Macroevolutionary Analyses Provide New Evidences of Phasmids Wings Evolution as a Reversible Process.

Giobbe Forni* [1], Jacopo Martelossi* [1], Pablo Valero [2], Frank H. Hennemann [3], Oskar Conle [4], Andrea Luchetti [1], Barbara Mantovani [1].

[1] Dip. Scienze Biologiche, Geologiche e Ambientali (BiGeA), University of Bologna, Italy.

[2] Aachrain 1, 87534 Oberstaufen, Germany.

[3] Tannenwaldallee 53, 61348 Bad Homburg, Germany.

[4] Am Freischütz 16, 47058 Duisburg, Germany.

* These authors contributed equally to this work.

Corresponding authors: giobbe.forni2@unibo.it, andrea.luchetti@unibo.it

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.

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.

Figures and tables captions:

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 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 branchlengths. 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 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.

Supplementary Table 2 - Primers sequence and PCR annealing condition.

Supplementary Table 3 - Maximum Likelihood Inference model selection performed by ModelFinder with the -MERGE option and in according to the BIC score.

Supplementary Table 4 - Fossil calibrations used for the divergence times analysis. The calibration nodes are reported in Sup. Fig. 2.

Supplementary Table 5 - AICc, -lnL, and ΔAICc values of the evolutionary models fitted on the Maximum Clade Credibility tree 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.

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

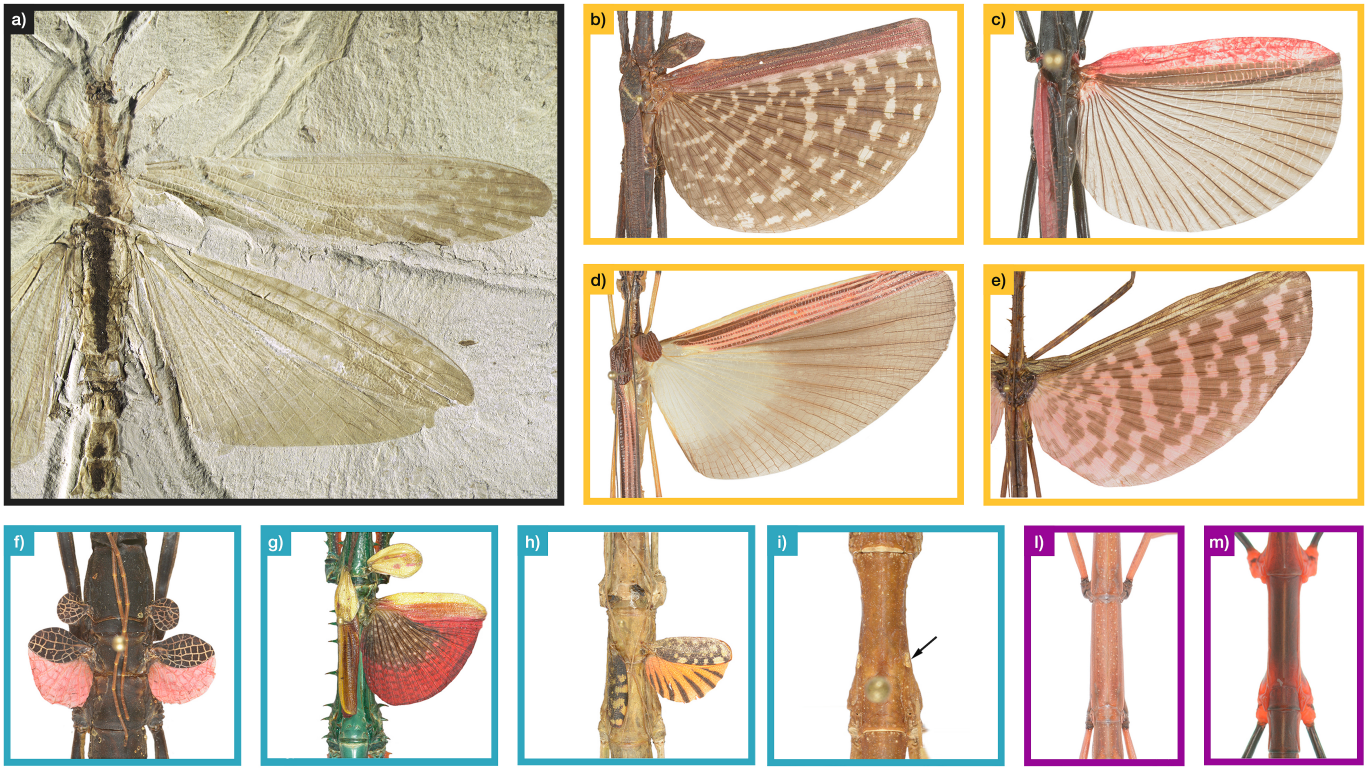
Supplementary Table 7 - (a) and (b) model selection results using *SSE models on respectively the original binary dataset and on the modified version with brachiptery coded as absence (b). Details on the evolutionary models can be found in the body of the text. In (c) are reported the diversification parameters of the best-fit model for each dataset: 1 =original binary dataset, 2 = modified version; 0 –1 =

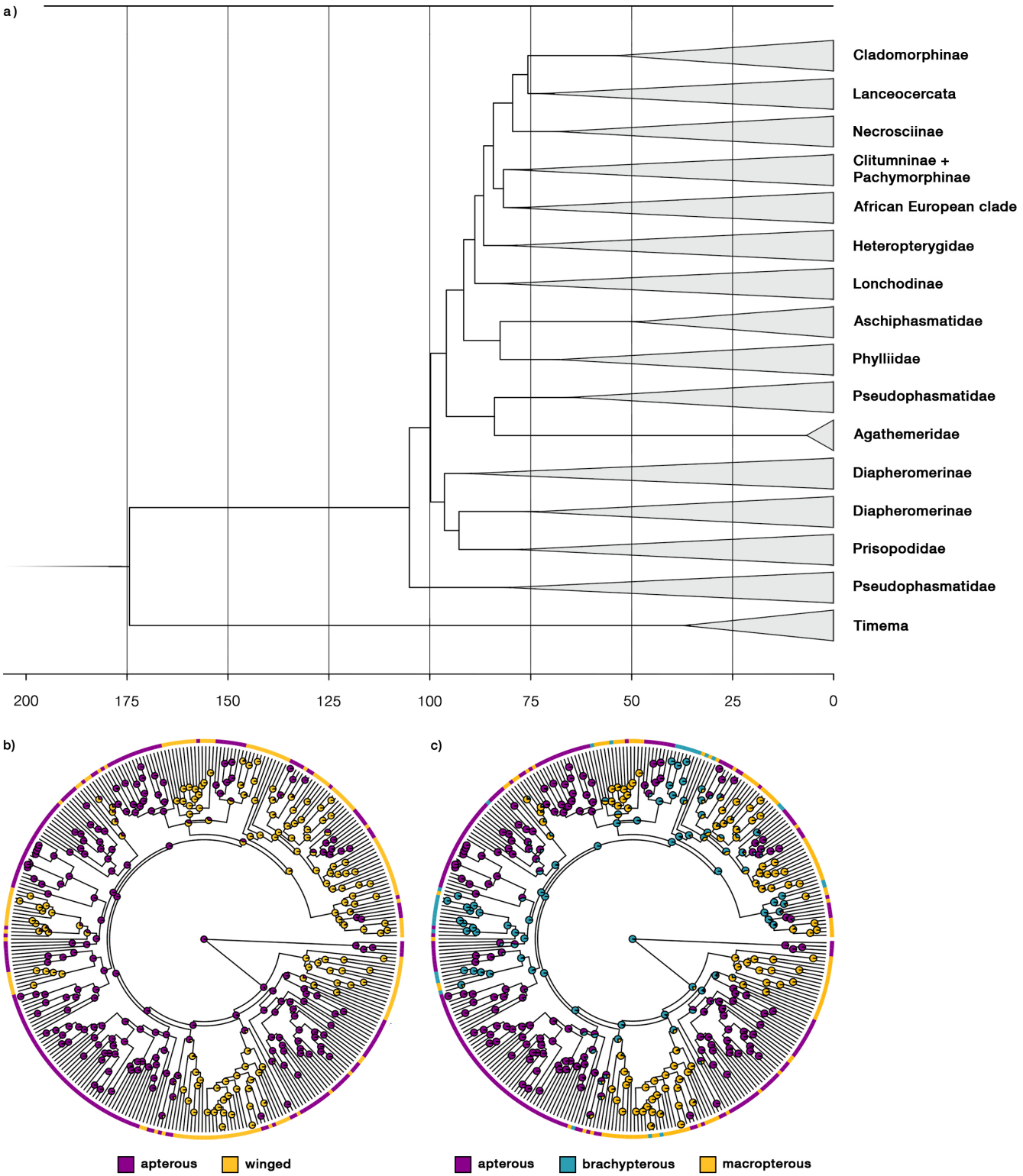
explicit character states (respectively wingless and winged); A-B = hidden states; r = net diversification rates (= $\lambda - \mu$); a = extinction fraction (= μ / λ).

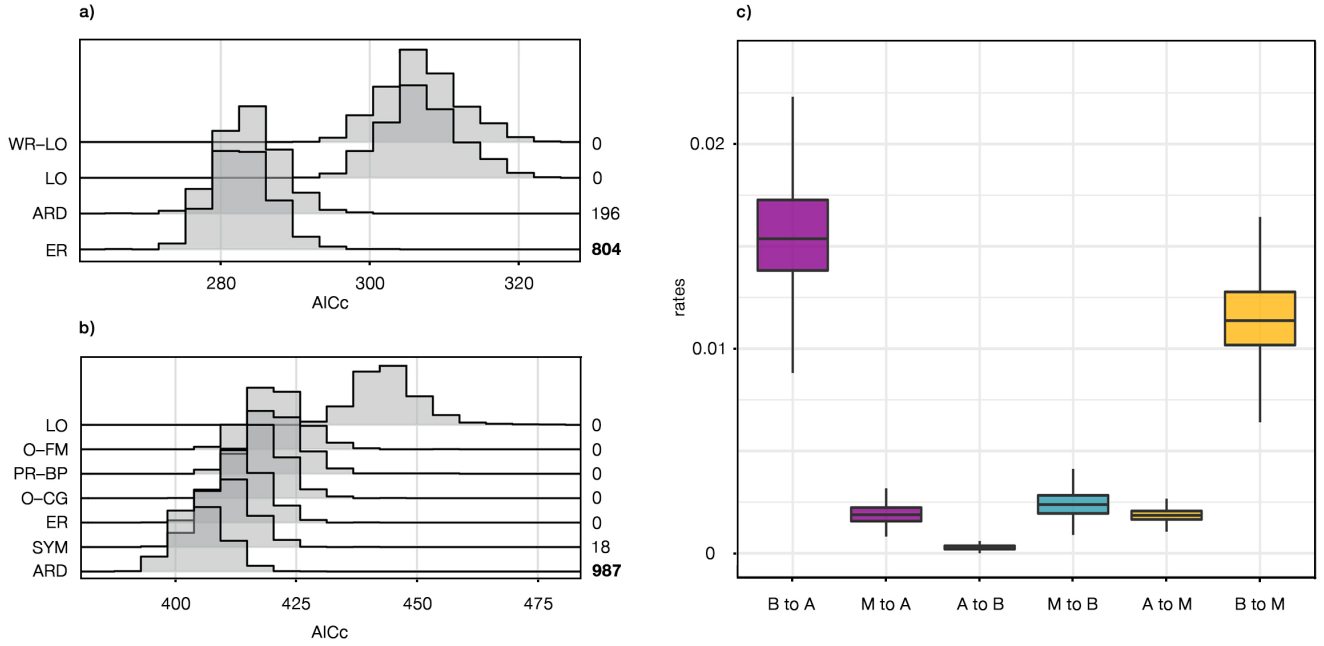
Supplementary data: on request.

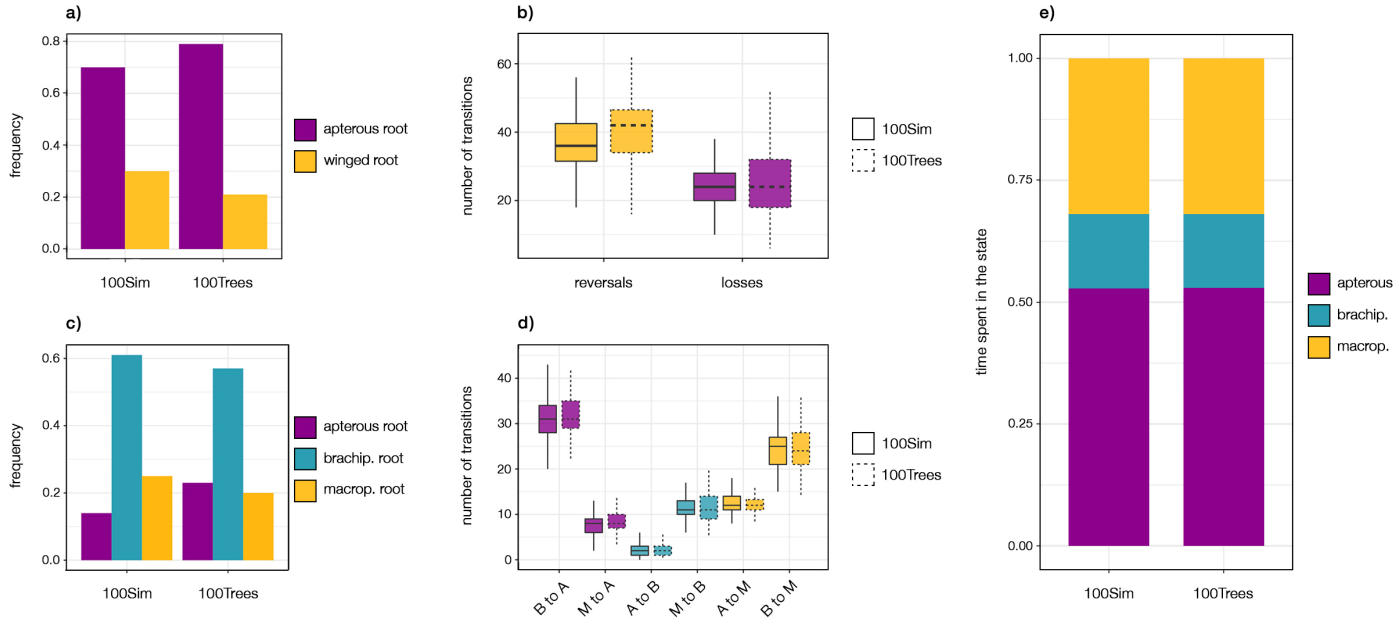
Aknowledgments:

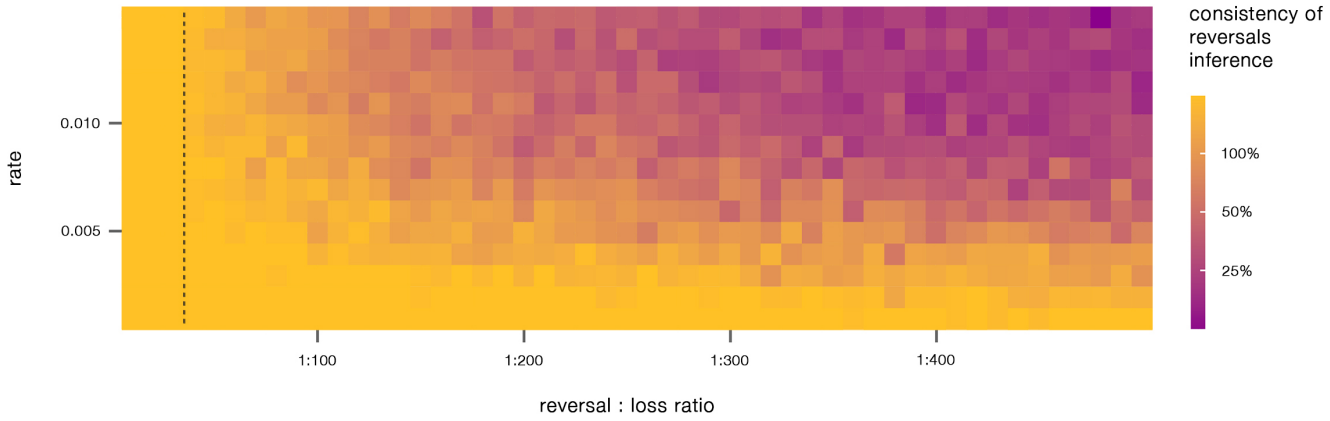
The authors would like to thank Mattia Ragazzini for the help in revising the carachter-coding tables. We also wish to thank Taiping Gao and Hongru Yang for permitting the usage of the photo of *Aclistophasma echinulatum* fossil.


a)
b)
c)
d)
e)
f)
g)
h)
i)
l)
m)


a)
Cladomorphinae
Lanceocercata
Necrosciinae
Clitumninae +
Pachymorphinae
African European clade
Heteropterygidae
Lonchodinae
Aschiphasmatidae
Phylliidae
Pseudophasmatidae
Agathemeridae
Diapheromerinae
Diapheromerinae
Prisopodidae
Pseudophasmatidae
Timema
200
175
150
125
100
75
50
25
0
b)
apterous
winged
c)
apterous
brachypterous
macropterous


a)
WR–LO
LO
ARD
ER
0
0
196
804
280
300
320
AICc
b)
LO
O–FM
PR–BP
O–CG
ER
SYM
ARD
0
0
0
0
0
18
987
400
425
450
475
AICc
c)
rates
0.02
0.01
0
B to A
M to A
A to B
M to B
A to M
B to M


a)
frequency
0.0
0.2
0.4
0.6
0.8
100Sim
100Trees
apterous root
winged root
b)
number of transitions
20
40
60
reversals
losses
100Sim
100Trees
e)
time spent in the state
0.00
0.25
0.50
0.75
1.00
100Sim
100Trees
apterous
brachip.
macrop.
c)
frequency
0.0
0.2
0.4
0.6
100Sim
100Trees
apterous root
brachip. root
macrop. root
d)
number of transitions
0
10
20
30
40
B to A
M to A
A to B
M to B
A to M
B to M
100Sim
100Trees


consistency of reversals inference
100%
50%
25%
rate
0.010
0.005
1:100
1:200
1:300
1:400
reversal : loss ratio