Insights into global planktonic diatom diversity: The importance of comparisons between phylogenetically equivalent units that account for time

Teofil Nakov¹ · Jeremy M. Beaulieu¹ · Andrew J. Alverson¹

Received: 6 March 2018 / Revised: 16 May 2018 / Accepted: 11 June 2018 / Published online: 16 July 2018
© International Society for Microbial Ecology 2018

Abstract

Metabarcoding has offered unprecedented insights into microbial diversity. In many studies, short DNA sequences are binned into consecutively lower Linnaean ranks, and ranked groups (e.g., genera) are the units of biodiversity analyses. These analyses assume that Linnaean ranks are biologically meaningful and that identically ranked groups are comparable. We used a metabarcode dataset for marine planktonic diatoms to illustrate the limits of this approach. We found that the 20 most abundant marine planktonic diatom genera ranged in age from 4 to 134 million years, indicating the non-equivalence of genera because some have had more time to diversify than others. However, species richness was largely independent of genus age, suggesting that disparities in species richness among genera were better explained by variation in rates of speciation and extinction. Taxonomic classifications often do not reflect phylogeny, so genus-level analyses can include phylogenetically nested genera, further confounding rank-based analyses. These results underscore the indispensable role of phylogeny in understanding patterns of microbial diversity.

With potentially millions of species occupying all the world's aquatic and terrestrial biomes, microbial diversity is notoriously difficult to discover and catalogue. Traditional approaches to species discovery are time and labour intensive, and they miss species that cannot be cultivated in the lab [1]. The phylogenetic diversity of this undiscovered “microbial dark matter” is often characterised through community DNA sequencing of barcode genes. A typical workflow includes DNA extraction from an environmental sample, PCR amplification of a DNA barcode region, and high-throughput sequencing of the amplicon [2]. Sequencing reads are clustered into operational taxonomic units (OTUs) that are subsequently binned into consecutively lower taxonomic ranks, and these ranked groups, in turn, are often the focus of biodiversity assessments [3].

Linnaean names and ranks are often taken to mean more than what they are: arbitrary taxon delimitations disconnected from evolutionary history. The treatment of

named groups as anything other than arbitrary implies that identically ranked taxa are somehow comparable, encouraging comparisons of their ecology, biogeography, and species richness [4–6]. The only meaningful comparisons involve groups with comparable evolutionary histories [7]. In this sense, monophyletic groups (clades) are more likely to be biologically cohesive units, and they should have comparable species richness if they are similar in age and have diversified at similar rates [8]. Comparison of monophyletic groups, while accounting for time, therefore provides a robust framework for detecting clades with exceptional species richness and comparing their functional, ecological, or biogeographic breadth [9].

The Tara Oceans Project sequenced 18S-V9 metabarcode fragments from plankton samples to characterise microbial communities and species richness across the world's oceans [10]. Strikingly, just 20 genera accounted for nearly 99% of all diatom sequencing reads, and comparisons among these genera revealed differences in relative abundance, cell size, habitat preference, geographical distribution, and species richness [3]. It was not clear, however, whether these patterns deviated from expectations. We focused our analyses on the genus-based patterns of species richness and expected that older genera would be more species rich because they have had more time to diversify

✉ Teofil Nakov
tnakov@uark.edu

¹ Department of Biological Sciences, University of Arkansas, SCEN 601, Fayetteville, AR 72701, USA

Fig. 1 Age and estimated taxon richness of the 20 most abundant marine planktonic diatom genera identified by the Tara Oceans metabarcode project [3]. Crown ages and uncertainty (grey bars) in million years ago (Mya) were estimated from 1000 bootstrap phylogenies [11]. **a** Taxon richness was estimated from the number of OTU swarms in the Tara Oceans dataset (**b**) and the number of accepted species names in DiatomBase [14] (**c**). Black curves in **b**, **c** delimit 95% confidence intervals of expected richness given the crown age of

a clade, empirical extinction fraction, and diatom-wide estimate of the net diversification rate (see [11] for details). Blue and green bars in **c** show the difference in species richness as measured by OTU swarms **b** and DiatomBase names **c**. Blue bars show which genera have fewer OTUs than DiatomBase names, suggesting that the number of OTUs might underestimate species richness, whereas green bars show which genera might have more species than described by traditional taxonomy

[8]. We calculated net diversification (i.e., speciation–extinction) using (1) the crown age of diatoms estimated from a 1151-taxon phylogeny of diatoms [11], (2) relative extinction (i.e., extinction/speciation) from Cenozoic fossil diatoms [12], and (3) a minimum approximation of total described and undescribed diatom diversity (30,000 species [13]). We then used the inferred net diversification rate to calculate upper and lower bounds of expected OTU richness [9] for the 20 most abundant genera of marine planktonic diatoms in the Tara Oceans survey.

The 20 diatom genera ranged in age from 4–134 million years (My), though OTU richness was only weakly correlated with clade age ($r = 0.36$, 95% CI = -0.1 – 0.7 , $df = 18$, $P = 0.12$). A total of 12 of the 20 most-abundant genera fell within expectation for OTU number given their age (Fig. 1). The most abundant and OTU-rich genus, *Chaetoceros*, was also the oldest (Fig. 1a). The birth–death diversification model predicted that the diversity of a clade as old as *Chaetoceros* could range between 57 and 7940 species—the Tara Oceans dataset recovered 644 *Chaetoceros* OTUs, consistent with expectations for a clade of this age (Fig. 1b). Some of the most diverse genera identified by metabarcoding (e.g., *Corethron* and *Pseudo-*

nitzschia) had OTU richness estimates that exceeded expectations (Fig. 1b, black curves). Assuming OTUs correspond to species and that our estimates of clade age are not heavily biased, these genera have either exceptionally high speciation or low extinction rates. Identifying the drivers of these patterns might offer new mechanistic insights into phytoplankton diversification. Comparisons between OTU richness (Fig. 1b) and number of accepted taxonomic names from DiatomBase [14] (Fig. 1c) showed expected discrepancies for lineages with substantial diversity in benthic or freshwater habitats that were not sampled during the Tara Oceans Expedition (e.g., *Navicula*; Fig. 1b, B and F annotations; Fig. 1c, blue bars). These discrepancies also highlight clades that might be under-described at the species level (Fig. 1c, green bars).

Metabarcoding identified *Thalassiosira* as one of the most abundant, OTU-rich, and geographically widespread genera of marine planktonic diatoms. A total of eight *Thalassiosirales* genera were detected in the Tara Oceans project (*Cyclotella*, *Lauderia*, *Minidiscus*, *Planktoniella*, *Porosira*, *Shionodiscus*, *Skeletonema*, and *Thalassiosira*), and these genera ranged in age from 4–63 My (Fig. 2). *Thalassiosirales* embodies many of the problems with

Fig. 2 The genus *Thalassiosira* encompasses at least ten marine (white circles) and four freshwater (black squares) planktonic diatom genera (including *Thalassiosira*) that range from 4–63 My in age. Topology and divergence times are based on Nakov et al. [11]

misappropriation of biological or evolutionary properties to taxa based on their names [15]. The name *Thalassiosira* applies to a polyphyletic set of species whose common ancestor dates to at least 63 million years ago (Mya) and gave rise to nearly the full phylogenetic breadth of *Thalassiosirales* diversity (Fig. 2, diamond). As a result, including *Thalassiosira* in genus-level analyses leads to highly biased comparisons involving a genus that, in reality, is more like a taxonomic order (Fig. 2). Moreover, four of the eight *Thalassiosirales* genera detected by metabarcoding are nested within *Thalassiosira*, highlighting a common source of non-independence in rank-based comparisons (Fig. 2, yellow branches). A phylogenetically based genus-level classification of *Thalassiosirales* may have revealed clade-specific habitat preferences or geographic distributions among the many distinct *Thalassiosira* lineages [16].

The problems with rank-based comparisons, including as they relate to diatoms, are well known [15–17]. A frequently cited advantage of metabarcoding is that it does not require taxonomic expertise. Still, the taxonomic affiliations of metabarcoding sequences often become the units of biodiversity analyses. Analyses that explicitly incorporate phylogenetic history and systematics—which invariably highlight the deficiencies of Linnaean classifications—ensure comparisons among biologically equivalent units that account for time.

Acknowledgements This work was supported by a grant from the Simons Foundation (403249, AJA). This material is also based upon work supported by the National Science Foundation (NSF) under Grant no. DEB-1353131. This research used computational resources available through the Arkansas High Performance Computing Center, which was funded through multiple NSF grants and the Arkansas Economic Development Commission.

Compliance with ethical standards

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

References

- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*. 2013;499:431–7.
- De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, et al. Eukaryotic plankton diversity in the sunlit ocean. *Science*. 2015;348:1261605.
- Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, et al. Insights into global diatom distribution and diversity in the world's ocean. *P Natl Acad Sci USA*. 2016;113:E1516–25.
- Pleijel F, Rouse GW. Ceci n'est pas une pipe: Names, clades and phylogenetic nomenclature. *J Zool Syst Evol Res*. 2003;41:162–74.
- Sundberg PER, Pleijel F. Phylogenetic classification and the definition of taxon names. *Zool Scr*. 1994;23:19–25.
- Cantino PD, de Queiroz K (2010). *PhyloCode: A Phylogenetic Code of Biological Nomenclature*. Ohio University. Athens, Ohio. <https://www.ohio.edu/PhyloCode/PhyloCode2a.pdf>

7. Harvey PH, Pagel MK. The Comparative Method in Evolutionary Biology. Oxford: Oxford University Press; 1991.
8. Stadler T, Rabosky DL, Ricklefs RE, Bokma F. On age and species richness of higher taxa. *Am Nat*. 2014;184:447–55.
9. Magallón S, Sanderson MJ. Absolute diversification rates in angiosperm clades. *Evolution*. 2001;55:1762–80.
10. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. *Science*. 2015;348:1261359.
11. Nakov T, Beaulieu JM and Alverson AJ (2018). Accelerated diversification is related to life history and locomotion in a hyperdiverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). *New Phytol*. <https://doi.org/10.1111/nph.15137>
12. Lazarus D, Barron J, Renaudie J, Diver P, Türke A. Cenozoic planktonic marine diatom diversity and correlation to climate change. *PLOS ONE*. 2014;9:e84857.
13. Mann DG, Vanormelingen P. An inordinate fondness? The number, distributions, and origins of diatom species. *J Eukaryot Microbiol*. 2013;60:414–20.
14. Kociolek JP, Balasubramanian K, Blanco S, Coste M, Ector L, Liu Y et al. (2018). DiatomBase. <http://www.diatombase.org> Accessed 18 Feb 2018.
15. Alverson AJ, Beszteri B, Julius ML, Theriot EC. The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*. *BMC Evol Biol*. 2011;11:125.
16. Wiese R, Renaudie J, Lazarus DB. Testing the accuracy of genus-level data to predict species diversity in Cenozoic marine diatoms. *Geology*. 2016;44:1051–4.
17. Kociolek JP. Taxonomy and ecology: further considerations. *Proc Calif Acad Sci*. 2005;56:99–106.