

Low support values and lack of reproducibility of molecular phylogenetic analysis of Copepoda orders

Низкие значения поддержки и отсутствие воспроизводимости результатов молекулярно-филогенетического анализа отрядов Copepoda

K.V. Mikhailov^{1,2}, V.N. Ivanenko³
К.В. Михайлов^{1,2}, В.Н. Иваненко³

¹ Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 19992, Russia. E-mail: kv.mikhailov@gmail.com

¹ Институт физико-химической биологии им. Белозерского МГУ им. М.В. Ломоносова, Москва 19992, Россия

² Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127051, Russia.

² Институт проблем передачи информации им. А.А. Харкевича РАН, Москва 127051, Россия

³ Department of Invertebrate Zoology, Biological Faculty, Lomonosov Moscow State University, Moscow 19992, Russia. E-mails: ivanenko.slava@gmail.com

³ Кафедра зоологии беспозвоночных Биологического факультета МГУ им. М.В. Ломоносова, Москва 19992, Россия.

KEY WORDS: reproducibility, molecular phylogeny, Copepoda, Crustacea.

КЛЮЧЕВЫЕ СЛОВА: воспроизводимость, молекулярная филогения, веслоногие ракообразные.

ABSTRACT. Reanalysis of the dataset used by Khodami *et al.* [2017] reveals low support values for the key nodes of the copepod (Crustacea) phylogeny and fails to reproduce the results obtained in the study. Maximum likelihood (ML) and Bayesian analyses with the dataset produce phylogenies that are inconsistent with the branching of copepod groups proposed by Khodami *et al.* [2017]. The proposed phylogeny is refuted by the approximately unbiased (AU) statistical test, which undermines several conclusions drawn from the original study.

How to cite this paper: Mikhailov K.V., Ivanenko V.N. 2021. Low support values and lack of reproducibility of molecular phylogenetic analysis of Copepoda orders // *Arthropoda Selecta*. Vol.30. No.1. P.39–42, Supplement. doi: 10.15298/arthscl.30.1.04

РЕЗЮМЕ. Повторный анализ данных, использованных в работе Khodami *et al.* [2017] выявил низкие значения поддержки ключевых узлов филогении копепод (Crustacea) и не воспроизвел результаты, представленные в этой работе. Анализ данных методом максимального правдоподобия и байесовским подходом выявил филогенетические отношения, несоответствующие ветвлению групп веслоногих ракообразных, показанного Khodami *et al.* [2017]. Предложенная авторами филогения также отвергнута приблизительно несмещенным статистическим тестом (AU test), что ставит под сомнение ряд выводов, приведенных в рассматриваемой работе.

Introduction

Copepoda is a large and diverse group of arthropods that encompasses roughly 10 well-recognized orders. To this day, the phylogenetic relationship between the copepod orders remains in dispute [Huys, Boxshall, 1991; Ho, 1994; Ferrari *et al.*, 2010; Ferrari, von Vaupel Klein, 2019; Khodami *et al.*, 2019]. Early molecular phylogenetic analyses of the copepod phylogeny based on rDNA sequences found it difficult to draw clear conclusions on the branching of several groups [Huys *et al.*, 2007]. In this context, the study by Khodami *et al.* [2017], which explored this question using additional phylogenetic marker genes (COI and histone H3) and a diverse sample of copepods, provides an important milestone for addressing the copepod phylogeny. The trees presented in the study featured high support values for almost all previously uncertain branches, thus portraying many of the open questions in copepod phylogeny as resolved. However, in our analyses using the dataset published by the authors, we have found it impossible to obtain high support values for many key branches of the tree or to simply replicate the inferred phylogeny. To confirm our negative results, we have conducted reanalysis of the original dataset made available by the authors using the methods indicated in the paper. Our analyses with both maximum likelihood and Bayesian approaches yield phylogenies that differ from the results presented in the study by Khodami *et al.* [2017]. Furthermore, the phylogeny of copepods proposed by the authors is firmly rejected by the AU test.

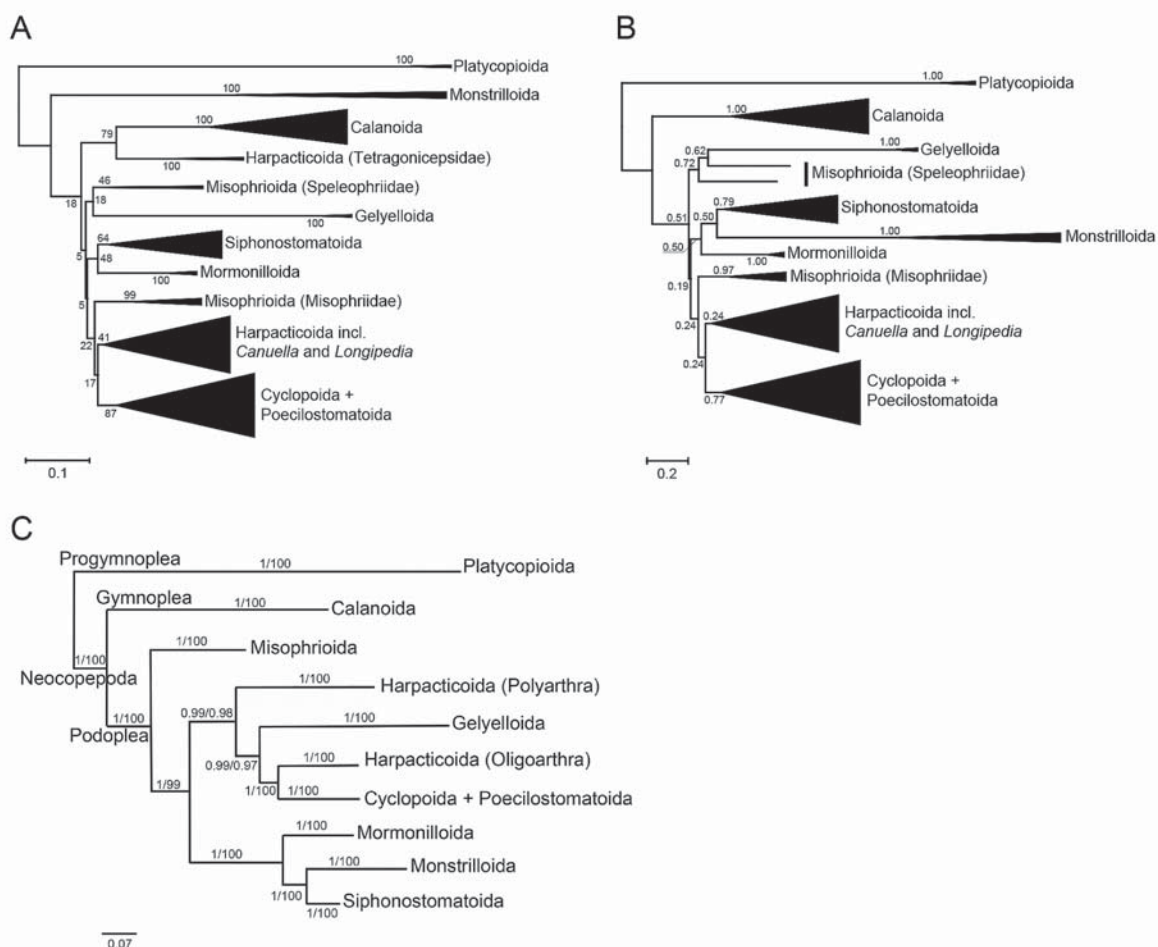


Fig. 1. A — phylogram of the copepod orders obtained in the RAxML reanalysis with the concatenated alignment of the four genes from the study by Khodami *et al.* [2017]. Node support is indicated by the bootstrap values. B — phylogram of the copepod orders recalculated with MrBayes using the concatenated alignment and summarized with a 25% burn-in after 20M generations. Node support is indicated by the posterior probabilities. C — phylogram of the copepod orders from Khodami *et al.* [2017: fig. 2]: “Order-level phylogram of 10 copepod orders. Phylogenetic relationship of 210 copepod species (collapsed to order level) based on Maximum Likelihood and Bayesian analysis of 18S and 28S rRNA, COI mtDNA and H3 histone protein. Nodal support is indicated by posterior probabilities and bootstrap values.”

Рис. 1. А — выравнивания четырех генов, использованных в работе Khodami *et al.* [2017]. Поддержка узлов обозначена значениями бутстрэп. В — филограмма отрядов веслоногих ракообразных, рассчитанная для конкатенированного выравнивания с помощью MrBayes в 20 миллионах поколений с отбрасыванием первых 25% точек выборки. Поддержка узлов соответствует значениям апостериорной вероятности. С — филограмма десяти отрядов веслоногих ракообразных из Khodami *et al.* [2017: fig. 2]: «Филограмма 10 отрядов веслоногих ракообразных на уровне отрядов. Филогенетические отношения 210 видов веслоногих ракообразных (свернутые до уровня отрядов) основанные на анализе максимального правдоподобия и байесовского анализа 18S и 28S рНК, COI мтДНК и гистонового белка H3. Поддержка узлов обозначает апостериорные вероятности и значения бутстрэп».

Methods

The phylogenetic analyses were performed with the concatenated alignment of four genes (18S rRNA, 28S rRNA, COI, and histone H3) sampled from 203 copepod species, made available by the authors of the original study [Khodami *et al.*, 2017] (<https://treebase.org/treebase-web/search/study/matrices.html?id=20470>, ID: M39845). Where possible, the tree reconstruction parameters were selected to parallel the settings used in the original study. The maximum likelihood analysis was carried out with RAxML using the GTRGAMMAI model optimized separately for the four gene parti-

tions. The tree search in RAxML 8.2.9 [Stamatakis, 2014] employed the random starting tree option (-d), and the tree node support values were evaluated with 10,000 bootstrap replicates. The Bayesian analysis was performed with MrBayes 3.2.6 [Ronquist *et al.*, 2012]. For rRNA gene partitions, we used the GTR model with 8 gamma-distributed rate categories and a proportion of invariants; for protein-coding gene partitions the codon model was used in conjunction with the GTR and 8 gamma-distributed rate categories. In the Bayesian analysis, the model parameters between the partitions were unlinked, and for the COI gene partition “metazoan mitochondrial” code was used. The analysis

Table 1. AU test of alternative tree topologies from Fig. 1: A — RaxML reconstruction, B — MrBayes reconstruction, C — tree topology from Khodami *et al.*, 2017.
 Таблица 1. Тест AU альтернативных топологий деревьев, приведенных на Рис. 1: А — реконструкция RaxML, В — реконструкция MrBayes, С — топология дерева из Khodami *et al.*, 2017.

Tree	Log Likelihood	log ₁₀ Δ with best tree	AU test <i>p</i> -value
A	-102862	0	0.948
B	-102923	-61	0.052
C	-103474	-612.2	2×10 ⁻⁶

was performed in four independent runs of four chains each with 20M generations, and summarized with a 25% burn-in. For the alternative topology test, we used CONSEL [Shimodaira, Hasegawa, 2001] with the site-wise likelihood values estimated by RAXML to compare the phylogenies obtained in this reanalysis with those of the original study. The site-wise likelihood values for the test were evaluated using the same model for RAXML as the one used for tree reconstruction. The CONSEL AU test was used for the estimation of *p*-values.

Results

The phylogenies reconstructed with the ML (Fig 1A and Supplementary Fig. S1) and Bayesian inference (Fig. 1B and Supplementary Fig. S2) are in discord with the phylogenetic tree presented by Khodami *et al.* [2017] in fig. 2 (Fig. 1C here). We do not observe the monophyly of Misophrioida in the reconstructed trees or the split of Harpacticoida into Polyarthra (Longipedidae and Canuellidae) and Oligoarthra. The Gelyelloida do not group sister to Harpacticoida and Cyclopoida. Position of the divergent clade Monstrilloida is uncertain in the analyses — their sister group position to the Siphonostomatoidea is only partially recovered in the Bayesian analysis. Generally, the phylogenetic trees obtained here have low support values, reflecting poor resolution of the relative branching of the copepod orders. This is in sharp contrast to high to maximal support values obtained in the ML and Bayesian analyses by Khodami *et al.* [2017: figs 2, 3, and S6].

To confirm that the phylogenies obtained here and poor support values were not simply a result of insufficient tree reconstruction efforts, we conducted a test of alternative topologies using the trees obtained here and the tree from the original study by Khodami *et al.* [2017: fig. S6]. The results of the AU test give preference to the ML tree topology over the alternatives (Table 1, A). The *p*-value for the topology from the Bayesian analysis (Table 1, B) is close to the 0.05 significance margin, scoring worse than the ML tree, presumably due to the differences in the evolutionary models used for tree inference. The tree topology from Khodami *et al.* [2017: fig. S6], however, scores significantly worse than the alternatives, and is rejected by

the AU test at the 0.05 and 0.01 significance levels (Table 1, C).

Discussion

Our findings demonstrate that the highly supported copepod phylogeny obtained by Khodami *et al.* [2017] is not reproducible. This indicates that the revision of copepod taxonomy undertaken by the authors of the original study is not supported by the molecular phylogenetic analyses. The rRNA genes (18S rRNA and 28S rRNA) are routinely used to propose phylogenetic relationships among copepod families or genera [Huys *et al.*, 2006, 2012; Blanco-Bercial *et al.*, 2011; Yeom *et al.*, 2018], but these analyses fall short in resolving the relationship between copepod orders. The extended dataset, which employs COI and histone H3 gene sequences, does not significantly improve the resolution of copepod phylogeny, furthermore, the maximum likelihood and Bayesian analyses with the dataset fail to consistently support the monophyly of recognized orders Harpacticoida and Misophrioida. This lack of resolution might also be exacerbated by incomplete sampling of genes — the concatenated alignment produced by Khodami *et al.* [2017] does not equally cover all representatives of the copepod orders.

Low support values and a lack of reproducibility are not unusual problems. For example, Regier *et al.* [2010], cited approvingly by Khodami *et al.* [2017], attempted to analyze relationships among a large number of arthropod groups. Yet Rota-Stabelli *et al.* [2013], using data from Regier *et al.* [2010], were unable to resolve satisfactorily the relationships among pancrustaceans using nucleotide or amino acid sequences. They concluded that these relationships should be considered unresolved.

The current reanalysis of data obtained by Khodami *et al.* [2017] further emphasizes that the issues with the phylogeny of Copepoda are far from being settled. Remarkably, similar problem with reproducibility was found for the results published by Khodami *et al.* [2019]. For more details see Mikhailov & Ivanenko [2019].

The results of reanalysis presented here have been available as a preprint since May 2019; the paper of Khodami *et al.* [2017] was retracted in October 2020

after many months of our communication with the authors and the editors of the journal.

Supplementary Figures are available online.

Supplementary Fig. S1. Maximum likelihood reconstruction with the concatenated alignment of the four genes (18S and 28S rRNA, COI mtDNA and H3 histone) from the study by Khodami *et al.* [2017]. The tree was reconstructed with RAxML using the GTRGAMMAI model optimized separately for the four gene partitions; tree node support values are bootstrap percentages evaluated with 10,000 replicates.

Supplementary Fig. S2. Bayesian inference with the concatenated alignment of the four genes (18S and 28S rRNA, COI mtDNA and H3 histone) from the study by Khodami *et al.* [2017]. The tree was reconstructed by MrBayes using the GTR model for the rDNA partitions and the codon models in conjunction with GTR for the protein coding gene partitions; tree node support values are posterior probabilities given as percentage values.

Acknowledgements. The reanalysis of the dataset was conducted with support of the Russian Foundation for Basic Research (grant #18-04-01192 on Evolution, phylogeny and ecology of crustacean copepods, symbionts of marine invertebrates). Frank Ferrari kindly commented on drafts of the manuscript.

Contributions. K.V.M. conducted data analysis and prepared manuscript; V.N.I. took part in planning, discussion of the results and preparing of manuscript.

Conflict of Interest. The authors declare that they have no conflict of interest.

References

- Blanco-Bercial L., Bradford-Grieve J., Bucklin A. 2011. Molecular phylogeny of the Calanoida (Crustacea: Copepoda) // *Molecular Phylogenetics and Evolution*. Vol.59. No.1. P.103–113.
- Ferrari F.D., Ivanenko V.N., Dahms H.-U. 2010. Body architecture and relationships among basal copepods // *Journal of Crustacean Biology*. Vol.30. No.3. P.465–477.
- Ferrari F.D., von Vaupel Klein J.C. 2019. Rhabdomoplea, a new superorder for the thaumatopsyllid copepods: the consequence of an alternative hypothesis of copepod phylogeny // *Crustaceana*. Vol.92. No.2. P.177–188.
- Ho J. 1994. Copepod phylogeny: a reconsideration of Huys & Boxshall's "parsimony versus homology" // *Hydrobiologia*. Vol.292–293. No.1. P.31–39.
- Huys R., Boxshall G.A. 1991. Copepod evolution. London: Ray Society. 468 p.
- Huys R., Fatih F., Ohtsuka S., Llewellyn-Hughes J. 2012. Evolution of the bomolochiform superfamily complex (Copepoda: Cyclopoida): New insights from ssrDNA and morphology, and origin of umazuracolids from polychaete-infesting ancestors rejected // *International Journal for Parasitology*. Vol.42. No.1. P.71–92.
- Huys R., Llewellyn-Hughes J., Conroy-Dalton S., Olson P.D., Spinks J.N., Johnston D.A. 2007. Extraordinary host switching in siphonostomatoid copepods and the demise of the Monstrilloida: Integrating molecular data, ontogeny and antennular morphology // *Molecular Phylogenetics and Evolution*. Vol.43. No.2. P.368–378.
- Huys R., Llewellyn-Hughes J., Olson P.D., Nagasawa K. 2006. Small subunit rDNA and Bayesian inference reveal *Pectenophilus ornatus* (Copepoda incertae sedis) as highly transformed Mytilicolidae, and support assignment of Chondracanthidae and Xariffiidae to Lichomolgoida (Cyclopoida) // *Biological Journal of the Linnean Society*. Vol.87. No.3. P.403–425.
- Khodami S., McArthur J.V., Blanco-Bercial L., Martinez Arbizu P. 2017. Molecular phylogeny and revision of Copepod orders (Crustacea: Copepoda) // *Scientific Reports*. Vol.7. Art.9164.
- Khodami S., Mercado-Salas N.F., Tang D., Martinez Arbizu P. 2019. Molecular evidence for the retention of the Thaumtopsyllidae in the order Cyclopoida (Copepoda) and establishment of four suborders and two families within the Cyclopoida // *Molecular Phylogenetics and Evolution*. Vol.138. P.43–52.
- Mikhailov K.V., Ivanenko V.N. 2019. Lack of reproducibility of molecular phylogenetic analysis of Cyclopoida // *Molecular Phylogenetics and Evolution*. Vol.139. Art.106574.
- 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*. Vol.463. P.1079–1083.
- Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space // *Systematic Biology*. Vol.61. No.3. P.539–542.
- Rota-Stabelli O., Lartillot N., Philippe H., Pisani D. 2013. Serine codon-usage bias in deep phylogenomics: Pancrustacean relationships as a case study // *Systematic Biology*. Vol.62. P.121–133.
- Shimodaira H., Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection // *Bioinformatics*. Vol.17. No.12. P.1246–1247.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies // *Bioinformatics*. Vol.30. No.9. P.1312–1313.
- Yeom J., Nikitin M.A., Ivanenko V.N., Lee W. 2018. A new minute ectosymbiotic harpacticoid copepod living on the sea cucumber *Eupentacta fraudatrix* in the East/Japan Sea // *PeerJ*. Vol.6. Art.4979.

Responsible editor K.G. Mikhailov