nature ecology & evolution

MATTERS ARISING

https://doi.org/10.1038/s41559-022-01871-3

Check for updates

Phylogenetic affiliation of mitochondria with Alpha-II and Rickettsiales is an artefact

Joran Martijn[1,2,3]✉, Julian Vosseberg[4], Lionel Guy[1], Pierre Offre[5] and Thijs J. G. Ettema[6]✉

ARISING FROM: Fan et al. *Nature Ecology and Evolution* https://doi.org/10.1038/s41559-020-1239-x (2020)

Fan et al. challenge the finding that mitochondria represent a sister group to Alphaproteobacteria[1], arguing that support for this position can be explained by unreliable site removal methods and outgroup attraction. The authors chose to reduce compositional heterogeneity by replacing AT-rich Rickettsiales and mitochondria with GC-rich alternatives and attempted to attenuate long branch attraction (LBA) effects by removing all fast-evolving lineages but one. The study suggests that mitochondria form a sister clade to Rickettsiales, within the 'Alpha-II' clade. Here we show that the association with Alpha-II is an artefact caused by a problematic taxon and that residual support for Rickettsiales-sister can be explained by convergent evolution of the selected mitochondria and Rickettsiales towards high %GC. We further argue that site removal methods are in fact reliable and that outgroup attraction is unlikely to explain the Alphaproteobacteria-sister tree.

The phylogeny key to Fan et al.'s hypothesis[2] was inferred from a phylogenomic dataset in which the AT-rich Rickettsiales and mitochondria were replaced with GC-rich alternatives and all other fast-evolving lineages were removed. Here, mitochondria branched sister to Rickettsiales within the so-called Alpha-II group (comprising MarineAlpha3, -9, -11, -12 and the Rhodospirillaceae) (Fan et al.'s Fig. 3h/Supplementary Fig. 35b). When observing that Rickettsiales-sister was preserved after removing the most heterogeneous sites (Fan et al.'s Supplementary Figs. 37b–41b), it was concluded that this topology probably reflected historical signal. We have several major concerns with this conclusion.

First, the link between mitochondria and Alpha-II is most probably an artefact (Supplementary Text 1): mitochondria branching within Alpha-II is not recovered in most other analyses that include both groups (Fan et al.'s Fig. 3b–f/Supplementary Figs. 31b–34b) and disappears as soon as the top biased sites are removed (Fan et al.'s Supplementary Figs. 37b–40b).

Second, the phylogenomics dataset on which these phylogenies were based lacked signal important to resolving the placement of mitochondria. We found it included 4 paralogues and was missing 20 orthologues (Supplementary Text 2). One marker gene in particular was missing 9 orthologues, including all 5 Rickettsiales representatives (Supplementary Fig. 1). After we removed the paralogues and added the missing orthologues, 3 out of 4 Bayesian chains and the maximum likelihood inference recovered markedly different trees in which the artefactual Alpha-II relationship was no longer observed (Supplementary Figs. 5–9). The fact that 3 out of our 4 chains resolved the artefact, but none of Fan et al.'s 4 chains did, strongly suggests that the update of the phylogenomics dataset added critical phylogenetic signal.

Third, the dataset includes MarineAlpha9 Bin5, a problematic taxon that attracts mitochondria into Alpha-II (Supplementary Text 3). MarineAlpha9 Bin5 has previously been associated with phylogenetic artefacts[1] and we suspected it was responsible for the artificial Alpha-II relationship. Indeed, once we had removed this taxon from the updated dataset, all 4 Bayesian chains converged to the same tree in which mitochondria branched away from Alpha-II as a sister group to Rickettsiales ('Rickettsiales-sister'; Supplementary Figs. 10 and 11).

Fourth, the internal branches in Fan et al.'s key phylogeny (Fan et al.'s Fig. 3h/Supplementary Fig. 35b) leading to the mitochondria, Rickettsiales and outgroups were relatively long and possibly made the tree inference more susceptible to LBA artefacts. The branches leading to mitochondria and Rickettsiales are long because of their very narrow phylogenetic scope: only the Embryophyta and Anaplasmataceae were sampled, respectively. The branches leading to the outgroups are long because of the lack of MarineProteo1 and the poor representation of Magnetococcales and Gammaproteobacteria. We also noticed that several deep-branching but 'slow-evolving' alphaproteobacteria (see Methods) that may be important were missing. To make the tree inference less susceptible to LBA, add potentially important phylogenetic information and make the dataset as a whole more representative of mitochondria-related lineages, we broke the long branches by supplementing the updated dataset with the most GC-rich representatives of other mitochondrial and rickettsial lineages, as well as the aforementioned outgroup lineages and deep-branching alphaproteobacteria. The main relationships and the Rickettsiales-sister topology remained broadly unaffected (Supplementary Figs. 20 and 21), suggesting that the long internal branches in the original analysis did not contribute to LBA artefacts.

Finally, Fan et al. do not investigate whether the observed support for the Rickettsiales-sister could be explained by convergent evolution towards high %GC in their selected mitochondria and Rickettsiales. Organelles and endosymbiotic bacteria generally evolve towards more AT-rich genomes[3,4] and this is typically true for mitochondria[6] and Rickettsiales[5] as well. Given that GC-rich Rickettsiales and mitochondria are rare and phylogenetically flanked by AT-rich close relatives, it is highly plausible that the chosen taxa did not retain the GC-rich composition of their free-living ancestors, as was assumed by Fan et al. (Fig. 1a), but rather had an independent AT-rich past (Fig. 1b). If so, convergent evolution towards GC-rich compositions

[1]Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [2]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada. [3]Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [4]Theoretical Biology and Bioinformatics, Department of Biology, Utrecht University, Utrecht, the Netherlands. [5]NIOZ Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, Den Burg, the Netherlands. [6]Laboratory of Microbiology, Wageningen University and Research, Wageningen, the Netherlands. ✉e-mail: joran.martijn@dal.ca; thijs.ettema@wur.nl

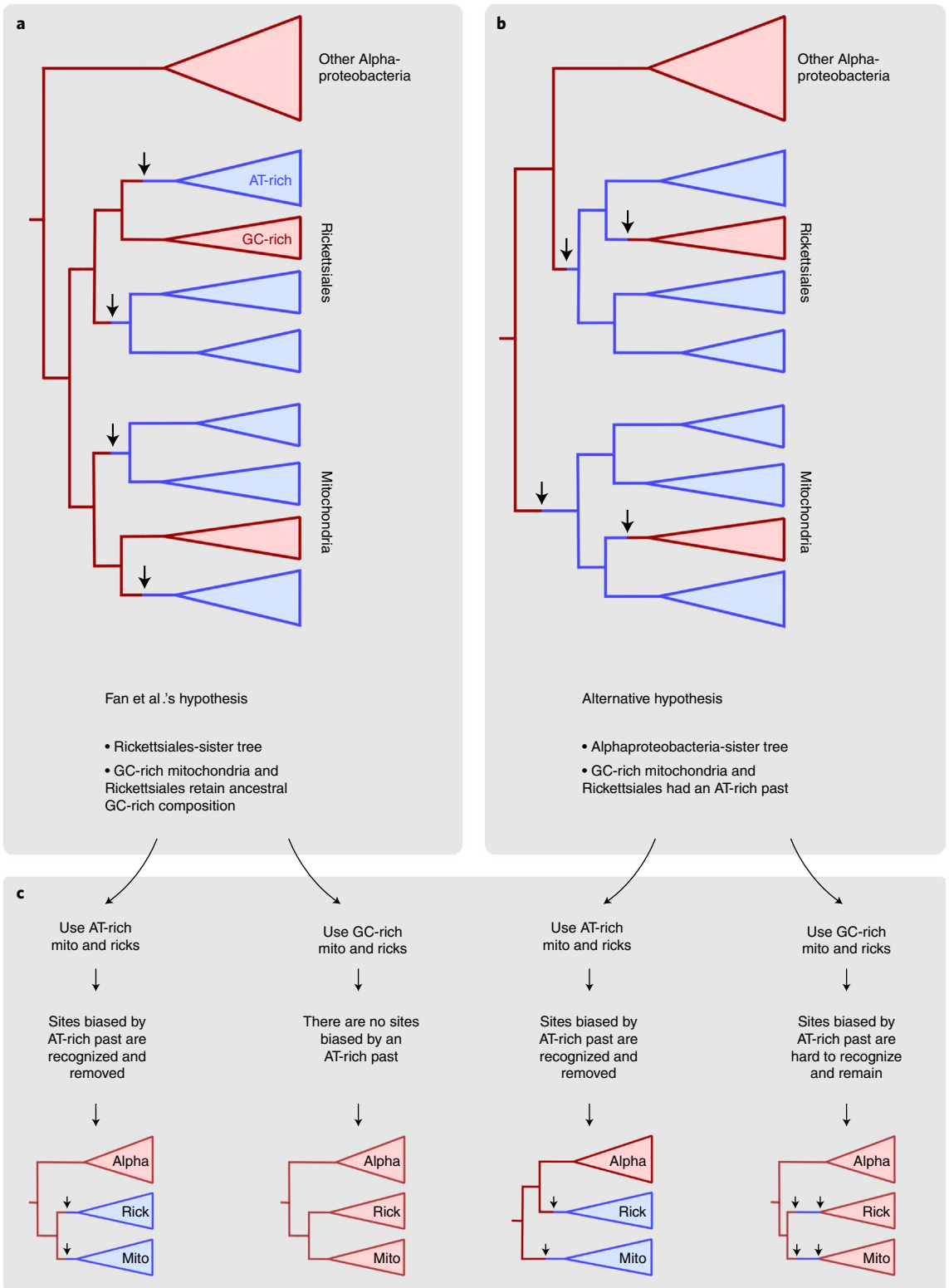

Fig. 1 | The position of mitochondria and evolution of sequence compositions. a,**b**, Fan et al.'s hypothesis (**a**) and the alternative hypothesis considered here (**b**). Selection of either GC-rich or AT-rich mitochondria and Rickettsiales affects site removal and tree inferences in different ways depending on the underlying tree (**c**). Under Fan et al.'s hypothesis, if AT-rich mitochondria and Rickettsiales are selected, then sites biased by an AT-rich past would be efficiently identified by the χ^2-score site removal strategy and the remaining unbiased sites are expected to support a Rickettsiales tree. If GC-rich representatives are selected instead, then there would be little to no biased sites as no compositional shifts occurred in their evolutionary history. Thus, sites remaining after site removal are expected to always support a Rickettsiales-sister tree. Under the alternative hypothesis, if AT-rich mitochondria and Rickettsiales are selected, then sites biased by an AT-rich past would be efficiently identified by the χ^2-score site removal strategy and the remaining unbiased sites are expected to support an Alphaproteobacteria-sister tree. If GC-rich representatives are selected instead, then sites biased by an AT-rich past would be less efficiently identified by the χ^2-score site removal strategy and biased sites that support a false Rickettsiales-sister tree remain. GC-rich taxa and lineages are indicated in red, AT-rich taxa and lineages in blue. Hypothetical ancestral compositional shifts are indicated with a black arrow.

may have induced false support for a Rickettsiales-sister relationship. To test whether this is the case, we would need to remove all those sites that have been affected by this convergent evolution (biased sites) but keep all sites that have not (unbiased sites). If support for Rickettsiales-sister falls or is replaced by another strongly supported topology, support for Rickettsiales-sister can be attributed to convergent evolution. Fan et al. use the χ^2-score strategy to estimate which sites are biased. When using AT-rich mitochondria and Rickettsiales alongside generally more GC-rich alphaproteobacteria, this is a sound strategy (see also Supplementary Text 5). Sites biased by convergent evolution towards AT-richness contribute much to the overall compositional heterogeneity of the data and are easily recognized (Fig. 1c). When using GC-rich mitochondria and Rickettsiales, however, sites biased by a secondary shift to GC-richness contribute relatively little to overall heterogeneity and will be less efficiently recognized. As a result, many biased sites remain and lend false support to a Rickettsiales-sister tree. To better estimate the sites biased by an AT-rich past, we applied the χ^2-score strategy on the supplemented dataset in which the GC-rich mitochondria and Rickettsiales were replaced with gene-rich and AT-rich alternatives (see Methods). As explained above, sites biased by the ancestral shift towards AT-richness will be among the most heterogeneous. Because the GC-rich and AT-rich mitochondria and Rickettsiales share this ancestral shift in their respective histories (Fig. 1b), we can infer that sites of the original 'GC-rich' dataset homologous to the most heterogeneous sites of the 'AT-rich' dataset are more likely to be biased by an AT-rich past. When removing up to 10% of these sites, maximum likelihood and Bayesian inferences yielded Rickettsiales-sister trees with significant support (Supplementary Figs. 32–35). Yet, from 20% removal onwards, support for Rickettsiales-sister dropped substantially and was exceeded by (albeit not significant) support for Alphaproteobacteria-sister (Supplementary Figs. 36–41 and Table 2). Importantly, the support dropped substantially more compared with the previous site removal assays despite removing the same number of sites. This result cannot be explained under a Rickettsiales-sister tree in which the mitochondria and Rickettsiales retained the ancestral %GC. In that case, all sites are expected to be unbiased (Fig. 1c) and removing the same number of sites, regardless of which they are, should yield approximately the same support for the Rickettsiales-sister tree. This strongly suggests that at least some of the support for Rickettsiales-sister that we observed in trees of the supplemented dataset can be explained by convergent evolution in the selected mitochondria and Rickettsiales.

We conclude that support for a mitochondrial sister relationship with Rickettsiales within the Alpha-II group can be explained by the combination of a suboptimal phylogenomics dataset, a long branch attraction induced by MarineAlpha9 Bin5 and convergent evolution towards high %GC. After accounting for these issues, the dataset harbours too little signal to confidently resolve the origin of mitochondria.

Data availability

The phylogenomic datasets and supermatrix alignments generated and analysed during the current study are available in the FigShare repository, https://doi.org/10.6084/m9.figshare.17108234.v1.

Code availability

The scripts count_tripartitions.py and parse_tripartition_counts.py were used to enumerate node support in ML and Bayesian phylogenies. An updated alignment_pruner.pl script was used to extract per-site heterogeneity scores ($\Delta\chi^2$ scores). All scripts are available at https://github.com/novigit/broCode/tree/master/nee_matters_arising

Received: 19 November 2021; Accepted: 3 August 2022;
Published online: 24 October 2022

References

1. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018).
2. Fan, L. et al. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-020-1239-x (2020).
3. Toft, C. & Andersson, S. G. E. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* **11**, 465–475 (2010).
4. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2012).
5. Darby, A. C., Cho, N.-H., Fuxelius, H.-H., Westberg, J. & Andersson, S. G. E. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* **23**, 511–520 (2007).
6. Rodríguez-Ezpeleta, N. & Embley, T. M. The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. *PLoS ONE* **7**, e30520 (2012).

Acknowledgements

We thank N. Lartillot for assistance in investigating the posterior predictive test reproducibility issue, J. Viklund for updating the alignment_pruner.pl script, and A. Roger and K. Williamson for helpful insights and discussions. Phylobayes-MPI computations were performed on resources provided by the PDC Center for High-Performance Computing (PDC) under Project SNIC 2020/5-473. This work was supported by grants from the European Research Council Consolidator (grant 817834), the Dutch Research Council (NWO; VICI grant VI.C.192.016) to T.J.G.E. and the Swedish Research Council (VR International PostDoc grant 2018-06727) to J.M.

Author contributions

J.M. provided the main constructive criticism of the Fan et al. manuscript and designed and carried out all analyses. J.V., L.G., P.O. and T.J.G.E. provided additional constructive criticism. J.M. and T.J.G.E. wrote, and all authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01871-3.

Correspondence and requests for materials should be addressed to Joran Martijn or Thijs J. G. Ettema.

Peer review information *Nature Ecology and Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work.