

# **Multispecies coalescent analysis unravels the non-monophyly and controversial relationships of Hexapoda**

Lucas A. Freitas, Beatriz Mello and Carlos G. Schrago\*

*Departamento de Genética, Universidade Federal do Rio de Janeiro, RJ, Brazil*

*\*Address for correspondence:*

Carlos G. Schrago

Universidade Federal do Rio de Janeiro

Instituto de Biologia, Departamento de Genética, CCS, A2-092

Rua Prof. Rodolpho Paulo Rocco, S/N

Cidade Universitária

Rio de Janeiro, RJ

CEP: 21.941-617

BRAZIL

Phone: +55 21 2562-6397

Email: [carlos.schrago@gmail.com](mailto:carlos.schrago@gmail.com)

*Running title:* Species tree estimation of Hexapoda

*Keywords:* incomplete lineage sorting, effective population size, Insecta, phylogenomics

## Abstract

With the increase in the availability of genomic data, sequences from different loci are usually concatenated in a supermatrix for phylogenetic inference. However, as an alternative to the supermatrix approach, several implementations of the multispecies coalescent (MSC) have been increasingly used in phylogenomic analyses due to their advantages in accommodating gene tree topological heterogeneity by taking account population-level processes. Moreover, the development of faster algorithms under the MSC is enabling the analysis of thousands of loci/taxa. Here, we explored the MSC approach for a phylogenomic dataset of Insecta. Even with the challenges posed by insects, due to large effective population sizes coupled with short deep internal branches, our MSC analysis could recover several orders and evolutionary relationships in agreement with current insect systematics. However, some phylogenetic relationships were not recovered by MSC methods. Most noticeable, a remiped crustacean was positioned within the Insecta. Additionally, the interordinal relationships within Polyneoptera and Neuropteroidea contradicted recent works, by suggesting the non-monophyly of Neuroptera. We notice, however, that these phylogenetic arrangements were also poorly supported by previous analyses and that they were sensitive to gene sampling.

## Introduction

Biological sciences have been experiencing a boom in the availability of genomic data, so that the number of loci employed to reconstruct phylogenies has increased significantly in the last decade. When dealing with multiple genomic regions, the most usual approach employed by authors consists in the concatenation of such aligned sequences into a supermatrix to build the phylogeny.

Thus far, the main microevolutionary phenomenon accounted for at species tree estimation methods is incomplete lineage sorting (ILS), which is mathematically treated as modeled by the multispecies coalescent (MSC) (Degnan and Rosenberg, 2009). MSC is a derivation of the coalescent theory (Kingman, 1982) that allows the coalescent process to occur independently along branches of the phylogeny. It models gene tree topology heterogeneity, as different *loci* are treated as independent trials of the coalescent processes within the phylogeny. The fact of taking into account the discordant information derived from a sample of gene trees is the principal advantage of MSC over supermatrix methods, which assume the same history for all genes (Edwards, 2009).

Recent studies applied the MSC to infer the relationships of several animal groups (Cannon et al., 2016; Jarvis et al., 2014; Lanier and Knowles, 2015; Song et al., 2012; Xi et al., 2014). However, due to the computational limitation of MSC algorithms, the datasets employed were limited to dozens of species and few hundreds of loci (Zimmermann et al., 2014). In this sense, the advent of faster MSC methods, such as ASTRAL (Mirarab et al., 2014), enables MSC-based phylogeny estimation using thousands of species/loci. Importantly, it allows the use of genomic data available for model and non-model organisms to estimate species trees, which is revolutionizing the field of molecular phylogenetics (Edwards, 2009).

Recently, a phylogenomic dataset of the class Insecta was generated to estimate the phylogeny and divergence times for the major orders of this lineage (Misof et al., 2014). However, gene tree heterogeneity was not considered when reconstructing the phylogenetic

relationships. Compared to other biological groups, the Insecta poses a major challenge to MSC-based inference. This is mainly assigned to the large effective population sizes (Romiguier et al., 2014) associated with short speciation intervals, which leads to high degrees of ILS (Maddison and Knowles, 2006; Pamilo and Nei, 1988). Further, this class still has several controversial relationships, e.g., the Strepsiptera position within Neuropteroidea (Boussau et al., 2014; Huelsenbeck and Biology, 2007; Whiting et al., 1997), the position of Paraneoptera (Kjer, 2004; Kjer et al., 2006), the interordinal relationships within Polyneoptera (Whitfield and Kjer, 2008), and the sister-group of Insecta (Regier et al., 2010; Trautwein et al., 2012).

Once the MSC approach was already successfully applied to resolve deep phylogenies (Song et al., 2012; Xi et al., 2014), we were prompted to test how this methodology would impact the interordinal relationships of Insecta. Since large ancestral population sizes coupled with consecutive speciation events may cause high discordance between gene and species trees, the species tree inferred by MSC should better reflect lineage history. Therefore, we expect that interordinal branching, which are deeper located in the phylogeny, would present a distinct pattern when compared with the supermatrix based approach. To test this, we estimated the phylogenetic relationships of Hexapoda using the MSC approach and a genomic dataset, focusing on the interordinal relationships within this subphylum. We found three main differences compared with the supermatrix approach: the non-monophyly of Hexapoda and the interordinal relationships within Polyneoptera and Neuropteroidea, including the non-monophyly of Neuroptera, which indicates that more studies are urged to better elucidate the evolutionary patterns of this lineage.

## Material and methods

We downloaded all amino acid alignments (1,478 genes) used to reconstruct the phylogenetic relationships of Hexapoda, as well as the concatenated supermatrix C from the dataset of (Misof et al., 2014), available from the Dryad repository (<http://dx.doi.org/10.5061/dryad.3c0f1>). Then we used RAxML v. 8.2.6 (Stamatakis, 2014) to estimate the ML (maximum likelihood) gene trees with the LG+G(4) model of amino acid substitution (Le and Gascuel, 2008; Yang, 1993). For each of the 1,478 genes, 100 bootstrap replicates were inferred using CIPRES REST API (Miller et al., 2015). (S1 file).

After gene tree estimation, ASTRAL-II v. 4.10.2 (Sayyari and Mirarab, 2016) was used to independently reconstruct two species trees. Firstly, the complete dataset with 1,478 gene trees was used to estimate the Hexapoda phylogeny. To investigate the effects of gene sampling, the species tree was estimated using 377 gene trees inferred from alignments containing more than 500 amino acids. This dataset with longer genes was composed to minimize the effects of sampling errors in gene tree topology estimation, which arguably lead to a higher probability of inferring the correct species tree (Song et al., 2012). In ASTRAL, 100 bootstrap replicates were estimated to obtain branch supports.

To evaluate the robustness of bootstrap values of the species tree, we used the ape package v. 3.5 (Paradis et al., 2004) of R language v. 3.2.0 (R Core Team, 2016) to draw a correlation between the bootstrap support value and the number of genes and sites in the alignments. Thus, we investigated whether nodes with low bootstrap support were the ones that included species with lower number of genes/sites. Additionally, Robinson-Foulds distances were computed comparing the inferred gene trees to Misof's phylogeny and to both species trees that we inferred in ASTRAL (with all genes and with genes that were longer than 500 amino acids).

## Results and Discussion

Our inferred species tree from the full dataset failed to support the monophyly of Hexapoda, as recovered by Misof et al. (Misof et al., 2014), due the placement of a crustacean species, the remipede *Speleonectes tulumensis*, as the sister-group of Insecta (**Figure 1**). This result was also recovered in the dataset containing gene alignments longer than 500 bp (**Figure S1**). The only interordinal difference within Hexapoda between both inferred species trees was that the Megaloptera was recovered as monophyletic only when all genes were used. Also, in both datasets, MSC failed to infer the monophyly of Neuroptera. *Conwentzia psociformis* was placed as sister group to the remaining Neuropterida (Raphidioptera + Megaloptera + Neuroptera) in the complete dataset and as sister-group to the Neuropterida (with exception of *Sialis*) in the dataset with longer alignments.

In Insecta, all orders except Megaloptera, Neuroptera and Blattodea were inferred as monophyletic by both data sets. Moreover, several major supra ordinal groups were also recovered as monophyletic, e.g., Holometabola, Condylgnatha, Polyneoptera and Palaeoptera. The phylogenetic relationships within Hexapoda orders were stable between data sets, with the exception of Coleoptera. In the complete dataset, *Priacma* was inferred as sister-clade of a group formed by *Carabus* and *Gyrinus*; this three-species clade was the sister-group of the remaining Coleoptera. When larger alignments were analyzed, *Priacma* was the sister-group of the remaining Coleoptera.

### *Non-monophyly of Hexapoda*

As (Misof et al., 2014), Our phylogenetic tree also recovered a clade formed by Protura and Collembola, which is the sister-group of all the remaining Hexapoda (Diplura, Insecta) in this same work. Our species tree did not recover Misof et al.'s hexapod arrangement, ((Protura, Collembola), (Diplura, Insecta)), because the Remipedia crustacean *Speleonectes tulumensis* was inferred as the sister-group of Insecta, whereas Diplura was the sister-group of *S. tulumensis* + Insecta. Von Reumont et al. (Von Reumont et al., 2012), using

phylogenomic data, also found a clade composed of Protura and Collembola, in their analysis, however, Diplura was the sister-group of Protura and Collembola, whereas Insecta was the sister-group of these three clades.

The majority of recent molecular phylogenetic surveys of Hexapoda proposed the arrangement (((Diplura, Protura), Collembola), Insecta) (Kjer et al., 2006; Meusemann et al., 2010; Rainford et al., 2016; Trautwein et al., 2012), although only Meusemann et al. (Meusemann et al., 2010) used a phylogenomic dataset to infer the phylogeny. An alternative higher order evolutionary relationship of hexapods was proposed by Sasaki et al. (Sasaki et al., 2013), who estimated the topology (((Collembola, Diplura), Insecta), Protura). When compared to Misof et al.'s (Misof et al., 2014) tree, our species tree presented few interordinal differences within Insecta, mainly associated with the Neuropteroidea and Polyneoptera supergroups.

#### *Weak support for Neuropteroidea*

In the consensus phylogeny of Neuropteroidea, this lineage is composed by Neuropterida – ((Megalloptera, Neuroptera), Raphidioptera), which is sister to the (Coleoptera, Strepsiptera) clade (Mckenna et al., 2015; Misof et al., 2014; Peters et al., 2014; Wiegmann et al., 2009). In our analysis, relationships within Neuropterida differed. While recent studies supported either the previous topology (Cameron et al., 2009; Mckenna et al., 2015; Misof et al., 2014; Peters et al., 2014; Wang et al., 2012; Wiegmann et al., 2009; Yan et al., 2014; Zhao et al., 2013), or the remaining two topological alternatives (Aspöck et al., 2012; Ishiwata et al., 2011; McKenna and Farrell, 2010), our results supported the non-monophyly of Neuroptera for the first time using molecular data, placing *Conwentzia psociformis* as sister-group to the remaining Neuropterida.

It is worth mentioning that in previous studies using morphology or molecules the monophyly of Neuroptera was undisputed (Aspöck et al., 2012). Most studies, however, used only one to three Neuroptera species, and a few genes that were concatenated into a single supermatrix

for phylogeny inference. Thus, poor taxonomic sampling – the highest number of Neuroptera species included in a single analysis was 10 (Yan et al., 2014), associated with low number of sampled genes and the handling of data likely led to differences regarding Neuroptera relationships. The bootstrap support value for the Neuropterida clade excluding *Conwentzia psociformis* was very low (31). Such a low support indicates that additional genomic data from other species of these groups are needed to better resolve this node (Lambert et al., 2015).

The second issue within Neuropteroidea was the position of the Strepsiptera order, which has been debated for a long time (Beutel and Pohl, 2006; Whiting et al., 1997). The consensus based on recent works put this order together with Coleoptera (Boussau et al., 2014; Cameron et al., 2009; Mckenna et al., 2015; Misof et al., 2014; Peters et al., 2014; Wang et al., 2012; Wiegmann et al., 2009; Yan et al., 2014; Zhao et al., 2013), however our results do not corroborate to this general view, since Strepsiptera was placed as sister-group to all others Neuropteroidea with high support (100). This way, our findings indicate that Strepsiptera problem is still open, bringing up more discussion on this relationship.

#### *Polyneoptera interordinal relationships*

Within the Polyneoptera superorder, differences between Misof et al.'s (Misof et al., 2014) and the species trees estimated in the present work were more relevant. In our species tree, Orthoptera was the sister-group of a clade formed by Mantophasmatodea, Grylloblattodea, Mantodea, Blattodea and Isoptera with bootstrap support value of 55. This six-order clade clustered with a group composed of Empioptera and Phasmatodea (100). In Misof et al.'s tree, the Orthoptera was placed in basal position within this lineage, with Mantophasmatodea and Grylloblattodea clustering with Embioptera and Phasmatodea. This four-order clade was clustered with Dictyoptera (Mantodea, Blattodea and Isoptera), which was also recovered in our species tree. The remaining differences within the Polyneoptera was the placement of Zoraptera and Demaptera. In Misof et al.'s tree these two orders were clustered together and



they were the sister-group of the rest of Polyneoptera, while in our species tree Demaptera was the sister to the remaining Polyneoptera.

A general comparison with previous works is complicated by the fact that the number of orders used varied significantly among studies, making a one-to-one evaluation of the phylogenetic placement of Polyneoptera orders difficult. Although previous studies employed a varied combination of taxon and genes, they estimated some of the clades found here; for instance, the Dictyoptera (Ishiwata et al., 2011; Ma et al., 2014; Misof et al., 2014; Sasaki et al., 2013; Tomita et al., 2012; Wan et al., 2012; Wu et al., 2014), the Embioptera + Phasmatodea (Ishiwata et al., 2011; Misof et al., 2014; Sasaki et al., 2013; Wu et al., 2014) and the Mantophasmatodea + Grylloblattodea (Ma et al., 2014; Misof et al., 2014; Sasaki et al., 2013).

#### *Support values of controversial relationships*

It is worth noting that most nodes with poor support lied within Neuropteroidea and Polyneoptera groups. As shown by Lambert et al. (Lambert et al., 2015), topologies estimated by MSC methods and concatenation will tend to disagree when short branches and low support values are present in the concatenated tree. Differences between our species tree and Misof et al.'s concatenated tree possibly recapitulated this pattern. In Misof et al.'s concatenated tree, there were 7 support values below 50, with 3 located at interordinal ancestral nodes. These three nodes resulted in interordinal discordances between Misof et al.'s (Misof et al., 2014) and our MSC trees, namely, (i) the support of (Protura, Collembola) as sister-group of (Diplura, Insecta) in Misof et al.'s tree was 14, whereas our support for *Speleonectes* as sister-group of Insecta was 80; (ii) the support for Dictyoptera as sister-group to the ((Mantophasmatodea, Grylloblattodea), (Embioptera, Phasmatodea)) clade in Misof et al.'s (Misof et al., 2014) tree was 34 and we inferred Dictyoptera as sister-group of (Mantophasmatodea, Grylloblattodea) with bootstrap support value of 45; (iii) the bootstrap support of for the relationship of Neuroptera and Megaloptera

as sister-groups was 15 in Misof et al.'s tree, whereas we recovered Neuroptera as non-monophyletic clade with bootstrap value of 31.

Additionally, we investigated whether a correlation existed between the number of genes (or sites) (S1 Table) and the bootstrap support values of the species tree, but no significant correlation was found for either the number of genes or sites. Because no correlation was found between the number of sites and the bootstrap support, it is unlikely that sequencing more genes alone will solve the remaining controversies on insect phylogeny. Instead, it may be possible that increasing the number of sampled species for the orders near poorly-supported clades would improve topological support. The addition of taxa to clarify the discordances was already recommended in other works as well.

Given the discordances between our MSC tree and Misof et al.'s tree, we were prompted to investigate which of the two topologies lied closer to the forest of inferred gene trees. To this end, we calculated the distribution of topological distances between Misof et al.'s and the MSC tree and the gene tree forest. The distribution of RF distances failed to indicate any departure between both topologies tested, implying that distance from the forest of gene trees does not favor any topological hypothesis (**Figure 2**).

### *Species tree*

Despite some criticism on the MSC method (Springer and Gatesy, 2016), it has clear advantages over concatenated methods, e. g. better performance in the presence of high degrees of ILS and short branches and the incorporation of gene tree heterogeneity (Edwards et al., 2016; Liu et al., 2015). Then, using a huge data (144 species and 595,033 amino acid sites) we could infer a deep-level phylogenetic tree under the MSC approach. The comparison of the recovered species tree with phylogenetic hypotheses based solely on concatenated methods showed three main disagreements: the non-monophyly of Hexapoda, which also revealed an incongruence at the root of Insecta, and the interordinal relationships within Polyneoptera and Neuroptera.

Even with the challenges caused by the great size of the effective population size associated with short branches in Insecta, our MSC analysis was able to recover several orders and relationships according to the literature of Insecta systematics. Further, the results and discussion presented here clearly show the most incongruent nodes are in the Insecta tree of life, for which more studies are necessary to achieve a better historical hypothesis for the group.

## Acknowledgments

This work is part of the Doctoral thesis of LAF funded by Brazilian Ministry of Education (CAPES) and the Brazilian Research Council (CNPq).

## Figure legends

**Figure 1:** Insect species tree inferred from the full dataset. Support values on nodes were obtained from a hundred bootstrap replicates. Distinct colors indicate ordinal clades. Gray boxes indicate the different interordinal relationships when compared to Misof et al. (2014)'s results.

**Figure 2:** Robinson-Foulds distance values computed between each gene tree and Misof et al. (2014) topology and Astral topology. The dashed blue line indicates the one to one relationship.

## References

- Aspöck, U., Haring, E., Aspöck, H., 2012. The phylogeny of the Neuropterida: Long lasting and current controversies and challenges (Insecta: Endopterygota). *Arthropod Syst. Phylogeny* 70, 119–129.
- Beutel, R.G., Pohl, H., 2006. Endopterygote systematics - where do we stand and what is the goal (Hexoptera, Arthropoda)? *Syst. Biol.* 31, 202–219. doi:10.1111/j.1365-3113.2006.00341.x
- Boussau, B., Walton, Z., Delgado, J.A., Collantes, F., Beani, L., Stewart, I.J., Cameron, S.A., Whitfield, J.B., Johnston, J.S., Holland, P.W.H., Bachtrog, D., Kathirithamby, J., Huelsenbeck, J.P., 2014. Strepsiptera, phylogenomics and the long branch attraction problem. *PLoS One* 9. doi:10.1371/journal.pone.0107709
- Cameron, S.L., Sullivan, J., Song, H., Miller, K.B., Whiting, M.F., 2009. A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. *Zool. Scr.* 38, 575–590. doi:10.1111/j.1463-6409.2009.00392.x
- Cannon, J.T., Vellutini, B.C., Smith, J., Ronquist, F., Jondelius, U., Hejnol, A., 2016. Xenacoelomorpha is the sister group to Nephrozoa. *Nature* 530, 89–93. doi:10.1038/nature16520
- de Queiroz, A., Gatesy, J., 2007. The supermatrix approach to systematics. *Trends Ecol. Evol.* 22, 34–41. doi:10.1016/j.tree.2006.10.002
- Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24, 332–340. doi:10.1016/j.tree.2009.01.009
- Edwards, S. V., 2009. Is a new and general theory of molecular systematics emerging? *Evolution* (N. Y.) 63, 1–19. doi:10.1111/j.1558-5646.2008.00549.x
- Edwards, S. V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Mol. Phylogenet. Evol.* 94, 447–462. doi:10.1016/j.ympev.2015.10.027
- Huelsenbeck, J.P., 2007. Systematic Bias in Phylogenetic Analysis: Is the Strepsiptera Problem Solved? *Syst. Biol.* 47, 519–537.
- Ishiwata, K., Sasaki, G., Ogawa, J., Miyata, T., Su, Z.H., 2011. Phylogenetic relationships among insect orders based on three nuclear protein-coding gene sequences. *Mol. Phylogenet. Evol.* 58, 169–180. doi:10.1016/j.ympev.2010.11.001
- Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., Fonseca, R.R. da, Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M.S., Zavidovych, V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W.C., Ray, D., Green, R.E., Bruford, M., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E.P., Bertelsen, M.F., Sheldon, F., Brumfield, R.T., Mello, C., Lovell, P. V., Wirthlin, M., Samaniego, J.A., Velazquez, A.M.V., Alfaro-Núñez, A., Campos, P.F., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D., Zhou, Q., Perelman, P., Driskell, A.C., Ruby, G., Shapiro, B., Xiong, Z.,

- Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F.E., Braun, M., Fjeldsa, J., Orlando, L., Barker, K., Jönsson, K.A., Johnson, W., Koepfli, K.-P., O'Brien, S., Haussler, D., Ryder, O.A., Rahbek, C., Willerslev, E., Graves, G.R., Glenn, T.C., McCormack, J., Burt, D., Ellegren, H., Alström, P., Edwards, S., Stamatakis, A., Mindell, D.P., Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T.P., Zhang, G., 2014. Whole Genome Analyses Resolve the Early Branches in the Tree of Life of Modern Birds. *Science* (80-. ). 346, 1126–1138. doi:10.1126/science.1253451
- Kingman, J.F.C., 1982. The coalescent. *Stoch. Process. their Appl.* 13, 235–248. doi:10.1016/0304-4149(82)90011-4
- Kjer, K.M., 2004. Aligned 18S and insect phylogeny. *Syst. Biol.* 53, 506–514. doi:10.1080/10635150490445922
- Kjer, K.M., Carle, F.L., Litman, J., Ware, J., 2006. A molecular phylogeny of Hexapoda. *Arthropod Syst. Phylogeny* 64, 35–44. doi:10.1126/science.283.5404.998
- Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56, 17–24. doi:10.1080/10635150601146041
- Lambert, S.M., Reeder, T.W., Wiens, J.J., 2015. Molecular Phylogenetics and Evolution When do species-tree and concatenated estimates disagree ? An empirical analysis with higher-level scincid lizard phylogeny 82, 146–155.
- Lanier, H.C., Knowles, L.L., 2015. Applying species-tree analyses to deep phylogenetic histories: Challenges and potential suggested from a survey of empirical phylogenetic studies. *Mol. Phylogenet. Evol.* 83, 191–199. doi:10.1016/j.ympev.2014.10.022
- Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320. doi:10.1093/molbev/msn067
- Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S. V., 2015. Estimating phylogenetic trees from genome-scale data. *Ann. N. Y. Acad. Sci.* 1360, 36–53. doi:10.1111/nyas.12747
- Ma, C., Wang, Y., Wu, C., Kang, L., Liu, C., 2014. The compact mitochondrial genome of *Zorotypus medoensis* provides insights into phylogenetic position of Zoraptera. *BMC Genomics* 15, 1156. doi:10.1186/1471-2164-15-1156
- Maddison, W.P., Knowles, L.L., 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting. *Syst. Biol.* 55, 21–30. doi:10.1080/10635150500354928
- McKenna, D.D., Farrell, B.D., 2010. 9-genes reinforce the phylogeny of holometabola and yield alternate views on the phylogenetic placement of Strepsiptera. *PLoS One* 5, 1–8. doi:10.1371/journal.pone.0011887
- Mckenna, D.D., Wild, A.L., Kanda, K., Bellamy, C.L., Beutel, R.G., Caterino, M.S., Farnum, C.W., Hawks, D.C., Ivie, M.A., Jameson, M.L., Leschen, R.A.B., Marvaldi, A.E., Mchugh, J. V., Newton, A.F., Robertson, J.A., Thayer, M.K., Whiting, M.F., Lawrence, J.F., Ślipiński, A., Maddison, D.R., Farrell, B.D., 2015. The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. *Syst. Entomol.* 40, 835–880. doi:10.1111/syen.12132
- Meusemann, K., Von Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kück, P., Ebersberger, I., Walz, M., Pass, G., Breuers, S., Achter, V., Von Haeseler, A., Burmester, T., Hadrys, H., Wägele, J.W., Misof, B., 2010. A phylogenomic approach to resolve the arthropod tree of life.

Mol. Biol. Evol. 27, 2451–2464. doi:10.1093/molbev/msq130

- Miller, M.A., Schwartz, T., Pickett, B.E., He, S., Klem, E.B., Scheuermann, R.H., Passarotti, M., Kaufman, S., Oleary, M.A., 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol. Bioinforma.* 11, 43–48. doi:10.4137/EBO.S21501
- Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., S. Swenson, M., Warnow, T., 2014. ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, 541–548. doi:10.1093/bioinformatics/btu462
- 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., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walz, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Xu, X., Yang, H., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* (80-. ). 346, 763–767. doi:10.1126/science.1257570
- Oliver, J.C., 2013. Microevolutionary processes generate phylogenomic discordance at ancient divergences. *Evolution* (N. Y.). 67, 1823–1830. doi:10.1111/evo.12047
- Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5, 568–583. doi:10.1093/oxfordjournals.molbev.a040517
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi:10.1093/bioinformatics/btg412
- Peters, R.S., Meusemann, K., Petersen, M., Mayer, C., Wilbrandt, J., Ziesmann, T., Donath, A., Kjer, K.M., Aspöck, U., Aspöck, H., Aberer, A., Stamatakis, A., Friedrich, F., Hünefeld, F., Niehuis, O., Beutel, R.G., Misof, B., 2014. The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. *BMC Evol. Biol.* 14, 52. doi:10.1186/1471-2148-14-52
- R Core Team, 2016. R: A language and environment for statistical computing. R A Lang. Environ. Stat. Comput.
- Rainford, J.L., Hofreiter, M., Mayhew, P.J., 2016. Phylogenetic analyses suggest that diversification and body size evolution are independent in insects. *BMC Evol. Biol.* 16, 8. doi:10.1186/s12862-015-0570-3
- Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463, 1079–1083. doi:10.1038/nature08742
- Romiguier, J., Lourenco, J., Gayral, P., Faivre, N., Weinert, L.A., Ravel, S., Ballenghien, M., Cahais, V., Bernard, A., Loire, E., Keller, L., Galtier, N., 2014. Population genomics of eusocial insects: The costs of a vertebrate-like effective population size. *J. Evol. Biol.* 27,



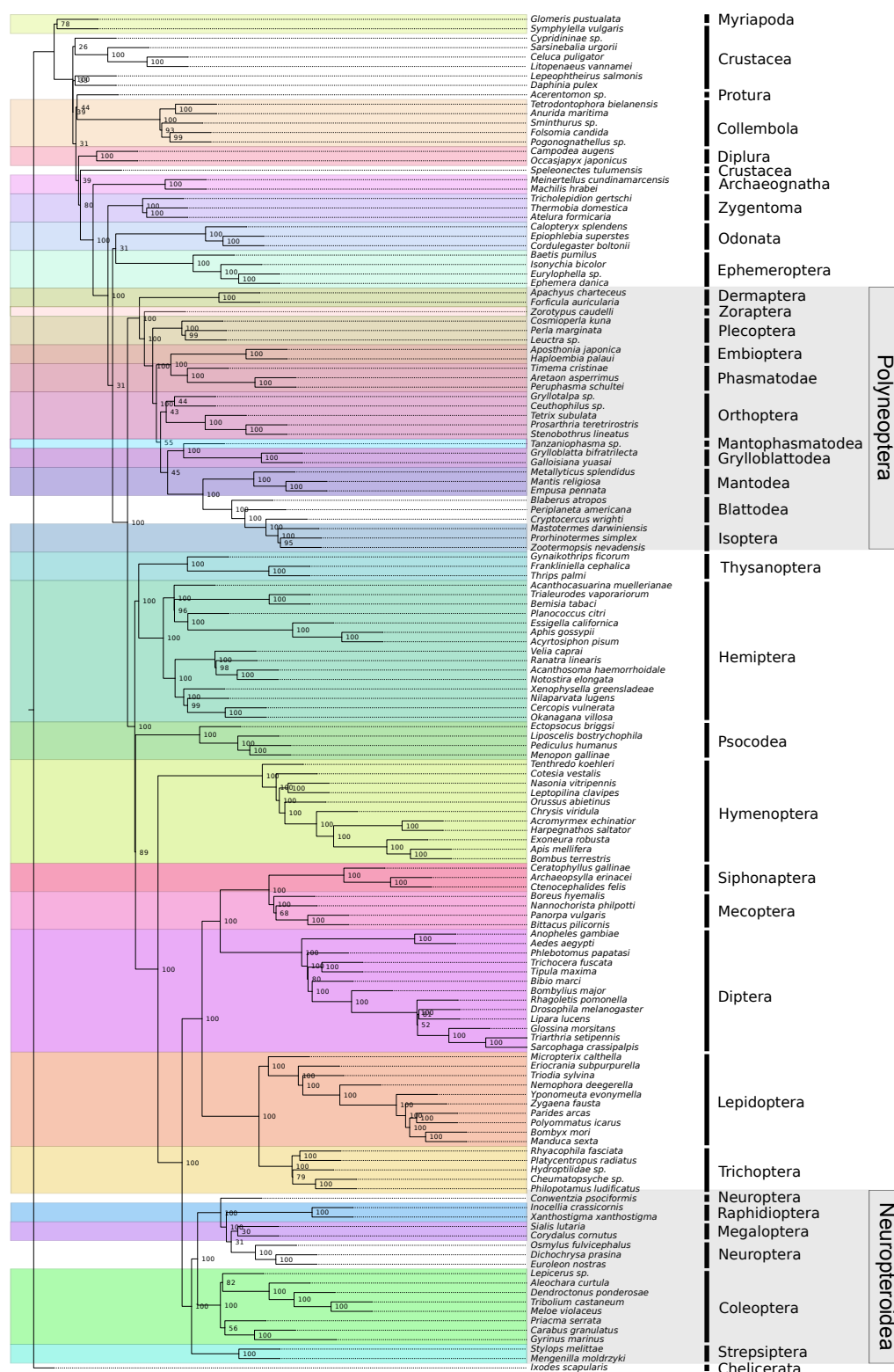
593–603. doi:10.1111/jeb.12331

- Sasaki, G., Ishiwata, K., Machida, R., Miyata, T., Su, Z.-H., 2013. Molecular phylogenetic analyses support the monophyly of Hexapoda and suggest the paraphyly of Entognatha. *BMC Evol. Biol.* 13, 236. doi:10.1186/1471-2148-13-236
- Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654–1668. doi:10.1093/molbev/msw079
- Song, S., Liu, L., Edwards, S. V., Wu, S., 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci.* 109, 14942–14947. doi:10.1073/pnas.1518753112
- Springer, M.S., Gatesy, J., 2016. The gene tree delusion. *Mol. Phylogenet. Evol.* 94, 1–33. doi:10.1016/j.ympev.2015.07.018
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033
- Tomita, S., Yukuhiro, K., Kômoto, N., 2012. The mitochondrial genome of a stick insect *Extatosoma tiaratum* (Phasmatodea) and the phylogeny of polyneopteran insects. *J. Insect Biotechnol. Sericology* 88, 79–88. doi:10.11416/jibs.80.3
- Trautwein, M.D., Wiegmann, B.M., Beutel, R., Kjer, K.M., Yeates, D.K., 2012. Advances in insect phylogeny at the dawn of the postgenomic era. *Annu. Rev. Entomol.* 57, 449–68. doi:10.1146/annurev-ento-120710-100538
- Von Reumont, B.M., Jenner, R.A., Wills, M.A., Dell’Ampio, E., Pass, G., Ebersberger, I., Meyer, B., Koenemann, S., Iliffe, T.M., Stamatakis, A., Niehuis, O., Meusemann, K., Misof, B., 2012. Pancrustacean phylogeny in the light of new phylogenomic data: Support for remipedia as the possible sister group of hexapoda. *Mol. Biol. Evol.* 29, 1031–1045. doi:10.1093/molbev/msr270
- Wan, X., Kim, M. Il, Kim, M.J., Kim, I., 2012. Complete mitochondrial genome of the free-living earwig, *Challia fletcheri* (dermaptera: Pygidicranidae) and phylogeny of polyneoptera. *PLoS One* 7. doi:10.1371/journal.pone.0042056
- Wang, Y., Liu, X., Winterton, S.L., Yang, D., 2012. The First Mitochondrial Genome for the Fishfly Subfamily Chauliodinae and Implications for the Higher Phylogeny of Megaloptera. *PLoS One* 7. doi:10.1371/journal.pone.0047302
- Whitfield, J.B., Kjer, K.M., 2008. Ancient rapid radiations of insects: challenges for phylogenetic analysis. *Annu. Rev. Entomol.* 53, 449–472. doi:10.1146/annurev.ento.53.103106.093304
- Whiting, M.F., Carpenter, J.C., Wheeler, Q.D., Wheeler, W.C., 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.* 46, 1–68. doi:10.1093/sysbio/46.1.1
- Wiegmann, B.M., Trautwein, M.D., Kim, J.-W., Cassel, B.K., Bertone, M.A., Winterton, S.L., Yeates, D.K., 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol.* 7, 34. doi:10.1186/1741-7007-7-34
- Wu, H.Y., Ji, X.Y., Yu, W.W., Du, Y.Z., 2014. Complete mitochondrial genome of the stonefly *Cryptoperla stilifera* Sivec (Plecoptera: Peltoperlidae) and the phylogeny of Polyneopteran insects. *Gene*. doi:10.1016/j.gene.2013.12.044

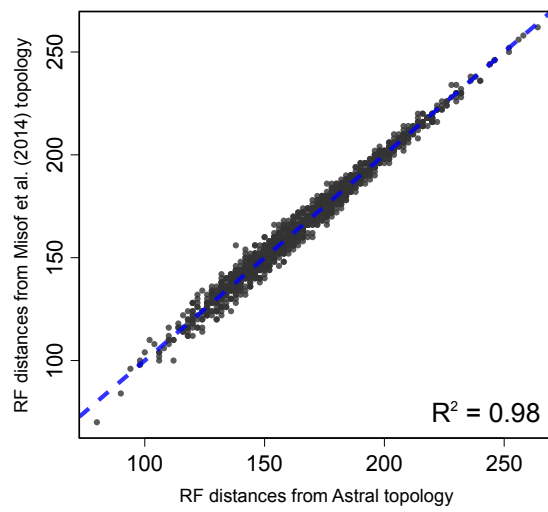


- Xi, Z., Liu, L., Rest, J.S., Davis, C.C., 2014. Coalescent versus concatenation methods and the placement of Amborella as sister to water lilies. *Syst. Biol.* 63, 919–932. doi:10.1093/sysbio/syu055
- Yan, Y., Wang, Y., Liu, X., Winterton, S.L., Yang, D., 2014. The first mitochondrial genomes of antlion (Neuroptera: Myrmeleontidae) and split-footed lacewing (Neuroptera: Nymphidae), with phylogenetic implications of myrmeleontiformia. *Int. J. Biol. Sci.* 10, 895–908. doi:10.7150/ijbs.9454
- Yang, Z., 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* 10, 1396–1401.
- Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* 13, 303–314. doi:10.1038/nrg3186
- Zhao, J., Li, H., Winterton, S.L., Liu, Z., 2013. Ancestral Gene Organization in the Mitochondrial Genome of *Thyridosmylus langii* (McLachlan, 1870) (Neuroptera: Osmylidae) and Implications for Lacewing Evolution. *PLoS One* 8, 1–12. doi:10.1371/journal.pone.0062943
- Zimmermann, T., Mirarab, S., Warnow, T., 2014. BBICA: Improving the scalability of \*BEAST using random binning. *BMC Genomics* 15. doi:10.1186/1471-2164-15-S6-S11

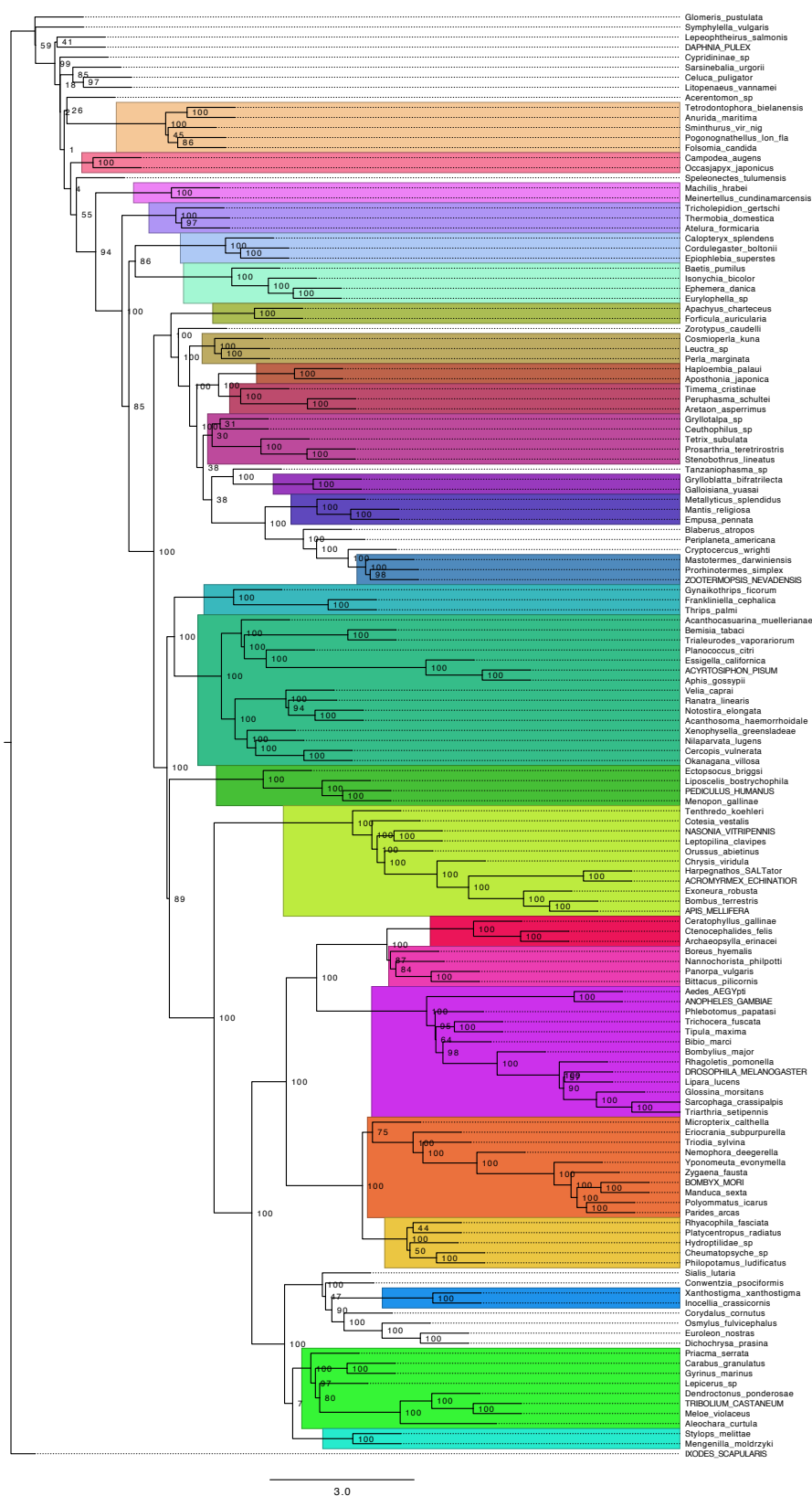
**Figure 1**



**Figure 2**



## Supplementary information



**Figure S1.** Phylogenetic tree inferred using 377 genes with more than 500 amino acids