



Letter to the Editor



Resolving the phylogenetic position of Hygrobiidae (Coleoptera: Adephaga) requires objective statistical tests and exhaustive phylogenetic methodology: a response to Cai et al. (2020)

1. Introduction

Cai et al. (2020) (hereafter CEA) investigated the phylogenetic position of the family Hygrobiidae by re-analyzing a phylogenomic data set that we had published (Vasilikopoulos et al., 2019). After trimming the original supermatrices to remove hypervariable sites and applying a site-heterogeneous model for phylogenetic reconstruction they suggested a sister group relationship between Dytiscidae and Hygrobiidae. They also concluded that the backbone phylogeny of Dytiscoidea is now more robustly resolved. Although a clade of Dytiscidae + Hygrobiidae was supported in some studies of morphological characters (e.g., Beutel et al., 2020), we disagree that the phylogenetic position of Hygrobiidae is robustly resolved. Here, we provide justification for our claims and point out some problems in the study of CEA.

2. Resolution of difficult phylogenetic questions requires integrative phylogenomic approaches by comparing concatenation-based and multi-species coalescent phylogenetic analyses

CEA reanalyzed the trimmed supermatrices by applying the site-heterogeneous mixture model CAT + GTR + Γ 4 implemented in the software Phylobayes (Lartillot et al., 2013). They suggest that the signal in support of Dytiscidae + Hygrobiidae is low, most likely due to a rapid diversification in the common ancestral lineage leading to this clade (Cai et al., 2020). Despite this, they only provided concatenation-based analyses and did not employ methods that account for gene-tree heterogeneity (Liu et al., 2015; Xu and Yang, 2016). Under conditions of rapid lineage diversification, a high level of incomplete lineage sorting might obscure reliable estimates of the topology in concatenation-based analyses (Kubatko and Degnan, 2007; Roch and Steel, 2015). Because evolutionary processes that generated the data under investigation are not known *a priori*, concatenation and multi-species coalescent approaches should be both applied for resolving highly controversial phylogenetic relationships when rapid diversification is suspected (Cloutier et al., 2019; Williams et al., 2020). Previous coalescent-based phylogenomic analyses of Adephaga have suggested Hygrobiidae either as sister to Amphizoidae + Aspdytidae or as sister to Dytiscidae + (Amphizoidae + Aspdytidae) (Gustafson et al., 2020; Vasilikopoulos et al., 2019). Therefore a clade Dytiscidae + Hygrobiidae as suggested by CEA has not yet been corroborated by coalescent-based phylogenomic analyses.

3. Claims about fit of phylogenetic models to the data should be accompanied by proper statistical tests of model fit

CEA reanalyzed our data both with maximum likelihood and Bayesian phylogenetic inference. For their maximum-likelihood phylogenetic analyses they used the LG4X mixture model (unpartitioned) (Le et al., 2012) without performing a proper statistical model-selection test before phylogenetic reconstructions (Posada and Buckley, 2004; Sullivan and Joyce, 2005). Selecting a less fitting model for phylogenetic reconstruction can affect both the inferred topology and branch lengths, but also the inferred branch support statistics (Buckley and Cunningham, 2002; Hoang et al., 2018; Sullivan and Joyce, 2005). Therefore, it remains unclear to what extent branch support statistics in their maximum-likelihood analyses are due to model misspecification.

CEA suggest that they applied “better-fitting” site-heterogeneous models in comparison to our previously applied “time-saving” site-homogeneous models. In our analyses we selected the best-fitting evolutionary models with the software ModelFinder and Partition-Finder based on objective statistical criteria (Sullivan and Joyce, 2005). In contrast, CEA concluded that CAT + GTR + Γ 4 is a better-fitting model than the LG4X model without performing a proper statistical test. Although we consider it very likely that the more complex CAT + GTR + Γ 4 fits the data better than LG4X, CEA did not provide statistical comparison between the LG4X and the CAT + GTR + Γ 4 model.

Our model selection procedure did not test for the relative fit of site-heterogeneous models. Despite this, we analyzed our data with maximum likelihood-based site-heterogeneous models (Wang et al., 2017). Although these models are less complex than the model CAT + GTR + Γ 4, they have been shown to effectively ameliorate long-branch attraction artifacts due to amino-acid site-pattern heterogeneity (Wang et al., 2017, 2019). Analyses under these models failed to recover Dytiscidae + Hygrobiidae and the same applies for the analysis of the original supermatrix H under the site-heterogeneous CAT + GTR + Γ 4 (Cai et al., 2020). Therefore, we maintain that the models we have previously applied are appropriate for the analysis of our data.

4. *A priori* selection of matrices for phylogenetic reconstruction should be based on objective statistical criteria

The clade of Dytiscidae + Hygrobiidae is only recovered in analyses of the trimmed matrices and only under the site-heterogeneous model CAT + GTR + Γ 4. However, CEA did not provide objective statistical criteria (e.g., calculation of information content, deviation from stationarity, reversibility and homogeneity, saturation plots, pairwise missing data) in order to justify their choice of these datasets for analyses. The proportion of missing data within each matrix that is provided by CEA is

not particularly informative, as it does not provide information on whether or not missing data are randomly distributed across taxa (Misof et al., 2014a). It is expected that the removal of hypervariable sites should alleviate some of the problems associated with fast evolving sites (Misof et al., 2001), but at the same time also removes phylogenetic information (Fan et al., 2020; Tan et al., 2015). Therefore a balance must be obtained between these two. The manual of the BMGE method suggests that the default value of $h = 0.5$ (maximum entropy-like score per site) should be used for obtaining biologically realistic results when removing fast-evolving sites (Criscuolo and Gribaldo, 2010). Despite this, CEA used a very stringent trimming approach with $h = 0.4$ and even $h = 0.3$ and in most cases combined with the very conservative BLOSUM95 matrix that should typically be used for the analysis of closely related species. However, the families of Dytiscoidea have diverged from each other at least 150 million years ago (Hawlitschek et al., 2012). Taken together, we do not think with the methodological approach for removing hypervariable sites or for selecting data matrices is well-justified.

5. Taxon-sampling is a potential source of phylogenetic error

Taxon sampling has been well-documented to affect phylogenomic analysis (e.g., Prasanna et al. 2020; Philippe et al. 2011). In our phylogenomic analyses of Dytiscoidea we have discussed the possibility that the inferred position of Hygrobiidae may not be robustly resolved due to the low number of taxa in the analyzed datasets (Vasilikopoulos et al., 2019). Additionally, Gustafson et al. (2020) reported that, in their study, the placement of Hygrobiidae changed primarily when taxon sampling was decreased. Despite this, CEA reanalyzed the same datasets of Vasilikopoulos et al. (2019), but did not recognize the potential negative effects of poor taxon sampling on their phylogenetic reconstructions suggesting a “well-resolved” phylogeny of Dytiscoidea (Cai et al., 2020).

6. The importance of critically evaluating branch support and phylogenetic conflict in phylogenomic analyses

An increasing number of studies have concluded that analysis of conflicting signals is critical to identify potentially inflated branch support or topological artefacts in concatenation-based analyses (Brown and Thomson, 2017; Johnson et al., 2018; Pease et al., 2018; Shen et al., 2017; Vasilikopoulos et al., 2020; Walker et al., 2018). Specifically, some of these studies have shown that heterogeneous phylogenetic signal at a few sites or genes of a supermatrix can distort phylogenomic inference resulting in strongly supported but incorrect phylogenies (Brown and Thomson, 2017; Shen et al., 2017; Walker et al., 2018). Some authors have explored alternative phylogenetic signals across taxon subsets (i.e., quartets or bipartitions) as alternative measures of phylogenetic support and in order to identify the relative contribution of various confounding factors on the phylogenetic results (Johnson et al., 2018; Kobert et al., 2016; Misof et al., 2014b; Pease et al., 2018). Concatenation-based analyses that do not examine the relative support for alternative hypotheses across alignment sites or across taxa in the data are potentially prone to erroneous inferences or inflated branch support (Johnson et al., 2018; Vasilikopoulos et al., 2020; Walker et al., 2018). In our analyses, we evaluated support for alternative phylogenetic hypotheses with concatenation-based quartet measures (i.e., four-cluster likelihood mapping) (Strimmer and von Haeseler, 1997) and with gene-tree-based quartet scores in our coalescent-based analyses (Sayyari and Mirarab, 2016). Based on these analyses we found that the inferred position of Hygrobiidae is affected by conflicting signal, and we suggested that the position of the family requires further investigation. In contrast, CEA suggest that the phylogenetic position of Hygrobiidae is “well-resolved” by relying only on the Bayesian posterior probabilities under the CAT + GTR + Γ 4 model. Despite this, they did not employ measures to investigate conflicting signals in the data. We acknowledge that existing measures of phylogenomic conflict or topology tests are

model-dependent and more work is needed to develop appropriate tests based on site-heterogeneous models (Feuda et al., 2017). Nevertheless, posterior probabilities are not appropriate for evaluating conflicting phylogenetic signal and may suggest overconfident estimates (Kapli et al., 2020), especially when the datasets contain conflicting or low phylogenetic signal (Simmons and Norton, 2014; Suzuki et al., 2002), as is probably the case for datasets used to infer the phylogeny of Dytiscoidea (Vasilikopoulos et al., 2019). For these reasons we deem the phylogenetic position of Hygrobiidae to be still not robustly resolved.

7. The importance of properly documenting convergence statistics in Bayesian phylogenetics

CEA used the software *bpcomp* to assess convergence of their runs based on the largest discrepancy observed across all bipartitions (maxdiff). This measure assesses convergence on the tree space but not for the summary variables of the model (e.g., total length of the tree, alpha parameter of the gamma distribution, number of occupied components of the mixture), that is typically assessed with the software *tracemap*. It is therefore unclear if a sufficiently large number of independent samples for each of these variables (i.e., effective sample size) had been drawn from the posterior distribution during each Markov chain Monte Carlo (MCMC) run. The reliability of results from Bayesian phylogenetic analyses critically relies on the fact that the chains have reached stationarity and that a large number of independent samples is drawn from the posterior distribution (Nascimento et al., 2017; Ronquist et al., 2009; Williams et al., 2020). We therefore highlight the importance of properly documenting all convergence statistics in future Bayesian phylogenetic analyses of Dytiscoidea.

8. The importance of cross-validating results between different types of molecular data

CEA suggested that amino-acid sequences should be preferred over nucleotide sequences for analyses of ancient divergences. This argument relies on the assumption that amino-acid sequences can alleviate substitution saturation related to highly saturated third codon positions and codon usage bias (Breinholt and Kawahara, 2013; Rota-Stabelli et al., 2013). Despite this, different codon positions show different degree of deviation from the model assumptions (Naser-Khdour et al., 2019), and in many cases phylogenomic analyses of nucleotide sequences, after removing highly saturated nucleotide sites, have produced very similar results to those obtained from analyzing amino acids (Breinholt and Kawahara, 2013; Misof et al., 2014b). In addition, substitution processes at the nucleotide sequence level may be under different evolutionary constraints than amino-acids and therefore analyses of nucleotide sequences constitute an additional basis for cross-validating the results of analyses of amino-acids. The CAT + GTR + Γ 4 model was generated to account for site-specific amino-acid preferences (Lartillot and Philippe, 2004) but it is unclear whether or not it can fully describe substitutions at the protein level which can be dependent on complex properties of protein sequences such as protein secondary structure or solvent accessibility (Le and Gascuel, 2010; Pandey and Braun, 2020). We therefore suggest that comparing phylogenetic results between different types of data is important for assessing the robustness of phylogenetic estimates.

An example of different types of data for phylogenomic analysis are coding and non-coding regions of the genome. Many phylogenomic analyses are performed utilizing non-coding (or partially non-coding) nucleotide sequences such as ultraconserved elements (UCEs). UCEs have been successfully used to infer phylogenetic relationships both at shallow and deep timescales (Faircloth et al., 2012; Gustafson et al., 2020) and are potentially subject to different substitution processes than protein-coding nucleotide sequences. Therefore, UCEs constitute an additional largely independent type of data for testing the validity of phylogenetic hypotheses (Reddy et al., 2017). Phylogenomic analyses of

UCEs that properly modeled among-site rate heterogeneity did not recover a clade of Dytiscidae + Hygrobiidae (Baca et al., 2017; Gustafson et al., 2020). We conclude from these observations that a clade of Dytiscidae + Hygrobiidae as suggested by CEA requires confirmation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Baca, S.M., Alexander, A., Gustafson, G.T., Short, A.E.Z., 2017. Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly of 'Hydradephaga'. *Syst. Entomol.* 42, 1–10. <https://doi.org/10.1111/syen.12244>.
- Beutel, R.G., Ribera, I., Fikáček, M., Vasilikopoulos, A., Misof, B., Balke, M., 2020. The morphological evolution of the Adephaga (Coleoptera). *Syst. Entomol.* 45, 378–395. <https://doi.org/10.1111/syen.12403>.
- Breinholt, J.W., Kawahara, A.Y., 2013. Phylotranscriptomics: Saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome Biol. Evol.* 5, 2082–2092. <https://doi.org/10.1093/gbe/evt157>.
- Brown, J.M., Thomson, R.C., 2017. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. *Syst. Biol.* 66, 517–530. <https://doi.org/10.1093/sysbio/syw101>.
- Buckley, T.R., Cunningham, C.W., 2002. The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. *Mol. Biol. Evol.* 19, 394–405. <https://doi.org/10.1093/oxfordjournals.molbev.a004094>.
- Cai, C., Tihelka, E., Pisani, D., Donoghue, P.C.J., 2020. Data curation and modeling of compositional heterogeneity in insect phylogenomics: a case study of the phylogeny of Dytiscoidea (Coleoptera: Adephaga). *Mol. Phylogenet. Evol.* 147, 106782. <https://doi.org/10.1016/j.ympev.2020.106782>.
- Cloutier, A., Sackton, T.B., Grayson, P., Clamp, M., Baker, A.J., Edwards, S.V., 2019. Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. *Syst. Biol.* 68, 937–955. <https://doi.org/10.1093/sysbio/syz019>.
- Criscuolo, A., Gribaldo, S., 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10, 210. <https://doi.org/10.1186/1471-2148-10-210>.
- Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* 61, 717–726. <https://doi.org/10.1093/sysbio/sys004>.
- Fan, L., Wu, D., Goremykin, V., Xiao, J., Xu, Y., Garg, S., Zhang, C., Martin, W.F., Zhu, R., 2020. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-020-1239-x>.
- Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., Wörheide, G., Pisani, D., 2017. Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Curr. Biol.* 27, 3864–3870. <https://doi.org/10.1016/j.cub.2017.11.008>.
- Gustafson, G.T., Baca, S.M., Alexander, A.M., Short, A.E.Z., 2020. Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance. *Syst. Entomol.* 45, 552–570. <https://doi.org/10.1111/syen.12413>.
- Hawiltschek, O., Hendrich, L., Balke, M., 2012. Molecular phylogeny of the squeak beetles, a family with disjunct Palearctic-Australian range. *Mol. Phylogenet. Evol.* 62, 550–554. <https://doi.org/10.1016/j.ympev.2011.09.015>.
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Le, S.V., 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. <https://doi.org/10.1093/molbev/msx281>.
- Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., Wipfler, B., Peters, R.S., Allen, J.M., Petersen, M., Donath, A., Walden, K.K.O., Kozlov, A.M., Podsiadlowski, L., Mayer, C., Meusemann, K., Vasilikopoulos, A., Waterhouse, R.M., Cameron, S.L., Weirauch, C., Swanson, D.R., Percy, D.M., Hardy, N.B., Terry, I., Liu, S., Zhou, X., Misof, B., Robertson, H.M., Yoshizawa, K., 2018. Phylogenomics and the evolution of hemipteroid insects. *Proc. Natl. Acad. Sci.* 115, 12775–12780. <https://doi.org/10.1073/pnas.1815820115>.
- Kapli, P., Yang, Z., Telford, M.J., 2020. Phylogenetic tree building in the genomic age. *Nat. Rev. Genet.* <https://doi.org/10.1038/s41576-020-0233-0>.
- Kobert, K., Salichos, L., Rokas, A., Stamatakis, A., 2016. Computing the internode certainty and related measures from partial gene trees. *Mol. Biol. Evol.* 33, 1606–1617. <https://doi.org/10.1093/molbev/msw040>.
- Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56, 17–24. <https://doi.org/10.1080/10635150601146041>.
- Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095–1109. <https://doi.org/10.1093/molbev/msh112>.
- Lartillot, N., Rodrigue, N., Stubbs, D., Richer, J., 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* 62, 11–615. <https://doi.org/10.1093/sysbio/syt022>.
- Le, S.Q., Dang, C.C., Gascuel, O., 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol. Biol. Evol.* 29, 2921–2936. <https://doi.org/10.1093/molbev/mss112>.
- Le, S.Q., Gascuel, O., 2010. Accounting for solvent accessibility and secondary structure in protein phylogenetics is clearly beneficial. *Syst. Biol.* 59, 277–287. <https://doi.org/10.1093/sysbio/syq002>.
- 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. <https://doi.org/10.1111/nyas.12747>.
- 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., Jeremiin, 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., 2014b. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767. <https://doi.org/10.1126/science.1257570>.
- Misof, B., Meusemann, K., von Reumont, B.M., Kück, P., Prohaska, S.J., Stadler, P.F., 2014a. A priori assessment of data quality in molecular phylogenetics. *Algorithms Mol. Biol.* 9, 22. <https://doi.org/10.1186/s13015-014-0022-4>.
- Misof, B., Rickert, A.M., Buckley, T.R., Fleck, G., Sauer, K.P., 2001. Phylogenetic signal and its decay in mitochondrial SSU and LSU rRNA gene fragments of Anisoptera. *Mol. Biol. Evol.* 18, 27–37. <https://doi.org/10.1093/oxfordjournals.molbev.a003717>.
- Nascimento, F.F., dos Reis, M., Yang, Z., 2017. A biologist's guide to Bayesian phylogenetic analysis. *Nat. Ecol. Evol.* 1, 1446–1454. <https://doi.org/10.1038/s41559-017-0280-x>.
- Naser-Khdour, S., Minh, B.Q., Zhang, W., Stone, E.A., Lanfear, R., Bryant, D., 2019. The prevalence and impact of model violations in phylogenetic analysis. *Genome Biol. Evol.* 11, 3341–3352. <https://doi.org/10.1093/gbe/evz193>.
- Pandey, A., Braun, E.L., 2020. Phylogenetic analyses of sites in different protein structural environments result in distinct placements of the metazoan root. *Biology* 9, 64. <https://doi.org/10.3390/biology9040064>.
- Pease, J.B., Brown, J.W., Walker, J.F., Hinchliff, C.E., Smith, S.A., 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am. J. Bot.* 105, 385–403. <https://doi.org/10.1002/ajb2.1016>.
- Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9, e1000602. <https://doi.org/10.1371/journal.pbio.1000602>.
- Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53, 793–808. <https://doi.org/10.1080/10635150490522304>.
- Prasanna, A.N., Gerber, D., Kijpornyongpan, T., Aime, M.C., Doyle, V.P., Nagy, L.G., 2020. Model choice, missing data, and taxon sampling impact phylogenomic inference of deep Basidiomycota relationships. *Syst. Biol.* 69, 17–37. <https://doi.org/10.1093/sysbio/syz029>.
- Reddy, S., Kimball, R.T., Pandey, A., Hosner, P.A., Braun, M.J., Hackett, S.J., Han, K.-L., Harshman, J., Huddleston, C.J., Kingston, S., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Witt, C.C., Yuri, T., Brown, E.L., 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Syst. Biol.* 66, 857–879. <https://doi.org/10.1093/sysbio/syx041>.
- Roch, S., Steel, M., 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theor. Popul. Biol.* 100, 56–62. <https://doi.org/10.1016/j.tpb.2014.12.005>.
- Ronquist, F., van der Mark, P., Huelsenbeck, J.P., 2009. Bayesian phylogenetic analysis using MrBayes. In: Lemey, P., Salemi, M., Vandamme, A.-M. (Eds.), *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press, New York, pp. 210–265.
- Rota-Stabelli, O., Lartillot, N., Philippe, H., Pisani, D., 2013. Serine codon-usage bias in deep phylogenomics: pancrustacean relationships as a case study. *Syst. Biol.* 62, 121–133. <https://doi.org/10.1093/sysbio/sys077>.
- Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654–1668. <https://doi.org/10.1093/molbev/msw079>.
- Shen, X.-X., Hittinger, C.T., Rokas, A., 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* 1, 0126. <https://doi.org/10.1038/s41559-017-0126>.
- Simmons, M.P., Norton, A.P., 2014. Divergent maximum-likelihood-branch-support values for polytomies. *Mol. Phylogenet. Evol.* 73, 87–96. <https://doi.org/10.1016/j.ympev.2014.01.018>.
- Strimmer, K., von Haeseler, a., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. U. S. A.* 94, 6815–6819. <https://doi.org/10.1073/pnas.94.13.6815>.

- Sullivan, J., Joyce, P., 2005. Model selection in phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 36, 445–466. <https://doi.org/10.1146/annurev.ecolsys.36.102003.152633>.
- Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16138–16143. <https://doi.org/10.1073/pnas.212646199>.
- Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., Dessimoz, C., 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* 64, 778–791. <https://doi.org/10.1093/sysbio/syv033>.
- Vasilikopoulos, A., Balke, M., Beutel, R.G., Donath, A., Podsiadlowski, L., Pflug, J.M., Waterhouse, R.M., Meusemann, K., Peters, R.S., Escalona, H.E., Mayer, C., Liu, S., Hendrich, L., Alarie, Y., Bilton, D.T., Jia, F., Zhou, X., Maddison, D.R., Niehuis, O., Misof, B., 2019. Phylogenomics of the superfamily Dytiscoidea (Coleoptera: Adephaga) with an evaluation of phylogenetic conflict and systematic error. *Mol. Phylogenet. Evol.* 135, 270–285. <https://doi.org/10.1016/j.ympev.2019.02.022>.
- Vasilikopoulos, A., Misof, B., Meusemann, K., Lieberz, D., Flouri, T., Beutel, R.G., Niehuis, O., Wappler, T., Rust, J., Peters, R.S., Donath, A., Podsiadlowski, L., Mayer, C., Bartel, D., Böhm, A., Liu, S., Kapli, P., Greve, C., Jepson, J.E., Liu, X., Zhou, X., Aspöck, H., Aspöck, U., 2020. An integrative phylogenomic approach to elucidate the evolutionary history and divergence times of Neuropterida (Insecta: Holometabola). *BMC Evol. Biol.* 20, 64. <https://doi.org/10.1186/s12862-020-01631-6>.
- Walker, J.F., Brown, J.W., Smith, S.A., 2018. Analyzing contentious relationships and outlier genes in phylogenomics. *Syst. Biol.* 67, 916–924. <https://doi.org/10.1093/sysbio/syy043>.
- Wang, H.-C., Minh, B.Q., Susko, E., Roger, A.J., 2017. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* 67, 216–235. <https://doi.org/10.1093/sysbio/syx068>.
- Wang, H.C., Susko, E., Roger, A.J., 2019. The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. *Syst. Biol.* 68, 1003–1019. <https://doi.org/10.1093/sysbio/syz021>.
- Williams, T.A., Cox, C.J., Foster, P.G., Szöllösi, G.J., Embley, T.M., 2020. Phylogenomics provides robust support for a two-domains tree of life. *Nat. Ecol. Evol.* 4, 138–147. <https://doi.org/10.1038/s41559-019-1040-x>.
- Xu, B., Yang, Z., 2016. Challenges in species tree estimation under the multispecies coalescent model. *Genetics* 204, 1353–1368. <https://doi.org/10.1534/genetics.116.190173>.

Alexandros Vasilikopoulos^{a,*}, Grey T. Gustafson^b, Michael Balke^c,
Oliver Niehuis^d, Rolf G. Beutel^e, Bernhard Misof^f

^a Center for Molecular Biodiversity Research, Zoological Research Museum
Alexander Koenig, 53121 Bonn, Germany

^b Department of Ecology and Evolutionary Biology, University of Kansas,
Lawrence, 66045 KS, USA

^c Department of Entomology, SNSB-Bavarian State Collections of Zoology,
81247 Munich, Germany

^d Department of Evolutionary Biology and Ecology, Institute of Biology I
(Zoology), Albert Ludwig University of Freiburg, 79104 Freiburg, Germany

^e Institut für Zoologie und Evolutionsforschung, Friedrich-Schiller-
Universität Jena, 07743 Jena, Germany

^f Zoological Research Museum Alexander Koenig, 53121 Bonn, Germany

* Corresponding author.

E-mail address: a.vasilikopoulos@leibniz-zfmk.de (A. Vasilikopoulos).