacillus atticus* and *Bacillus rossius* gonad; numbers in the tiles are Zsummary and medianRank, respectively at the top and bottom; for Zsummary, the higher the metric, the more preserved the module; for medianRank, the lower the metric, the more preserved the module; (C) male gonad genes Spearman correlations between gene total connectivity in the bisexual *Bacillus grandii* co-expression network and the LogFC found between reproductive and non-reproductive tissues for each *Bacillus* species and sex; points colour is based on differential expression FDR: >0.05=gray; <0.05=coloured.

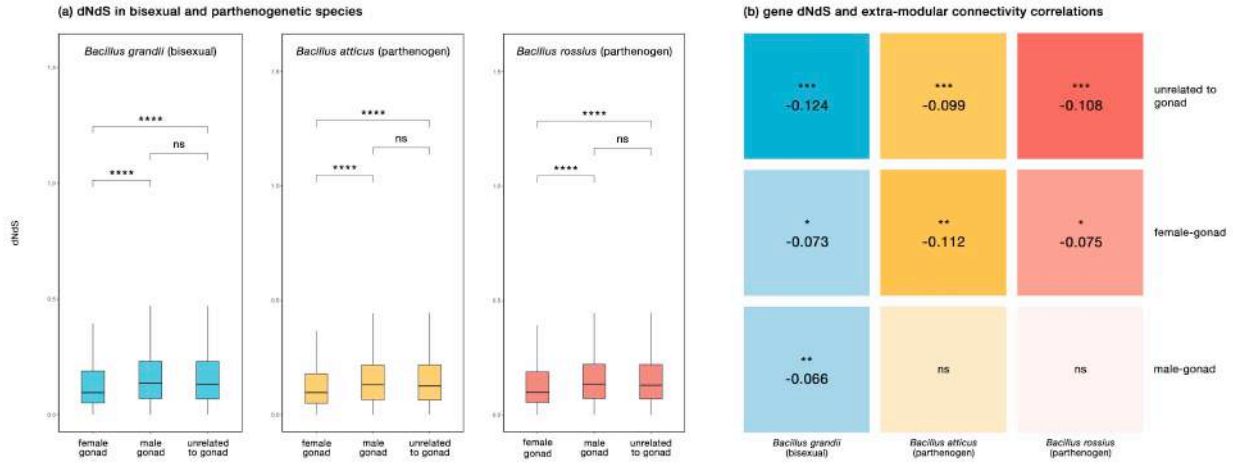
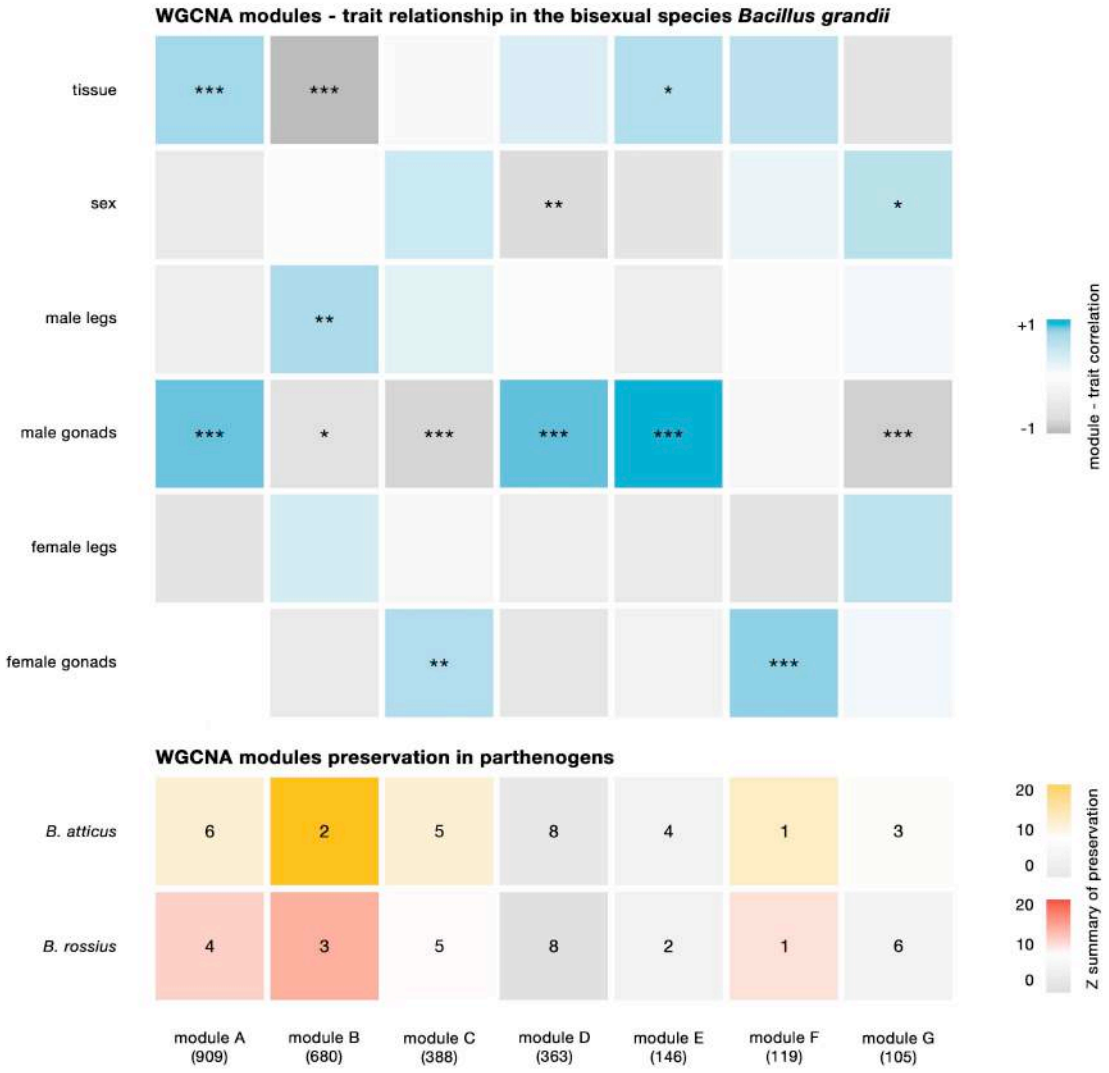
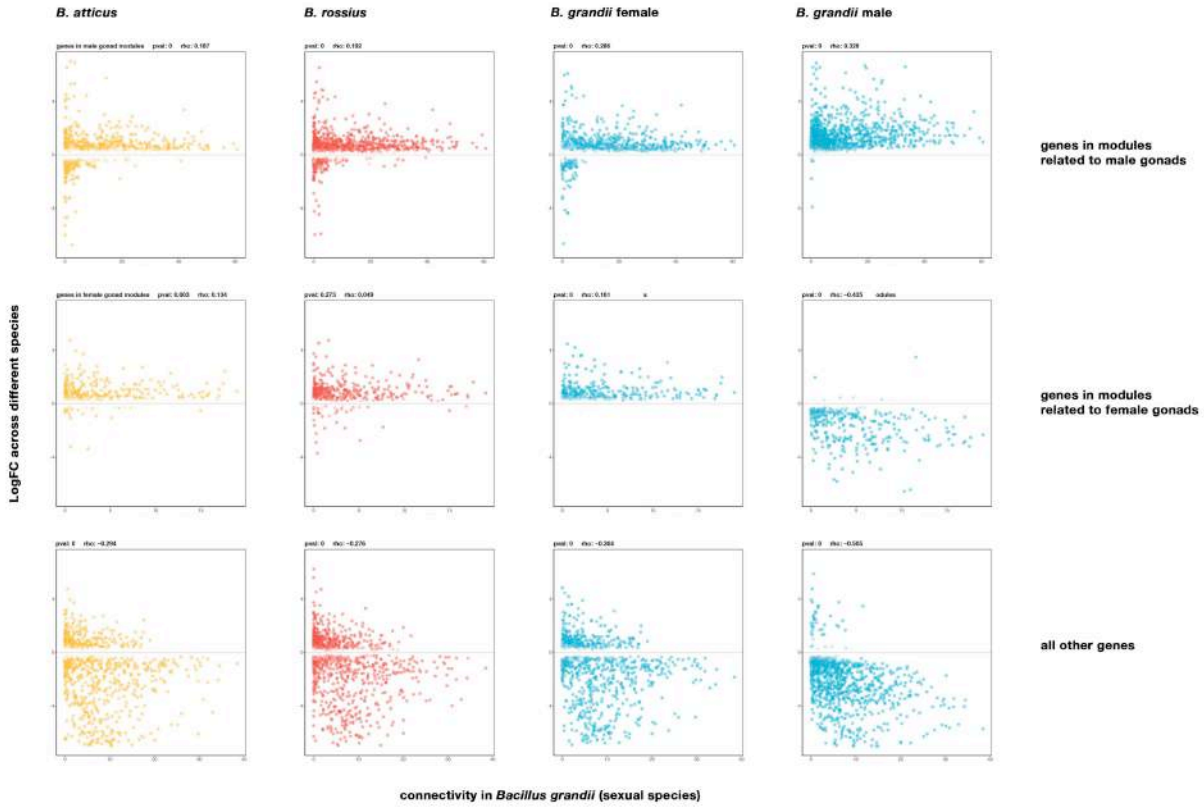


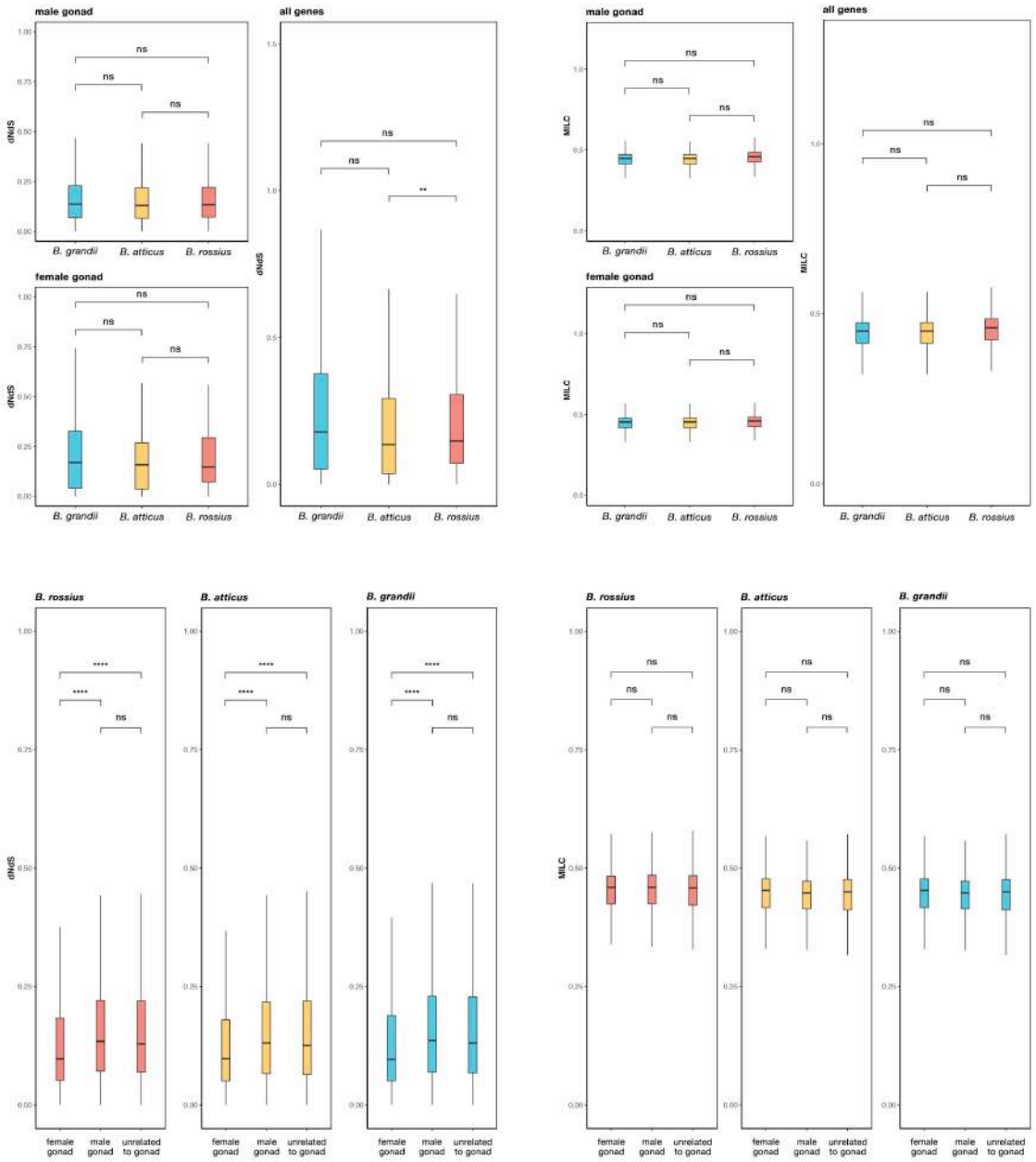
Figure 3 - Sequence preservation of male gonad-associated genes in parthenogens: colors used reflect species coloring as in Figure 1; (A) comparison of dNdS ratios inferred in the terminal branches leading to the three *Bacillus* species for male gonad-associated, female gonad-associated and gonad-unrelated genes in the bisexual species *Bacillus grandii*; statistical comparisons were carried out using Wilcoxon tests and the resulting *p* values are summarized using asterisks (*** p<0.0005; ** p<0.005; * p<0.05; ns > 0.05); (B) Spearman correlations between gene extra-modular connectivity in *Bacillus grandii* network of co-expression and dNdS in each *Bacillus* terminal branch, for male gonad-associated, female gonad-associated or gonad-unrelated genes in *Bacillus grandii*; in each tile asterisks summarize *p* values while numbers inside tiles represent Spearman *r* values; tile colour intensity is proportional to *r* values.



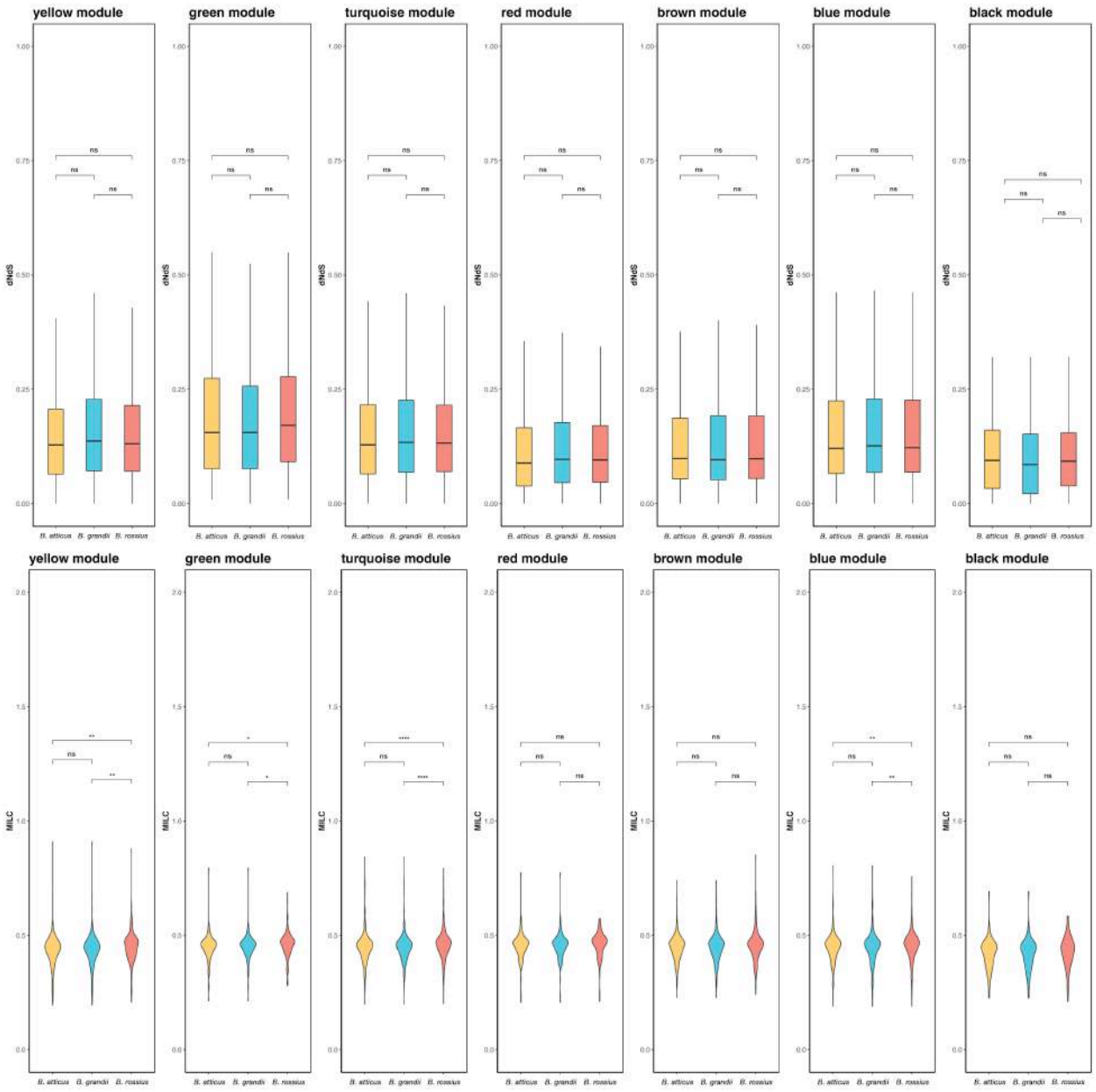
Suppl. Figure 1: (A) co-expression modules in the bisexual species *Bacillus grandii* and their correlation with correlation with different traits. Tiles colour intensity is promportional to module-trait correlations, with relative *p* values expressed as asterisks. (B) gonad-associated modules preservation in the two parthenogens *Bacillus atticus* and *Bacillus rossius* gonad; numbers in the tiles are medianRank valeues, while tiles color intensity is proportional to Zsummary.



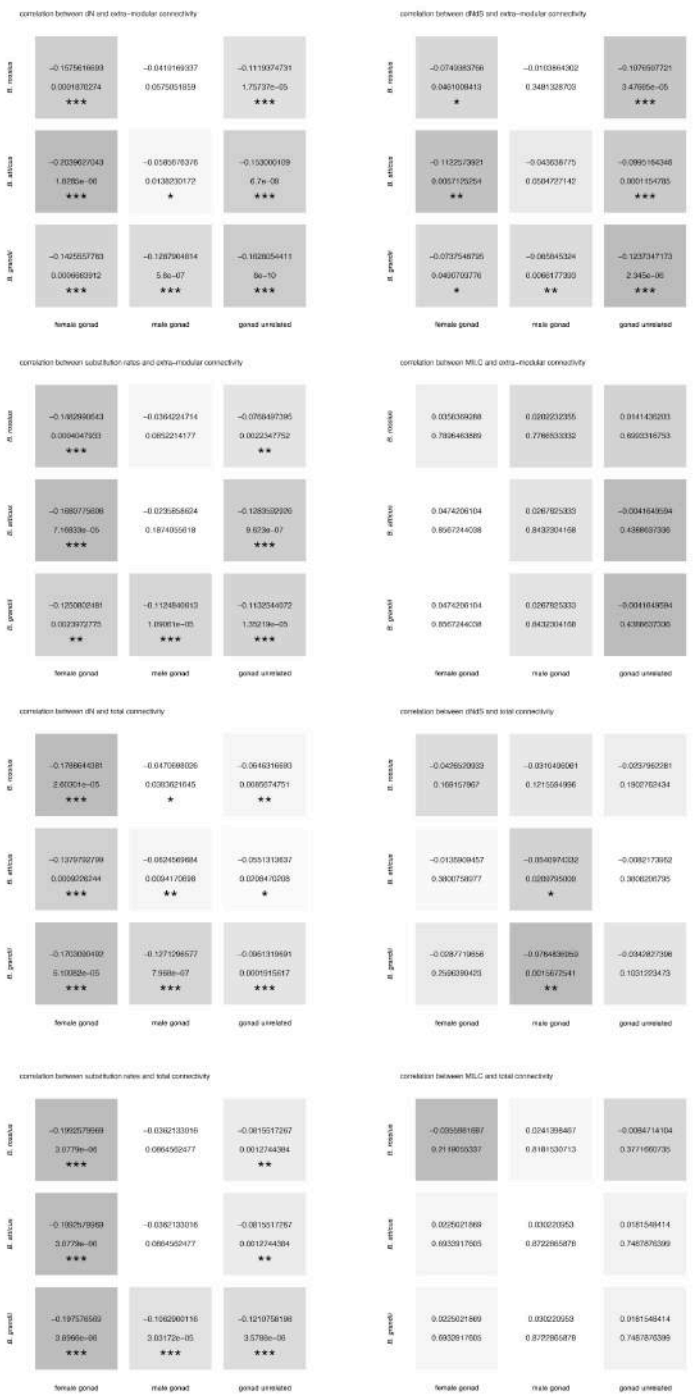
Suppl. Figure 2 - Spearman correlations between gene total connectivity in the bisexual *Bacillus grandii* co-expression network and the LogFC found between reproductive and non-reproductive tissues for each *Bacillus* species and sex; points colour is based on differential expression FDR: >0.05=gray and <0.05=coloured



Suppl. Figure 3: comparison of selection strength metrics inferred in the terminal branches leading to the three *Bacillus* species for male gonad-associated, female gonad-associated and gonad-unrelated genes in the bisexual species *Bacillus grandii*; statistical comparisons were carried out using Wilcoxon tests and the resulting *p* values are summarized using asterisks.



Suppl. Figure 4: comparison of selection strength metrics inferred in the terminal branches leading to the three *Bacillus* species for genes associated to each co-expression module found in *Bacillus grandii* network; statistical comparisons were carried out using Wilcoxon tests and the resulting *p* values are summarized using asterisks.



Suppl. Figure 5: genes correlations between gene connectivity (total and extra-modular) and each *Bacillus* terminal branch metrics (dN, substitution rates, dNdS and MILC) for for male gonad-associated, female gonad-associated and gonad-unrelated genes in the bisexual species *Bacillus grandii*; in each tile asterisks summarize *p* values and numbers are *r* values.

Black module

GO.ID	Term	Annotated	Significant	Expected	Elim_Fisher	Weight_Fisher
GO:0001731	formation of translation preinitiation c...	10	4	0.38	0.00035	0.00035
GO:0006122	mitochondrial electron transport ubiqui…	5	3	0.19	0.00050	0.00050
GO:0002182	cytoplasmic translational elongation	11	4	0.42	0.00053	100.000
GO:0046395	carboxylic acid catabolic process	40	7	1.52	0.00062	100.000
GO:0000028	ribosomal small subunit assembly	6	3	0.23	0.00098	0.00098
GO:0006633	fatty acid biosynthetic process	26	5	0.99	0.00249	0.01125
GO:0009206	purine ribonucleoside triphosphate biosy...	8	3	0.3	0.00259	0.03734
GO:1990542	mitochondrial transmembrane transport	17	4	0.65	0.00319	0.00319
GO:0006627	protein processing involved in protein t...	3	2	0.11	0.00418	0.00418
GO:0042254	ribosome biogenesis	170	16	6.45	0.00641	0.02174
GO:0006438	valyl-tRNA aminoacylation	4	2	0.15	0.00815	0.00815
GO:1990145	maintenance of translational fidelity	4	2	0.15	0.00815	0.00815
GO:2000766	negative regulation of cytoplasmic trans...	4	2	0.15	0.00815	100.000
GO:0003333	amino acid transmembrane transport	22	4	0.84	0.00846	0.00846
GO:0015807	L-amino acid transport	23	4	0.87	0.00995	0.20314
GO:0015980	energy derivation by oxidation of organi...	57	9	2.16	0.01124	0.35360
GO:0033539	fatty acid beta-oxidation using acyl-CoA...	5	2	0.19	0.01325	0.01325
GO:0045900	negative regulation of translational elo...	5	2	0.19	0.01325	0.01325
GO:0044242	cellular lipid catabolic process	25	4	0.95	0.01341	100.000
GO:1901606	alpha-amino acid catabolic process	25	4	0.95	0.01341	100.000
GO:0006091	generation of precursor metabolites and ...	110	12	4.18	0.01408	100.000
GO:0034440	lipid oxidation	15	3	0.57	0.01735	100.000
GO:0019395	fatty acid oxidation	15	3	0.57	0.01735	0.31672
GO:0055114	oxidation-reduction process	211	17	8.01	0.01742	0.55368
GO:0009060	aerobic respiration	27	4	1.03	0.01756	0.01756
GO:0042776	mitochondrial ATP synthesis coupled prot...	6	2	0.23	0.01938	100.000
GO:0018206	peptidyl-methionine modification	6	2	0.23	0.01938	0.01938
GO:0009063	cellular amino acid catabolic process	28	4	1.06	0.01990	100.000
GO:1902600	proton transmembrane transport	16	3	0.61	0.02077	0.29034
GO:0009062	fatty acid catabolic process	16	3	0.61	0.02077	0.31682
GO:1901605	alpha-amino acid metabolic process	59	6	2.24	0.02284	0.54486
GO:1902475	L-alpha-amino acid transmembrane transpo...	17	3	0.65	0.02454	100.000
GO:0006754	ATP biosynthetic process	7	2	0.27	0.02647	100.000
GO:0006749	glutathione metabolic process	7	2	0.27	0.02647	0.02647
GO:0015985	energy coupled proton transport down el…	7	2	0.27	0.02647	100.000
GO:0015986	ATP synthesis coupled proton transport	7	2	0.27	0.02647	100.000
GO:0002183	cytoplasmic translational initiation	42	8	1.59	0.02776	100.000
GO:0006119	oxidative phosphorylation	24	6	0.91	0.03084	100.000
GO:0015813	L-glutamate transmembrane transport	8	2	0.3	0.03443	100.000
GO:0015909	long-chain fatty acid transport	8	2	0.3	0.03443	0.03443
GO:0006448	regulation of translational elongation	8	2	0.3	0.03443	100.000

GO:0006120	mitochondrial electron transport NADH t…	8	2	0.3	0.03443	0.03443
GO:0009152	purine ribonucleotide biosynthetic proce...	33	4	1.25	0.03443	0.25310
GO:0008202	steroid metabolic process	49	5	1.86	0.03630	0.14975
GO:0016042	lipid catabolic process	34	4	1.29	0.03792	100.000
GO:0009150	purine ribonucleotide metabolic process	50	5	1.9	0.03916	100.000
GO:0006164	purine nucleotide biosynthetic process	35	4	1.33	0.04160	100.000
GO:0072329	monocarboxylic acid catabolic process	21	3	0.8	0.04303	100.000
GO:0046128	purine ribonucleoside metabolic process	9	2	0.34	0.04319	100.000
GO:1990822	basic amino acid transmembrane transport	9	2	0.34	0.04319	100.000
GO:1901570	fatty acid derivative biosynthetic proce...	9	2	0.34	0.04319	0.23571
GO:0015908	fatty acid transport	9	2	0.34	0.04319	100.000
GO:0006520	cellular amino acid metabolic process	92	9	3.49	0.04363	100.000
GO:1901566	organonitrogen compound biosynthetic pro...	375	43	14.24	0.04369	0.60044
GO:0009260	ribonucleotide biosynthetic process	36	4	1.37	0.04548	100.000
GO:0072522	purine-containing compound biosynthetic ...	36	4	1.37	0.04548	100.000

Blue Module

GO.ID	Term	Annotated	Significant	Expected	Elim_Fisher	Weight_Fisher
GO:0006732	coenzyme metabolic process	21	11	4.4	0.0013	100.000
GO:0009225	nucleotide-sugar metabolic process	8	6	1.68	0.0015	0.00153
GO:0006040	amino sugar metabolic process	4	4	0.84	0.0019	0.00189
GO:0006022	aminoglycan metabolic process	6	5	1.26	0.0019	0.00194
GO:0098542	defense response to other organism	7	5	1.47	0.0056	0.03027
GO:0034032	purine nucleoside bisphosphate metabolic...	10	6	2.1	0.0078	100.000
GO:0048589	developmental growth	10	6	2.1	0.0078	0.01943
GO:0033875	ribonucleoside bisphosphate metabolic pr...	10	6	2.1	0.0078	100.000
GO:0042558	pteridine-containing compound metabolic ...	5	4	1.05	0.0079	0.00790
GO:0071774	response to fibroblast growth factor	3	3	0.63	0.0091	0.00913
GO:0090101	negative regulation of transmembrane rec...	3	3	0.63	0.0091	0.00913
GO:0090288	negative regulation of cellular response...	3	3	0.63	0.0091	0.00913
GO:0060401	cytosolic calcium ion transport	3	3	0.63	0.0091	0.00913
GO:0042402	cellular biogenic amine catabolic proces...	3	3	0.63	0.0091	0.00913
GO:0046364	monosaccharide biosynthetic process	3	3	0.63	0.0091	100.000
GO:0001525	angiogenesis	3	3	0.63	0.0091	100.000
GO:0046879	hormone secretion	3	3	0.63	0.0091	100.000
GO:0001676	long-chain fatty acid metabolic process	3	3	0.63	0.0091	0.00913
GO:0050877	nervous system process	8	5	1.68	0.0125	100.000
GO:1901617	organic hydroxy compound biosynthetic pr...	8	5	1.68	0.0125	0.19235
GO:0098660	inorganic ion transmembrane transport	11	6	2.31	0.0143	100.000
GO:0098662	inorganic cation transmembrane transport	11	6	2.31	0.0143	0.01427
GO:0003008	system process	14	7	2.94	0.0145	100.000
GO:0017144	drug metabolic process	34	13	7.13	0.0146	0.02163

GO:1901605	alpha-amino acid metabolic process	24	10	5.03	0.0164	100.000
GO:0034220	ion transmembrane transport	21	9	4.4	0.0185	0.27816
GO:0022603	regulation of anatomical structure morph...	15	7	3.15	0.0224	0.60339
GO:1901606	alpha-amino acid catabolic process	15	7	3.15	0.0224	0.02238
GO:0009259	ribonucleotide metabolic process	25	10	5.24	0.0224	0.03198
GO:0043648	dicarboxylic acid metabolic process	12	6	2.52	0.0236	0.02362
GO:0098655	cation transmembrane transport	12	6	2.52	0.0236	100.000
GO:0016054	organic acid catabolic process	22	9	4.61	0.0257	100.000
GO:0046395	carboxylic acid catabolic process	22	9	4.61	0.0257	100.000
GO:0006811	ion transport	36	15	7.55	0.0261	0.35207
GO:0055085	transmembrane transport	33	12	6.92	0.0284	100.000
GO:0019693	ribose phosphate metabolic process	26	10	5.45	0.0299	100.000
GO:0046173	polyol biosynthetic process	4	3	0.84	0.0308	0.03083
GO:0030203	glycosaminoglycan metabolic process	4	3	0.84	0.0308	100.000
GO:0035966	response to topologically incorrect prot...	4	3	0.84	0.0308	100.000
GO:0035967	cellular response to topologically incor...	4	3	0.84	0.0308	0.20841
GO:0009150	purine ribonucleotide metabolic process	23	9	4.82	0.0347	0.49977
GO:0006163	purine nucleotide metabolic process	23	9	4.82	0.0347	0.37001
GO:0045333	cellular respiration	7	4	1.47	0.0387	0.19357
GO:0006997	nucleus organization	7	4	1.47	0.0387	0.03871
GO:0006575	cellular modified amino acid metabolic p...	7	4	1.47	0.0387	0.03871
GO:0035383	thioester metabolic process	7	4	1.47	0.0387	0.03068
GO:0006637	acyl-CoA metabolic process	7	4	1.47	0.0387	100.000
GO:0022900	electron transport chain	7	4	1.47	0.0387	0.19357
GO:0071495	cellular response to endogenous stimulus	20	8	4.2	0.0401	100.000
GO:0001568	blood vessel development	5	5	1.05	0.0429	0.04287
GO:0051480	regulation of cytosolic calcium ion conc...	2	2	0.42	0.0438	100.000
GO:0046503	glycerolipid catabolic process	2	2	0.42	0.0438	0.04384
GO:0060840	artery development	2	2	0.42	0.0438	100.000
GO:0070189	kynurenine metabolic process	2	2	0.42	0.0438	0.04384
GO:1901342	regulation of vasculature development	2	2	0.42	0.0438	100.000
GO:1901343	negative regulation of vasculature devel...	2	2	0.42	0.0438	100.000
GO:0046113	nucleobase catabolic process	2	2	0.42	0.0438	0.04384
GO:0045765	regulation of angiogenesis	2	2	0.42	0.0438	100.000
GO:0007179	transforming growth factor beta receptor...	2	2	0.42	0.0438	0.04384
GO:0016101	diterpenoid metabolic process	2	2	0.42	0.0438	100.000
GO:0002040	sprouting angiogenesis	2	2	0.42	0.0438	0.04384
GO:0007611	learning or memory	2	2	0.42	0.0438	0.04384
GO:0030239	myofibril assembly	2	2	0.42	0.0438	100.000
GO:0035265	organ growth	2	2	0.42	0.0438	100.000
GO:0050890	cognition	2	2	0.42	0.0438	100.000
GO:0060402	calcium ion transport into cytosol	2	2	0.42	0.0438	100.000
GO:0019695	choline metabolic process	2	2	0.42	0.0438	0.04384

GO:0007606	sensory perception of chemical stimulus	2	2	0.42	0.0438	0.04384
GO:0042417	dopamine metabolic process	2	2	0.42	0.0438	100.000
GO:0000038	very long-chain fatty acid metabolic pro...	2	2	0.42	0.0438	0.04384
GO:0071560	cellular response to transforming growth...	2	2	0.42	0.0438	100.000
GO:0071559	response to transforming growth factor b...	2	2	0.42	0.0438	100.000
GO:0017156	calcium-ion regulated exocytosis	2	2	0.42	0.0438	100.000
GO:0017158	regulation of calcium ion-dependent exoc...	2	2	0.42	0.0438	0.04384
GO:0042537	benzene-containing compound metabolic pr...	2	2	0.42	0.0438	100.000
GO:1901071	glucosamine-containing compound metaboli...	2	2	0.42	0.0438	100.000
GO:2000181	negative regulation of blood vessel morp...	2	2	0.42	0.0438	100.000
GO:0018958	phenol-containing compound metabolic pro...	2	2	0.42	0.0438	100.000
GO:0007204	positive regulation of cytosolic calcium...	2	2	0.42	0.0438	100.000
GO:0009712	catechol-containing compound metabolic p...	2	2	0.42	0.0438	100.000
GO:0031033	myosin filament organization	2	2	0.42	0.0438	100.000
GO:0031034	myosin filament assembly	2	2	0.42	0.0438	100.000
GO:0003012	muscle system process	2	2	0.42	0.0438	0.04384
GO:0006551	leucine metabolic process	2	2	0.42	0.0438	0.04384
GO:0071688	striated muscle myosin thick filament as...	2	2	0.42	0.0438	0.04384
GO:0006584	catecholamine metabolic process	2	2	0.42	0.0438	100.000
GO:0006577	amino-acid betaine metabolic process	2	2	0.42	0.0438	100.000
GO:0006515	protein quality control for misfolded or...	2	2	0.42	0.0438	0.04384
GO:0006094	gluconeogenesis	2	2	0.42	0.0438	0.04384
GO:0001523	retinoid metabolic process	2	2	0.42	0.0438	0.04384
GO:0046883	regulation of hormone secretion	2	2	0.42	0.0438	0.04384
GO:0046349	amino sugar biosynthetic process	2	2	0.42	0.0438	100.000
GO:0046339	diacylglycerol metabolic process	2	2	0.42	0.0438	0.04384
GO:0035927	RNA import into mitochondrion	2	2	0.42	0.0438	0.04384
GO:0006047	UDP-N-acetylglucosamine metabolic proces...	2	2	0.42	0.0438	100.000
GO:0044344	cellular response to fibroblast growth f...	2	2	0.42	0.0438	100.000
GO:0019319	hexose biosynthetic process	2	2	0.42	0.0438	100.000
GO:0030072	peptide hormone secretion	2	2	0.42	0.0438	100.000
GO:0030073	insulin secretion	2	2	0.42	0.0438	0.04384
GO:0016525	negative regulation of angiogenesis	2	2	0.42	0.0438	0.04384
GO:0051188	cofactor biosynthetic process	17	7	3.57	0.0462	100.000
GO:0009063	cellular amino acid catabolic process	17	7	3.57	0.0462	100.000
GO:1901137	carbohydrate derivative biosynthetic pro...	39	13	8.18	0.0467	0.57176

Brown Module

GO.ID	Term	Annotated	Significant	Expected	Elim_Fisher	Weight_Fisher
GO:0006091	generation of precursor metabolites and ...	16	8	2.61	0.0017	0.1808
GO:0051224	negative regulation of protein transport	3	3	0.49	0.0043	0.0043
GO:0006739	NADP metabolic process	3	3	0.49	0.0043	0.1613

GO:0070861	regulation of protein exit from endoplas...	3	3	0.49	0.0043	0.0043
GO:0097190	apoptotic signaling pathway	12	6	1.96	0.0067	0.0271
GO:0043408	regulation of MAPK cascade	9	5	1.47	0.0078	10.000
GO:0055114	oxidation-reduction process	17	7	2.77	0.0123	0.5799
GO:0070988	demethylation	4	3	0.65	0.0150	10.000
GO:0070972	protein localization to endoplasmic reti...	4	3	0.65	0.0150	0.0262
GO:0032872	regulation of stress-activated MAPK casc...	4	3	0.65	0.0150	0.0150
GO:0048813	dendrite morphogenesis	4	3	0.65	0.0150	0.0150
GO:0008214	protein dealkylation	4	3	0.65	0.0150	10.000
GO:0043410	positive regulation of MAPK cascade	4	3	0.65	0.0150	0.0262
GO:0070302	regulation of stress-activated protein k...	4	3	0.65	0.0150	10.000
GO:0006482	protein demethylation	4	3	0.65	0.0150	0.1614
GO:0051403	stress-activated MAPK cascade	4	3	0.65	0.0150	10.000
GO:0097193	intrinsic apoptotic signaling pathway	7	4	1.14	0.0159	10.000
GO:0080135	regulation of cellular response to stres...	22	8	3.58	0.0173	0.7493
GO:0080134	regulation of response to stress	31	10	5.05	0.0198	10.000
GO:2001233	regulation of apoptotic signaling pathwa...	11	5	1.79	0.0217	10.000
GO:0015780	nucleotide-sugar transmembrane transport	2	2	0.33	0.0264	10.000
GO:0009581	detection of external stimulus	2	2	0.33	0.0264	10.000
GO:0009582	detection of abiotic stimulus	2	2	0.33	0.0264	10.000
GO:0009583	detection of light stimulus	2	2	0.33	0.0264	10.000
GO:0051156	glucose 6-phosphate metabolic process	2	2	0.33	0.0264	10.000
GO:0032874	positive regulation of stress-activated ...	2	2	0.33	0.0264	10.000
GO:0005977	glycogen metabolic process	2	2	0.33	0.0264	0.0264
GO:1904152	regulation of retrograde protein transpo...	2	2	0.33	0.0264	10.000
GO:0071214	cellular response to abiotic stimulus	2	2	0.33	0.0264	10.000
GO:0046328	regulation of JNK cascade	2	2	0.33	0.0264	10.000
GO:0104004	cellular response to environmental stimu...	2	2	0.33	0.0264	10.000
GO:0006112	energy reserve metabolic process	2	2	0.33	0.0264	10.000
GO:1903513	endoplasmic reticulum to cytosol transpo...	2	2	0.33	0.0264	10.000
GO:0090317	negative regulation of intracellular pro...	2	2	0.33	0.0264	10.000
GO:0032511	late endosome to vacuole transport via m...	2	2	0.33	0.0264	0.0264
GO:0007254	JNK cascade	2	2	0.33	0.0264	10.000
GO:0043467	regulation of generation of precursor me...	2	2	0.33	0.0264	0.0264
GO:0032509	endosome transport via multivesicular bo...	2	2	0.33	0.0264	10.000
GO:0090481	pyrimidine nucleotide-sugar transmembran...	2	2	0.33	0.0264	0.0264
GO:0030970	retrograde protein transport ER to cyto…	2	2	0.33	0.0264	10.000
GO:0070371	ERK1 and ERK2 cascade	2	2	0.33	0.0264	10.000
GO:0070372	regulation of ERK1 and ERK2 cascade	2	2	0.33	0.0264	0.0264
GO:0007602	phototransduction	2	2	0.33	0.0264	0.0264
GO:0044042	glucan metabolic process	2	2	0.33	0.0264	10.000
GO:0071482	cellular response to light stimulus	2	2	0.33	0.0264	0.0264
GO:0071478	cellular response to radiation	2	2	0.33	0.0264	10.000

GO:0070304	positive regulation of stress-activated ...	2	2	0.33	0.0264	10.000
GO:0015693	magnesium ion transport	2	2	0.33	0.0264	0.0264
GO:0006740	NADPH regeneration	2	2	0.33	0.0264	10.000
GO:0010960	magnesium ion homeostasis	2	2	0.33	0.0264	0.0264
GO:0042182	ketone catabolic process	2	2	0.33	0.0264	0.0264
GO:0070863	positive regulation of protein exit from...	2	2	0.33	0.0264	10.000
GO:0006098	pentose-phosphate shunt	2	2	0.33	0.0264	0.0264
GO:0032387	negative regulation of intracellular tra...	2	2	0.33	0.0264	10.000
GO:0071985	multivesicular body sorting pathway	2	2	0.33	0.0264	10.000
GO:0007050	cell cycle arrest	2	2	0.33	0.0264	0.0264
GO:0006073	cellular glucan metabolic process	2	2	0.33	0.0264	10.000
GO:0009894	regulation of catabolic process	20	7	3.26	0.0318	0.4070
GO:0010941	regulation of cell death	20	7	3.26	0.0318	10.000
GO:0016358	dendrite development	5	3	0.81	0.0330	10.000
GO:0031098	stress-activated protein kinase signalin...	5	3	0.81	0.0330	10.000
GO:0032990	cell part morphogenesis	16	6	2.61	0.0330	0.2949
GO:0042981	regulation of apoptotic process	16	6	2.61	0.0330	10.000
GO:0009895	negative regulation of catabolic process	9	4	1.47	0.0438	0.0328
GO:0009628	response to abiotic stimulus	17	6	2.77	0.0443	10.000
GO:0034976	response to endoplasmic reticulum stress	17	6	2.77	0.0443	10.000
GO:0043067	regulation of programmed cell death	17	6	2.77	0.0443	10.000
GO:0048667	cell morphogenesis involved in neuron di...	13	5	2.12	0.0459	0.5037
GO:0065008	regulation of biological quality	75	18	12.22	0.0478	0.8195

Green Module

GO.ID	Term	Annotated	Significant	Expected	Elim_Fisher	Weight_Fisher
GO:0090502	RNA phosphodiester bond hydrolysis endo...	4	3	0.24	0.00081	
GO:0030951	establishment or maintenance of microtub...	2	2	0.12	0.00360	0.0036
GO:0007314	oocyte anterior/posterior axis specifica...	2	2	0.12	0.00360	0.0036
GO:0035107	appendage morphogenesis	6	3	0.36	0.00370	0.1661
GO:0045596	negative regulation of cell differentiat...	8	3	0.48	0.00950	0.1143
GO:1904029	regulation of cyclin-dependent protein k...	3	2	0.18	0.01039	10.000
GO:0035108	limb morphogenesis	3	2	0.18	0.01039	0.0104
GO:0060173	limb development	3	2	0.18	0.01039	10.000
GO:0018200	peptidyl-glutamic acid modification	3	2	0.18	0.01039	0.0104
GO:0000478	endonucleolytic cleavage involved in rRN...	3	2	0.18	0.01039	10.000
GO:0000479	endonucleolytic cleavage of tricistronic...	3	2	0.18	0.01039	0.0104
GO:0008298	intracellular mRNA localization	3	2	0.18	0.01039	0.0104
GO:0000079	regulation of cyclin-dependent protein s...	3	2	0.18	0.01039	0.0104
GO:0044092	negative regulation of molecular functio...	16	4	0.97	0.01276	0.1132
GO:0071900	regulation of protein serine/threonine k...	10	3	0.6	0.01864	10.000
GO:0046785	microtubule polymerization	4	2	0.24	0.01997	0.0200

GO:0000469	cleavage involved in rRNA processing	4	2	0.24	0.01997	10.000
GO:0071901	negative regulation of protein serine/th...	4	2	0.24	0.01997	0.0200
GO:0008285	negative regulation of cell population p...	4	2	0.24	0.01997	0.0200
GO:0090304	nucleic acid metabolic process	157	18	9.5	0.02153	0.5292
GO:0008283	cell population proliferation	11	3	0.67	0.02454	10.000
GO:0006974	cellular response to DNA damage stimulus	29	5	1.75	0.02646	0.0639
GO:0044085	cellular component biogenesis	111	12	6.71	0.02817	0.8180
GO:0051241	negative regulation of multicellular org...	12	3	0.73	0.03134	0.3058
GO:0006403	RNA localization	12	3	0.73	0.03134	10.000
GO:0043086	negative regulation of catalytic activit...	12	3	0.73	0.03134	10.000
GO:0045787	positive regulation of cell cycle	5	2	0.3	0.03199	0.0320
GO:0098813	nuclear chromosome segregation	13	3	0.79	0.03901	0.0609
GO:0065009	regulation of molecular function	43	6	2.6	0.04006	10.000
GO:0007049	cell cycle	55	7	3.33	0.04245	0.4128
GO:0051726	regulation of cell cycle	33	5	2	0.04387	10.000
GO:0033673	negative regulation of kinase activity	6	2	0.36	0.04613	10.000
GO:0006469	negative regulation of protein kinase ac...	6	2	0.36	0.04613	10.000
GO:0051961	negative regulation of nervous system de...	6	2	0.36	0.04613	10.000
GO:0050768	negative regulation of neurogenesis	6	2	0.36	0.04613	10.000
GO:0010721	negative regulation of cell development	6	2	0.36	0.04613	10.000
GO:0045665	negative regulation of neuron differenti...	6	2	0.36	0.04613	0.0461
GO:0060271	cilium assembly	6	2	0.36	0.04613	0.0461
GO:0051276	chromosome organization	56	7	3.39	0.04624	0.6330
GO:0007059	chromosome segregation	14	3	0.85	0.04755	10.000

Red module

GO.ID	Term	Annotated	Significant	Expected	Elim_Fisher	Weight_Fisher
GO:0006518	peptide metabolic process	42	6	1.52	0.0030	0.3686
GO:0006898	receptor-mediated endocytosis	4	2	0.14	0.0073	0.0073
GO:1902235	regulation of endoplasmic reticulum stre...	4	2	0.14	0.0073	0.0073
GO:0009987	cellular process	878	37	31.69	0.0114	0.2659
GO:0042157	lipoprotein metabolic process	5	2	0.18	0.0118	10.000
GO:0042158	lipoprotein biosynthetic process	5	2	0.18	0.0118	10.000
GO:0006497	protein lipidation	5	2	0.18	0.0118	0.0118
GO:0043038	amino acid activation	14	3	0.51	0.0120	10.000
GO:0043039	tRNA aminoacylation	14	3	0.51	0.0120	10.000
GO:0006418	tRNA aminoacylation for protein translat...	14	3	0.51	0.0120	0.0120
GO:0016192	vesicle-mediated transport	47	7	1.7	0.0132	10.000
GO:0019538	protein metabolic process	197	13	7.11	0.0151	10.000
GO:0048193	Golgi vesicle transport	16	3	0.58	0.0175	0.0175
GO:0099504	synaptic vesicle cycle	7	2	0.25	0.0238	10.000
GO:0099003	vesicle-mediated transport in synapse	7	2	0.25	0.0238	10.000

GO:0043170	macromolecule metabolic process	343	18	12.38	0.0371	10.000
GO:0043043	peptide biosynthetic process	37	4	1.34	0.0404	10.000
GO:0006412	translation	37	4	1.34	0.0404	10.000
GO:0044770	cell cycle phase transition	10	2	0.36	0.0476	10.000
GO:0044772	mitotic cell cycle phase transition	10	2	0.36	0.0476	10.000

Yellow module

GO.ID	Term	Annotated	Significant	Expected	Elim_Fisher	Weight_Fisher
GO:0072657	protein localization to membrane	11	6	1.51	0.0016	0.4370
GO:0051346	negative regulation of hydrolase activit...	3	3	0.41	0.0026	0.0026
GO:0043122	regulation of I-kappaB kinase/NF-kappaB ...	4	3	0.55	0.0092	0.0092
GO:0045937	positive regulation of phosphate metabol...	15	6	2.06	0.0105	0.1374
GO:0010562	positive regulation of phosphorus metabo...	15	6	2.06	0.0105	10.000
GO:0043085	positive regulation of catalytic activit...	16	6	2.2	0.0150	10.000
GO:0072655	establishment of protein localization to...	8	4	1.1	0.0154	0.0204
GO:0044089	positive regulation of cellular componen...	8	4	1.1	0.0154	0.0154
GO:0006605	protein targeting	12	5	1.65	0.0162	10.000
GO:0051591	response to cAMP	2	2	0.28	0.0188	0.0188
GO:0008156	negative regulation of DNA replication	2	2	0.28	0.0188	0.0188
GO:0046890	regulation of lipid biosynthetic process	2	2	0.28	0.0188	10.000
GO:0042304	regulation of fatty acid biosynthetic pr...	2	2	0.28	0.0188	0.0188
GO:0098840	protein transport along microtubule	2	2	0.28	0.0188	10.000
GO:0014074	response to purine-containing compound	2	2	0.28	0.0188	10.000
GO:0033523	histone H2B ubiquitination	2	2	0.28	0.0188	0.0188
GO:0006613	cotranslational protein targeting to mem...	2	2	0.28	0.0188	0.0188
GO:0099118	microtubule-based protein transport	2	2	0.28	0.0188	10.000
GO:0042073	intraciliary transport	2	2	0.28	0.0188	0.0188
GO:0043010	camera-type eye development	2	2	0.28	0.0188	0.0188
GO:0046683	response to organophosphorus	2	2	0.28	0.0188	10.000
GO:0002097	tRNA wobble base modification	2	2	0.28	0.0188	10.000
GO:0002098	tRNA wobble uridine modification	2	2	0.28	0.0188	0.0188
GO:1903441	protein localization to ciliary membrane	2	2	0.28	0.0188	0.0188
GO:0120034	positive regulation of plasma membrane b...	2	2	0.28	0.0188	10.000
GO:0071806	protein transmembrane transport	5	3	0.69	0.0207	0.0207
GO:0006470	protein dephosphorylation	5	3	0.69	0.0207	0.0207
GO:0042327	positive regulation of phosphorylation	13	5	1.79	0.0235	10.000
GO:0070585	protein localization to mitochondrion	9	4	1.24	0.0248	10.000
GO:0044782	cilium organization	9	4	1.24	0.0248	0.5169
GO:0015833	peptide transport	38	10	5.23	0.0269	0.6324
GO:0010638	positive regulation of organelle organiz...	18	6	2.48	0.0275	0.0880
GO:0031401	positive regulation of protein modificat...	18	6	2.48	0.0275	0.2419
GO:0051130	positive regulation of cellular componen...	28	8	3.85	0.0290	10.000

GO:0045184	establishment of protein localization	39	10	5.36	0.0320	10.000
GO:0072594	establishment of protein localization to...	14	5	1.93	0.0325	10.000
GO:0016311	dephosphorylation	10	4	1.38	0.0370	0.4418
GO:0044770	cell cycle phase transition	10	4	1.38	0.0370	10.000
GO:0044772	mitotic cell cycle phase transition	10	4	1.38	0.0370	10.000
GO:0007018	microtubule-based movement	10	4	1.38	0.0370	0.5173
GO:0009893	positive regulation of metabolic process	51	12	7.02	0.0371	10.000
GO:0099111	microtubule-based transport	6	3	0.83	0.0372	10.000
GO:0010970	transport along microtubule	6	3	0.83	0.0372	0.4429
GO:0042886	amide transport	40	10	5.5	0.0377	10.000
GO:0009605	response to external stimulus	30	8	4.13	0.0430	0.8453
GO:0044419	interspecies interaction between organis...	15	5	2.06	0.0436	0.6867
GO:0044267	cellular protein metabolic process	183	33	25.17	0.0443	0.5886
GO:0065003	protein-containing complex assembly	64	14	8.8	0.0450	0.9610
GO:0044093	positive regulation of molecular functio...	20	6	2.75	0.0454	10.000
GO:0050790	regulation of catalytic activity	34	11	4.68	0.0466	0.0214
GO:0019538	protein metabolic process	197	35	27.1	0.0469	0.5969
GO:0015031	protein transport	36	9	4.95	0.0478	10.000

Turquoise m.

GO.ID	Term	Annotated	Significant	Expected	Elim_Fisher	Weight_Fisher	
1	GO:0016458	gene silencing	9	8	2.92	0.00074	0.0328
2	GO:0040029	regulation of gene expression epigeneti…	9	8	2.92	0.00074	10.000
3	GO:1902275	regulation of chromatin organization	11	9	3.56	0.00102	0.0105
4	GO:0000077	DNA damage checkpoint	6	6	1.94	0.00112	0.0011
5	GO:0016071	mRNA metabolic process	34	20	11.01	0.00113	0.5313
6	GO:0006355	regulation of transcription DNA-templat…	28	17	9.07	0.00166	10.000
7	GO:0006333	chromatin assembly or disassembly	10	8	3.24	0.00265	0.0162
8	GO:0006913	nucleocytoplasmic transport	10	8	3.24	0.00265	10.000
9	GO:0006281	DNA repair	18	12	5.83	0.00276	10.000
10	GO:0034728	nucleosome organization	5	5	1.62	0.00349	0.0035
11	GO:0018022	peptidyl-lysine methylation	5	5	1.62	0.00349	0.0035
12	GO:0032543	mitochondrial translation	5	5	1.62	0.00349	10.000
13	GO:0010558	negative regulation of macromolecule bio...	19	12	6.15	0.00533	0.2403
14	GO:0051052	regulation of DNA metabolic process	7	6	2.27	0.00569	10.000
15	GO:0032774	RNA biosynthetic process	39	25	12.63	0.00588	10.000
16	GO:0045934	negative regulation of nucleobase-contai...	13	9	4.21	0.00682	10.000
17	GO:0034470	ncRNA processing	29	16	9.39	0.00851	0.0238
18	GO:0034629	cellular protein-containing complex loca...	4	4	1.3	0.01087	0.0109
19	GO:0090304	nucleic acid metabolic process	157	90	50.85	0.01223	6.1e-05
20	GO:0006351	transcription DNA-templated	34	22	11.01	0.01367	0.0062
21	GO:0045935	positive regulation of nucleobase-contai...	14	9	4.53	0.01371	10.000

22	GO:0006401	RNA catabolic process	14	9	4.53	0.01371	10.000
23	GO:0097659	nucleic acid-templated transcription	36	23	11.66	0.01451	10.000
24	GO:0051253	negative regulation of RNA metabolic pro...	12	8	3.89	0.01510	10.000
25	GO:0051254	positive regulation of RNA metabolic pro...	12	8	3.89	0.01510	10.000
26	GO:0035194	post-transcriptional gene silencing by R...	6	5	1.94	0.01536	10.000
27	GO:0035195	gene silencing by miRNA	6	5	1.94	0.01536	0.0154
28	GO:0016441	posttranscriptional gene silencing	6	5	1.94	0.01536	10.000
29	GO:0031047	gene silencing by RNA	6	5	1.94	0.01536	10.000
30	GO:1901659	glycosyl compound biosynthetic process	6	5	1.94	0.01536	0.3216
31	GO:1905269	positive regulation of chromatin organiz...	6	5	1.94	0.01536	10.000
32	GO:1903507	negative regulation of nucleic acid-temp...	10	7	3.24	0.01619	0.1554
33	GO:1902679	negative regulation of RNA biosynthetic ...	10	7	3.24	0.01619	10.000
34	GO:0051321	meiotic cell cycle	8	6	2.59	0.01656	10.000
35	GO:0031056	regulation of histone modification	8	6	2.59	0.01656	10.000
36	GO:1903046	meiotic cell cycle process	8	6	2.59	0.01656	0.0166
37	GO:0051168	nuclear export	8	6	2.59	0.01656	0.0153
38	GO:0022613	ribonucleoprotein complex biogenesis	36	18	11.66	0.01929	0.6176
39	GO:0051276	chromosome organization	56	33	18.14	0.01992	0.1089
40	GO:2000113	negative regulation of cellular macromol...	17	10	5.51	0.02133	0.1268
41	GO:0006259	DNA metabolic process	36	24	11.66	0.02403	10.000
42	GO:0071826	ribonucleoprotein complex subunit organi...	22	12	7.13	0.02482	0.0248
43	GO:0051252	regulation of RNA metabolic process	39	24	12.63	0.02812	10.000
44	GO:0006397	mRNA processing	20	11	6.48	0.02924	10.000
45	GO:0048285	organelle fission	20	11	6.48	0.02924	0.2328
46	GO:0031327	negative regulation of cellular biosynth...	20	11	6.48	0.02924	10.000
47	GO:0019222	regulation of metabolic process	154	73	49.88	0.02968	0.4251
48	GO:0006396	RNA processing	52	28	16.84	0.03157	10.000
49	GO:0031503	protein-containing complex localization	11	7	3.56	0.03230	10.000
50	GO:0071103	DNA conformation change	11	7	3.56	0.03230	0.0323
51	GO:0002520	immune system development	11	7	3.56	0.03230	0.0892
52	GO:0034404	nucleobase-containing small molecule bio...	11	7	3.56	0.03230	10.000
53	GO:0006261	DNA-dependent DNA replication	11	7	3.56	0.03230	0.5146
54	GO:0065003	protein-containing complex assembly	64	28	20.73	0.03290	0.5308
55	GO:0051171	regulation of nitrogen compound metaboli...	113	56	36.6	0.03338	0.8352
56	GO:1903313	positive regulation of mRNA metabolic pr...	3	3	0.97	0.03377	0.0338
57	GO:0033260	nuclear DNA replication	3	3	0.97	0.03377	0.0338
58	GO:0034968	histone lysine methylation	3	3	0.97	0.03377	10.000
59	GO:0006342	chromatin silencing	3	3	0.97	0.03377	10.000
60	GO:0045814	negative regulation of gene expression …	3	3	0.97	0.03377	10.000
61	GO:0034661	ncRNA catabolic process	3	3	0.97	0.03377	0.0338
62	GO:0070192	chromosome organization involved in meio...	3	3	0.97	0.03377	10.000
63	GO:0044786	cell cycle DNA replication	3	3	0.97	0.03377	10.000
64	GO:0010608	posttranscriptional regulation of gene e...	18	10	5.83	0.03445	10.000

65	GO:0009451	RNA modification	18	10	5.83	0.03445	0.5380
66	GO:0000280	nuclear division	18	10	5.83	0.03445	0.5380
67	GO:0008380	RNA splicing	18	10	5.83	0.03445	0.3845
68	GO:0044085	cellular component biogenesis	111	45	35.95	0.03468	0.7848
69	GO:1903311	regulation of mRNA metabolic process	9	6	2.92	0.03626	10.000
70	GO:0006323	DNA packaging	9	6	2.92	0.03626	0.3215
71	GO:0045892	negative regulation of transcription DN…	9	6	2.92	0.03626	10.000
72	GO:0034655	nucleobase-containing compound catabolic...	23	12	7.45	0.03711	0.6852
73	GO:0016070	RNA metabolic process	123	69	39.84	0.03866	10.000
74	GO:0031497	chromatin assembly	7	5	2.27	0.03953	10.000
75	GO:0006405	RNA export from nucleus	7	5	2.27	0.03953	10.000
76	GO:0140013	meiotic nuclear division	7	5	2.27	0.03953	10.000
77	GO:0034401	chromatin organization involved in regul...	7	5	2.27	0.03953	10.000
78	GO:0097549	chromatin organization involved in negat...	7	5	2.27	0.03953	10.000
79	GO:0022402	cell cycle process	41	19	13.28	0.04023	10.000
80	GO:0006354	DNA-templated transcription elongation	5	4	1.62	0.04039	10.000
81	GO:0006302	double-strand break repair	5	4	1.62	0.04039	10.000
82	GO:0048232	male gamete generation	5	4	1.62	0.04039	0.0404
83	GO:0043628	ncRNA 3'-end processing	5	4	1.62	0.04039	10.000
84	GO:0009163	nucleoside biosynthetic process	5	4	1.62	0.04039	10.000
85	GO:0048522	positive regulation of cellular process	101	41	32.71	0.04221	0.8527
86	GO:0016570	histone modification	21	11	6.8	0.04402	0.6686
87	GO:0022618	ribonucleoprotein complex assembly	21	11	6.8	0.04402	0.5379
88	GO:0034622	cellular protein-containing complex asse...	60	26	19.43	0.04443	0.7973
89	GO:0060255	regulation of macromolecule metabolic pr...	134	65	43.4	0.04529	0.7598
90	GO:0031323	regulation of cellular metabolic process	129	61	41.78	0.04548	10.000
91	GO:0000398	mRNA splicing via spliceosome	14	8	4.53	0.04786	0.0561
92	GO:0000377	RNA splicing via transesterification re…	14	8	4.53	0.04786	10.000

Suppl. Table 1: GO-terms enrichment for each module inferred in the co-expression network of the bisexual species *Bacillus grandii*.

4. Methods development.

Complete mitochondrial genomes from transcriptomes: assessing pros and cons of data mining for assembling new mitogenomes

Giobbe Forni, Guglielmo Puccio, Thomas Bourguignon, Theodore Evans, Barbara Mantovani, Omar Rota-Stabelli and Andrea Luchetti.

SCIENTIFIC REPORTS
nature research

OPEN

Complete mitochondrial genomes from transcriptomes: assessing pros and cons of data mining for assembling new mitogenomes

Received: 5 March 2019
Accepted: 25 September 2019
Published online: 15 October 2019

Giobbe Forni[1], Guglielmo Puccio[1], Thomas Bourguignon[2,3], Theodore Evans[4], Barbara Mantovani[1], Omar Rota-Stabelli[5] & Andrea Luchetti[1]

Thousands of eukaryotes transcriptomes have been generated, mainly to investigate nuclear genes expression, and the amount of available data is constantly increasing. A neglected but promising use of this large amount of data is to assemble organelle genomes. To assess the reliability of this approach, we attempted to reconstruct complete mitochondrial genomes from RNA-Seq experiments of *Reticulitermes* termite species, for which transcriptomes and conspecific mitogenomes are available. We successfully assembled complete molecules, although a few gaps corresponding to tRNAs had to be filled manually. We also reconstructed, for the first time, the mitogenome of *Reticulitermes banyulensis*. The accuracy and completeness of mitogenomes reconstruction appeared independent from transcriptome size, read length and sequencing design (single/paired end), and using reference genomes from congeneric or intra-familial taxa did not significantly affect the assembly. Transcriptome-derived mitogenomes were found highly similar to the conspecific ones obtained from genome sequencing (nucleotide divergence ranging from 0% to 3.5%) and yielded a congruent phylogenetic tree. Reads from contaminants and nuclear transcripts, although slowing down the process, did not result in chimeric sequence reconstruction. We suggest that the described approach has the potential to increase the number of available mitogenomes by exploiting the rapidly increasing number of transcriptomes.

The NCBI GenBank database counts more than 78,000 mitochondrial genome entries from more than 30,000 different species. Due to their relatively small size, ease of sequencing and clear orthology, mitochondrial genomes are currently the most widely used genomic markers for animal systematics and phylogenetic studies, particularly in insects[1]. Mitogenome-based phylogenies have helped resolving inter-ordinal[2], intra-ordinal[3] and intra-familial[4,5] relationships. Many biogeographic, population genetics, and museum genomics studies also rely on mitochondrial genome sequencing[6,7].

Most mitochondrial genomes are now obtained with high-throughput sequencing (HTS) of DNA (either including a long PCRs step or by direct HTS sequencing of total DNA) and a wide variety of methods have been developed for sequencing, assembly and annotation[8,9]. As pointed out by Smith[10], a considerable amount of information for organelle genomics may be obtained from RNA-Seq data and, taking into account the rate at which transcriptome data accumulate, they can be further used to recover mitochondrial genomes from taxa in which they are not available yet. The NCBI SRA database (Sequence Read Archive; last accessed January 2019) stores transcriptome raw reads of more than 1,790 different insect species: for about 1,000 of these species there are no mitogenomes in the Nucleotide archive (last accessed January 2019). Considering that more than 4,800 insect mitogenomes are actually present in Genbank databases, mining mitochondrial genomes from transcriptomes has the potential to increase the number of insects mitogenomes by ~20%.

[1]Department of Biological, Geological and Environmental Sciences - University of Bologna, via Selmi 3, 40126, Bologna, Italy. [2]Okinawa Institute of Science & Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa, 904–0495, Japan. [3]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic. [4]School of Animal Biology, University of Western Australia, Perth, WA, 6009, Australia. [5]Agrarian Entomology, Research and Innovation Centre, Fondazione Edmund Mach (FEM), Via E. Mach 1, 38010, San Michele all'Adige, TN, Italy. Correspondence and requests for materials should be addressed to A.L. (email: andrea.luchetti@unibo.it)

The use of RNA-Seq data for mitochondrial genomes reconstruction provides several advantages. Genes from organelles have higher expression levels than nuclear genes and, therefore, a large portion of raw reads generated from eukaryotic RNA-Seq experiments are of organelle origin[11]. Because organelle genomes are pervasively transcribed as polycistronic RNAs, it is possible to recover almost complete mitogenomes from transcriptomes[12]. On the other hand, the use of RNA-Seq data could still retain some of the drawbacks of DNA-based approaches. As an example, despite the low chance of a NUMT (nuclear mitochondrial DNA; *i.e.* mitochondrial DNA copies migrated into the nuclear genome) to be integrated in a nuclear transcript, there are some evidences of the presence of nuclear mitochondrial pseudogenes in transcriptomic data[13].

Nonetheless, literature indicates that transcriptome data is not yet routinely used to assemble mitogenomes. A few studies, for example, report on successfully assembled mitochondrial genomes either by mapping reads on conspecific references and assembling them[14,15], or by mining mitochondrial contigs from de novo assembled transcriptomes[16,17]. However, most of these studies did not provide new mitochondrial genomes and, in more general terms, all of them were unable to assemble regions, such as tRNAs or the control region, due to low reads coverage. This latter finding, besides being of fundamental importance to obtain fully assembled mitogenomes, suggests that some pitfalls may occur when dealing with RNA-Seq data for mitogenomic studies.

In this report we investigate potential pros and cons of de novo assembling mitogenomes from transcriptome data in order to get information about genome structure and nucleotide variability, in a systematic framework. For this aim, we mined mitochondrial DNA from transcriptome data of the subterranean termite genus *Reticulitermes* (Blattodea; Termitoidae) as a case study. RNA-Seq data and mitogenome sequences are already available for several species of *Reticulitermes*, allowing direct comparison of transcriptome-derived mitogenome sequences with those obtained through traditional methods (long-PCR + Sanger sequencing or high throughput genome sequencing). Moreover, we assembled for the first time the mitochondrial genome of *Reticulitermes banyulensis*. The genus *Reticulitermes* has been the subject of many phylogenetic studies during the last two decades[18,19], making it a suitable group to test the phylogenetic accuracy of the mitogenomes reconstructed from transcriptomes.

Overall, we demonstrate the validity of the iterative reference mapping and de novo assembly of mitogenomes from transcriptomes, recovering reliable complete molecules from the assayed species. On the other hand, some pitfalls may emerge, such as in the case of contaminants, similarity with nuclear transcripts or for tRNA reconstruction.

Results and Discussion

Reconstruction process analysis. Overall, the reconstruction process led to the assembly of either complete or nearly complete mitogenomes for all the six analyzed species (Supplementary Fig. S1).

For *R. flavipes*, *R. grassei*, *R. banyulensis* and *R. lucifugus*, the number of iterations necessary to mine all mitochondrial reads (i.e., in our pipeline, to reach a stationary number of mapping reads) varied depending on the phylogenetic relatedness of the initial reference (congeneric as blue triangles; intra-familial as orange circles in Fig. 1). As it could be expected, the use of intra-familial reference led to recover fewer reads during the first iterations than using congeneric references (Fig. 2). However, all assemblies, whether based on congeneric or on intra-familial references, converged to a plateau within the 10th iteration; at this point, the total number of mitochondrial reads covered between 7.5% and 13.9% of the whole transcriptomes. We observed fewer and longer contigs as the number of iterations increased (Fig. 2) and this trend was not influenced by the choice of the initial reference. Overall, the number of contigs at the 10th iteration varied between six and twelve, and their average length varied between 1,376 and 2,544 bp. In contrast, no plateau was reached within the 10th iteration for both *R. labralis* and *R. speratus*, independently of the initial references (Fig. 1). With respect to the other four species, the iteration process for *R. labralis* and *R. speratus* resulted in a far larger number of contig (33,906–34,439 and 157–262, respectively; Table 1). Moreover, we did not observe a reduction in number and an increase in length of the contigs throughout the iterations (Fig. 2). In *R. labralis*, the percentage of recruited reads reached 57%, with both intra-familial and congeneric starting references. In *R. speratus*, the percentage of recruited reads reached 18% and 16% using the intra-familial and congeneric starting references, respectively. We analyzed more in details the contigs obtained after the 10th iteration in these two species, filtering for possible contaminants and nuclear genes matches. In both species, mitochondrial contigs represented the minority of assembled ones: less than the 1% in *R. labralis* and less than the 14% in *R. speratus*. The majority of contigs were, therefore, contaminant or nuclear transcripts, with a different relative contribution of the two kinds of leakage in the two species (Table 1). The contaminant leakage played a major role in *R. labralis*, with fraction of contigs significantly matching also to sequences of non-hexapod taxa (Table 1 and Suppl. Table S1). On the other hand, the nuclear leakage appeared predominant in *R. speratus* (Table 1 and Suppl. Table S1).

Excluding contaminant and nuclear transcripts, all contigs obtained at the end of the automated process in the six assayed species matched the relative mitochondrial genomes. On the other hand, a few gaps were still present at the end of the iterative reconstruction process. These gaps encompassed mainly tRNA regions and/or the control region, with the only exception of a fragment of the *R. speratus atp6* gene (Supplementary Fig. S1a). Only *R. banyulensis* and *R. labralis* resulted completely reconstructed (except the control region) at the end of the iterative process. The use of intra-familial references in the first iteration, instead of congeneric ones, led to similar results but with slightly larger gap regions and no mitogenomes completely reconstructed at the end of the process (Supplementary Fig. S1b). We recovered missing regions using *blastn* search of the transcriptome reads against the homolog portion of the reference mitochondrial genomes and filled the gaps. Control regions remained only partially assembled, mainly because they are composed of tandem repeats, thus very difficult to de novo assemble properly, and their expression is low or completely lacking. Moreover, we obtained partial control regions only when mitochondrial genomes of congeneric species were used as initial references (Supp. Table S2).

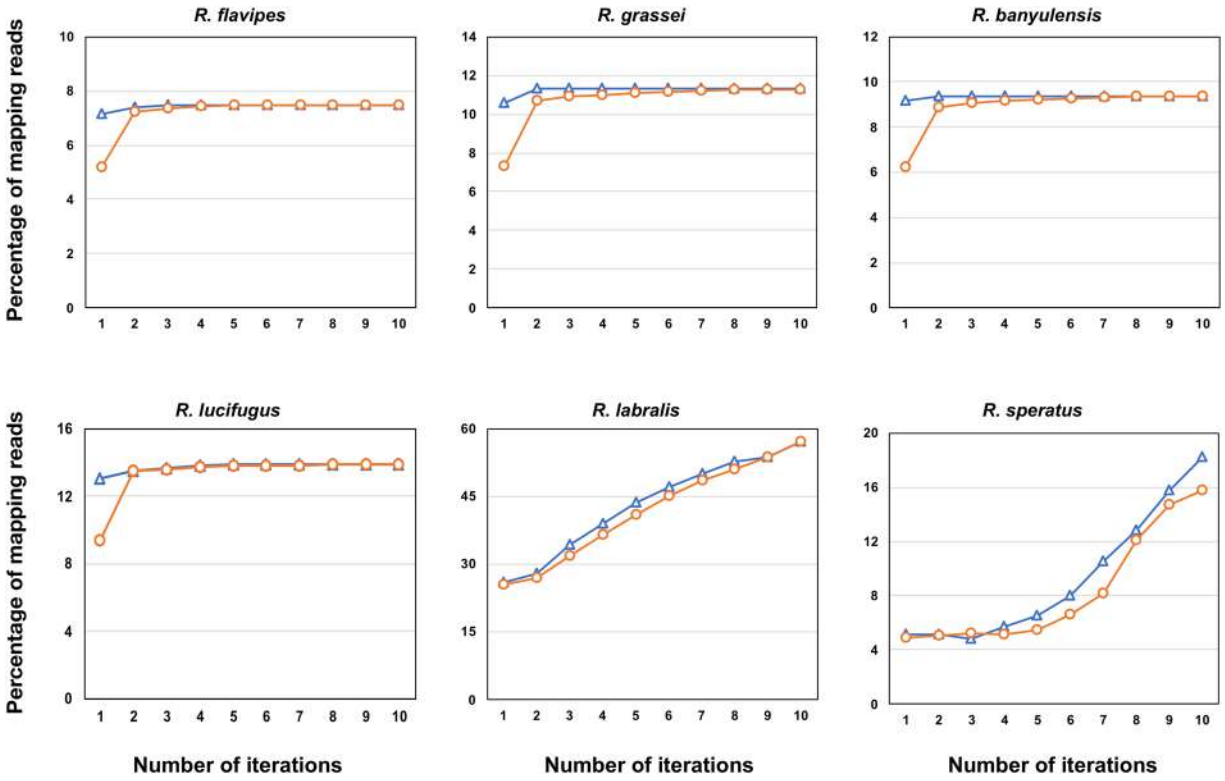


Figure 1. Short reads recruitment across iterations. Relationships between the number of iterations of the process (*x* axis) and mapped reads (% of the total reads; *y* axis) using congeneric (blue triangles) or intra-familial (orange circles) references.

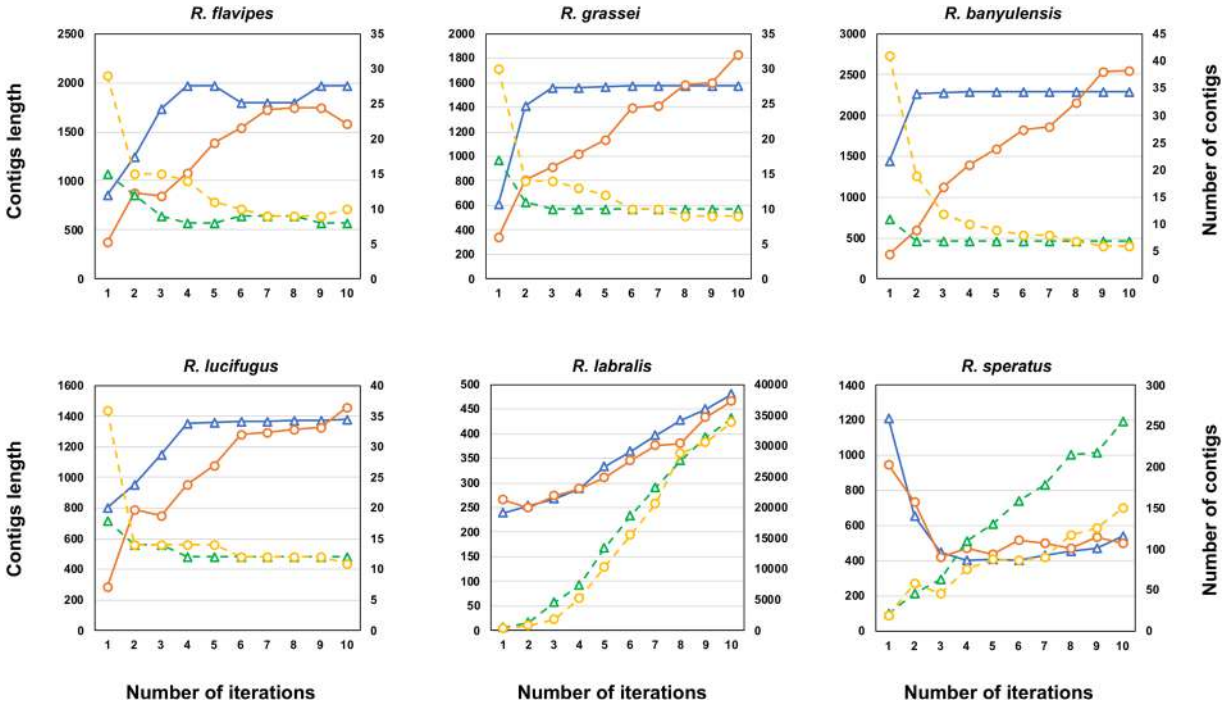


Figure 2. Contigs analysis across iterations. Relationship between the number of iteration process (*x* axis) and contigs length (bp, primary *y* axis; solid lines) and contigs number (secondary *y* axis; dashed lines) using congeneric (blue triangles) or intra-familial (orange circles) references.

Contigs at the 10th iteration	*Reticulitermes labralis*		*Reticulitermes speratus*	
	Congeneric references	Intra-familial references	Congeneric references	Intra-familial references
Total	34,439	33,906	262	157
Target mitochondrial contigs	138 (0.40%)	180 (0.53%)	32 (12.2%)	22 (14.0%)
Contaminant leakage	23,819 (69.2%)	23,572 (69.5%)	113 (43.1%)	66 (42.0%)
Nuclear leakage	10,482 (30.4%)	10,154 (29.9%)	117 (44.7%)	69 (43.9%)

Table 1. Non-mitochondrial reads leakage analysis in *Reticulitermes labralis* and *R. speratus* RNA-Seq experiments. Numbers refer to contigs number (percentage) that were found of contaminant or nuclear DNA origin.

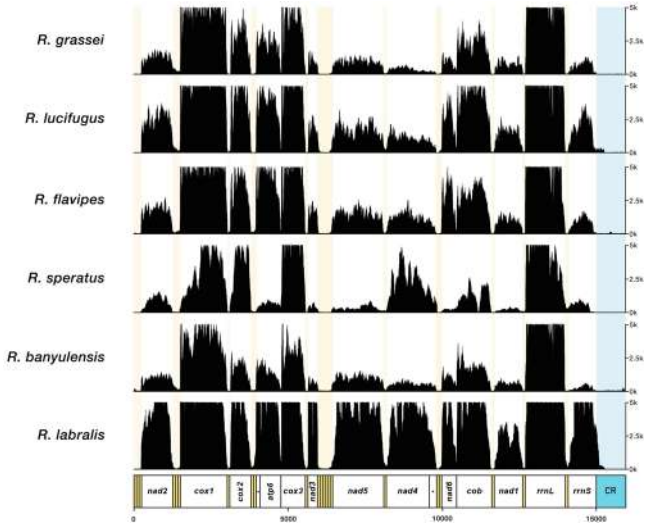


Figure 3. Reads coverage along mitogenomes reconstructed using congeneric references. Coverage has been capped at 5,000 × for graphical purposes; shaded areas indicate tRNAs and the control region.

Mitogenomes completeness and accuracy assessment. The six mitochondrial genomes we recovered were comparable in length and content to that of other insect mitogenomes, containing 13 PCGs, 22 tRNAs and 2 rRNAs, with gene order consistent to that of other *Reticulitermes* mitochondrial genomes. No gap or stop codon were detected within open reading frames and no significant variation of tRNA cloverleaf structures was found between the RNA-Seq assembled mitogenomes and those already available from DNA sequencing (Supplementary Fig. S2).

Sequence coverage was highly variable among genes, but in all cases PCGs and rRNAs were well covered (Fig. 3). tRNA coverage was comparatively much lower than that of the other genes, and this holds for all six *Reticulitermes* species. In particular, manually reconstructed tRNAs showed a drop in the fold coverage below the 4×. This is could be a consequence of their short sequence length: during the size-selection step of the Illumina library preparation protocol it is likely that their mature forms are removed. However, being also part of the non-mature polycistron, we successfully assembled and annotated most of them from RNA-Seq experiments with the automated approach. Overall, the gene coverage profiles obtained (Fig. 3) nicely matche to the gene expression profiles observed in other insects[14,20] and vertebrate mitogenomes[16,21], suggesting a similar pattern of differential gene expression also in termites.

Mitogenomes obtained using congeneric and intra-familial initial references were nearly identical: up to seven nucleotide positions show polymorphism over the entire sequence length (Suppl. Table S3). This variation might indicate the presence of heteroplasmy, similarly to what has been previously reported in many organisms[22,23]. Alternatively, mitogenomes sequence identity can be affected by the choice of assembling tools and approaches (de novo or reference-based) used to generate the final assembly[24]. Mitogenomes recovered from RNA-Seq data show divergences ranging from 0.0% to 3.5% with conspecific references (Suppl. Table S4): most of nucleotide substitutions (ranging from 68.9% to 76.9%) occur at the third codon position of protein-coding genes, compatibly with the genetic variability among individuals of the same species (Suppl. Table S5).

When our pipeline was run on references with artificial rearrangements (Supplementary Fig. S3), results do not differ from those obtained using the original references: it can be clearly seen that no chimeric contigs are generated with the different gene order of the references and the correct genes order is obtained in all instances. This can be explained by taking into account the way the pipeline works: mapping on the initial references, and on reconstructed contigs in the subsequent iterations, is only used to recover mitochondrial reads but the assembly

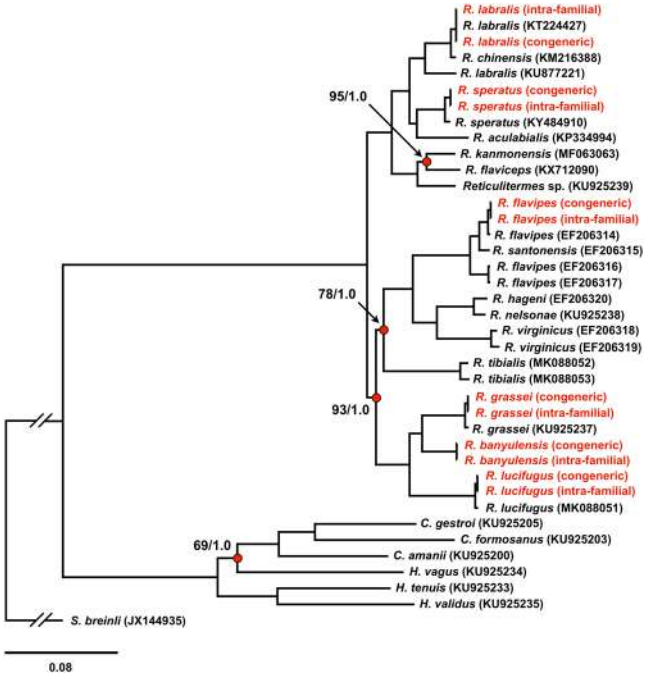


Figure 4. Mitogenomes phylogenetic analysis. Schematic drawing of Maximum Likelihood ($-lnL = 78672.19$) and Bayesian inference ($-lnL = 78489.80$). Maximum bootstrap (=100%) and posterior probability (=1.0) values are omitted, while they are reported on nodes showing supports lower than the maximum (indicated with red dots). Mitogenomes assembled from RNA-Seq are highlighted in red.

is performed de novo at every iteration. Therefore, the gene order resulting at the end of the process (and at the end of every iteration) is given only by the information stored within collected reads.

Phylogenetic validation of *Reticulitermes* mitogenomes. In order to assess the accuracy of the mitogenomes obtained from RNA-Seq experiments we tested them in a phylogenetic framework. Our Bayesian and Maximum Likelihood trees (Fig. 4) were congruent with published phylogenetic trees[18,19]. All mitogenomes isolated from RNA-Seq experiments clustered with conspecific DNA-obtained sequences, with maximum node support. Moreover, mitogenomes assembled using different initial references (congeneric or intra-familial) grouped together, with branch lengths equal or close to zero (Fig. 4).

Conclusions

Although the use of a straight-forward approach, such as the one we applied in this study, appears enough to carry out the task of reconstructing mitochondrial genomes from transcriptome data, some limitations emerged. Some tRNAs and control regions were not reconstructed, but these gaps could be filled by a manual procedure. As evidenced by the coverage analysis, these regions showed a decidedly lower coverage ($<4\times$) with respect to other mitogenome regions, as also observed in other systems[14,16,20,21,25], and this may have led to fragmented reconstruction. Although we easily overcome the problem by manually assembling gap regions, this may cast some doubts on the scalability of the method when trying to obtain tRNAs and the control region. Most systematics and phylogenetic studies rely on PCGs and rRNA only, but the potential utility of tRNAs should be not overlooked as their inclusion in these analyses may help to improve the inference[1]. Moreover, the reconstruction of gene order and of small non-coding regions can be also important to mark rare genomic changes that could be useful for identifying clades in absence of (or in addition to) the phylogenetic signal of substitutions[26,27]. Finally, information on the control region sequence may be relevant for explaining phylogenetic biases or artefacts obtained in mitochondrial genomes analyses[28].

Our approach has been tested in the context of insect mitochondrial genomes, which are well known for their mostly conserved structure and the compact size (typically 15–18 kb)[1]. A wide variety of mitochondrial genome size can be found, especially outside metazoan, but some evidences indicated that even mitogenomes rich of non-coding sequences are fully transcribed[12,15]. While this would suggest that it can be possible to recover their complete mitogenome, further experiments are required to evaluate these peculiar conditions.

In this view, possible future development of pipelines running on RNA-Seq data for mitogenome assembly should take into account such challenges. However, we successfully recover the six whole mitogenomes, even in those samples where contaminants and/or random similarities with nuclear transcripts outnumber genuine mitochondrial contigs. In these two instances, namely in *R. speratus* and *R. labralis*, the only issue was a small increase of computational time across the whole process.

Overall, we showed that mitochondrial genome sequences can be accurately reconstructed from transcriptome data using an iterative reference mapping and de novo assembly approach. None of the RNA-Seq data

Species	SRA acc. no.	Gbp	Reads length	Conspecific mitogenomes
Reticulitermes flavipes	SRR1325101	2.4	51 (single end)	KY484910
				EF206314
				EF206315[a]
				EF206316
				EF206317
Reticulitermes grassei	SRR1325103	1.5	51 (single end)	KU925237
Reticulitermes banyulensis	SRR5253660	1.3	51 (single end)	—
Reticulitermes lucifugus	SRR1325112	2.3	51 (single end)	MK088051
Reticulitermes labralis	SRR5808263	9.1	150 (pair end)	KT224427
				KU877221
Reticulitermes speratus	DRR030843	3.4	93 (pair end)	KY484910

Table 2. Data information for Illumina RNA-Seq experiments and conspecific mitogenomes obtained from NCBI Genbank database. [a]This was attributed to *R. santonensis*, which is synonym species of *R. flavipes*.

employed in this study were initially generated to investigate mitochondrial aspects[29,30]; the proposed approach has the potential to generate a large number of new mitochondrial genomes for non-model species, by exploiting the increasing amount of publicly available RNA high-throughput sequencing data. This approach can be relevant for molecular taxonomy and systematics but also to investigate various aspects of mitochondrial genome biology[10,31], such as transcription[32] and polyadenylation profiles[21].

Material and Methods

Data information. The 12 complete mitochondrial genomes of *Reticulitermes* species sequenced to date have been downloaded from GenBank (accessed on January 2019), alongside the complete mitochondrial genomes of 44 *Coptotermes* and 28 *Heterotermes* (Table 2; Suppl. Table S6). We used the mitogenome of *Schedorhinotermes breinli* (Genbank accession number: JX144935) as outgroup to root phylogenetic trees[18,19,27].

We selected the RNA-Seq Illumina reads of six species of *Reticulitermes* from which we attempted to extract mitogenomes (Table 2). For five species (*R. flavipes*, *R. grassei*, *R. lucifugus*, *R. labralis* and *R. speratus*), a conspecific mitochondrial genome sequence was already available, which we used to assess the accuracy and completeness of our assemblies. For *R. banyulensis*, no mitogenome reference was available, and we here present the first complete mitochondrial genome.

Mitochondrial genome reconstruction pipeline. Raw reads were downloaded with Fastq-dump, quality-checked with FastQC[33] and trimmed and clipped with Trimmomatic[34], using parameters ILLUMINACLIP:TruSeq3-PE.fa:1:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:25:33 MINLEN:45 for reads of over 50 bp length and parameters ILLUMINACLIP:TruSeq3-PE.fa:1:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:25:33 MINLEN:90 for reads of over 90 bp length. In order to check the assembly process also during intermediate steps, we set up an iteration-based pipeline (available at https://github.com/mozoo/mitoRNA), summarized as follow (Fig. 5):

(i) in the first step, all transcriptome reads are mapped to reference mitochondrial genomes with Bowtie2[35] in local mode with parameters set to–very-sensitive-local, whose default settings are -D 20 -R 3 -N 0 -L 20 -i S,1,0.50, but allowing one mismatch in the seed alignment (−N 1) and by reducing the length of the seed substrings to align (−L 10). This increases the sensitivity of the mapping.
(ii) in the second step, mapped reads are de novo assembled by Trinity[36] with the settings *--no_normalize_reads --min_contig_length* 150 bp, which remove the in silico reads normalization step and set the minimum length for a contig to be assembled at 150 bp.

Generated contigs are then used as a new reference for successive iterations. In order to test whether reference mitogenomes affect the reconstruction of the RNA-Seq-derived mitogenomes, we excluded conspecific sequences from the initial reference list used in the first step of the first iteration (*i.e.* when analyzing the RNA-Seq reads of a given species, the mitogenome(s) of the same species was excluded from the list of reference genomes). In addition, to evaluate the effect of phylogenetic relatedness of the starting reference mitogenomes, we used either *Reticulitermes* mitogenomes (congeneric references) or *Coptotermes* + *Heterotermes* mitogenomes (intra-familial) during the first step of the first iteration.

For each transcriptome analyzed, we evaluated the progress of this iterative approach using both congeneric and intra-familial starting references, stopping the process after 10 iterations. At the 10th iteration, then, we recovered a variable number of contigs which were validated and scaffolded with *blastn*[37] using default parameters against the closest, non-conspecific reference mitogenome. We then merged the scaffolded contigs using Aliview[38]. Where necessary (see Results) gaps in reconstructed molecule were filled by BLASTing transcriptome reads against the homolog portion of the reference mitochondrial genomes, then reads were assembled with CAP3[39] and aligned to the scaffold using MAFFT v.7[40].

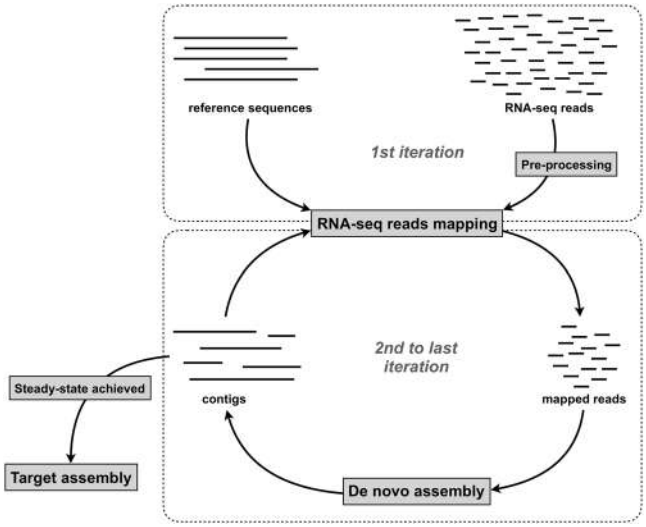


Figure 5. Conceptual map of the pipeline implemented for mitogenome reconstruction from RNA-Seq raw reads.

In order to assess the impact of using references showing a different gene order with respect to the target mitogenome, we generated artificial rearrangements of the congeneric references (Supplementary Fig. S3) and we then re-run our pipeline using the rearranged molecules as initial references.

We annotated the 22 tRNA genes, the 13 protein coding genes (PCGs), and the two ribosomal RNA genes manually, aided by congeneric published reference sequences.

Quality assessment and phylogenetic analysis. During the mapping process, it is possible that some reads derived from either contaminant (if any) or nuclear DNA can leak into the iterative process, because of random homology, and assembled into contigs. Therefore, contigs generated at the 10th iteration were analyzed to check if they derive from contaminant reads (contaminants leakage) or from nuclear reads (nuclear leakage). To estimate the contaminants leakage, we identified contigs that did not match with insects using DIAMOND[41]. Then, to estimate the nuclear leakage, we identified contigs matching to insects but not to the reconstructed mitochondrial genomes by means of a *blastn* search.

The coverage of all newly reconstructed mitogenomes was determined by mapping reads with Bowtie2 and analysing the output with SAMtools[42]. Newly obtained mitogenomes were annotated based on homologies with previously published ones, and all protein coding genes (PCGs) were manually inspected for open reading frame correctness. We also estimated tRNAs secondary structure for both DNA-derived and RNA-derived mitogenomes using Mitos2[43] (available at: http://mitos2.bioinf.uni-leipzig.de/index.py).

Phylogenetic and nucleotide divergence analyses have been carried out using PCGs and rRNAs. Each gene was aligned separately using MAFFT with the option *--auto* for protein coding gene (PGC) and with the option *--X-INS-i* for the two rRNA genes. We omitted control regions from the final matrix as they were only partially assembled. After concatenation, the final matrix included 37 sequences spanning 13,547 nucleotide positions. Phylogenetic trees were reconstructed using Bayesian inference and Maximum Likelihood approaches. Model selection and phylogenetic inference were carried out through CIPRES Science Gateway (www.phylo.org)[44]. For both Bayesian and Maximum Likelihood approaches, the best-fit models of nucleotide substitution were identified using IQ-TREE Model Selection[45], using the edge-linked parameter and the TESTNEWMERGE flag (Suppl. Tables S7 and S8). The ML search was run with 1,000 ultrafast bootstraps replicates using IQ-TREE v1.6.1[46]. The Bayesian inference was carried out using MrBayes v. 3.2.6[47]: two Markov chains Monte Carlo (MCMC) were run simultaneously for 10,000,000 generations and were sampled every 1,000 generations. Burn-in was set at a conservative threshold of 25%. Average deviation of split frequencies fell below 0.01 within 1 million generations, indicating the chain reached convergence.

Data Availability

Mitogenomes assembled from RNA-Seq experiments are available on Fig Share under the https://doi.org/10.6084/m9.figshare.7181969.

References

1. Cameron, S. L. Insect mitochondrial genomics: implications for evolution and phylogeny. *Ann. Rev. Entomol.* **59**, 95–117 (2014).
2. Song, N., Li, H., Song, F. & Cai, W. Molecular phylogeny of Polyneoptera (Insecta) inferred from expanded mitogenomic data. *Sci. Rep.* **6**, 36175 (2016).
3. Bourguignon, T. *et al.* The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Mol. Biol. Evol.* **32**, 406–421 (2015).
4. Ometto, L. *et al.* Linking genomics and ecology to investigate the complex evolution of an invasive *Drosophila* pest. *Genome Biol. Evol.* **5**, 745–757 (2013).

5. Bourguignon, T. *et al*. Mitochondrial phylogenomics resolves the global spread of higher termites, ecosystem engineers of the tropics. *Mol. Biol. Evol.* **34**, 589–597 (2017).
6. Ma, C. *et al*. Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. *Mol. Ecol.* **21**, 4344–4358 (2012).
7. Mikheyev, A. S. *et al*. Museum genomics confirms that the Lord Howe Island stick insect survived extinction. *Curr. Biol.* **27**, 3157–3161 (2017).
8. Cameron, S. L. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Syst. Entomol.* **39**, 400–411 (2014).
9. Sullivan, K. A. M., Platt, R. N. II, Bradley, R. D. & Ray, D. A. Whole mitochondrial genomes provide increased resolution and indicate paraphyly in deer mice. *BMC Zool.* **2**, 11 (2017).
10. Smith, D. R. RNA-Seq data: a goldmine for organelle research. *Brief. Funct. Genomics* **12**, 454–456 (2013).
11. Smith, D. R. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief. Funct. Genomics* **15**, 47–54 (2016).
12. Sanitá Lima, M. & Smith, D. R. Pervasive transcription of mitochondrial, plastid, and nucleomorph genomes across diverse plastid-bearing species. *Genome Biol. Evol.* **9**, 2650–2657 (2017).
13. Williams, S. T. *et al*. Curius bivalves: systematic utility and unusual properties of anomalodesmatan mitochondrial genomes. *Mol. Phylogenet. Evol.* **110**, 60–72 (2017).
14. Wang, H. L. *et al*. The characteristics and expression profiles of the mitochondrial genome for the Mediterranean species of the *Bemisia tabaci* complex. *BMC Genomics* **14**, 401 (2013).
15. Tian, Y. & Smith, D. R. Recovering complete mitochondrial genome sequences from RNA-Seq: A case study of *Polytomella* non-photosynthetic green algae. *Mol. Phylogenet. Evol.* **98**, 57–62 (2016).
16. Moreira, D. A., Furtado, C. & Parente, T. E. The use of transcriptomic next-generation sequencing data to assembly mitochondrial genomes of *Ancistrus* spp. (Loricariidae). *Gene* **573**(1), 171–175 (2015).
17. Song, N., An, S., Yin, X., Cai, W. & Li, H. Application of RNA-seq for mitogenome reconstruction, and reconsideration of long-branch artifacts in Hemiptera phylogeny. *Sci. Rep.* **6**, 33465 (2016).
18. Bourguignon, T. *et al*. Oceanic dispersal, vicariance and human introduction shaped the modern distribution of the termites *Reticulitermes*, *Heterotermes* and *Coptotermes*. *Proc. R. Soc. B* **283**, 20160179 (2016).
19. Dedeine, F. *et al*. Historical biogeography of *Reticulitermes* termites (Isoptera: Rhinotermitidae) inferred from analyses of mitochondrial and nuclear loci. *Mol. Phylogenet. Evol.* **94**, 778–790 (2016).
20. Perera, P. O., Walsh, T. K. & Luttrell, L. G. Complete mitochondrial genome of *Helicoverpa zea* (Lepidoptera: Noctuidae) and expression profiles of mitochondrial-encoded genes in early and late embryos. *J. Insect Sci.* **16**, 1–10 (2016).
21. Sun, Y., Kurisaki, M., Hashiguchi, Y. & Kumazawa, Y. Variation and evolution of polyadenylation profiles in sauropsid mitochondrial mRNAs as deduced from the high-throughput RNA sequencing. *BMC Genomics.* **18**, 665 (2017).
22. Ghiselli, F. *et al*. The complete mitochondrial genome of the grooved carpet shell, *Ruditapes decussatus* (Bivalvia, Veneridae). *PeerJ.* **5**, e3692 (2017).
23. Konrad, A. *et al*. Mitochondrial mutation rate, spectrum and heteroplasmy in *Caenorhabditis elegans* spontaneous mutation accumulation lines of differing population size. *Mol. Biol. Evol.* **34**, 1319–1334 (2017).
24. Timbó, R. V., Togawa, R. C., Costa, M. M. C., Andow, A. D. & Paula, D. P. Mitogenome sequence accuracy using different elucidation methods. *PLoS One.* **12**, e0179971 (2017).
25. Mercer, T. R. *et al*. The human mitochondrial transcriptome. *Cell* **146**(4), 645–658 (2011).
26. Boore, J. L. & Brown, W. M. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Biol.* **8**, 668–674 (1998).
27. Cameron, S. L., Lo, N., Bourguignon, T., Svenson, G. J. & Evans, T. A. A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): robust support for interfamilial relationships and molecular synapomorphies define major clades. *Mol. Phylogenet. Evol.* **65**, 163–173 (2012).
28. Zhang, D. *et al*. Mitochondrial architecture rearrangements produce asymmetrical nonadaptive mutational pressures that subvert the phylogenetic reconstruction in Isopoda. *Genome Biol. Evol.* **11**, 1797–1812 (2019).
29. Romiguier, J. *et al*. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263 (2014).
30. Mitaka, Y., Kobayashi, K. & Matsuura, K. Caste-, sex-, and age-dependent expression of immune-related genes in a Japanese subterranean termite, *Reticulitermes speratus*. *PLoS One* **12**, e0175417–22 (2017).
31. Sanitá Lima, M., Woods, L. C., Cartwright, M. W. & Smith, D. R. The (in)complete organelle genome: exploring the use and nonuse of available technologies for characterizing mitochondrial and plastid chromosomes. *Mol. Ecol. Res.* **16**, 1279–1286 (2016).
32. Castandet, B., Hotto, A. M., Strickler, S. R. & Stern, D. B. ChloroSeq, an optimized chloroplast RNA-Seq bioinformatic pipeline, reveals remodeling of the organellar transcriptome under heat stress. G3-Genes Genom Genet. *G3-GENES GENOM. GENET.* **6**, 2817–2827 (2016).
33. Andrews, S. Fast QC, A Quality Control tool for High Throughput Sequence Data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed January 2019 (2014).
34. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
35. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie2. *Nat. Methods* **9**, 357–359 (2012).
36. Haas, B. J. *et al*. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
37. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
38. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
39. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).
40. Katoh, K., Kuma, K. I., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**(2), 511–518 (2005).
41. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
42. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
43. Bernt, M. *et al*. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet. Evol.* **69**, 313–319 (2013).
44. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 November 2010. New Orleans, LA. pp 1–8 (2010).
45. Kalyaanamoorthy, S. *et al*. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).
46. Nguyen, L.-T., Schmidt, H. A., Haeseler, von, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol.* **32**, 268–274 (2014).
47. Ronquist, F. *et al*. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

Acknowledgements
This work was supported by Canziani funding to A.L. and B.M. T.B. was supported by the Czech Science Foundation (Project No. 15-07015Y). We wish to thank Fabrizio Ghiselli for his precious suggestions and two anonymous Reviewers for their useful and constructive comments.

Author Contributions
G.F. and G.P. set up the bioinformatic pipeline and wrote the script. G.F. and A.L. carried out analyses and wrote the manuscript. A.L., T.B., T.E. and O.R.S. produced part of the reference mitogenomes. A.L., T.B., T.E., O.R.S. and B.M. supervised the work. G.F., A.L. and B.M. conceived and developed the idea. All authors reviewed the manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-51313-7.

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Complete mitochondrial genomes from transcriptomes: assessing pros and cons of data mining for assembling new mitogenomes.

Giobbe Forni, Guglielmo Puccio, Thomas Bourguignon, Theodore Evans, Barbara Mantovani, Omar Rota-Stabelli, Andrea Luchetti

Table S1. Contig filtering for contaminants, based on NCBI *taxid* (indicated in parentheses).

	R. labralis **congeneric**	***R. labralis*** **intra-familial**	***R. speratus*** **congeneric**	***R. speratus*** **intra-familial**
Total contigs	34439	33906	262	157
Total hits with e-value ≤ 0.001	11285	11013	157	95
No hits	67.23%	67.52%	40.08%	39.49%
Bacteria (2)	74	83	0	0
Viridiplantae (33090)	161	123	0	0
Fungi (4751)	89	125	0	0
Vertebrata (7742)	216	192	4	4
Hexapoda (6960) - excluding termites (1912919)	880	830	22	8
Insecta (50557)	9427	9264	134	75
Reticulitermes (36988)	187	210	36	20
Contaminants or unassigned	72.63%	72.68%	48.85%	52.23%

Table S2. Control regions (CR) length in reference and RNA-Seq derived mitogenomes (using congeneric or intra-familial reference in the first step of the first iteration).

Species	Genbank acc. nos.	CR length (bp)	CR RNA-Seq from congeneric ref. (bp)	CR RNA-Seq from intra-familial ref. (bp)
Reticulitermes flavipes	KY484910	1099	1194	107
	EF206314	1751		
	EF206315	1751		
	EF206316	1751		
	EF206317	1751		
Reticulitermes grassei	KU925237	14	926	n\a
Reticulitermes banyulensis	-	n\a	1024	n\a
Reticulitermes lucifugus	MK088051	1471	1340	293
Reticulitermes labralis	KT224427	1308	1127	736
	KU877221	1098		
Reticulitermes speratus	KY484910	1098	955	113

Table S3. Nucleotide substitutions between RNA-Seq derived mitogenomes obtained with either congeneric or intrafamilial starting references.

Species	Number of differences[a]	Position (nt)	Gene	Strand	Congeneric variant	Intra-familial variant	Outcome congeneric	Outcome intra-familial	Coverage congeneric	Coverage intra-familial
Reticulitermes flavipes	0	-	-		-	-	-	-	-	-
Reticulitermes lucifugus	4	9722	ND4L	-	G	A	AGC (s)	AGT (s)	713	732
		9800			A	T	TTT (f)	TTA (l)	50	61
		9803			A	C	TGT (c)	TGG (w)	42	52
		3031	tRNA-Leu	+	T	C			421	148
Reticulitermes grassei	4	14825	12S	+	A	T			1366	1331
		388	ND2	+	A	G	ACA (t)	ACG (t)	1617	1559
		548			G	A	GGG (g)	GGA (g)	1859	1616
		6307	tRNA-Glu	+	A	C			8	7
Reticulitermes banyulensis	4	14793	12S	+	A	C			399	661
		14795			C	A			337	640
		14796			T	C			318	633
		14797			C	T			304	619
Reticulitermes speratus	7	14576	12S	+	A	T			1087	1052
		14579			-	A			-	1031
		14580			-	A			-	1033
		14272			-	C			-	1004
		6479	ND5	-	A	C	TTT (f)	TTG (l)	635	998
		6538			G	C	CAT (h)	GAT (f)	359	836
		6241	tRNA-Ser	+	C	G			8	11
Reticulitermes labralis	6(4)	14246	12S	+	T	-			182	-
		14247			C	A			3701	3827
		11580	beteween CYTB and tRNA-Ser2		C	G			34	24

[a] differences excluding CR in parentheses

Table S4. Nucleotide substitutions (% divergence) between conspecific reference sequences drawn from Genbank versus RNA-Seq derived mitogenomes obtained with either congeneric or intrafamilial references in the first step of the first iteration.

Species	**Conspecific reference**	***vs* RNA Seq (congeneric)**	***vs* RNA Seq (intrafamilial)**
Reticulitermes flavipes	EF206314	46 (0.3%)	47 (0.3%)
	EF206315 [a]	208 (1.3%)	189 (1.3%)
	EF206316	325 (2.2%)	326 (2.2%)
	EF206317	323 (2.2%)	324 (2.2%)
	KU925236 [b]	248 (2.5%)	248 (2.5%)
Reticulitermes grassei	KU925237	69 (0.5%)	72 (0.5%)
Reticulitermes lucifugus	MK088051	58 (0.4%)	55 (0.4%)
Reticulitermes labralis	KT224427	3 (0.0%)	3 (0.0%)
	KU877221	511 (3.5%)	511 (3.5%)
Reticulitermes speratus	KY484910	115 (0.8%)	112 (0.8%)

[a] *R. santonensis* is synonym species of *R. flavipes*

[b] incomplete sequence

Table S5. Number of nucleotide substitutions in protein-coding genes between RNA-Seq derived mitogenomes obtained with either congeneric or intra-familial starting references and the closest mitogenome obtained by DNA-sequencing.

	***vs* Closest reference**	**1st**	**2nd**	**3rd**
PCGs - all *Reticulitermes*		1059.4	497.7	2985.9
R. flavipes (congeneric)	EF206314	5	4	30
R. flavipes (intra-familial)	EF206314	6	4	30
R. grassei (congeneric)	KU925237	11	4	43
R. grassei (intra-familial)	KU925237	11	4	45
R. lucifugus (congeneric)	MK088051	9	5	34
R. lucifugus (intra-familial)	MK088051	9	5	31
R. labralis (congeneric)	KT224427	0	0	0
R. labralis (intra-familial)	KT224427	0	0	0
R. speratus (congeneric)	KY484910	16	9	63
R. speratus (intra-familial)	KY484910	15	9	62

Table S6. Mitochondrial genomes used as reference (those used for the phylogenetic analysis are marked with an asterisk).

Species	Genbank accession numbers
Coptotermes acinaciformis	KU925196
Coptotermes acinaciformis	KU925197
Coptotermes acinaciformis	KU925198
Coptotermes acinaciformis	KU925199
Coptotermes amanii	KU925200*
Coptotermes elisae	KU925201
Coptotermes formosanus	KU925203*
Coptotermes formosanus	AB626145
Coptotermes frenchi	KU925204
Coptotermes gestroi	KU925205*
Coptotermes heimi	KU925206
Coptotermes heimi	KU925207
Coptotermes heimi	KU925208
Coptotermes kalshoveni	KU925209
Coptotermes kalshoveni	KU925210
Coptotermes lacteus	KU925211
Coptotermes lacteus	JX144934
Coptotermes michaelseni	KU925212
Coptotermes sepangensis	KU925215
Coptotermes sjoestedti	KU925216
Coptotermes sjoestedti	KU925217
Coptotermes testaceus	KU925218
Coptotermes testaceus	KU925219
Coptotermes testaceus	KR872938
Coptotermes travians	KU925221
Coptotermes travians	KU925222
Heterotermes cf. occiduus	KU925229
Heterotermes cf. occiduus	KU925230
Heterotermes cf. paradoxus	KU925223
Heterotermes cf. paradoxus	KU925224
Heterotermes cf. paradoxus	KU925225
Heterotermes crinitus	KU925226
Heterotermes malabaricus	KU925227
Heterotermes nr. tenuis	KU925228
Heterotermes platycephalus	KU925231
Heterotermes tenuior	KU925232
Heterotermes tenuis	KU925233*
Heterotermes vagus	KU925234*

Heterotermes validus	KU925235*
Reticulitermes aculabilis	KP334994*
Reticulitermes chinensis	KM216388*
Reticulitermes flaviceps	KX712090*
Reticulitermes flavipes	EF206316*
Reticulitermes flavipes	EF206317*
Reticulitermes flavipes	EF206314*
Reticulitermes grassei	KU925237*
Reticulitermes hageni	EF206320*
Reticulitermes kanmonensis	MF063063*
Reticulitermes labralis	KU877221*
Reticulitermes labralis	KT224427*
Reticulitermes lucifugus	MK088051*
Reticulitermes nelsonae	KU92523*
Reticulitermes santonensis	EF206315[a]*
Reticulitermes sp.	KU925239*
Reticulitermes speratus	KY484910*
Reticulitermes tibialis	MK088052*
Reticulitermes tibialis	MK088053*
Reticulitermes virginicus	EF206318*
Reticulitermes virginicus	EF206319*

[a] *R. santonensis* is synonym species of *R. flavipes*

Table S7. Best partition scheme and substitution models for the Maximum Likelihood search in IQ-TREE.

Partition	Subset	Best Model
1	rrnS+rrnL	GTR+F+I+G4
2	atp6_1st+cox2_1st+cox3_1st	TIM+F+G4
3	atp6_2nd+nad1_2nd+nad4_2nd+nad4L_2nd+nad5_2nd	TIM3+F+I+G4
4	atp6_3rd+atp8_3rd+cox2_3rd+cox3_3rd+cob_3rd+nad3_3rd+nad6_3rd	HKY+F+I+G4
5	atp8_1st+atp8_2nd+cob_1st+nad2_1st+nad3_1st+nad6_1st	TIM2+F+I+G4
6	cox1st	TIM2e+G4
7	cox1nd	HKY+F+I
8	cox1_3rd+nad2_3rd	TVM+F+G4
9	cox2_2nd+cox3_2nd+cob_2nd+nad2_2nd+nad3_2nd+nad6_2nd	TPM3+F+I+G4
10	nad1_1st+nad4_1st+nad4L_1st+nad5_1st	HKY+F+G4
11	nad1_3rd+nad4_3rd+nad4L_3rd+nad5_3rd	GTR+F+I+G4

Table S8. Best partition scheme and substitution models for the Bayesian Inference in MrBayes.

Partition	Subset	Best Model
1	rrnS+rrnL	GTR+F+I+G4
2	atp6_1st+cox1_1st+cox2_1st+cox3_1st	GTR+F+G4
3	atp6_2nd+nad1_2nd+nad4_2nd+nad4L_2nd+nad5_2nd	GTR+F+I+G4
4	atp6_3rd+atp8_3rd+cox2_3rd+cox3_3rd+cob_3rd+nad3_3rd+nad6_3rd	HKY+F+I+G4
5	atp8_1st+atp8_2nd+nad6_1st	HKY+F+G4
6	cox1nd	HKY+F+I
7	cox1_3rd+nad2_3rd	GTR+F+G4
8	cox2_2nd+cox3_2nd+cob_2nd+nad2_2nd+nad3_2nd+nad6_2nd	HKY+F+I+G4
9	cob_1st+nad2_1st+nad3_1st	HKY+F+I+G4
10	nad1_1st+nad4_1st+nad4L_1st+nad5_1st	HKY+F+G4
11	nad1_3rd+nad4_3rd+nad4L_3rd+nad5_3rd	GTR+F+I+G4

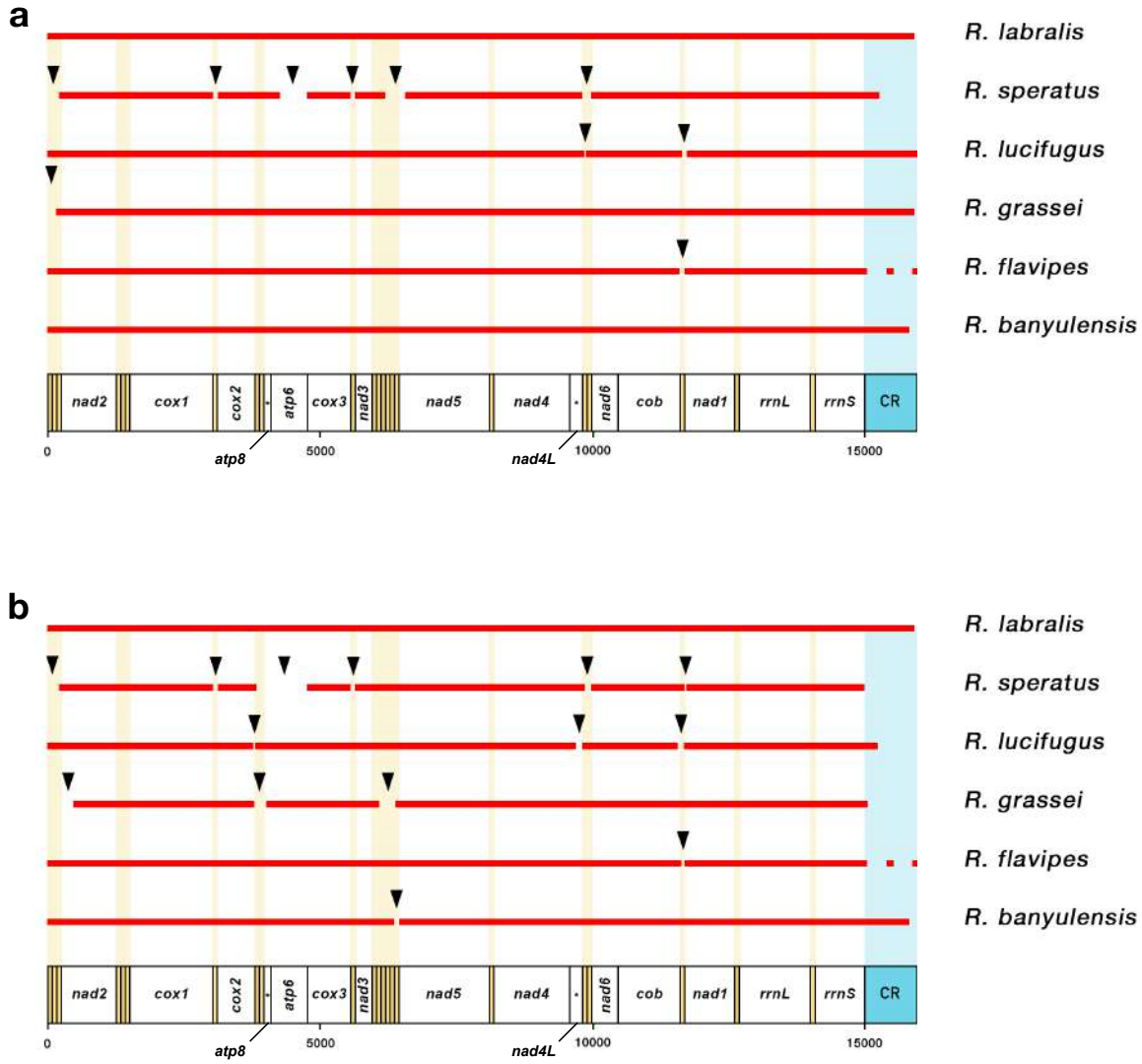
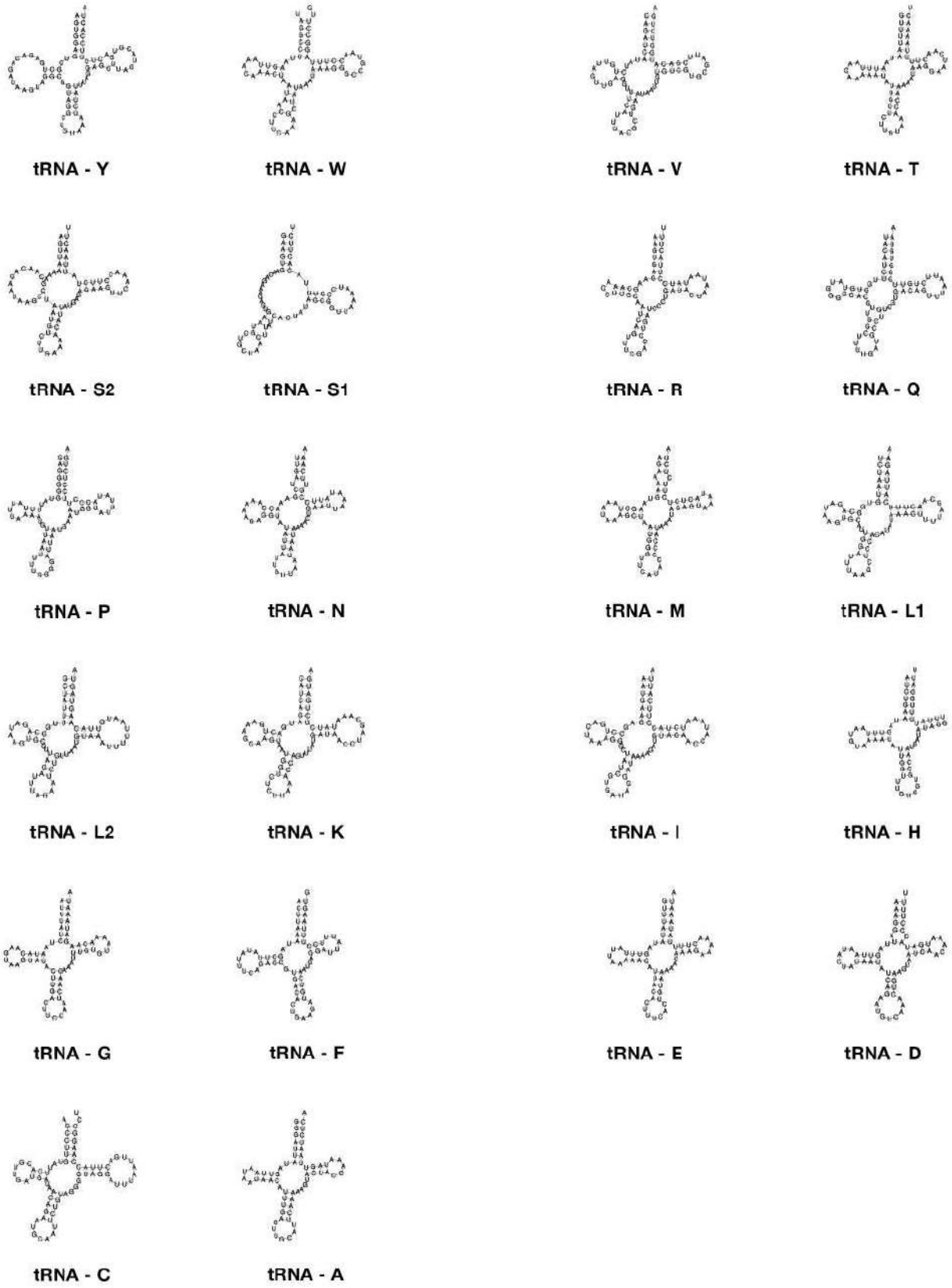


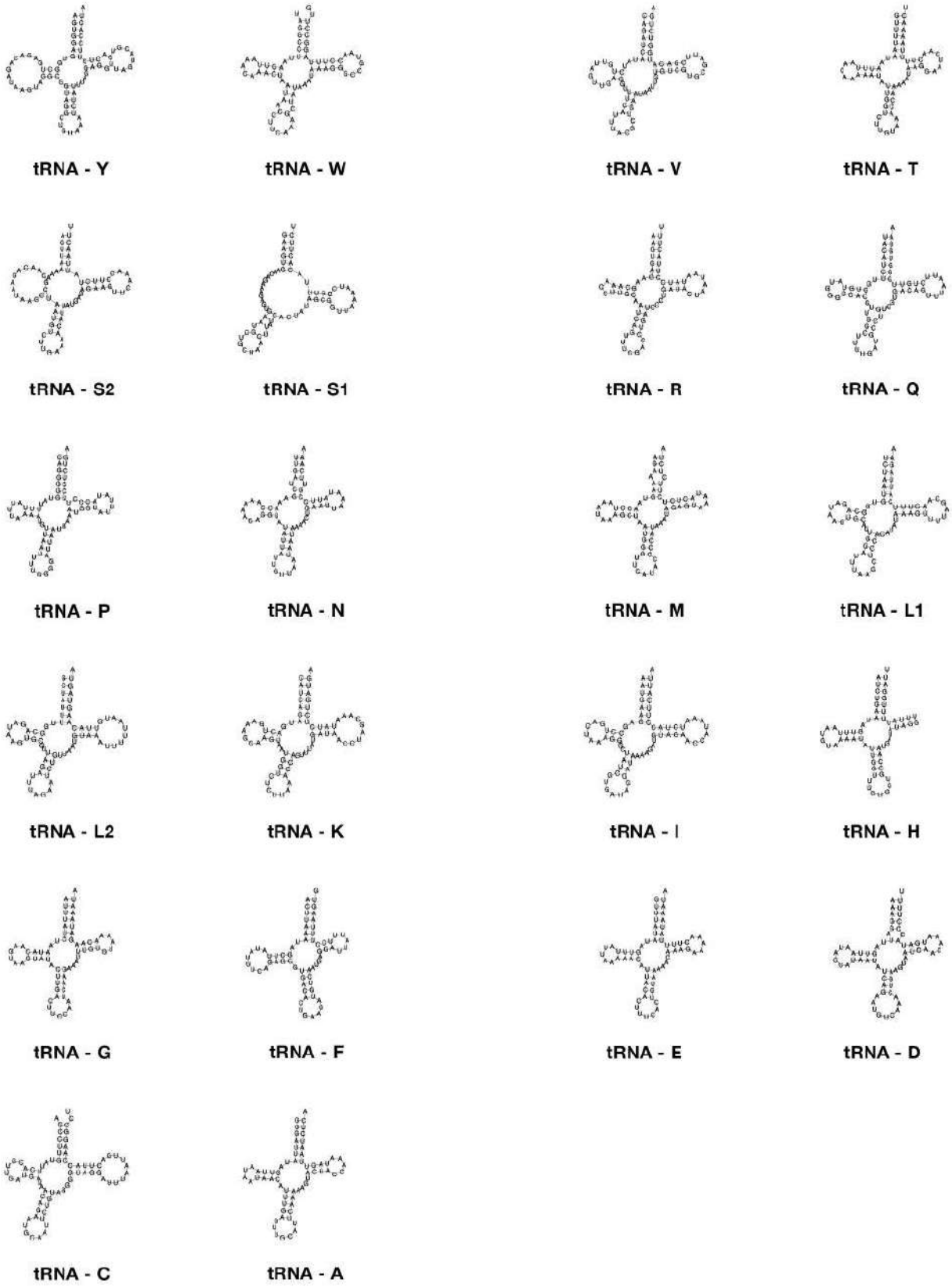
Figure S1. Localization of gap regions, after the 10th iteration, within reconstructed mitogenomes using a) congeneric and b) intra-familial references

Figure S2. tRNAs secondary structure of reconstructed and reference mitogenomes



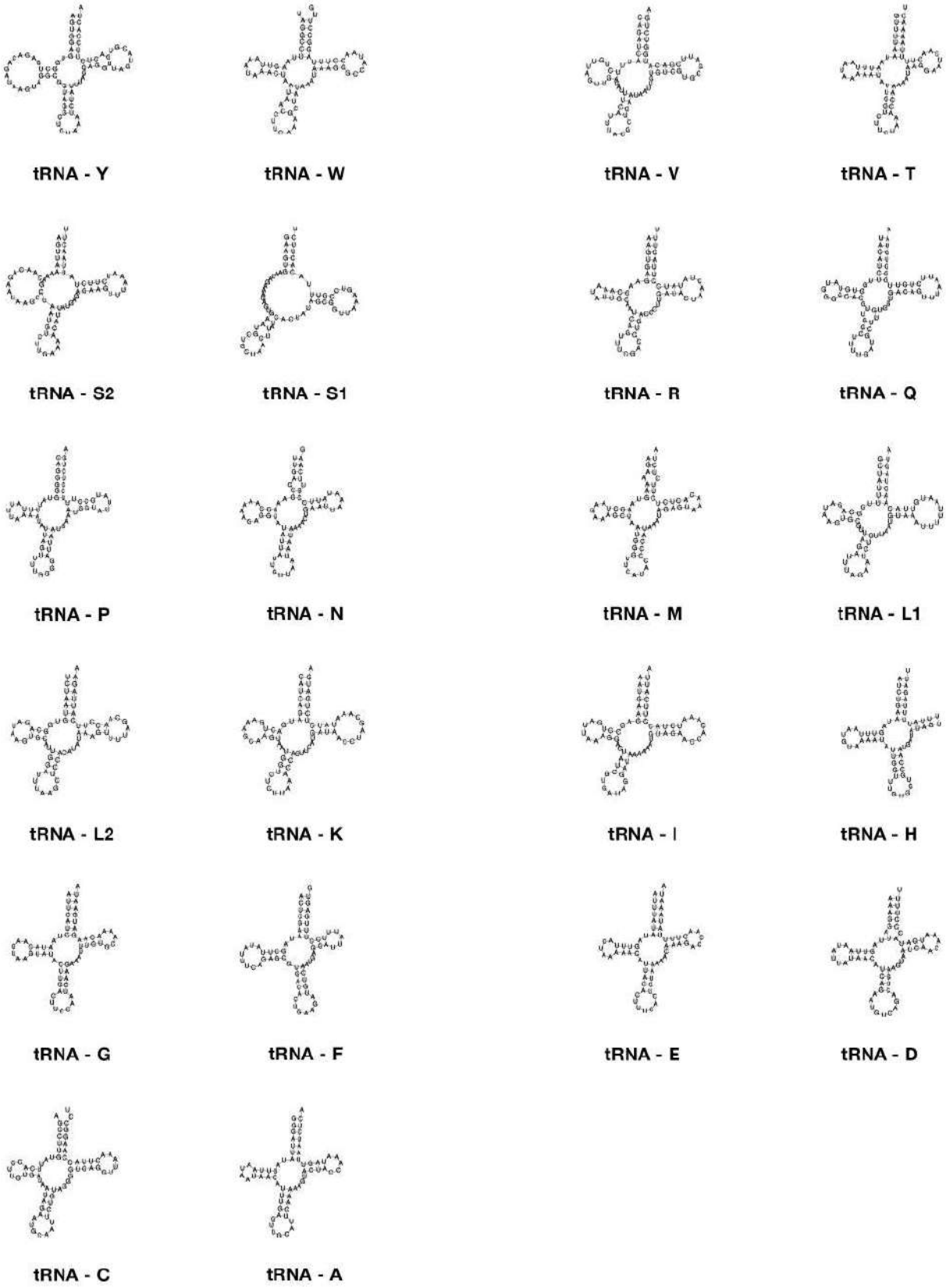
Reticulitermes banyulensis - intrafamiliar starting references

Figure S2. Continued



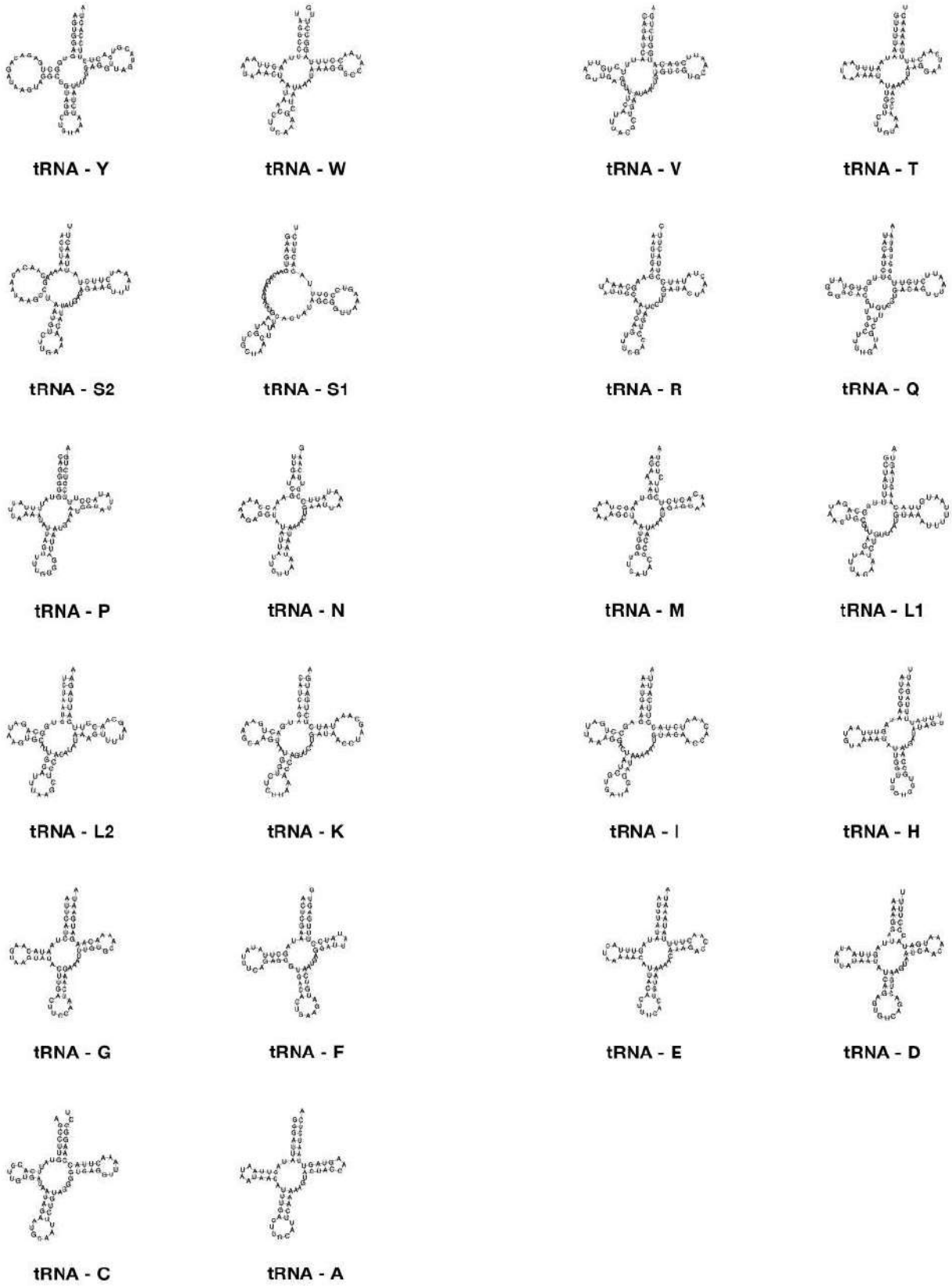
Reticulitermes banyulensis - congeneric starting references

Figure S2. Continued



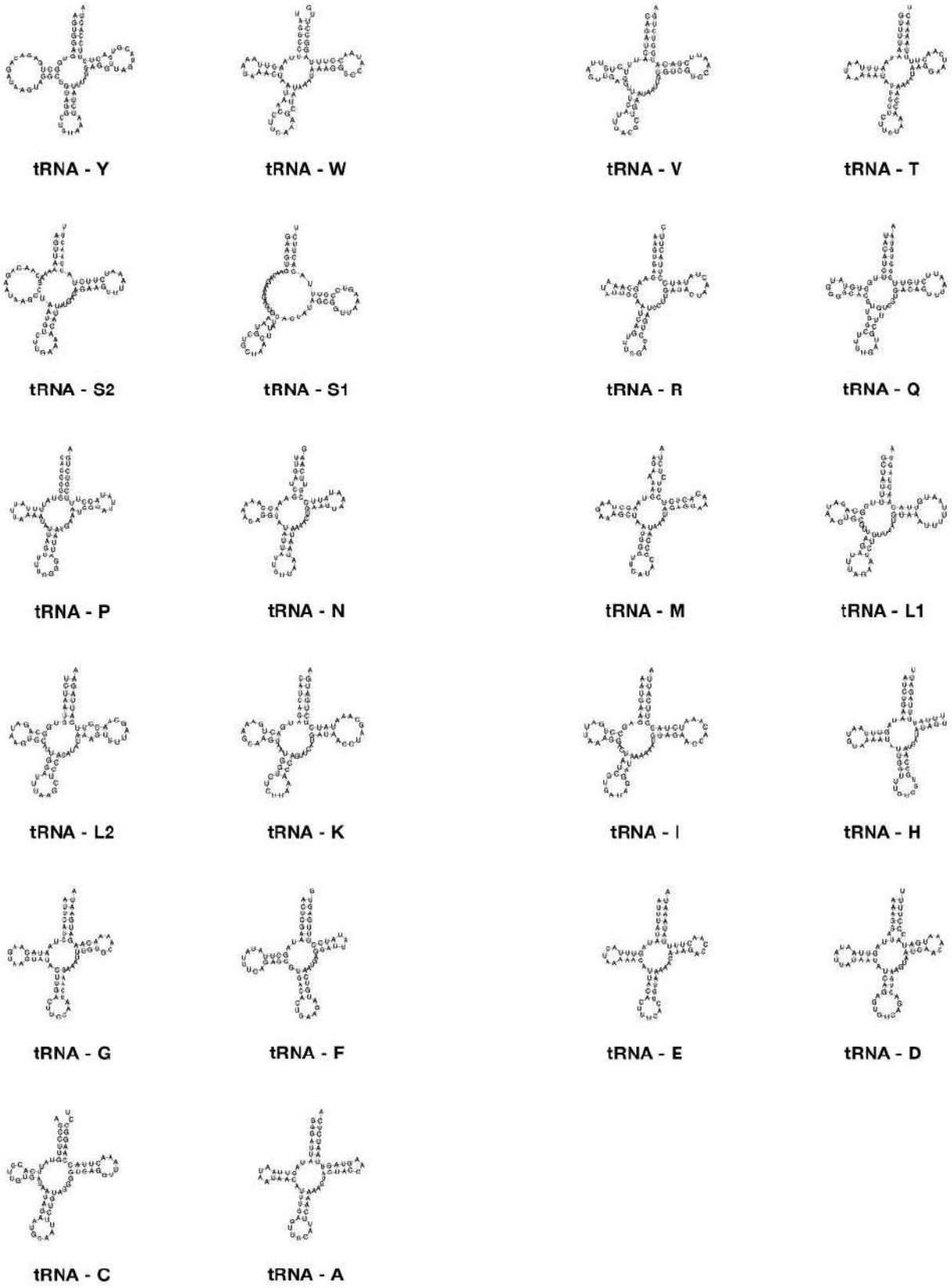
Reticulitermes flavipes EF206314.1

Figure S2. Continued



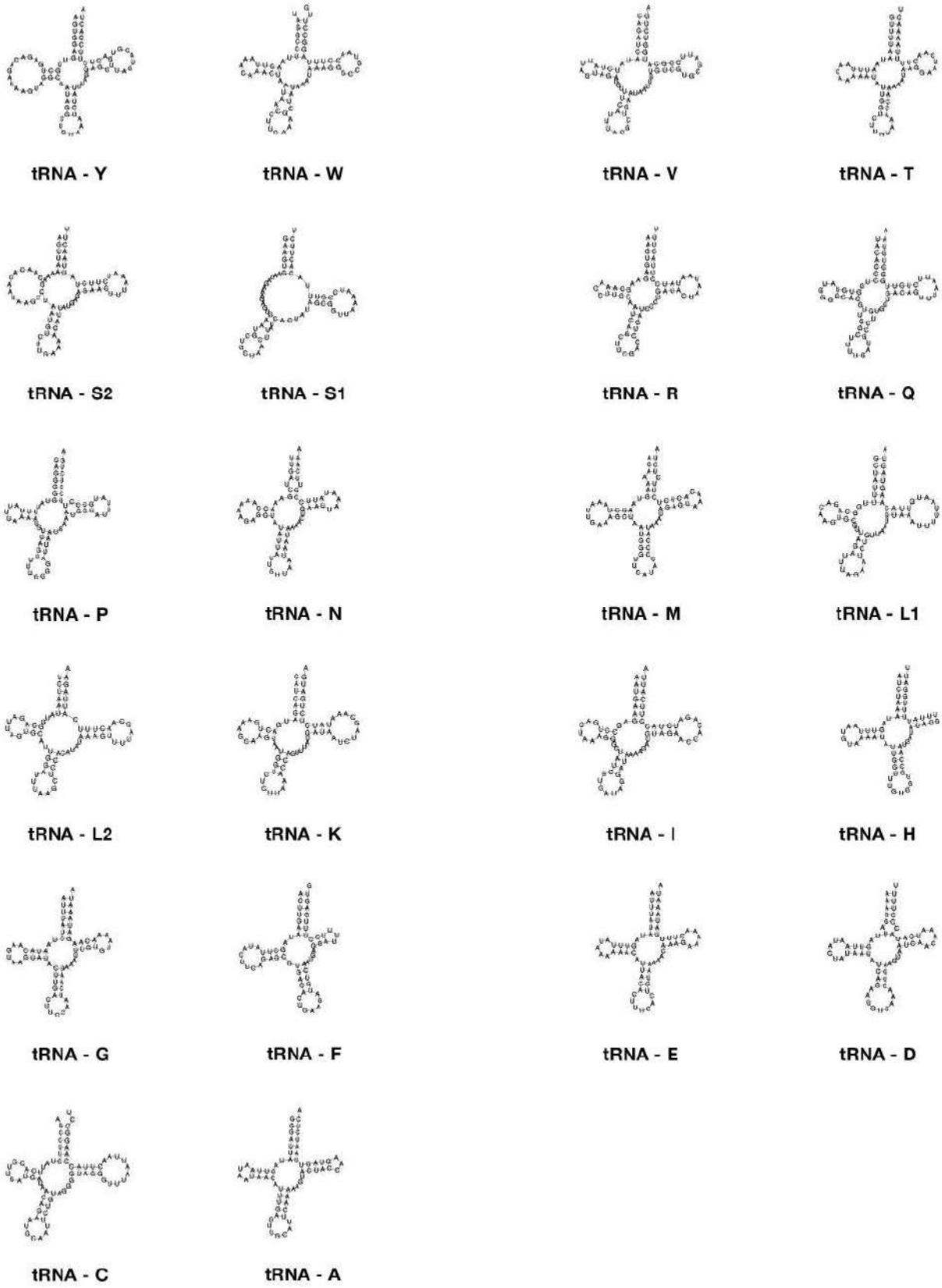
Reticulitermes flavipes - intrafamiliar starting references

Figure S2. Continued



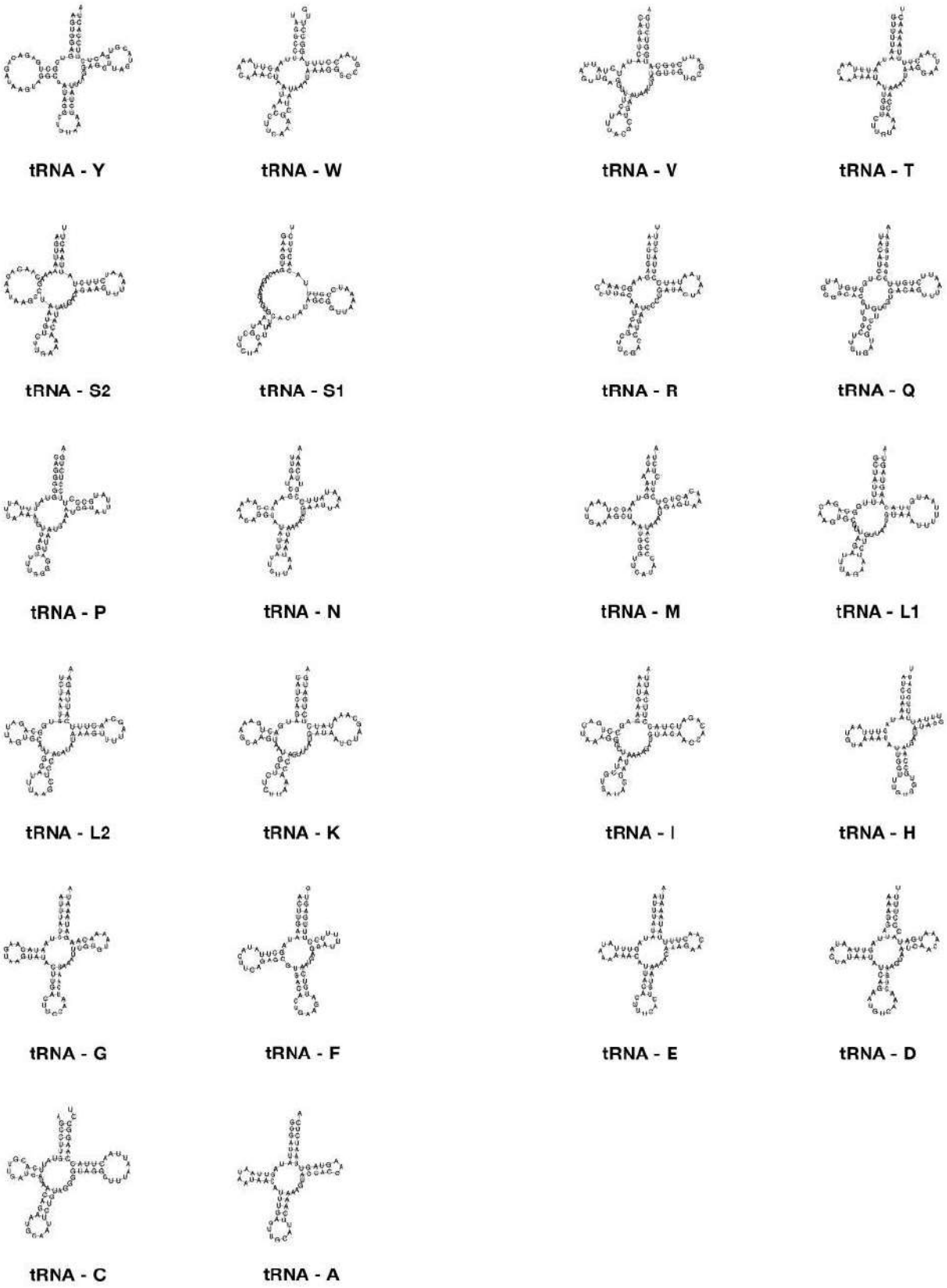
Reticulitermes flavipes - congeneric starting references

Figure S2. Continued



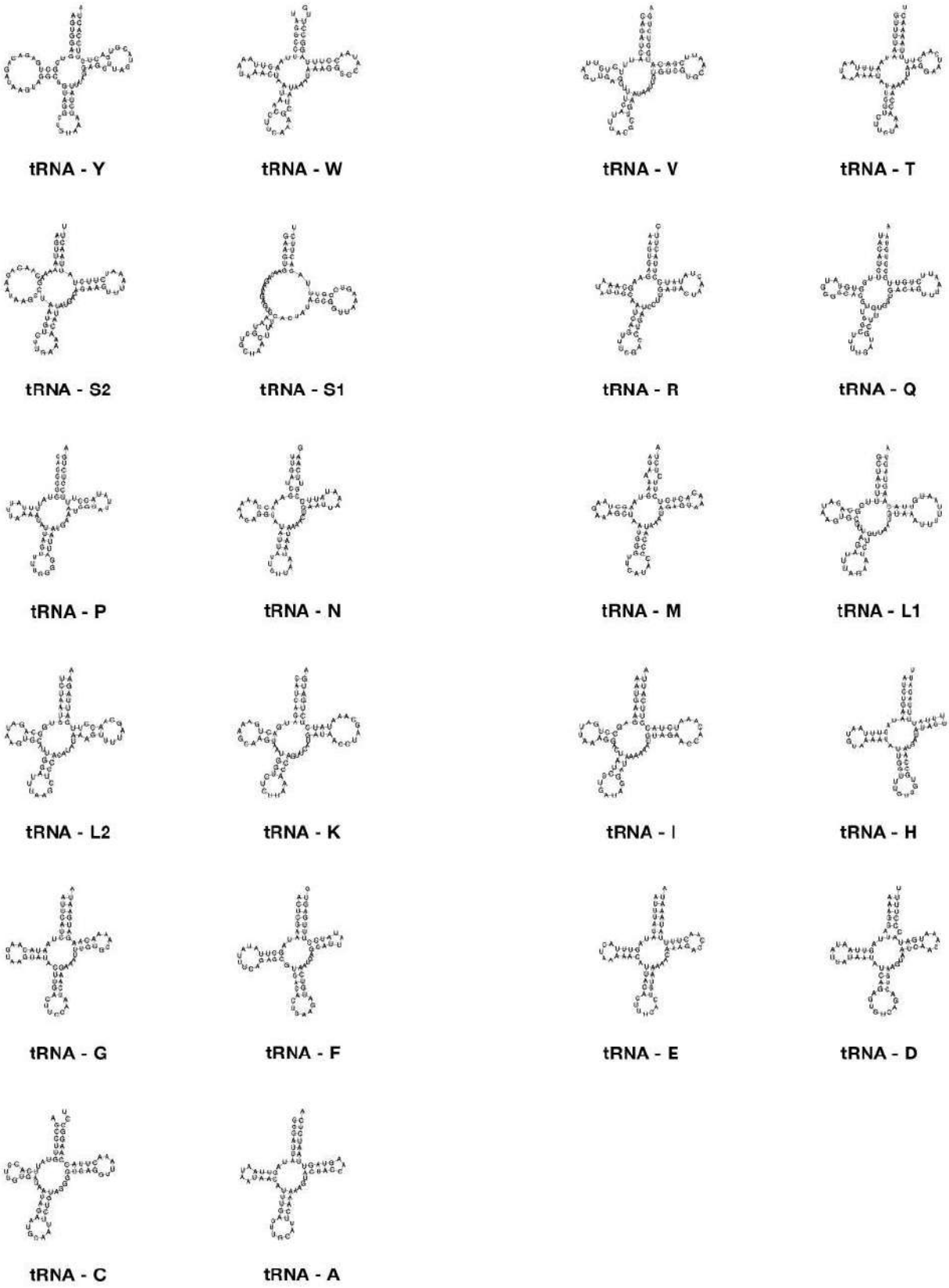
Reticulitermes grassei KU925237.1

Figure S2. Continued



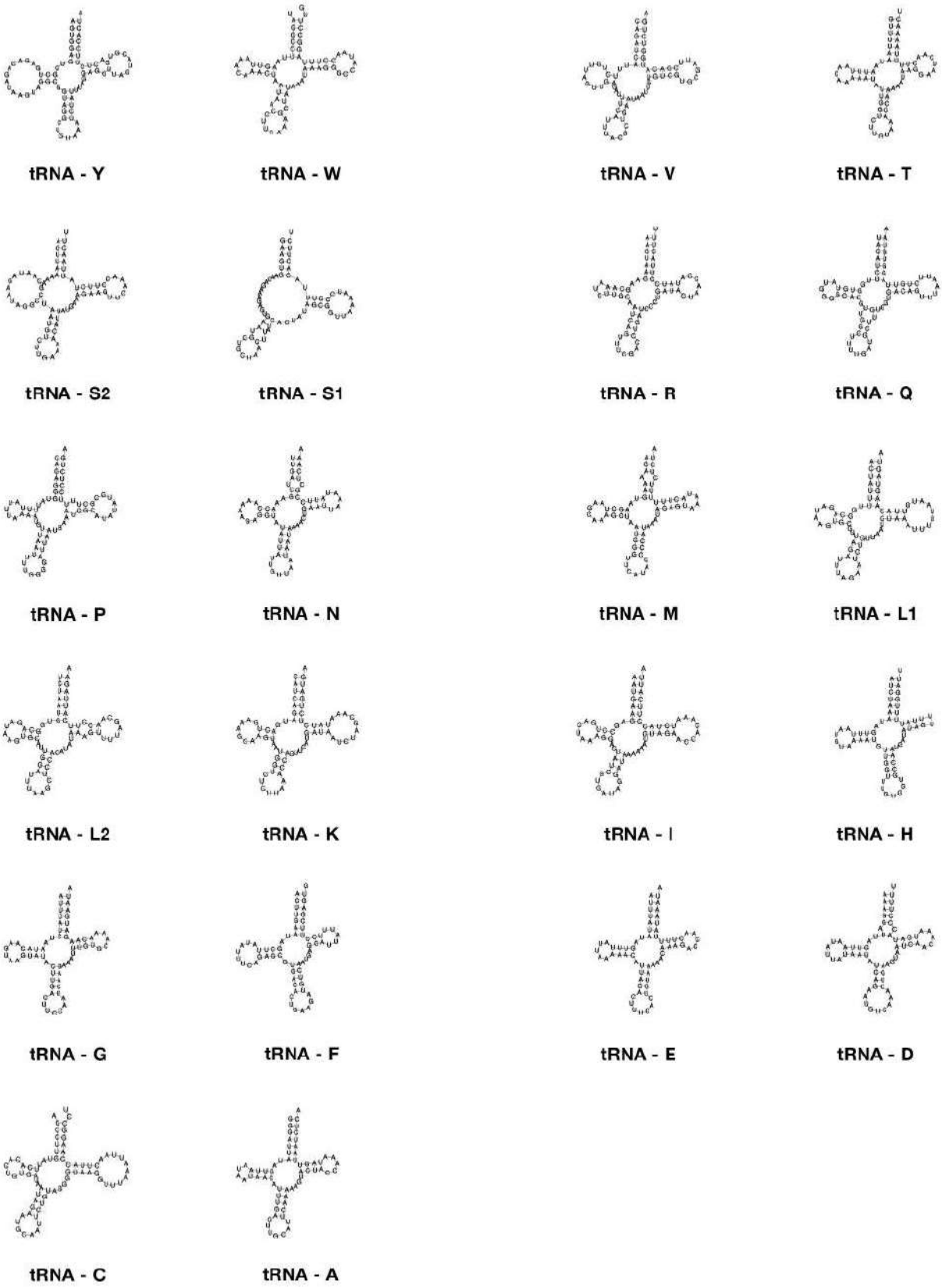
Reticulitermes grassei - intrafamiliar starting references

Figure S2. Continued



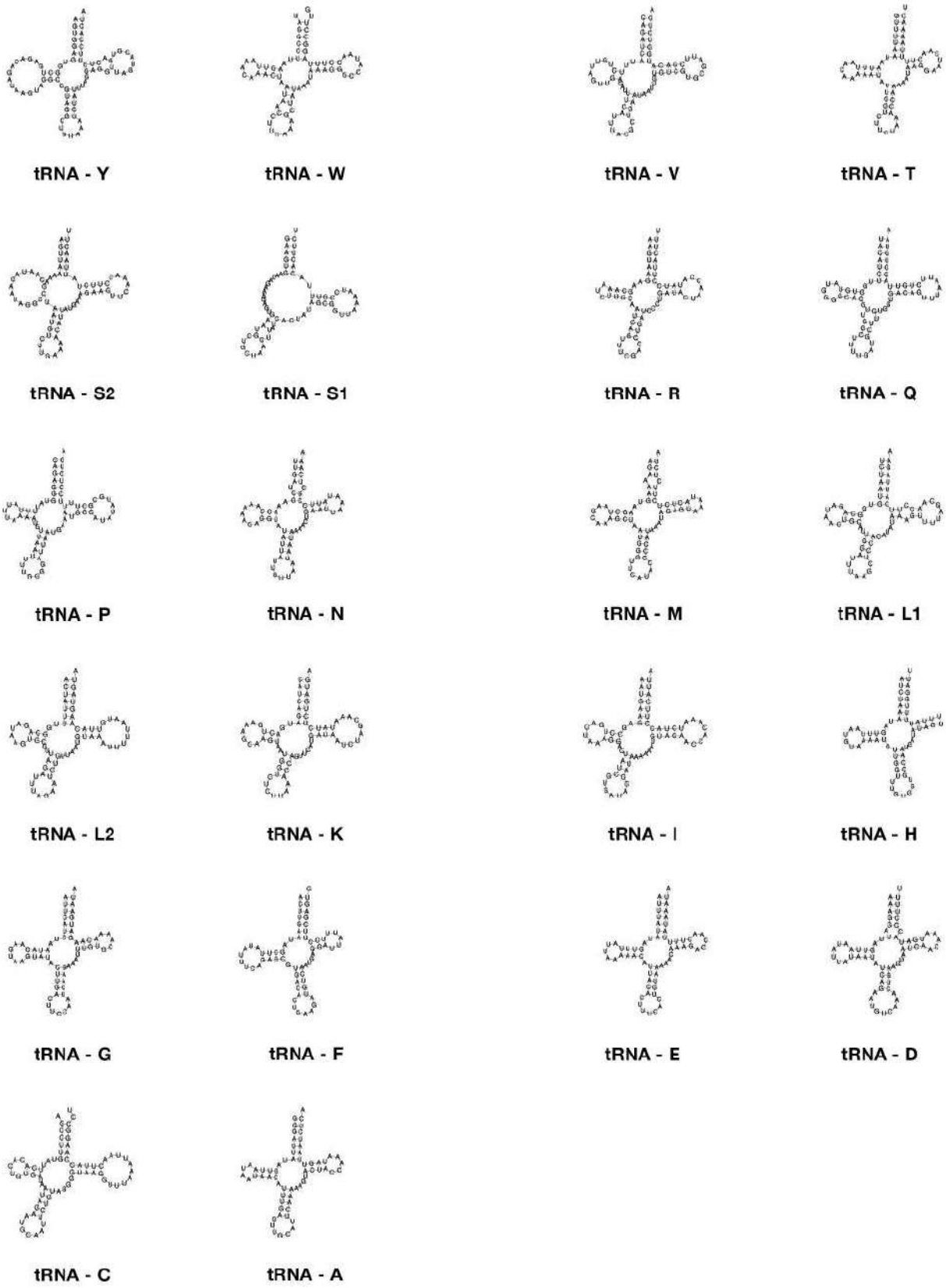
Reticulitermes grassei - congeneric starting references

Figure S2. Continued



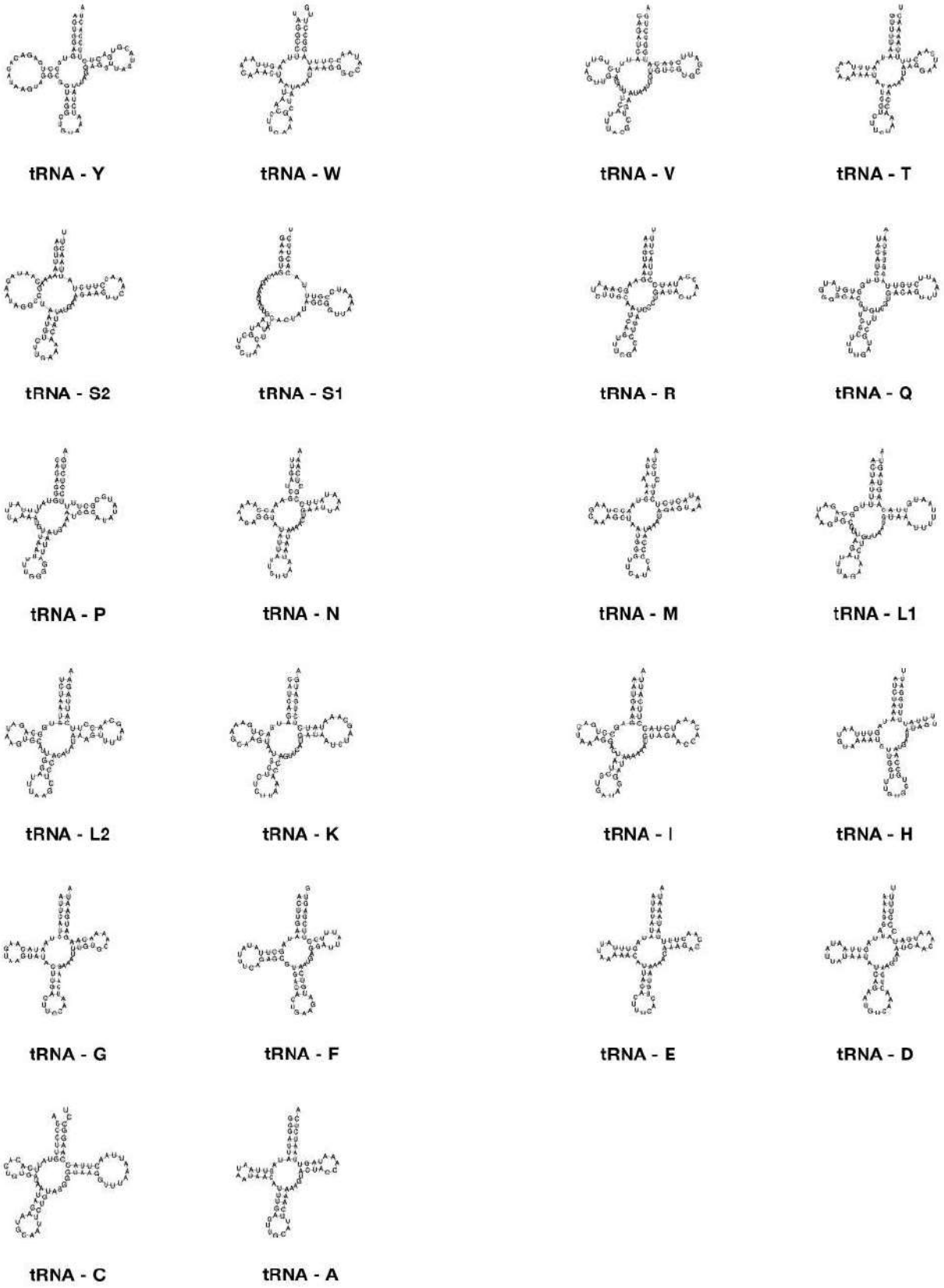
Reticulitermes labralis KT224427.1

Figure S2. Continued



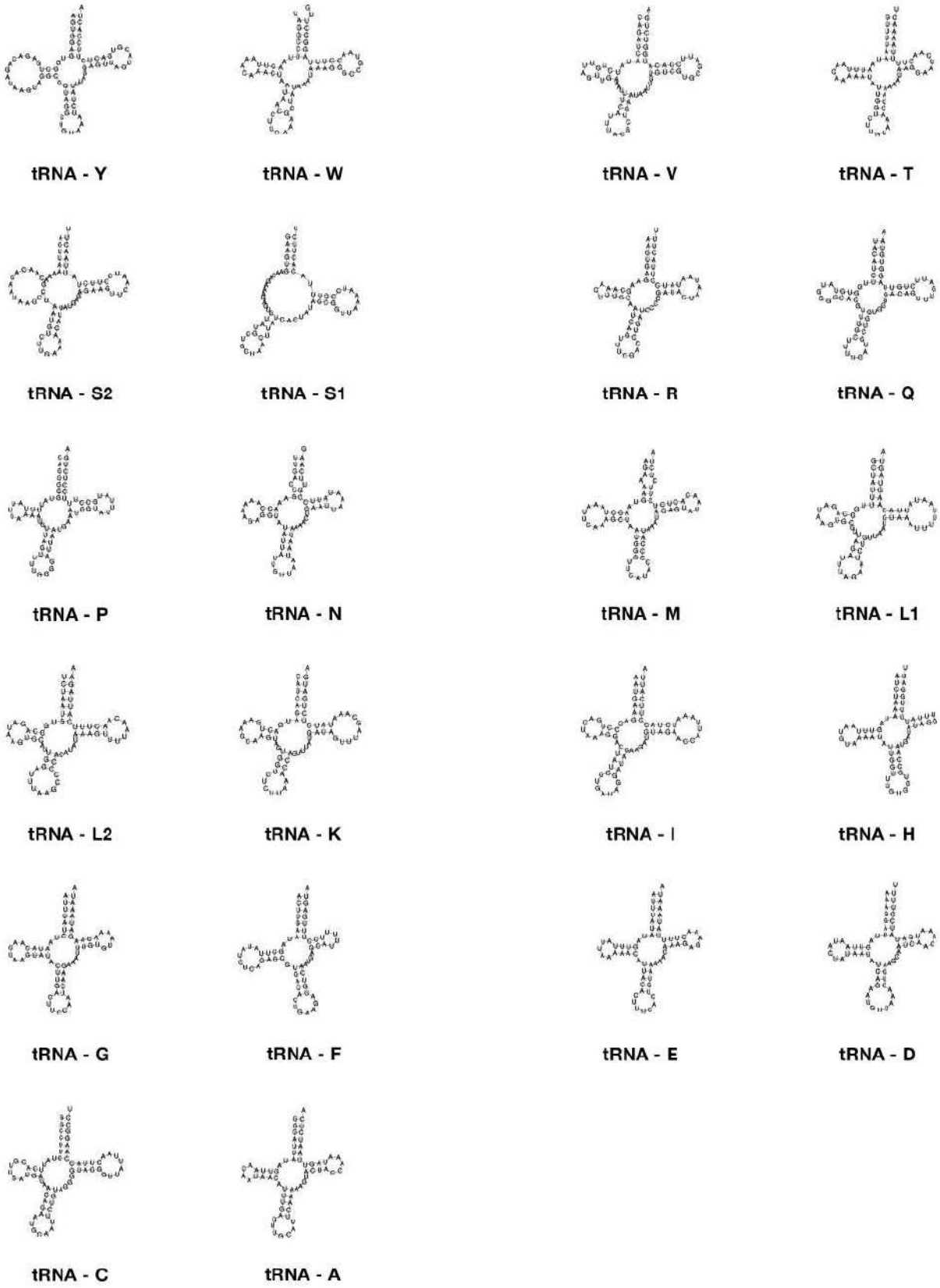
Reticulitermes labralis - intrafamiliar starting references

Figure S2. Continued



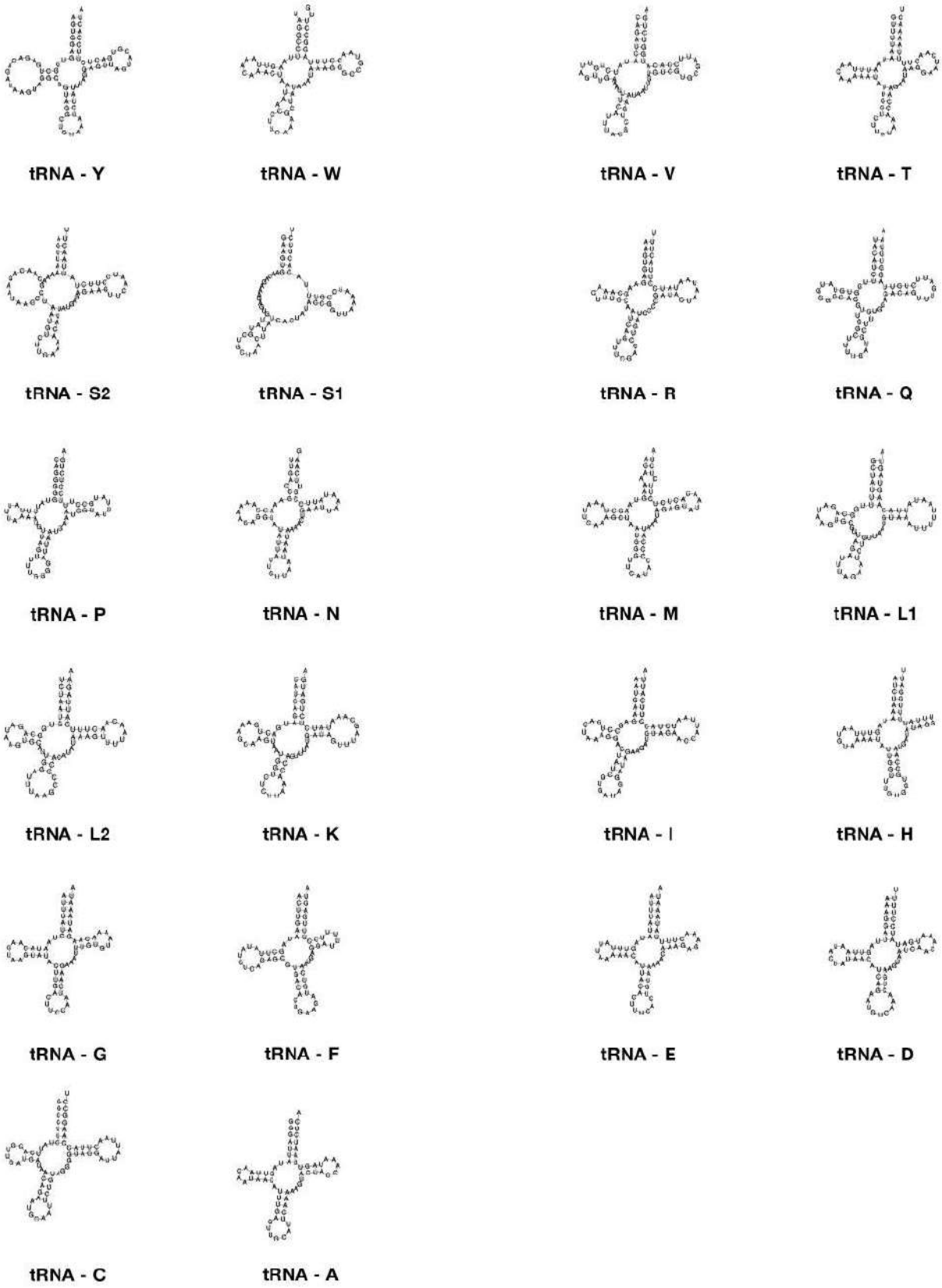
Reticulitermes labralis - congeneric starting references

Figure S2. Continued



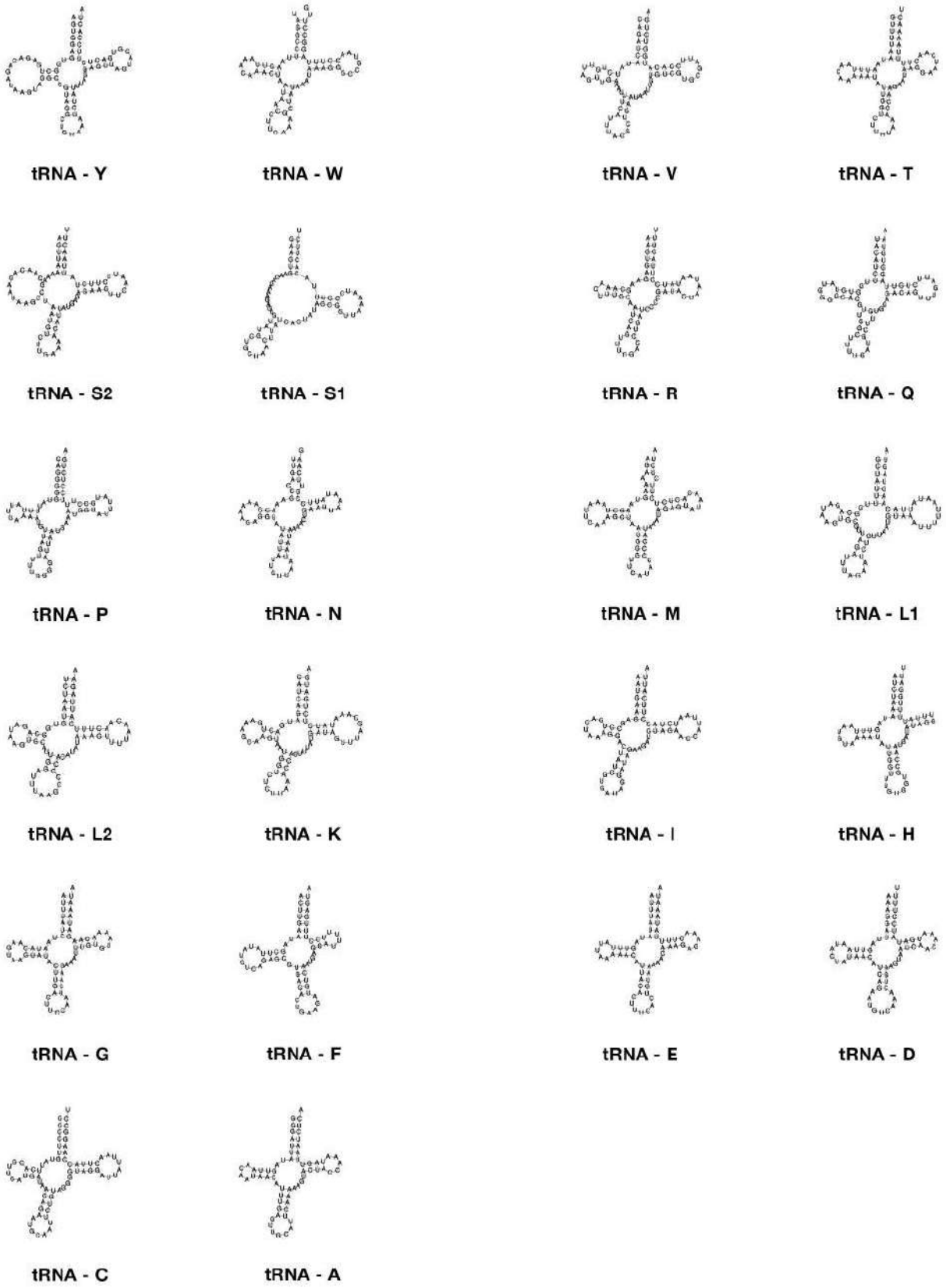
Reticulitermes lucifugus MK088051.1

Figure S2. Continued



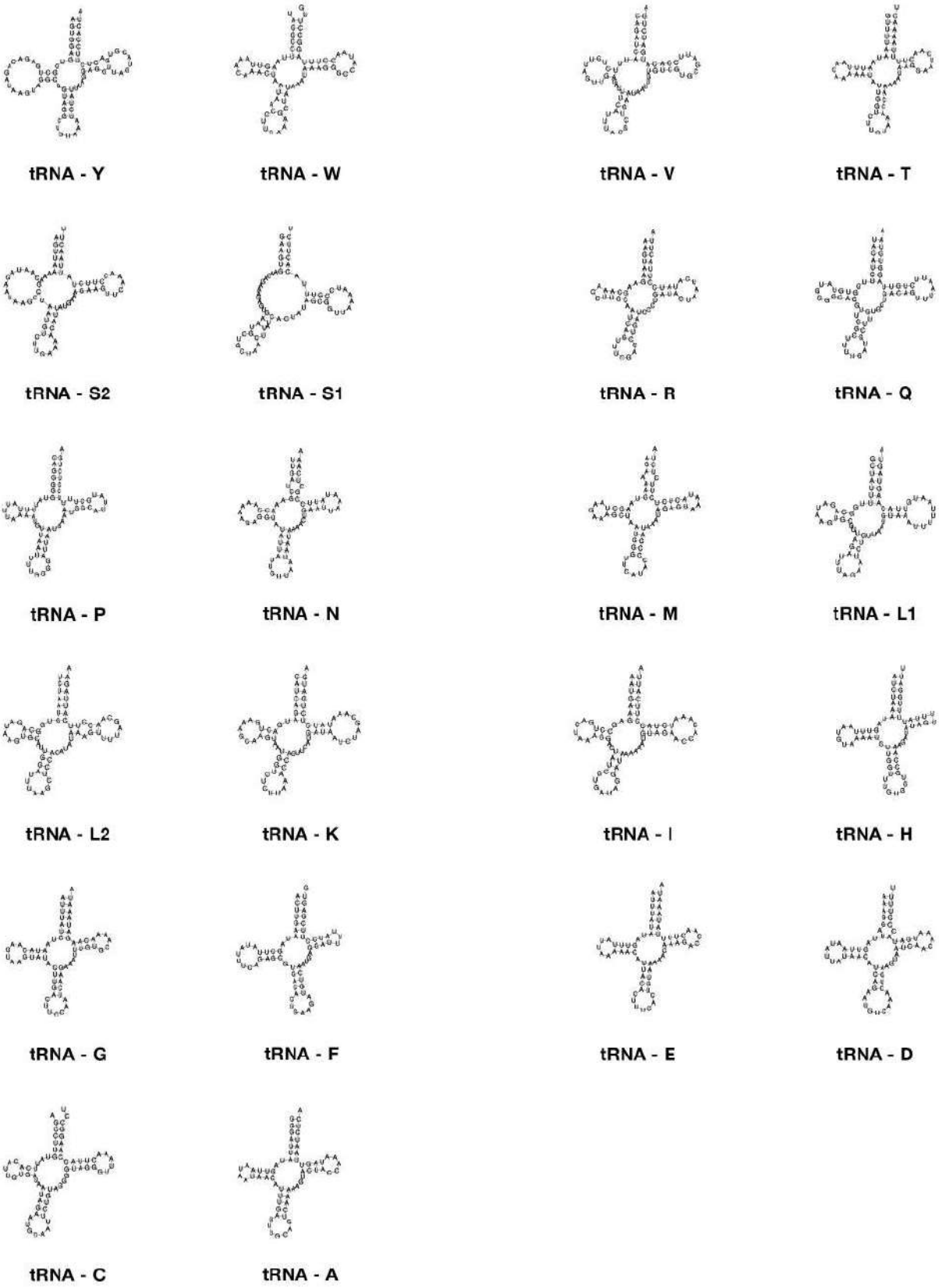
Reticulitermes lucifugus - intrafamiliar starting references

Figure S2. Continued



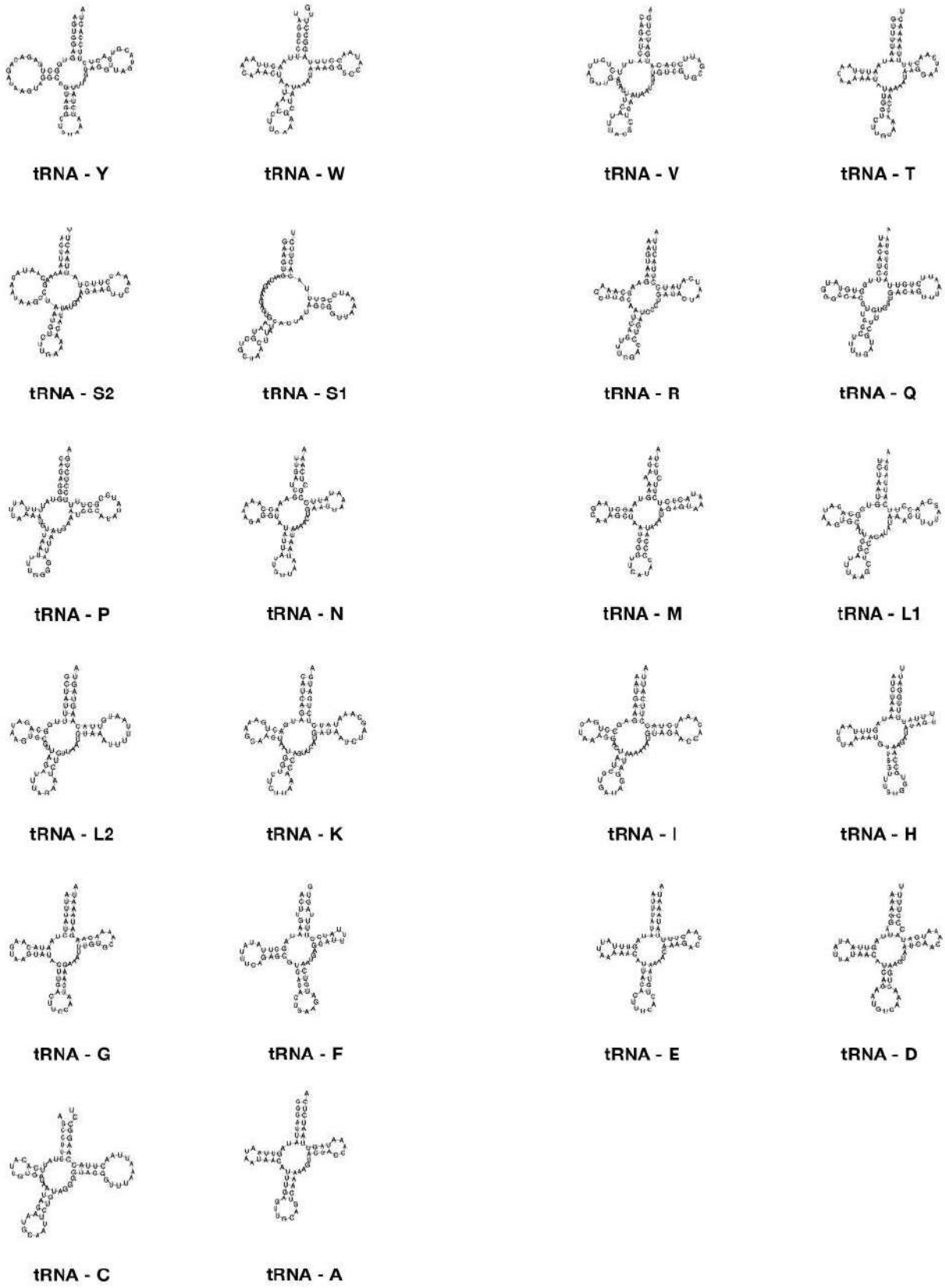
Reticulitermes lucifugus - congeneric starting references

Figure S2. Continued



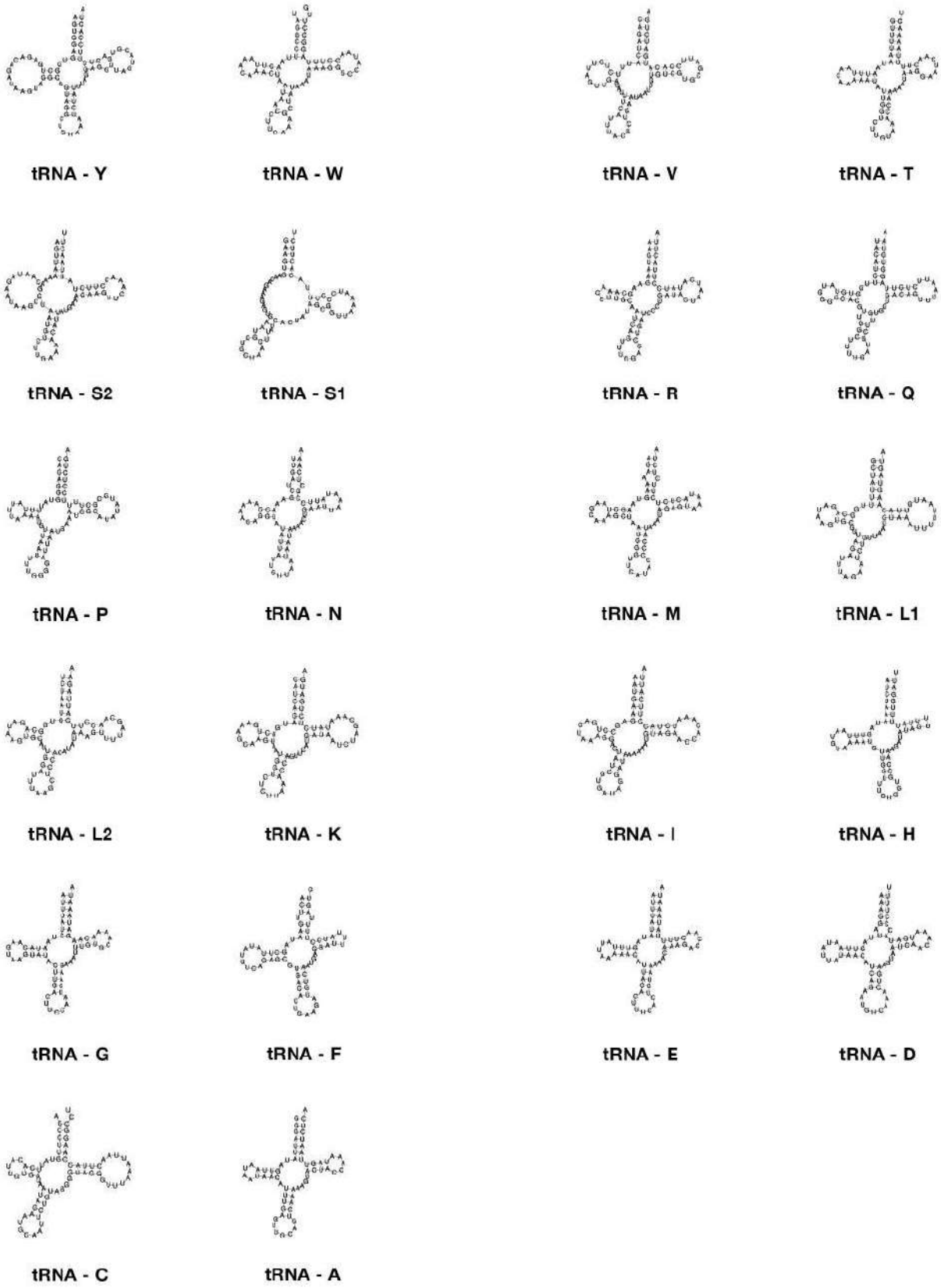
Reticulitermes speratus KY484910.1

Figure S2. Continued



Reticulitermes speratus - intrafamiliar starting references

Figure S2. Continued



Reticulitermes speratus - congeneric starting references

Figure S3. Alignment of contigs reconstructed from RNA-Seq experiments using the reference with rearranged gene order, to the correct mitogenome sequence. Artificial rearrangements are indicated by shaded areas.

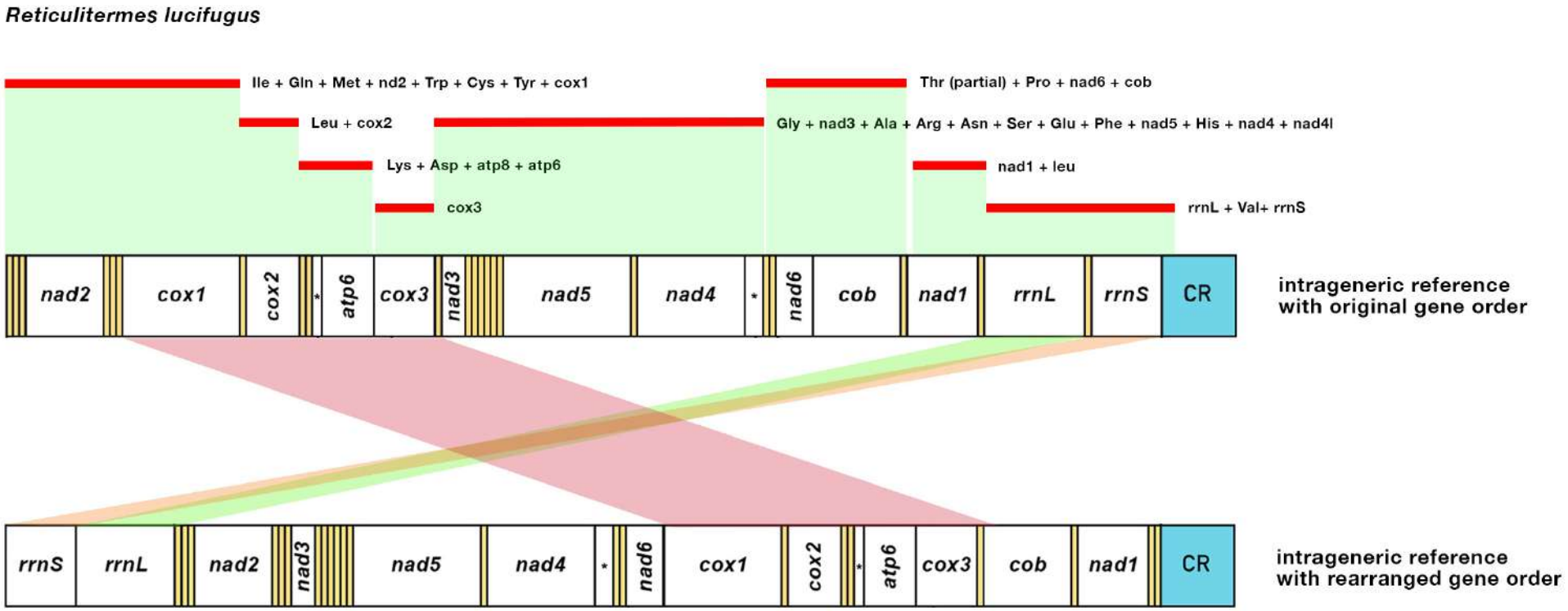


Figure S3. Continued

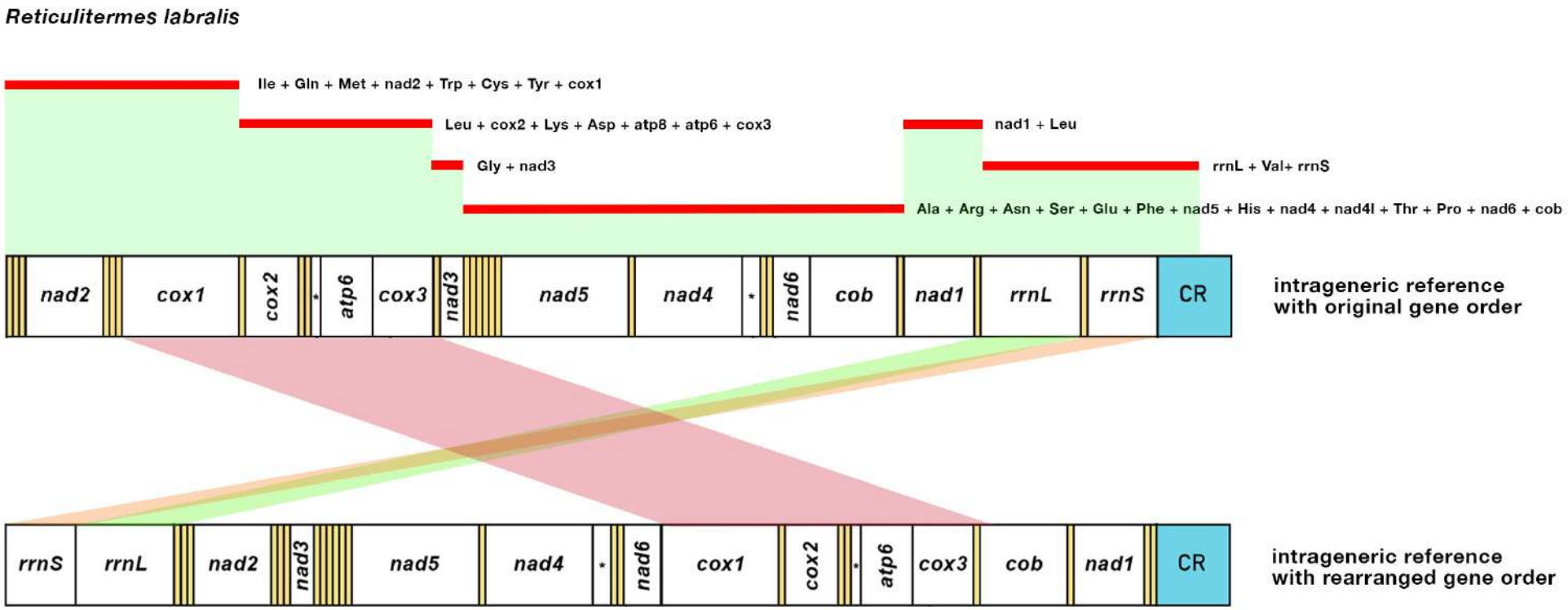


Figure S3. Continued

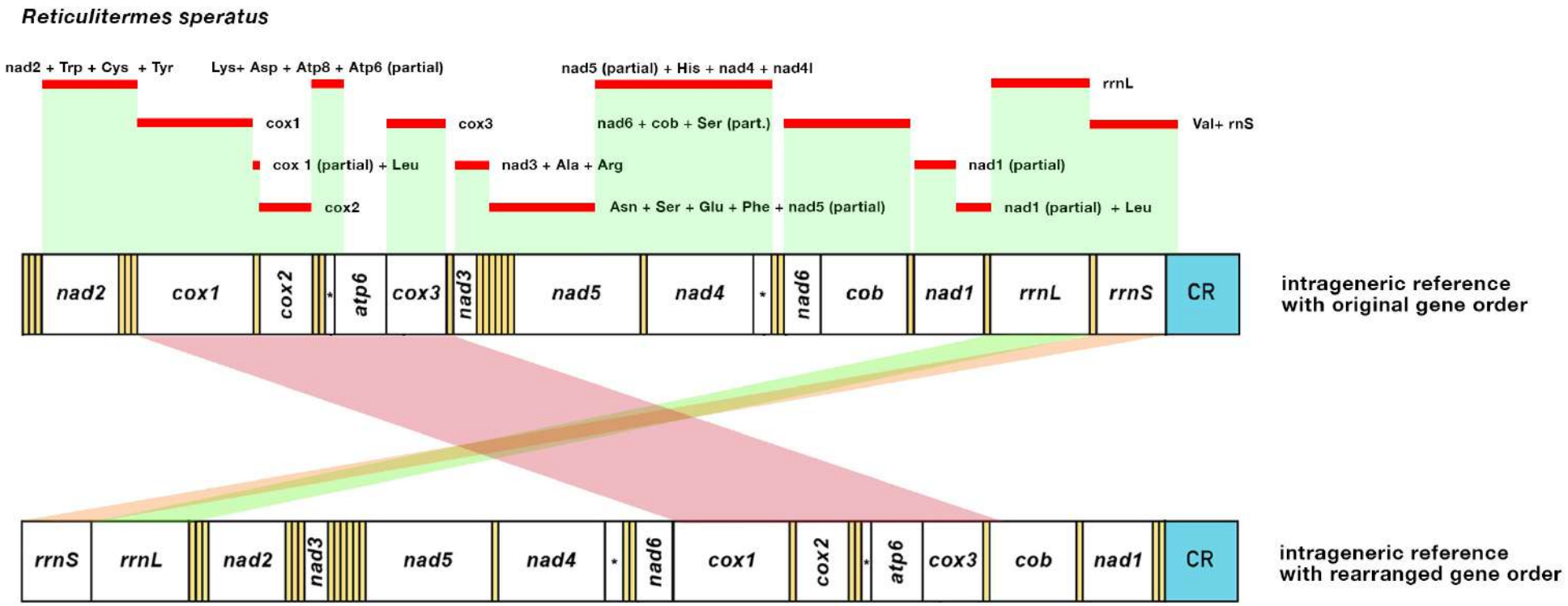


Figure S3. Continued

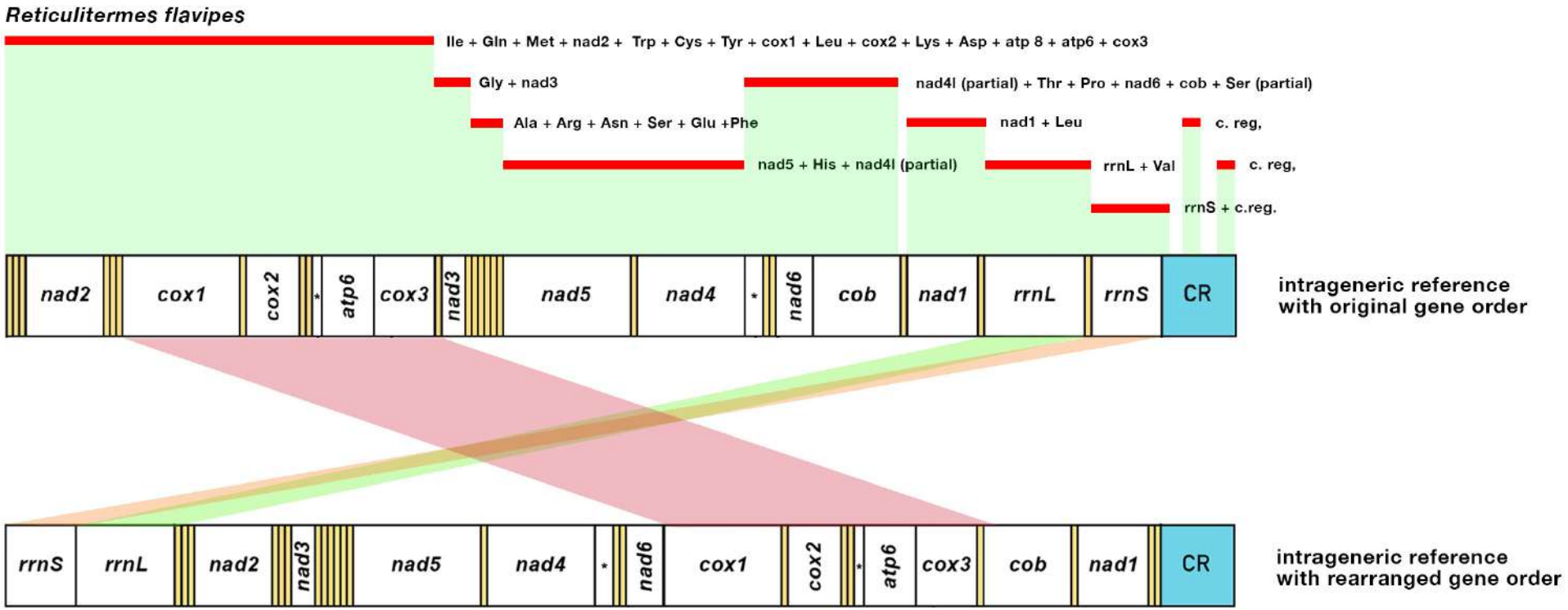


Figure S3. Continued

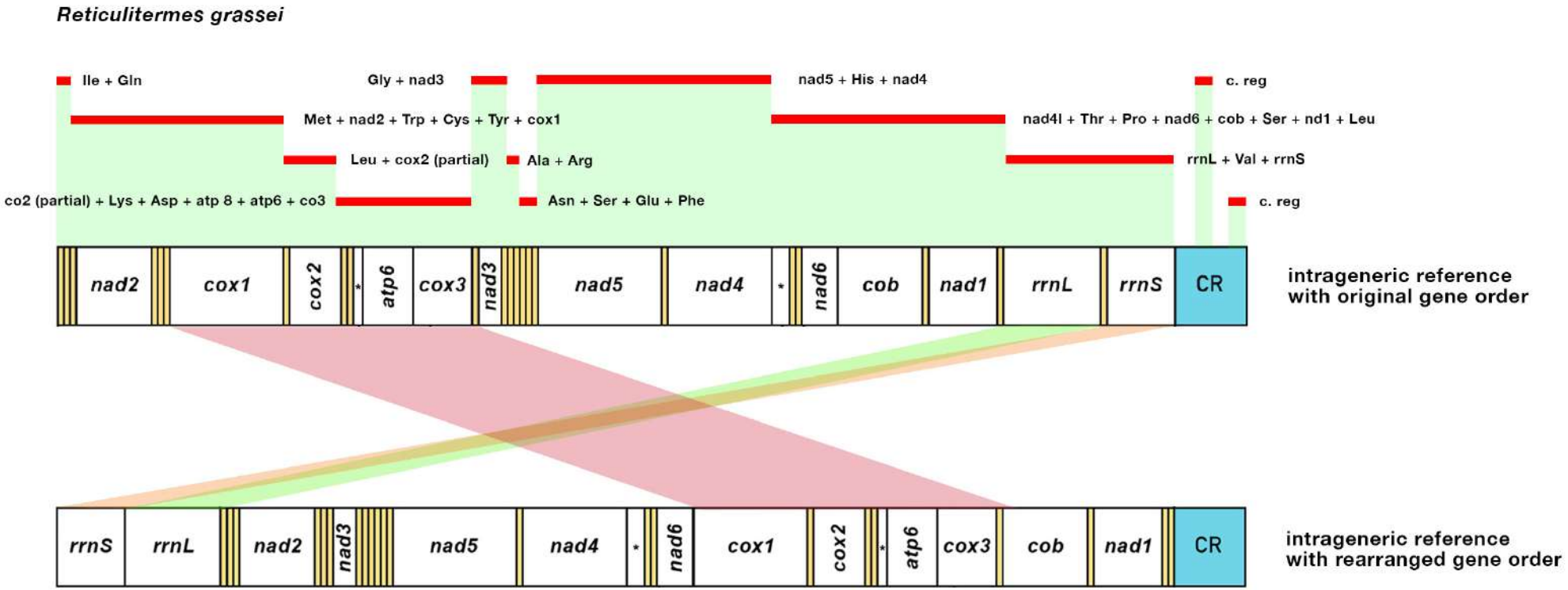
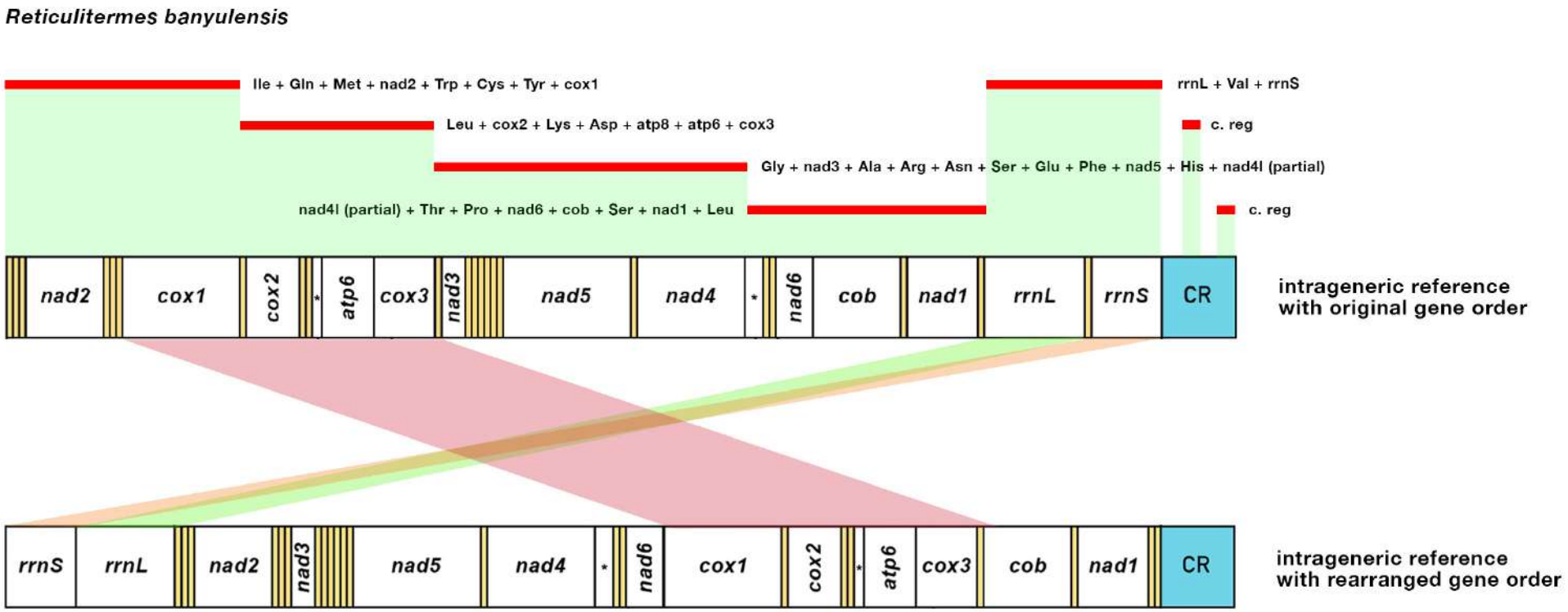


Figure S3. Continued



BASE: a novel workflow to integrate non-ubiquitous genes in comparative genomics analyses for selection.

Giobbe Forni, Angelo Alberto Ruggeri, Giovanni Piccinini and Andrea Luchetti.

Abstract – Inferring the selective forces that ortholog genes underwent across different lineages can make us understand the evolutionary processes which shaped their extant diversity. The more widespread metric to estimate coding sequences selection regimes across across their sites and species phylogeny is the ratio of nonsynonymous to synonymous substitutions (dN/dS, also known as ω). Nowadays, modern sequencing technologies and the large amount of already available sequence data allow the retrieval of thousands of genes orthology groups across large numbers of species. Nonetheless, the tools available to explore selection regimes are not designed to automatically process all orthogroups and practical usage is often restricted to those consisting of single-copy genes which are ubiquitous across the species considered (*i.e.* the subset of genes which is present in all the species considered). This approach limits the scale of the analysis to a fraction of single-copy genes, which can be as lower as an order of magnitude in respect to non-ubiquitous ones (*i.e.* those which are not present across all the species considered). Here we present a workflow named BASE that - leveraging the CodeML framework - ease the inference and interpretation of selection regimes in the context of comparative genomics. Although a number of bioinformatics tools have already been developed to facilitate this kind of analyses, BASE is the first to be specifically designed to ease the integration of non-ubiquitous genes orthogroups. The workflow - along with all the relevant documentation - is available at github.com/for-giobbe/BASE.

Keywords – molecular evolution; ubiquitous and non-ubiquitous orthogroups; selection; dN/dS; omega.

Introduction:

Selection can drive the evolution of genes by constraining changes in their sequences (purifying selection) or by favoring new adaptive variants (positive selection). Quantifying its mode and strength is a key step to understand the diverse evolutionary histories of ortholog genes across different clades. Statistical models of molecular evolution have proved as a fundamental approach to investigate such processes and can be divided into those based on comparing divergence and segregating polymorphism - such as the MK test and its extensions (McDonald and Kreitman 1991) - and those based on multi-species sequence divergence - also known as codon models. The two approaches use different conceptual frameworks and are better applied for analyses at different timescales, with the first approaches more suited to investigate recent processes and the latter ones more apt to infer older events (Mugal et al 2014).

Approaches based on the sequence divergence among multiple species are cornerstones in the estimations of patterns of sequence evolution and selection regimes. After the first models were developed to infer the strength of selection on protein-coding sequences globally across their sites and species phylogeny (Goldman and Yang, 1994; Muse and Gaut, 1994), subsequent elaborations allowed for variation across lineages (branch models: Yang, 1998), sites (sites models; Nielsen and Yang, 1998; Yang et al. 2000; Anisimova et al. 2001) and both (branch-site models; Yang and Nielsen 2002; Zhang et al. 2005). Pairwise comparisons between models can be performed using likelihood-ratio tests to understand which one better reflects the molecular evolution of a group of ortholog genes (Anisimova et al. 2001). The interpretation of all these models is largely based on the dN/dS parameter (Kimura, 1977; also known as ω), which consists in the ratio of nonsynonymous substitution rates (non-synonymous substitutions over non-synonymous sites; dN) to synonymous substitution rates (synonymous substitutions over synonymous sites; dS). This metric is fundamental to investigate the extent to which selection modulates sequence evolution of the protein-coding portions of genes. While d*S* are assumed to evolve neutrally, d*N* are expected to be exposed to selection, as they change the aminoacidic structure of proteins. Despite some of these assumption have been challenged (Davydov et al., 2019; He et al., 2020), analyses based on codon models have proved themselves as key approaches in comparative genomics, such as investigating positive selection connected to evolutionary innovations (Parker et al. 2013; Zhang et al. 2014; Li et al. 2014) or testing the relaxation of selective constraints after trait decay (Liu et al. 2019; Policarpo et al. 2020). In other instances, these approaches have been used to observe genome-wide effects linked to events such as shifts in environmental niches or the loss of recombination in asexual genomes (Plazzi et al., 2017; Bast et al. 2018).

Several tools have been developed to infer codon models for coding sequences: Selecton (Stern et al. 2007), HyPhy (Pond et al., 2005), TreeSAAP (Woolley et al 2003) and the CodeML program in the PAML package (Yang, 2007). The latter program was also subject to several implementations, such as IDEA (Egan et al., 2008), SlimCodeML (Valle et al 2014), IMPACT_S (Maldonado et al. 2014), LMAP (Maldonado et al. 2016), ete-evol in the ETE3 package (Huerta-Cepas et al. 2016), VESPA (Webb et al. 2017), BlastPhyMe (Schott et al. 2019) and also some graphical front ends such as PAMLX (Xu and Yang 2013) or EasyCodeML (Gao et al., 2019). With the increment of genomics and transcriptomics studies, it has become rather common to analyze thousands of genes for up to hundreds of species and - despite CodeML still appearing to be the most widely used piece of software - all of its aforementioned implementations try to increment its ease of use in the context of comparative genomics.

Our focus developing BASE has been mainly directed to facilitate the integration of an often overlooked - yet incredibly large - portion of genomes. Orthogroups (*i.e.* genes descended from a single gene in the last common ancestor of all species considered; abbreviated as OGs) can differ in many aspects, such as the inclusion of single-copy or multi-copy genes. OGs can also be made up of ubiquitous (*i.e.* the subset of genes which is shared by all species considered) or non-ubiquitous (*i.e.* those which are lacking in some of the species considered) genes. In comparative genomics datasets, the majority of single-copy genes OGs consists of non-ubiquitous genes: as an exploratory example, we analyzed 20 recent datasets and found that - among single-copy genes OGs - the average proportion of those including non-ubiquitous genes is 73.4% (**Fig. 1**). Non-ubiquitous genes OGs are mostly left behind in selection analyses - which are typically based only on single copy and ubiquitous genes - due to the lack of automated approaches for their inclusion. Nonetheless, overlooking such a vast portion of genes may potentially conceal important evolutionary processes and therefore we developed a novel workflow specifically for this purpose.

Implementation:

The BASE workflow is written in BASH and R and has been tested on Linux operating systems, such as centOS 8. As it extensively leverages GNU utilities, its usage is restricted to Linux distributions. It consists of two main steps: in the first one (“analyze”), evolutionary model parameters are inferred across alignment sites and tree branches for the different OGs, while the subsequent step (“extract”) allows to retrieve the different metrics associated to specific branches or clades in the species tree. CodeML provides the statistical and computational framework to perform these analyses and is at the core of the workflow, whose general description is reported in **Fig. 2**.

The inputs required to start an analysis are: a) a species tree in the Newick format; the tree can be multifurcating but has to include all the species present across the different OGs; b) a folder containing aligned OGs, in fasta format with headers matching exactly with the species names found in the species tree; c) two CodeML control files describing two nested models - where one is a specific case of the other; all the parameters in control files can be customized, with branch-site, branch and site models all supported; d) a file including the target branches or clades in the species tree for which the different metrics will be retrieved, defined by their associated species.

In the “analyze” step, the workflow initally checks the alignments for the presence of stop codons using transeq of the EMBOSS package (Rice et al. 2000), on the basis of the genetic code specified in the control files; genes which include stop codons are reported and excluded from subsequent analyses. Branch length optimization is then carried out for each OG using RAxML (Stamatakis 2014) with a codon-aware GTR substitutions model; subsequently the two CodeML runs configured with the general and the alternative models are performed and compared through a Likelihood Ratio Test (LRT) using R (R Core Team 2013). LRTs results are summarized in a table and the output relative to the best-fit model is selected for each OG. The default behavior of BASE is to process all OGs, whether consisting of ubiquitous or non-ubiquitous genes, but the user can limit the analysis to just ubiquitous ones. When the analysis is configured to consider also OGs of non-ubiquitous genes, the species tree will be pruned on the basis of the species present in each OG using ape R package (Paradis et al. 2004), prior to branch length optimization. It is also possible to label specific clades (or branches) to test hypothesis on precise lineages of the phylogeny: to do so a file listing all the species associated to the clade need to be specified, so that the relative labels will be added to the tree using phangorn R package (Schliep 2011).

The “extract” step can be carried out subsequently to the “analyze” one, using as input the folder generated by the previous step; otherwise, CodeML outputs generated by means other than BASE can be used as well. In a first place this step will annotate internal nodes of each OG tree to match the output of CodeML and will list all species associated to each branch of the phylogeny. Subsequently, the pipeline will create a table for each branch/clade specified by the user, containing the dN/dS, dN, and dS values relative to the best-fit model for each OG.

Most - if not all - comparative genomics / phylogenomic datasets do not include the complete set of species for the group considered and represent partial sampling of extant lineages; nonetheless, specific branches and clades can still be identified in a phylogeny when some of the relative species are absent (**Fig. 3**). For example, a clade - and its associated branch - made up by tens of species can be considered to be still present when a few species are missing from the phylogeny, either out of the clade itself or within. Sustained by this concept, the "extract" step allows to include non-ubiquitous genes OGs with two approaches: (a) including OGs of non-ubiquitous genes just

relatively to species external to the target clades - *i.e.* OGs where all species in the target clades are present but some of those out of them are missing; (2) include OGs of non-ubiquitous genes relatively to all the phylogeny - *i.e.* OGs where species can be missing both inside and outside of the target clades. In the latter scenario, the user can configure a cutoff for missing species relatively to the target clades, which can be specified by either an absolute number or a proportion (for example: if 0.8 is specified, at least the 80% of the clade's species need to be present in a given OG in order for it to be included in the analysis). If representatives of a given clade do not meet the selected criteria, this will be reported in the final output and no associated metrics will be present in it. BASE can subsequently retrieve the metrics associated to the target clades (or branches), defined as the ones which contains the highest number of species whithin the original user-defined target clades and non of the others. When the "extract" step is carried out in the verbose mode, the output will also contain all target clade species present in each OG.

Even if the focal feature is the integration of single-copy non-ubiquitous genes OGs into selection analyses, BASE provides other features that ease the inference, comparison and interpretation of codon models. Labeling and/or retrieving model parameters for specific branches/clades in large phylogenies can be a tiresome process, which is instead made easy by the approach implemented in our workflow. BASE can also process batches of OGs simultaneously - substantially cutting down processing times - and implements a large number of error-messages which can definitively ease the user experience.

Conclusions:

BASE is a workflow for analyses on selection regimes that integrates several popular pieces of software, with CodeML at its core. It has been conceived to ease the integration of non-ubiquitous genes OGs into comparative genomics analyses for selection, yet it implements many other features and quality-of-life improvements. BASE can represent a useful tool to address theoretical and biological questions, such as the impact of sampling on dN/dS estimates or whether ubiquitous genes tend to have different evolutionary constraints with respect to non-ubiquitous ones. We hope that our effort proves to be a useful tool for studying molecular evolution and that it generates some interest towards the integration of non-ubiquitous genes OGs in selection analyses. BASE is an ongoing project, and we welcome bug reports, feedback and suggestions for feature implementations.

All the documentation, including detailed tutorials to explore BASE functionality can be found at github.com/for-giobbe/BASE.

Author contributions:

G.F. conceived the idea; All authors implemented ideas and design; G.F. led the writing of the manuscript and on-line resources; All authors contributed to the drafts and gave final approval for publication.

References:

Anisimova, M., Bielawski, J. P., and Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular biology and evolution, 18(8),*1585-92.

Bast, J., Parker, D. J., Dumas, Z., Jalvingh, K. M., Tran Van, P., Jaron, K. S., Figuet, E., Brandt, A., Galtier, N., and Schwander, T. (2018). Consequences of asexuality in natural populations: insights from stick insects. *Molecular biology and evolution*, *35*(7), 1668-1677.

Davydov, I. I., Salamin, N., & Robinson-Rechavi, M. (2019). Large-scale comparative analysis of codon models accounting for protein and nucleotide selection. *Molecular biology and evolution*, *36*(6), 1316-1332.

Egan, A., Mahurkar, A., Crabtree, J., Badger, J. H., Carlton, J. M., and Silva, J.C., 2008. IDEA: interactive display for evolutionary analyses. *BMC bioinformatics, 9(1),* 524.

Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., and Ho, S.Y. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecology and evolution, 9(7),* 3891-3898.

Goldman, N., Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular biology and evolution, 11(5),* 725–36.

He, Z., Chen, Q., Yang, H., Chen, Q., Shi, S., & Wu, C. I. (2020). Two decades of suspect evidence for adaptive DNA-sequence evolution-Failure in consistent detection of positive selection. bioRxiv, 417717.

Huerta-Cepas, J., Serr,a F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular biology and evolution*, *33*(6), 1635-1638.

Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature, 267(5608),* 275-276.

Kryazhimskiy, S., and Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genetics, 4:12,* 1000304.

Li, C., Zhang, Y., Li, J., Kong, L., Hu, H., Pan, H., Xu, L., Deng, Y., Li, Q., Jin, L., and Yu, H. (2014). Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *GigaScience*, *3*(1), 2047-217X.

Liu, A., He, F., Shen, L., Liu, R., Wang, Z., and Zhou, J. (2019). Convergent degeneration of olfactory receptor gene repertoires in marine mammals. *BMC genomics, 20(1),* 977.

Maldonado, E., Almeida, D., Escalona, T., Khan, I., Vasconcelos, V., and Antunes, A., 2016. LMAP: lightweight multigene analyses in PAML. *BMC bioinformatics*, *17*(1), pp.1-11.

Maldonado, E., Sunagar, K., Almeida, D., Vasconcelos, V., and Antunes, A. (2014). IMPACT_S: integrated multiprogram platform to analyze and combine tests of selection. *PloS one, 9(10),* 96243.

McDonald, J. H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, *351*(6328), 652-654.

Mugal, C. F., Wolf, J. B., and Kaj, I. (2014). Why time matters: codon evolution and the temporal dynamics of dN/dS. *Molecular biology and evolution, 31(1),* 212-231.

Muse, S. V., and Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular biology and evolution*, *11*(5), 715-724.

Nielsen R., and Yang Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope Gene. *Genetics, 148(3)*, 929-36.

Paradis E., Claude J., and Strimmer K (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics, 22;20(2),* 289-90.

Parker, J., Tsagkogeorga, G., Cotton, J. A., Liu, Y., Provero, P., Stupka, E., and Rossiter, S. J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*, *502*(7470), 228-231.

Plazzi, F., Puccio, G., and Passamonti, M. (2017). Burrowers from the past: mitochondrial signatures of Ordovician bivalve infaunalization. *Genome Biology and Evolution*, *9*(4), 956-967.

Policarpo, M., Fumey, J., Lafargeas, P., Naquin, D., Thermes, C., Naville, M., Dechaud, C., Volff, J.N., Cabau, C., Klopp, C., and Møller, P.R. (2020). Contrasting gene decay in subterranean vertebrates: insights from cavefishes and fossorial mammals. *Molecular Biology and Evolution*.

Pond, S. L. K., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics, 21,* 676–679.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in Genetics, 16(6)*, 276-7.

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592-593.

Schott, R. K., Gow, D., and Chang, B. S. (2019). BlastPhyMe: A toolkit for rapid generation and analysis of protein-coding sequence datasets. *bioRxiv*, 059881.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Stern, A., Doron-Faigenboim, A., Erez, E., Martz, E., Bacharach, E., and Pupko, T. (2007). Selecton 2007: Advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Research, 35*, W506–W511.

Sun, C., Huang, J., Wang, Y., Zhao, X., Su, L., Thomas, G., Zhao, M., Zhang, X., Jungreis, I., Kellis, M., Vicario, S., Sharakhov, I. V., Bondarenko, S. M., Hasselmann, M., Kim, C. N., Paten, B., Penso-Dolfin, L., Wang, L., Chang, Y., Gao, Q., … Mueller, R. L. (2020). Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. *Molecular biology and evolution*, msaa240

Team, R.C. (2013). R: A language and environment for statistical computing.

Valle, M., Schabauer, H., Pacher, C., Stockinger, H., Stamatakis, A., Robinson-Rechavi, M. and Salamin, N. (2014). Optimization strategies for fast detection of positive selection on phylogenetic trees. *Bioinformatics, 30(8),* 1129-1137.

Webb, A. E., Walsh, T. A. and O'Connell, M. J. (2017). VESPA: very large-scale evolutionary and selective pressure analyses. *PeerJ Computer Science*, *3*, 118.

Woolley, S., Johnson, J., Smith, M. J., Crandall, K. A., & McClellan, D. A. (2003). TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics, 19(5),* 671-672.

Xu, B., and Yang, Z. (2013). PAMLX: a graphical user interface for PAML. *Molecular biology and evolution*, *30*(12), 2723-2724.

Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution, 24, 15(5),* 568-73.

Yang, Z. (2007). PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24,*1586-1591

Yang, Z., and Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular biology and evolution*, *19*(6), 908-917.

Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, *155*(1), 431-449.

Yang, Z., Wong, W. S., and Nielsen, R (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular biology and evolution, 22(4),* 1107-18.

Zhang J., Nielsen R., and Yang Z (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular biology and evolution, 22(12),* 2472-9.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., and Ödeen, A. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science, 346(6215),* 1311-1320.

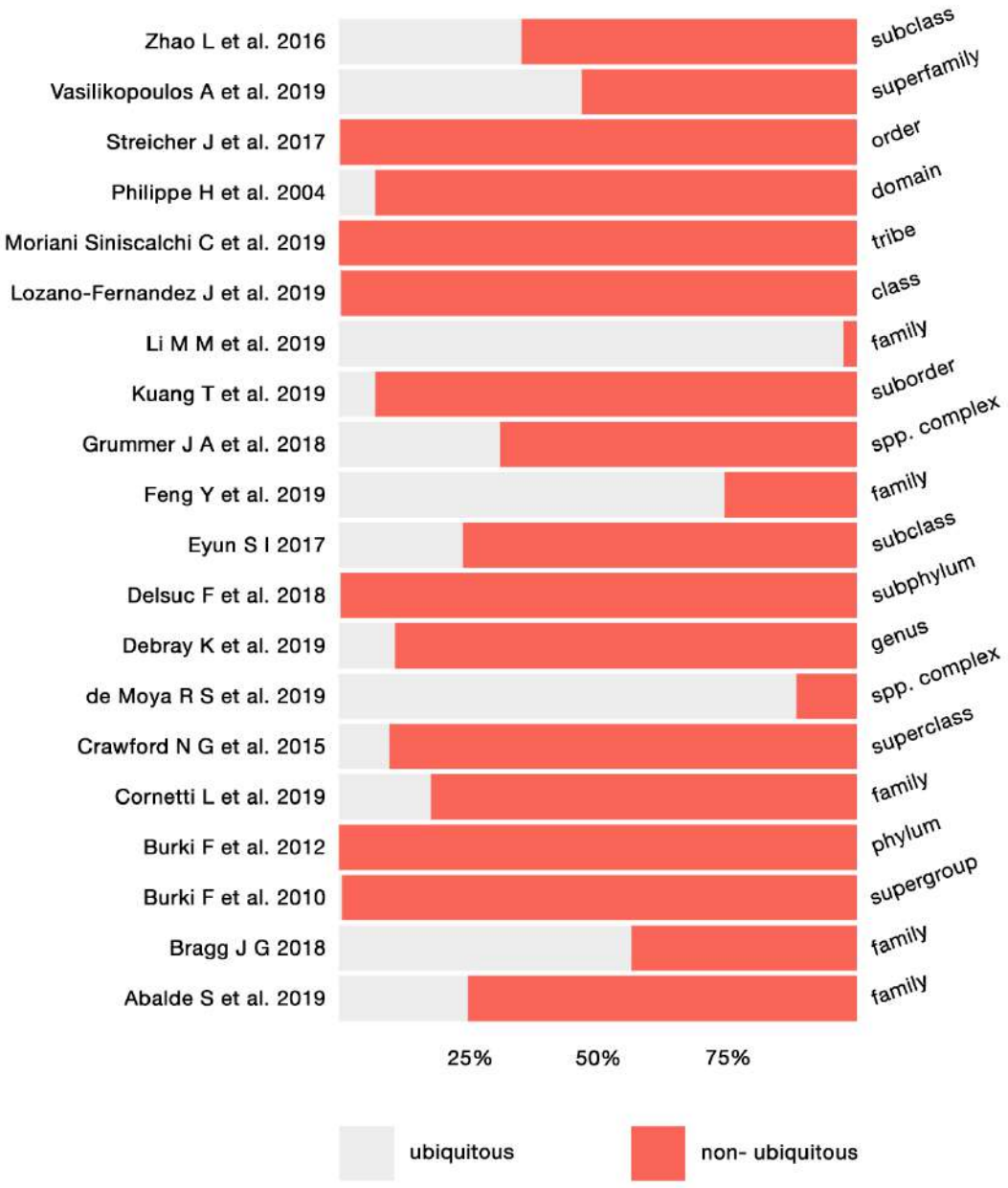


Fig. 1 – The proportion of ubiquitous and non-ubiquitous genes was calculated in 20 datasets, varying in the taxonomic level considered (from family to phylum). The average percentage of non-ubiquitous genes is 73,36 % while for ubiquitous ones is 26,6%.

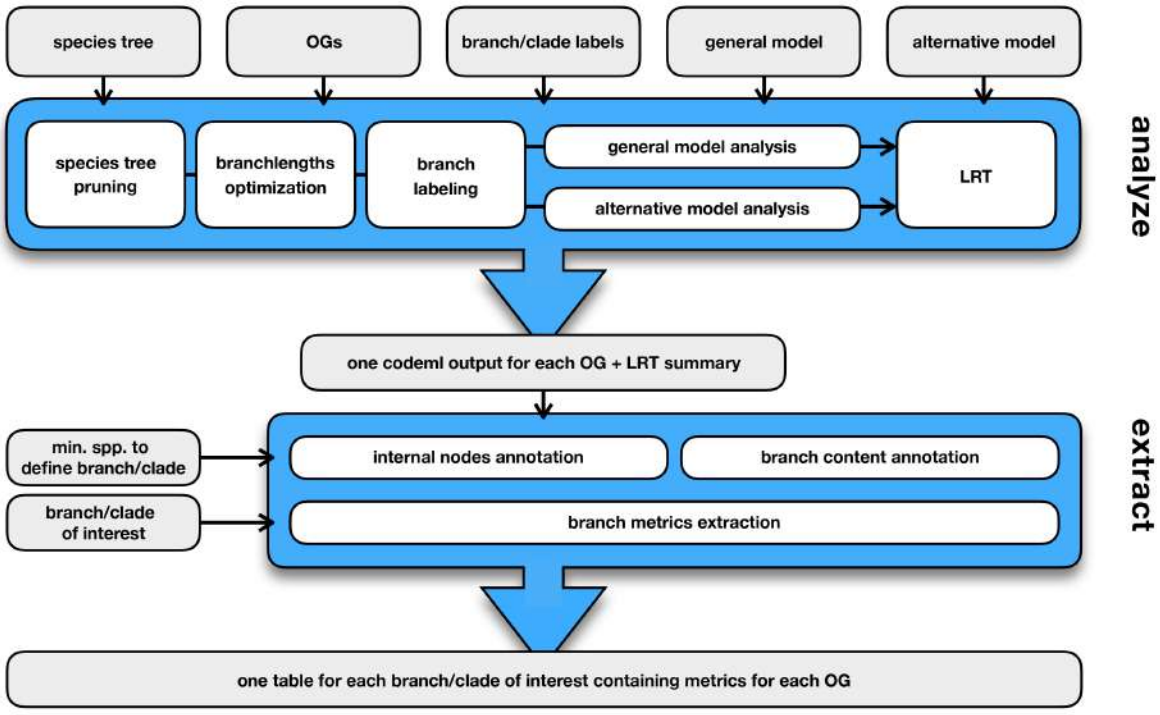


Fig. 2 – BASE workflow consists of the two steps: "analyze" and "extract". The two steps can be carried out independently and CodeML output generated outside of the workflow can be processed in the "extract" step.

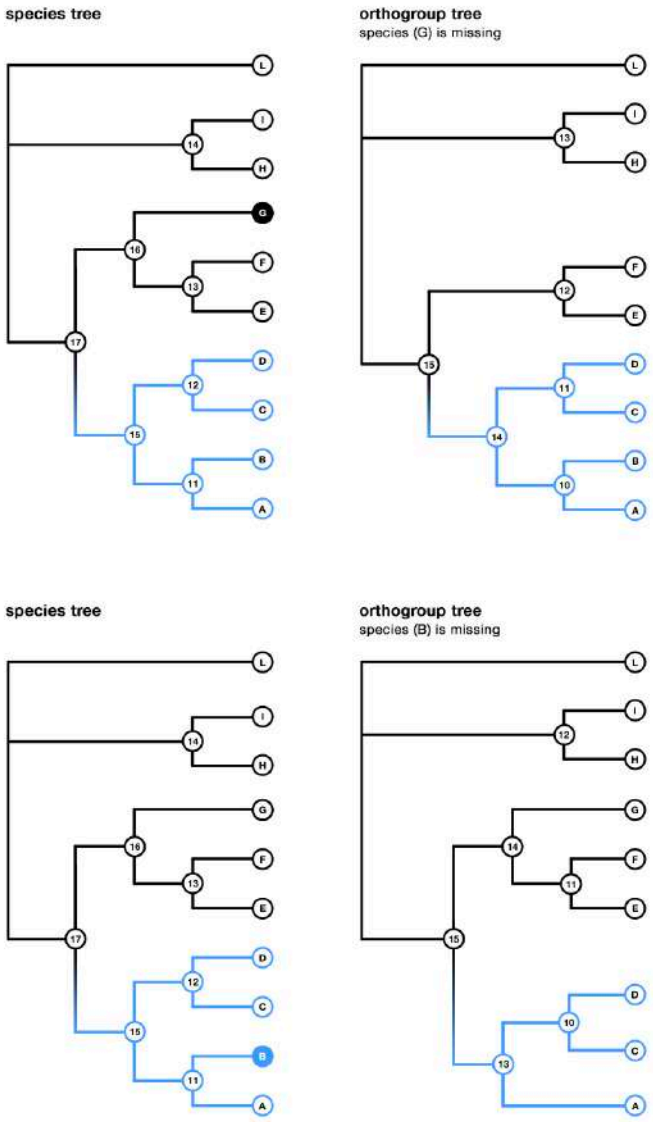


Fig. 3 – Trees of single-copy non-ubiquitous genes OGs differ in their structure from the species tree. In the upper figure a species external to the target clade (species G - higlighted with a black circle) is missing in the orthogroup tree while in the bottom figure a species in the target clade is missing in the orthogroup tree (species B - higlighted with a blue circle); in both figures the target clade is higlighted in blue. BASE can: (a) include OGs of non-ubiquitous genes only relatively species external to the group of interest, so that the the upper orthogroup tree is processed and the lower is discarded or (b) include OGs of non-ubiquitous genes relatively to both species internal and external to the group of interest, so that both the upper and lower orthogroup trees are processed. Despite the different annotations of internal nodes (circled numbers), BASE can then retrieve the metrics associated to the target clade(s) or the associated branch(es).

5. Conclusions.

The outcome of this thesis contributes to unravel the species richness and to improve the reliability of the systematic framework in phasmids. This effort has also proven to be a key step to define the major questions I tried to answer during my PhD and provides clear-cut examples of how overlooked parts of the tree of life can conceal interesting phenomena (*e.g.* the *Candovia* genus). Some results relative to phasmids systematic relationships appeared to be solid and most likely will result in adjustments to the clade taxonomy; nonetheless, I had special care in recognizing instances of uncertainty, and especially to not contributing to an already confounding scenario. I tried to elucidate the major phenomena playing a role in this uncertainty, which can be mainly ascribed to a lack of phylogenetic signal - from a technical point of view - and to an ancient and rapid evolutionary radiation - from a biological one. Moreover, a different perspective on the radiation of the Phasmatodea larger clade - Euphasmatodea - is proposed. The latter is retrieved at the Triassic-Jurassic boundary and such an hypothesis implies two things: 1) vicariance - and not dispersal - shaped the present species distribution and 2) the concurrent mass extinction event could have been an important driver of phasmids diversification, which happened within a broad species turnover pattern caused by ecological opportunity.

Throughout the thesis I leveraged widely different frameworks to elucidate the evolution of two distinct traits. Studying wings evolutionary patterns, I tested the hypothesis that a complex morphological trait could evolve in a reversible manner, going back and forth from ancestral and derived states. I found a strong support for a reversible evolution of wings, even when using all proposed measures to prevent a false positive result. If such a complex character as wings can be lost and reacquired after several million years, one question instantly comes to the mind: how is it possible to retain the potential to express a trait (*i.e.* its genetic groundplan) despite its absence at the phenotypical level? To answer this question, I investigated the effect of trait loss on the associated molecular underpinnings: for this purpose, I employed an RNA-seq experiment of two parthenogenetic *Bacillus* species and a sexual congeneric. The latter has been used to construct a gene regulatory network and to identify the genes associate to male gonads; subsequently I characterized the modifications these genes underwent in the parthenogenetic lineages, where males are not present. Transcriptional and genetic components associated to male gonads appear to be partially preserved in parthenogens, with pleiotropy being the most likely driver of the process. I also studied automixis from a different perspective, focusing on its enstablishment as a novel reproductive strategy. I characterized the relative contribution of novel and pre-existing genes along with the degree of convergent changes associated with the two independent shift to automixis in *Bacillus*. Parthenogens transcriptional programme is largely assembled from genes that were already present before the establishment of the novel reproductive strategies and while

convergent changes seem to have played a prominent role at the transcriptional level, they appear to have had a marginal one in sequence evolution.

While studying these phenomena, I leveraged different analytical frameworks: on one side macroevolutionary approaches - *i.e.* comparative methods - through which I explored large-scale patterns of characters evolution along phylogenies; on the other side I employed comparative molecular approaches, consisting in expression-based (differential expression, gene co-expression networks) and sequence-based ones (inferences on selection regime strength and evolutionary rates). Despite the vastly different frameworks, the results presented here are indeed tightly connected from a conceptual point of view. If we consider the findings as a whole, they can contribute to our understanding of the interplay between genes and trait evolution. Traits are made by the interplay of large number of genes: when a trait is established - or modified - previously unrelated genes can be modified to concur in its development and functioning; when a trait is lost, the more pleiotropic genes associated to it can largely retain their sequence identity and transcriptional programs - being under selection for other traits - potentially allowing the possibility to re-enstablished the former. These complex interactions underlie the interplay between the phenotypic and molecular levels and can result in outstanding phenomena - such as reversals to once lost traits - which challenge our understanding of trait evolution.