ORIGINAL ARTICLE

Revisiting global diversity and biogeography of freshwater diatoms: New insights from molecular data

Teofana Chonova[1,2] | Frédéric Rimet[1] | Agnès Bouchez[1] | Maria Kahlert[3] |
Susanne C. Schneider[4] | Bonnie Bailet[3] | Anne Eulin-Garrigue[5] | Gilles Gassiole[6] |
Olivier Monnier[7] | Allassane Ouattara[8] | Stéphanie Rey[9] | Mathieu Rhoné[10] |
François Keck[1,11] 🔵

[1]UMR CARRTEL, INRAE, Université de Savoie Mont-Blanc, Thonon-les-Bains, France

[2]Department Environmental Chemistry, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

[3]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden

[4]Norwegian Institute for Water Research, Oslo, Norway

[5]Hydreco Guyane, Bureau d'Etudes et de Recherches en Environnement, Laboratoire Environnement de Petit Saut, Kourou Cedex, France

[6]MicPhyc, Salazie, France

[7]Direction de la Recherche et de l'Appui Scientifique, OFB, Vincennes, France

[8]Environmental Science and Management Teaching and Research Unit, Environmental and Aquatic Biology Laboratory, Université Nangui Abrogoua (UNA), Abidjan, Côte d'Ivoire

[9]DEAL Guyane, Cayenne, France

[10]Office de l'Eau de Guyane, Cayenne, France

[11]Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

Correspondence
Teofana Chonova, Frédéric Rimet and François Keck UMR CARRTEL, INRAE, Université de Savoie Mont-Blanc, 74200 Thonon-les-Bains, France.
Email: teofana.chonova@gmail.com; Frederic.rimet@inrae.fr and Francois.keck@gmail.com

Abstract

The distribution of microorganisms has long been assumed to be cosmopolitan and primarily controlled by the environment, but recent studies suggest that microbes may also exhibit strong biogeographical patterns driven by dispersal limitation. Past attempts to study the global biogeography of freshwater diatoms have always encountered the great difficulty of collecting taxonomically harmonized large-scale data. However, developments in molecular techniques and DNA metabarcoding provide a unique opportunity to overcome these limitations and to disclose diatom biodiversity at an unprecedented scale and resolution. Here, we assembled DNA metabarcoding data of freshwater benthic diatom communities sampled in seven geographic regions across the world to investigate how diatom diversity varies along latitude and to assess the proportion of genetic variants of these microorganisms which are exclusive or shared across regions. We observed significant differences in assemblages among climate zones and found that genetic richness is not affected by latitude, but by an island effect. The genetic resolution directly impacts the proportion of variants shared across regions; however, the majority of taxa remained specific to a single geographic region. Freshwater diatoms disperse over long distances and across oceans but at

rates that allow the appearance of local genetic variants and the regionalization of assemblages. Future work should focus on putting these diversity dynamics into a temporal context, an approach that should be possible by bringing together new sequencing techniques and phylogeography.

1 | INTRODUCTION

Biogeography studies how life is distributed through time and space. It aims at understanding the diversity and dispersion patterns of organisms across the Earth and at disentangling the dominant factors that drive those events. The underlying processes driving species biogeography have intrigued scientists for a long time (MacArthur & Wilson, 1967; Wallace, 1876). The first theories mainly focused on macro-organisms, which were easier to study through direct observation. Several extensive works revealed clear patterns in the distribution of animals and plants around the globe related to the origin and the history of organisms (Finlay, 2002) and reported an increasing species diversity towards the tropics for several taxonomic groups (Hillebrand & Azovsky, 2001, Miraldo et al., 2016, Schluter & Pennell, 2017, De Kort et al., 2021). In islands, usually lower diversity and higher endemism were observed and were related to their geographic isolation and organism dispersal limitation (Kier et al., 2009; MacArthur & Wilson, 1967).

Studying the biogeography of microbes appeared more challenging, because microbial diversity and dispersal capacities were more difficult to assess. The supposed high-dispersal rates of microorganisms have driven the expectation that they have cosmopolitan geographic distribution and restricted global richness (Finlay, 2002). However, studies that have tried to confirm global distribution of microorganisms have found both seemingly globally distributed species as well as geographically limited ones (Fontaneto, 2019). In support of the Baas-Becking hypothesis "everything is everywhere, but the environment selects," some authors explained restricted microbial geographic distributions by species sorting related to habitat-specific environmental conditions (Baas-Becking, 1934). In turn, more recent studies suggested that the particular biogeography of microorganisms could be directly influenced by the species origin, historical dynamics, and dispersal limitation both at global scale (De Luca et al., 2019, 2021; Martiny et al., 2006; Metz et al., 2019) and local scale (Bottin et al., 2014; Pérez-Burillo et al., 2021). Therefore, it is still debated to what extent general ecological theories are valid for microorganisms and if they are globally distributed.

Diatoms are long-time used model organisms to study microbial biogeography. They are unicellular eukaryotic microalgae with large taxonomic diversity and marked species-specific ecological preferences. The ability of diatoms to disperse through the air or through animal transportation is critical for the colonization of new locations (Manning et al., 2021). Still, not much is known about the geographic range and rate of diatoms' dispersal and their effects on the dynamics of communities across landscapes. As for other microorganisms, some authors found that diatoms, at least some species, have high dispersal capacities (Cermeño & Falkowski, 2009, Evans & Mann, 2009, Mann et al., 2009, Trobajo et al., 2010) and cosmopolitanism could therefore prevail in this group. However, an increasing body of recent work suggests that endemism can be important in diatoms, and species may have restricted dispersal capacity (Bennett et al., 2010; Soininen et al., 2016; Verleyen et al., 2009, 2021; Vyverman et al., 2007). Thus, species origin and history are now expected to play a key role in diatom biogeography (Verleyen et al., 2021). Over the last decades, the biogeography of diatoms has been intensively discussed and fueled by numerous results obtained at local and regional scales, and more and more results based on the compilation of combined large-scale data have been published (e.g. Jamoneau et al., 2022; Soininen et al., 2016). Still, technical limitations hindered scientists from fully understanding geographical patterns at larger scales and their explanation. Most studies on diatom dispersion and endemism at global scale are based on traditional microscopy data and therefore suffer from drawbacks associated with optical microscopy such as identification errors, limited taxonomic resolution, and differences in taxonomic systems between regions requiring manual harmonization (Kahlert et al., 2009).

Better understanding of biogeographical patterns requires extensive datasets with good coverage of each studied spatial scale and harmonized taxonomy. New approaches offering higher resolution and standardized comparison across samples in large-scale datasets are essential to expand our understanding of microbial biogeography. Since the last decade, environmental genetic tools are increasingly used. The recent improvements in DNA metabarcoding technologies allow for better taxonomic resolution compared to microscopy and open the access to total biodiversity information, including cryptic species and taxa new to specialists. Bioinformatics tools now enable reliable and standardized comparisons of large datasets originating from distant geographic regions, overcoming long-standing issues related to biases in species identification (Soininen & Teittinen, 2019). Genetic information from different sequencing runs can now be easily harmonized and merged using denoised sequences for genetic variants where even one nucleotide difference can be detected (Callahan et al., 2016). Moreover, we have the possibility to combine genetic amplicon

sequence variants (ASVs) and clustering methods (OTUs), revealing information at different genetic and taxonomic depths (Antich et al., 2021). We are now reaching the point where new levels of genetic information enable a more holistic understanding of fundamental ecological and biogeographical questions in microorganisms. Still, the application of genetic tools to elucidate biodiversity patterns at a very large scale has been essentially limited to marine and terrestrial environments (Bahram et al., 2018; Malviya et al., 2016; Sunagawa et al., 2015).

In this paper, we use molecular data to study the global diversity and biogeographic patterns of freshwater diatom communities along a broad latitudinal gradient. We analyzed a large dataset of 711 river diatom communities sampled in four continents across the subpolar, temperate, and tropical climate zones. This dataset was obtained via DNA metabarcoding and analyzed to generate genetic variants and clustered sequences, which allowed a standardized comparison of diatom diversity and community structure at different levels of genetic precision. Based on this novel molecular-based information, we aim to revisit two long-standing biogeographic questions: (i) To what extent is the genetic diatom diversity affected by the latitudinal gradient and the islands effect? And (ii) Are genetic variants of freshwater diatoms cosmopolitan and if not, which level of genetic resolution is needed to study regional variants?

2 | MATERIALS AND METHODS

2.1 | Study sites and sampling

A total of 711 unique river sites were sampled to collect benthic diatoms. Sampling covered seven geographic regions from four continents (Africa, Europe, Oceania, and South America). They are located in the subpolar (Fennoscandia), temperate (France Mainland), and tropical (West Africa, French Guiana, New Caledonia, Tahiti island, and Mayotte island) climate zones, with latitude ranging from 68°N to 22°S. Geographic regions comprise four mainland zones (Fennoscandia, France Mainland, West Africa, and French Guiana) and three tropical islands (New Caledonia, Tahiti, and Mayotte). The sites are routinely sampled in the framework of local monitoring networks. The location of these sites was decided

by the monitoring agencies to be representative of the ecological quality of each region; however, there is no background environmental data available for the samples obtained in the frame of this project. Further details on each dataset are available in Table 1 and Table S1 in Supporting Information. Diatoms were sampled during stable water seasons. At each site, biofilm was collected from at least five different stones by brushing their upper surfaces using a toothbrush (CEN, 2014). Samples were preserved in ethanol (70%) until sequencing (CEN, 2018). Samples (ethanol, biofilm, and DNA extracts) are vouchered at UMR CARRTEL (Thonon-les-Bains, France).

2.2 | DNA extraction, amplification and sequencing

Pellets of biofilms were prepared by centrifuging (17,949 **g** for 30 min) between 2 and 4 mL of the initial biofilm suspension. Total genomic DNA was isolated from the pellets using the NucleoSpin® Soil kit (Macherey-Nagel) following the manufacturer's instructions.

A 312 bp fragment of the rbcL plastid gene was amplified using Takara LA Taq® polymerase and an equimolar mix of the forward primers Diat_rbcL_708F_1, 708F_2, 708F_3, and the reverse primers R3_1, R3_2, as described in Vasselon et al. (2017). A unique combination of forward and reverse adapters was included at the 5′-extremities of the forward and reverse primers to enable the multiplexing of all PCR products in a sequencing library. PCR reactions for each DNA sample were performed in triplicate using 1 μL of the extracted DNA in a final volume of 25 μL following the procedure described in Vasselon et al. (2017).

Equimolar concentrations of PCR products were pooled for each sample and sent to the sequencing platform where PCR amplicons were purified and libraries were prepared and quantified as described by Bailet et al. (2020). Sequencing was performed on an Illumina MiSeq platform using paired-end sequencing for 500 cycles with Standard kit v2. Samples were sequenced in six separate runs.

2.3 | Bioinformatics

Bioinformatics data processing was performed as described by Tapolczai et al. (2019) using the software package DADA2 and

TABLE 1 Details for each sampled geographic region.

Region	Land type	Climate zone	Region size (km²)	Number of sites	Sampling years
Fennoscandia	Continent	Subpolar	1,173,926	26	2011–2015, 2019
France Mainland	Continent	Temperate	551,695	411	2016–2017
West Africa	Continent	Tropical	568,320	16	2019
French Guiana	Continent	Tropical	83,534	85	2013–2017
New Caledonia	Island	Tropical	18,576	64	2016
Tahiti	Island	Tropical	1044	15	2018
Mayotte	Island	Tropical	374	19	2014–2015

1508 | WILEY | Environmental DNA
Dedicated to the study and use of environmental DNA for basic and applied sciences
Open Access

CHONOVA ET AL.

following the methods described by Callahan et al. (2016). For further details, see Supplementary methods. Taxonomic assignment was performed using Diat.barcode version 9 (Rimet et al., 2019) with a bootstrap confidence threshold of 60% and the Naïve Bayesian Classifier algorithm (Wang et al., 2007).

Taxonomic filtering was applied to remove ASVs assigned to groups different from Bacillariophyta (diatoms) and ASVs containing less than 10 reads across all samples were excluded from further analyses. To compare samples without statistical bias, sample size was normalized to 18,000 reads (the lowest read number per sample) after verification that rarefaction curves exhibited saturation (Figure S1). Remaining high-quality rbcL rRNA gene sequences represented a total of 6548 unique ASVs.

2.4 | Data analysis

Quantitative (relative abundances) and qualitative (presence-absence) sample-by-sequence ASVs tables were used to analyze diatom communities. Unless stated otherwise, statistical analyses were performed with R software 3.6.3 (R Core Team, 2017) and graphical representations with the ggplot2 package (Wickham, 2016).

Genetic alpha diversity was determined by calculating richness (number of ASVs based on presence-absence data) and diversity (Shannon index based on relative abundance data). Statistical differences were tested between climate zones (subpolar, temperate, or tropical) and land types (mainland or island) using Wilcoxon signed-rank test with Benjamini & Hochberg correction for multiple comparisons (Benjamini & Hochberg, 1995). Generalized estimating equation models using Poisson distribution (Liang & Zeger, 1986) were applied using the geepack R package (Højsgaard et al., 2006) to test differences in the number of ASVs in function of latitude while allowing a correlation structure within regions (exchangeable matrix). The effect of land type (mainland or island) was also tested.

Nonmetric multidimensional scaling analysis (NMDS) on Bray–Curtis dissimilarity matrix (relative abundance data) was used to derive a two-dimensional configuration of ASVs community structures. Differences between geographic regions were tested with PERMANOVA with 999 permutations (Anderson, 2001). Shannon diversity and multivariate analyses were calculated with the package vegan (Oksanen et al., 2007).

The percentage of exclusive and shared ASVs (presence-absence data) was calculated for geographic regions, land types, climate zones, and continents. To control for differences related to the sampling effort, mean of shared and unique ASVs was calculated 1000 times after randomized subsampling to the lowest sample number (15 for regions, 26 for climate zones, and 35 for continents). Percentages of exclusive and shared ASVs were shown in barplots representing the size of different sets (Lex et al., 2014). The number of exclusive and shared ASVs was presented in a global map to visualize the distribution patterns of genetic variants.

To study the effect of genetic resolution on the proportion of exclusive and shared variants between regions, the computations described above were performed on clustered operational taxonomic units (OTUs) and species. To obtain OTUs, ASVs were clustered with the farthest neighbor clustering method and similarity threshold from 99% to 80% using the R package bioseq (Keck, 2020).

3 | RESULTS

A total of 6548 genetic variants were detected by DNA metabarcoding across all studied regions (Figure 1, Table 1). Sequence-based interpolation (rarefaction) showed that sequencing depth was sufficient to saturate the curve and therefore the actual diversity is well represented (Figure S1). Genetic variants were clustered in OTUs, whose number ranged from 4126 (99% similarity) to 692 (93% similarity) depending on the applied similarity threshold. Diat.barcode v9 reference DNA library enabled taxonomic affiliation of 393 species from 103 genera encompassing 3143 (48%) and 4911 (75%) of the sequence variants, for species and genera respectively.

### 3.1 | Global genetic alpha diversity patterns

We studied genetic richness (number of ASVs) and diversity (Shannon) across climate zones and land types (mainland and island). At first sight, we observed lower diversity in the tropics compared to other climate zones and in islands compared to mainlands (Wilcoxon rank sum test, $p < 0.001$) (Figure 2; Figure S2). When islands were excluded from the analysis, however, the richness did not differ between climate zones, indicating that there is no general effect of latitude on benthic diatom diversity (Wilcoxon rank sum test, $p > 0.5$). This observation was confirmed with GEE modeling, which showed that, when land type was included in the model, the effect of latitude was not significant (Wald $= 0.29$, $p > 0.59$) while the effect of land type was highly significant (Wald $= 381.56$, $p < 0.001$).

3.2 | Spatial beta-diversity

Multivariate analysis of beta-diversity was conducted on relative abundances of ASVs in samples using Bray–Curtis dissimilarities and displayed as NMDS (Figure 3; Figure S3). Permutational multivariate analysis of variance (PERMANOVA) indicated significant differences in ASVs composition between regions, climate zones, continents, and land types (PERMANOVA, 999 permutations, $p < 0.001$). Samples from subpolar and temperate zones overlapped with each other, but not with tropical samples. Within the tropics, a gradient was observed from French Guiana to the islands (New Caledonia, Mayotte, and Tahiti) with West Africa located in the middle. Tropical islands overlapped with each other and New Caledonia showed the highest heterogeneity among samples.

—WILEY—
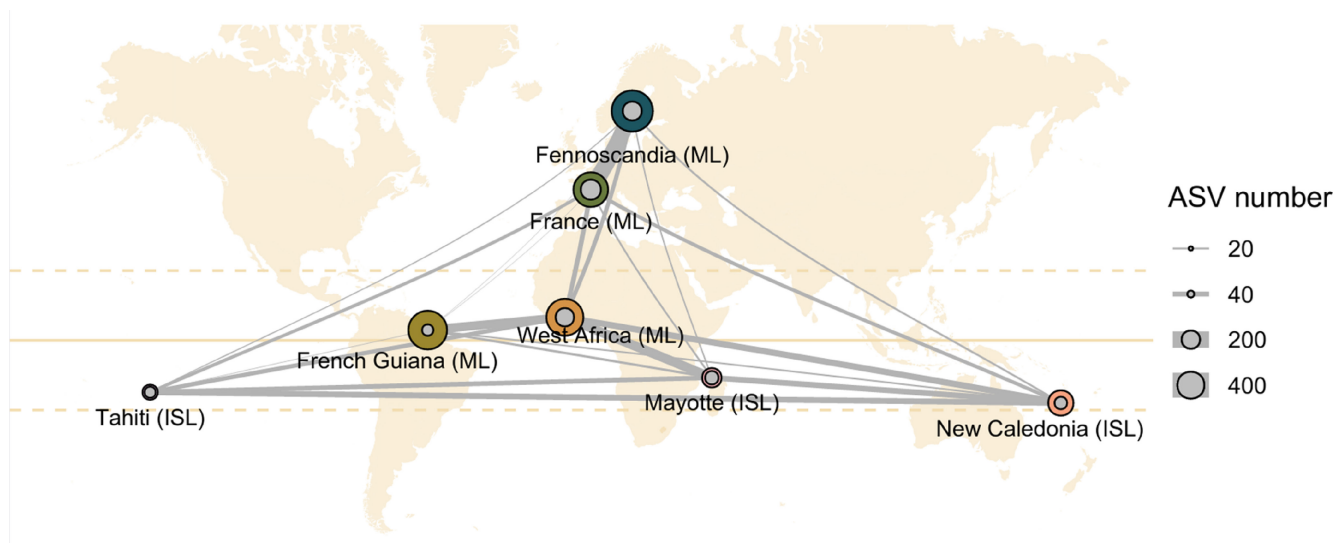
FIGURE 1 Map of sampling locations. Outer colored circles delineate the total number of genetic variants recorded in each geographic region. Inner gray circles delineate how many of these genetic variants are shared with other regions. Line widths show the number of genetic variants shared between two regions. To control for differences related to the sampling effort, the mean of 1000 randomizations was calculated using 15 random sites (i.e. the number of sites of the least sampled region - Tahiti). Key: ISL, island; ML, mainland.
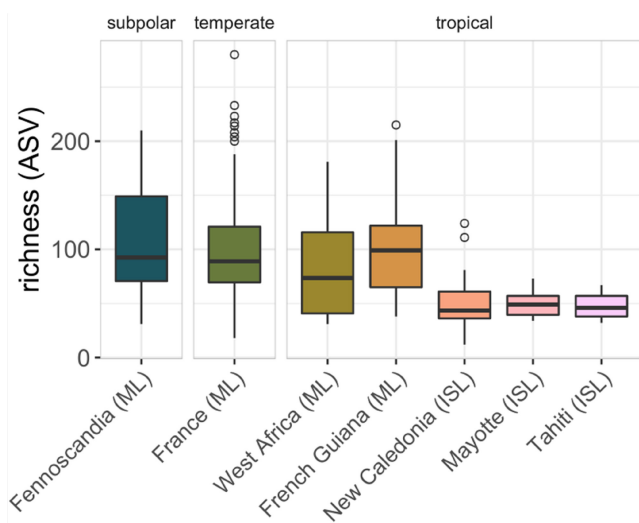
FIGURE 2 Distribution of diatom richness calculated from genetic variants of benthic diatoms collected in the subpolar (Fennoscandia), temperate (France), and tropical (French Guiana, West Africa, New Caledonia, Mayotte, and Tahiti) climate zones. Key: ISL, island; ML, mainland.

3.3 | Genetic variant distribution

The percentages of genetic variants specific for a given region and shared between two or more regions are given in Figure 4a–d, where on each panel, vertical bars indicate the frequency of a particular set of regions depicted by the intersecting dots below each bar.

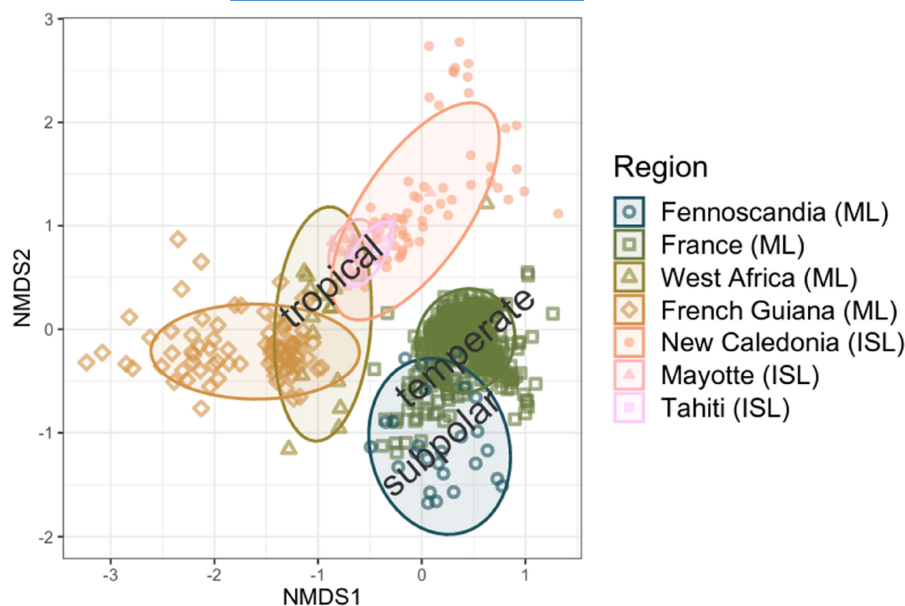Figure 4a–d shows that most variants were specific to a single region. 70% of the variants occurred exclusively in mainlands and 25% in islands (Figure 4a). Most genetic variants were also climate-specific (89%), with the highest number occurring in the tropical zone (37%), followed by the subpolar (33%) and the temperate zone (19%). The remaining 12% were found in more than one climate zone, with the highest percentage shared between the subpolar and temperate zones (10%) (Figure 4b). The cumulated proportion of genetic variants specific to a single continent (i.e. the sum of the dark bars in Figure 4c) was 92%, with the highest number recorded in South America (33%), followed by Europe (28%), Africa (19%) and Oceania (12%). Variants found in more than one continent corresponded to 8% only, whereby 2% were shared between Oceania and Africa and 2% between Africa and South America. Less than 0.5% of the variants were present in all continents (Figure 4c). Zooming into the geographic regions, 86% of the genetic variants were found in one region only, with highest percentage recorded in French Guiana (22%) and Fennoscandia (21%), followed by West Africa (17%) and France (13%), and finally by the islands New Caledonia (8%), Mayotte (3%), and Tahiti (2%) (Figures 1 and 4d). The remaining 14% are present in more than one region, with the highest number shared between Fennoscandia and France Mainland (5%), West Africa and French Guiana (2%), and West Africa and Mayotte (1%).

A relatively high proportion of the genetic variants were not specific to islands (30%, 54%, and 60% for New Caledonia, Mayotte, and Tahiti, respectively). This proportion for mainlands varied between 11% (French Guiana) and 37% (France) (Figure 1).

3.4 | Genetic resolution

In order to compare dispersal trends at different levels of genetic similarity, the proportion of region-specific taxonomic units and

1510 | WILEY Environmental DNA
Dedicated to the study and use of environmental DNA for basic and applied sciences
Open Access

CHONOVA ET AL.

FIGURE 3 Non-metric multidimensional scaling (NMDS) plot of observed similarities between genetic variant profiles of benthic diatom communities collected in the subpolar (Fennoscandia), temperate (France), and tropical (French Guiana, West Africa, New Caledonia, Mayotte, and Tahiti) climate zones (stress value = 0.13, 95% confidence ellipses are represented for each group). Key: ISL, island; ML, mainland.

those shared between several regions was calculated for OTUs (99%–91% similarity) and species (ASVs assigned with the reference library Diat.barcode) in addition to ASVs (Figure 4e). Relaxing the clustering threshold from 100% (ASVs) to 97% and 95% similarity (i.e. values typically used as a proxy for diatom species, Evans et al., 2007) resulted in a decrease in region-specific units from 86% to 69% and 56%, respectively, while units detected in two regions remained between 12 and 22%. Fractions calculated on species showed that only 37% were specific to a single region (see also Figure S4).

4 | DISCUSSION

Using a large DNA metabarcoding dataset, we studied genetic diversity and biogeography of stream diatoms from four continents and three different climate zones. Based on our results, three key findings were observed: (i) genetic richness of diatoms is not affected by a latitudinal gradient, but by an islands effect, (ii) genetic variants of diatoms are highly specific to bioclimatic regions and continents, and (iii) the proportion of region-specific variants increases when increasing the genetic resolution.

Before discussing each of these results in more detail, it is important to highlight that the data used here are not without limitations. In particular, the dataset is limited to certain regions of the world, and some of these regions are better represented than others. These limitations reflect the many technical, financial, and legal constraints inherent in collecting a large dataset of environmental DNA from rivers. Yet the metabarcoding dataset presented here is the first ever assembled at this scale for freshwater diatoms and helps to advance our understanding of diatom biogeography. Importantly, sites were carefully chosen to represent regional variations in environmental conditions and statistical rarefaction was systematically used to control for unbalanced designs.

4.1 | Diatom genetic diversity is affected by island effect but not latitude

Our results based on molecular data indicated significantly lower diversity in islands compared to continental regions. The small size and remoteness of islands is typically reflected by reduced species diversity of higher organisms (MacArthur & Wilson, 1967) and microbes (Peay et al., 2007). This trend was also shown for diatoms in the Antarctic region where morphological data showed that species diversity decreases with isolation (Verleyen et al., 2021). Bolgovics et al. (2016) studied lakes and ponds considering them as aquatic islands on the terrestrial landscape, and demonstrated a positive relationship between the area size and the diversity of benthic diatoms. Teittinen and Soininen (2015) on the other hand did not find such relation in Finish spring diatoms. Finally, a recent study (Jamoneau et al., 2022) also found a higher diversity of stream diatoms in continents compared to islands. The authors explain this difference by environmental characteristics of the islands (e.g. current velocity) rather than by their size and isolation. This is a hypothesis that we cannot explore here because of the lack of environmental data. It should nevertheless be noted that the diversity studied by Jamoneau et al. (2022) is a cumulative richness over 15 sites (i.e. a regional gamma diversity integrating space), whereas we focus here on alpha diversity (average richness per site).

Following typical patterns of a latitudinal diversity gradient (LDG), a decrease in species richness across organismal groups and habitat types would be expected from the tropics to the poles (Hillebrand & Azovsky, 2001). Different hypotheses have been proposed to explain these patterns suggesting that higher diversity at lower latitudes may be related to the older and historically larger tropical environments or to higher speciation or lower extinction rates in the tropics (Mittelbach et al., 2007). Compared to larger organisms, microbes show weaker LDG that may be
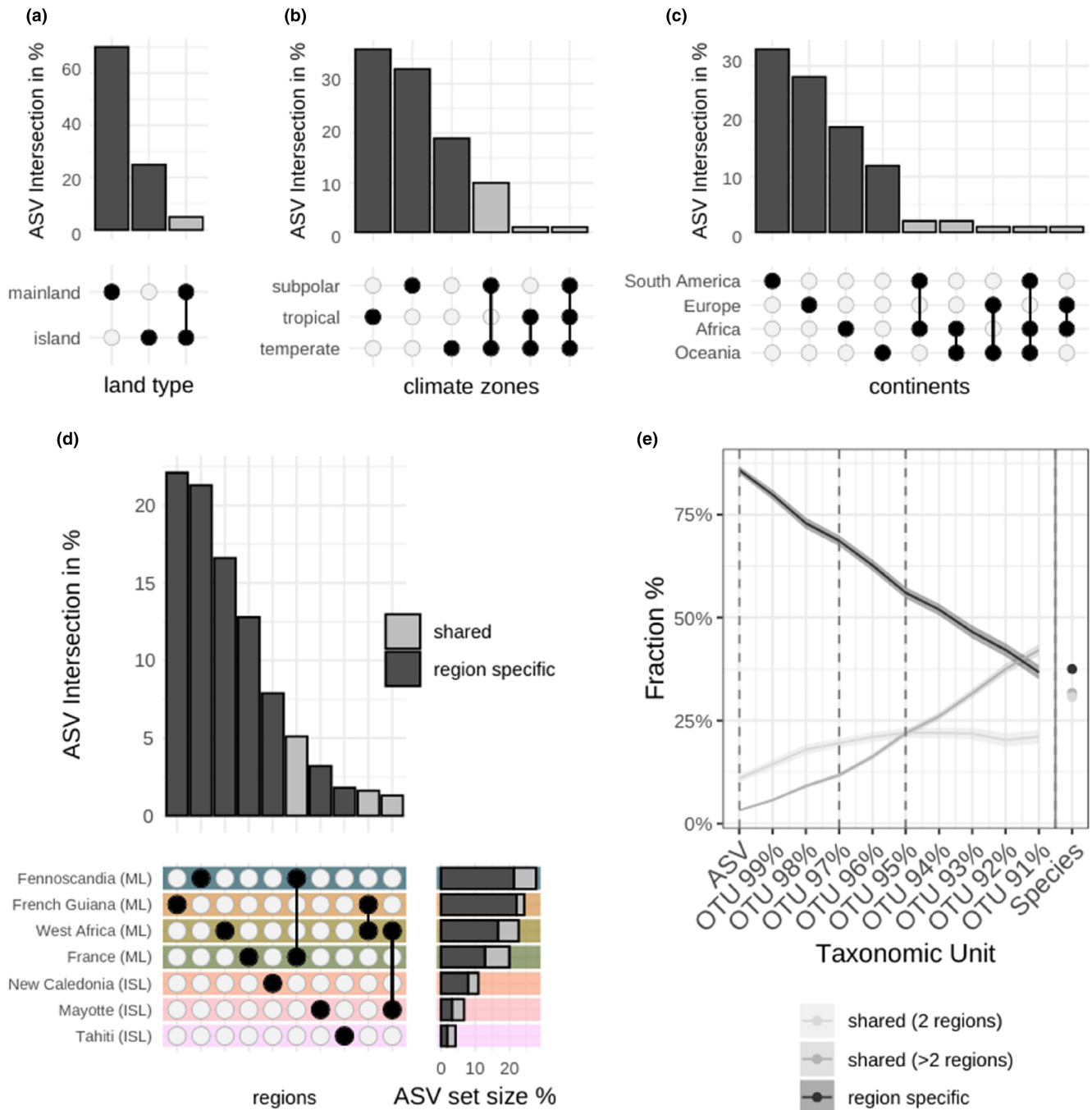
FIGURE 4 Barplots of set size presenting percentage of shared and region-specific genetic variants in (a) land types, (b) climate zones, and (c) continents. (d) Barplots of set size presenting percentage of shared and region-specific genetic variants in each geographic region and horizontal barplots presenting percentage of the set size of each region from the total genetic variants. (e) Percentage of ASVs, OTU (similarity threshold from 99% to 91%), and species richness unique for each geographic region and shared between two or more regions. The transparent area displays the standard deviation. ML stands for mainland and ISL – for island. To control for differences related to the sampling effort, mean shared and unique taxonomic units were calculated using the lowest sample number per category (98 for land type, 35 for continents, 26 for climate zones, and 15 for regions), chosen randomly 1000 times. For a–d, bars representing <1% are not displayed. Key: ASVs, amplicon sequence variant; ISL, island; ML, mainland.

related to their small body size and large populations resulting in higher chances for dispersal across long distances (Soininen & Teittinen, 2019). Previous studies on diatoms reported weak (Hillebrand & Azovsky, 2001) or non-significant (Schiaffino et al., 2016) LDG. Vyverman et al. (2007) reported strongly asymmetric LDG

between both hemispheres and found that diatom diversity is negatively correlated to the degree of isolation. Diatom diversity patterns opposite to the classical gradient were also observed. Findings obtained using fossil records of marine planktonic diatoms indicated higher diversity at the poles compared to the

1512 | **WILEY**— Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

CHONOVA ET AL.

equator (Powell & Glazier, 2017). A large-scale study on river diatom diversity based on morphology also reported a reversed gradient with decreasing richness near the equator (Soininen et al., 2016). It is worth noting that in the latter study, sampling sites at lower latitude are islands, while the remaining sites include continental areas. In addition to the dispersal abilities of diatoms, high nutrient, and light concentrations in (sub)polar regions that are important factors for diatom development may promote diversification in higher latitudes resulting in vanishing of typical LDG (Heino et al., 2018). The higher presence of wetlands in Fennoscandia may further stimulate diatom biodiversity due to supply of bioavailable iron and humic substances (Soininen et al., 2016). Finally, we must recall that, to avoid biases related to isolation and dispersion limitations of islands, it is highly relevant to take into account the land size when studying latitudinal effect on diversity. Taking into account the difference between continents and islands, our results did not suggest differences in diatom diversity along the latitudinal gradient. Previously observed patterns show contrasting trends that are possibly related to differences in habitats (islands, connectivity, availability of resources) which covaried to different degrees with the latitudinal gradient.

4.2 | Region-specific variants are common in diatoms

The majority of genetic variants (86%) was specific to a single geographic region. Such a high level of genetic specificity was hardly expected even if it is in line with the conclusions of Vyverman et al. (2007) who estimated that physical barriers are important to diatom dispersion and are comparable to those intervening for macro-organisms. However, it is important to note that this study, although covering a large scale, remains limited in terms of geographical coverage. Therefore, the notion of genetic specificity as discussed here is relative to the regions that were actually sampled. To precisely delineate the geographic range of genetic variants would require a high spatial resolution sampling, covering the entire globe.

Overall, islands show lower levels of exclusive variants compared to mainlands. Such a tendency may be a result of processes related to restricted colonization opportunities, higher extinction, and lower speciation rates in isolated environments. Some diatom taxa have higher dispersion possibilities and/or higher tolerance to variation in organic pollution and salinity, and these properties may be decisive for crossing large barriers to isolated environments (Soininen & Teittinen, 2019). Only a few species with such favorable characteristics possibly managed to colonize islands despite their high geographic distance. A parallel can be drawn from a remote Saharan oasis (south Algeria) isolated from other waterbodies, where diatoms detected via microscopy and sequencing were all cosmopolitan opportunistic species (Gastineau et al., 2021). In addition, smaller areas have lower carrying capacity and their populations are more vulnerable and prone to extinction due to demographic, environmental, or

genetic reasons (Frankham, 1997). It is therefore likely that only a few of the species that reach isolated islands are able to survive and that the remaining ones lose levels of genetic variations (Mittelbach et al., 2007). In addition, speciation rates in smaller areas are expected to be lower (Mittelbach et al., 2007) and genotypes that developed in islands and are endemic to them are more likely to become extinct compared to non-endemic ones (Frankham, 1997). These dynamics typical for islands would result in lower diversity and endemism.

Compared to other islands, communities in New Caledonia were more heterogeneous and showed a higher proportion of exclusive genetic variants. This can be particularly explained by the geological conditions of New Caledonia that has a large deposit of peridotites (Moser et al., 1998). Such conditions are only present in few places on earth and may be a reason for the transfer of a very particular chemistry to freshwater bodies leading to higher levels of endemism (40% according to Moser et al., 1998) based on morphology, and 30% according to our results). Such higher endemism is an indication for limitations in dispersal and adapted development to the specific water chemistry.

At continental level, Oceania exhibits the lowest number of specific variants and richness. In our study, Oceania comprises samples from tropical islands only. The small size and the isolation of these lands can possibly explain the low species richness and proportion of exclusive variants observed in this continent (see Verleyen et al. (2021) and our discussion above). On the contrary, South America, also located in the tropical climate zone, is the continent with the highest level of exclusive variants, despite the fact that it is represented by a single region (French Guiana). This result confirms previous findings showing that Amazonian regions are hotspots of diatom diversity and endemism (Carayon et al., 2020). Finally, Europe has the second highest number of specific variants, which is probably related to the presence of different climates – temperate and subpolar. Subpolar diatom flora is reported to be distinct from temperate flora (Keck et al., 2018; Verleyen et al., 2021). Our data also support this statement, since there is a high proportion of exclusive variants in Fennoscandia, despite the fact that it is located in the same continent and close to the temperate zone (France). Such high number of specific variants in Fennoscandia may also be related to the presence of specific environmental processes and conditions (many streams with relatively low pH or high amount of humic acids), which would result in a number of specific ecological niches hence increased gamma diversity (Keck et al., 2018; Soininen et al., 2016).

Finally, the proportion of shared genetic variants is generally low. Cosmopolitanism is particularly limited between continents (less than 3%), whereas when considering climate zones, the subpolar (Fennoscandia) and temperate (France) zones – both located in Europe, shared 10% of the genetic variants. Such higher similarity is related to their close geographic location and strong connection by land that facilitate the transportation of freshwater diatoms (Manning et al., 2021). The case of the tropical continental samples (South America and Africa) is interesting: these two continents, which

belong to the same climate zone, are separated by ca. 4000 km of sea and share only two percent of genetic variants. All these observations confirm that even if cosmopolitan genotypes exist, geographical barriers are a key limitation for continental freshwater diatoms (Vyverman et al., 2007).

4.3 | Genetic resolution impacts the proportion of variants shared across regions

Genetic variants give access to diversity at the finest possible scale. Our observations based on genetic variants show that freshwater diatoms are strongly region-specific. However, the choice of genetic resolution is important and may affect the observed proportion of specific variants. As recently discussed by Antich et al. (2021), clustering and denoising algorithms are complementary techniques that can reveal different information. Thus, it is possible to apply clustering algorithms on denoised sequences to obtain OTUs, in order to reduce genetic resolution and therefore approach different taxonomic levels.

When we decreased the genetic resolution (100%–91% similarity), the number of units detected in two geographic regions remained relatively stable. However, we observed a decrease from 86% (genetic variant, 100% similarity) to 37% (OTU clustered at 91% similarity) for units detected in only one region. Similar findings were previously reported for fungi (Tipton et al., 2022). Even if there is no clear barcoding gap between diatom species, 97%–96% similarity between sequences is the limit commonly accepted for the rbcL gene. Below 96%, sequences are usually belonging to different species and below 94%–91% – to different genera (Evans et al., 2007). Our results clearly show that the regional specificity is high at sub-species level (above 97%). Such genetic levels can also correspond to different species that are morphologically cryptic and might be unable to reproduce (Evans et al., 2007). For other taxa, such levels can correspond to populations that are still able to mate with each other (Trobajo et al., 2009). When we consider interspecific level (below 97%), an important fraction of diatoms remained region-specific, but this was not the case when considering genus level (below 90%). Similar observations were made on morphological data, showing that endemic genera (e.g. *Eunophora*, *Sabbea*, *Sinoperonia*, *Celebesia*, *Kurtkrammeria*) are rare due to their older divergence from mother lineages (e.g. Kapustin et al., 2018, Kociolek, 2018).

AUTHOR CONTRIBUTIONS

TC, FR, AB, and FK designed the study. TC analyzed and interpreted the data. FR, AB, MK, SCS, BB, AEG, GG, OM, AO, SR, and MR contributed to the data. TC wrote the manuscript with contributions from all other coauthors.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Chonova et al. Dataset from: Revisiting global diversity and biogeography of freshwater microalgae: new insights from molecular data. https://doi.org/10.5281/zenodo.6344975

ORCID

François Keck https://orcid.org/0000-0002-3323-4167

REFERENCES

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*, 32–46.

Antich, A., Palacin, C., Wangensteen, O. S., & Turon, X. (2021). To denoise or to cluster, that is not the question: Optimizing pipelines for COI metabarcoding and metaphylogeography. *BMC Bioinformatics*, *22*, 177.

Baas-Becking, L. G. M. (1934). *Geobiologie of inleiding tot de milieukunde*. WP Van Stockum & Zoon.

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., Bengtsson-Palme, J., Anslan, S., Coelho, L. P., Harend, H., Huerta-Cepas, J., Medema, M. H., Maltz, M. R., Mundra, S., Olsson, P. A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., ... Bork, P. (2018). Structure and function of the global topsoil microbiome. *Nature*, *560*, 233–237.

1514 | **WILEY** Environmental DNA
Dedicated to the study and use of environmental DNA for basic and applied sciences
Open Access

CHONOVA ET AL.

Bailet, B., Apothéloz-Perret-Gentil, L., Baričević, A., Chonova, T., Franc, A., Frigerio, J. M., Kelly, M., Mora, D., Pfannkuchen, M., Proft, S., Ramon, M., Vasselon, V., Zimmermann, J., & Kahlert, M. (2020). Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. *Science of the Total Environment*, *745*, 140948.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300.

Bennett, J. R., Cumming, B. F., Ginn, B. K., & Smol, J. P. (2010). Broad-scale environmental response and niche conservatism in lacustrine diatom communities: Scale and environment in diatom communities. *Global Ecology and Biogeography*, *19*, 724–732.

Bolgovics, Á., Ács, É., Várbíró, G., Görgényi, J., & Borics, G. (2016). Species area relationship (SAR) for benthic diatoms: A study on aquatic islands. *Hydrobiologia*, *764*, 91–102.

Bottin, M., Soininen, J., Ferrol, M., & Tison-Rosebery, J. (2014). Do spatial patterns of benthic diatom assemblages vary across regions and years? *Freshwater Science*, *33*, 402–416.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581–583.

Carayon, D., Eulin-Garrigue, A., Vigouroux, R., & Delmas, F. (2020). A new multimetric index for the evaluation of water ecological quality of French Guiana streams based on benthic diatoms. *Ecological Indicators*, *113*, 106248.

CEN. (2014). Water quality - Guidance for the routine sampling and preparation of benthic diatoms from rivers and lakes. 22.

CEN. (2018). Water quality - Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses. 8.

Cermeño, P., & Falkowski, P. G. (2009). Controls on diatom biogeography in the ocean. *Science*, *325*, 1539–1541.

De Kort, H., Prunier, J. G., Ducatez, S., Honnay, O., Baguette, M., Stevens, V. M., & Blanchet, S. (2021). Life history, climate and biogeography interactively affect worldwide genetic diversity of plant and animal populations. *Nature Communications*, *12*, 516.

De Luca, D., Kooistra, W. H., Sarno, D., Gaonkar, C. C., & Piredda, R. (2019). Global distribution and diversity of *Chaetoceros* (Bacillariophyta, Mediophyceae): Integration of classical and novel strategies. *PeerJ*, *7*, e7410.

De Luca, D., Piredda, R., Sarno, D., & Kooistra, W. H. (2021). Resolving cryptic species complexes in marine protists: Phylogenetic haplotype networks meet global DNA metabarcoding datasets. *The ISME Journal*, *15*, 1931–1942.

Evans, K. M., & Mann, D. G. (2009). A proposed protocol for nomenclaturally effective DNA barcoding of microalgae. *Phycologia*, *48*, 70–74.

Evans, K. M., Wortley, A. H., & Mann, D. G. (2007). An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist*, *158*, 349–364.

Finlay, B. J. (2002). Global dispersal of free-living microbial eukaryote species. *Science*, *296*, 1061–1063.

Fontaneto, D. (2019). Long-distance passive dispersal in microscopic aquatic animals. *Movement Ecology*, *7*, 10.

Frankham, R. (1997). Do Island populations have less genetic variation than mainland populations? *Heredity*, *78*, 311–327.

Gastineau, R., Hamedi, C., Baba Hamed, M. B., Abi-Ayad, S. M. E. A., Bąk, M., Lemieux, C., Turmel, M., Dobosz, S., Wróbel, R. J., Kierzek, A., Lange-Bertalot, H., & Witkowski, A. (2021). Morphological and molecular identification reveals that waters from an isolated oasis in Tamanrasset (extreme South of Algerian Sahara) are colonized by opportunistic and pollution-tolerant diatom species. *Ecological Indicators*, *121*, 107104.

Heino, J., Melo, A. S., Jyrkänkallio-Mikkola, J., Petsch, D. K., Saito, V. S., Tolonen, K. T., Bini, L. M., Landeiro, V. L., Silva, T. S. F., Pajunen, V., Soininen, J., & Siqueira, T. (2018). Subtropical streams harbour higher genus richness and lower abundance of insects compared to boreal streams, but scale matters. *Journal of Biogeography*, *45*, 1983–1993.

Hillebrand, H., & Azovsky, A. I. (2001). Body size determines the strength of the latitudinal diversity gradient. *Ecography*, *24*, 251–256.

Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, *15*, 1–11.

Jamoneau, A., Soininen, J., Tison-Rosebery, J., Boutry, S., Budnick, W. R., He, S., Marquié, J., Jyrkänkallio-Mikkola, J., Pajunen, V., Teittinen, A., Tupola, V., Wang, B., Wang, J., Blanco, S., Borrini, A., Cantonati, M., Valente, A. C., Delgado, C., Dörflinger, G., … Passy, S. I. (2022). Stream diatom biodiversity in islands and continents—A global perspective on effects of area, isolation and environment. *Journal of Biogeography*, *49*, 2156–2168.

Kahlert, M., Albert, R. L., Anttila, E. L., Bengtsson, R., Bigler, C., Eskola, T., Gälman, V., Gottschalk, S., Herlitz, E., Jarlman, A., Kasperoviciene, J., Kokociński, M., Luup, H., Miettinen, J., Paunksnyte, I., Piirsoo, K., Quintana, I., Raunio, J., Sandell, B., … Weckström, J. (2009). Harmonization is more important than experience—Results of the first Nordic–Baltic diatom intercalibration exercise 2007 (stream monitoring). *Journal of Applied Phycology*, *21*, 471–482.

Kapustin, D. A., Kulikovskiy, M., & Kociolek, J. P. (2018). *Celebesia* gen. Nov., a new cymbelloid diatom genus from the ancient Lake Matano (Sulawesi Island, Indonesia). *Nova Hedwigia Beihefte*, *146*, 147–155.

Keck, F. (2020). Handling biological sequences in R with the bioseq package (A Mahon, Ed.). *Methods in Ecology and Evolution*, *11*, 1728–1732.

Keck, F., Franc, A., & Kahlert, M. (2018). Disentangling the processes driving the biogeography of freshwater diatoms: A multiscale approach. *Journal of Biogeography*, *45*, 1582–1592.

Kier, G., Kreft, H., Lee, T. M., Jetz, W., Ibisch, P. L., Nowicki, C., Mutke, J., & Barthlott, W. (2009). A global assessment of endemism and species richness across Island and mainland regions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9322–9327.

Kociolek, J. P. (2018). A worldwide listing and biogeography of freshwater diatom genera: A phylogenetic perspective. *Diatom Research*, *33*, 509–534.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*, 1983–1992.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22.

MacArthur, R. H., & Wilson, E. O. (1967). *The theory of Island biogeography*. Princeton University Press.

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A., & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, E1516–E1525.

Mann, D. G., Evans, K. M., Chepurnov, V. A., & Nagai, S. (2009). Morphology and formal description of *Sellaphora bisexualis* sp. nov. (Bacillariophyta). *Fottea*, *9*, 199–209.

Manning, F. S., Curtis, P. J., Walker, I. R., & Pither, J. (2021). Potential long-distance dispersal of freshwater diatoms adhering to waterfowl plumage. *Freshwater Biology*, *66*, 1136–1148.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J.

A., Kuske, C. R., Morin, P. J., Naeem, S., Øvreås, L., Reysenbach, A. L., Smith, V. H., & Staley, J. T. (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews. Microbiology*, *4*, 102–112.

Metz, S., Singer, D., Domaizon, I., Unrein, F., & Lara, E. (2019). Global distribution of Trebouxiophyceae diversity explored by high-throughput sequencing and phylogenetic approaches. *Environmental Microbiology*, *21*, 3885–3895.

Miraldo, A., Li, S., Borregaard, M. K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K. A., & Nogués-Bravo, D. (2016). An Anthropocene map of genetic diversity. *Science*, *353*, 1532–1535.

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., Harrison, S. P., Hurlbert, A. H., Knowlton, N., Lessios, H. A., McCain, C. M., McCune, A. R., McDade, L. A., McPeek, M. A., Near, T. J., Price, T. D., Ricklefs, R. E., Roy, K., Sax, D. F., … Turelli, M. (2007). Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecology Letters*, *10*, 315–331.

Moser, G., Lange-Bertalot, H., & Metzeltin, D. (1998). Insel der Endemiten Geobotanisches Phänomen Neukaledonien (Island of endemics. New Caledonia-a botanical phenomenon). *Bibliotheca Diatomologica*, *38*, 464.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H., Oksanen, M. J., & Suggests, M. A. (2007). The vegan package. *Community Ecology Package*, *10*, 719.

Peay, K. G., Bruns, T. D., Kennedy, P. G., Bergemann, S. E., & Garbelotto, M. (2007). A strong species-area relationship for eukaryotic soil microbes: Island size matters for ectomycorrhizal fungi. *Ecology Letters*, *10*, 470–480.

Pérez-Burillo, J., Trobajo, R., Leira, M., Keck, F., Rimet, F., Sigró, J., & Mann, D. G. (2021). DNA metabarcoding reveals differences in distribution patterns and ecological preferences among genetic variants within some key freshwater diatom species. *Science of the Total Environment*, *798*, 149029.

Powell, M. G., & Glazier, D. S. (2017). Asymmetric geographic range expansion explains the latitudinal diversity gradients of four major taxa of marine plankton. *Paleobiology*, *43*, 196–208.

R Development Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rimet, F., Gusev, E., Kahlert, M., Kelly, M. G., Kulikovskiy, M., Maltsev, Y., Mann, D. G., Pfannkuchen, M., Trobajo, R., Vasselon, V., Zimmermann, J., & Bouchez, A. (2019). Diat.barcode, an open-access curated barcode library for diatoms. *Scientific Reports*, *9*, 15116.

Schiaffino, M. R., Lara, E., Fernández, L. D., Balagué, V., Singer, D., Seppey, C. C. W., Massana, R., & Izaguirre, I. (2016). Microbial eukaryote communities exhibit robust biogeographical patterns along a gradient of Patagonian and Antarctic lakes: Biogeography of freshwater microeukaryotes. *Environmental Microbiology*, *18*, 5249–5264.

Schluter, D., & Pennell, M. W. (2017). Speciation gradients and the distribution of biodiversity. *Nature*, *546*, 48–55.

Soininen, J., Jamoneau, A., Rosebery, J., & Passy, S. I. (2016). Global patterns and drivers of species and trait composition in diatoms: Global compositional patterns in stream diatoms. *Global Ecology and Biogeography*, *25*, 940–950.

Soininen, J., & Teittinen, A. (2019). Fifteen important questions in the spatial ecology of diatoms. *Freshwater Biology*, *64*, 2071–2083.

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., … Velayoudon, D. (2015). Structure and function of the global ocean microbiome. *Science*, *348*, 1261359.

Tapolczai, K., Keck, F., Bouchez, A., Rimet, F., Kahlert, M., & Vasselon, V. (2019). Diatom DNA Metabarcoding for biomonitoring: Strategies to avoid major taxonomical and bioinformatical biases limiting molecular indices capacities. *Frontiers in Ecology and Evolution*, *7*, 409.

Teittinen, A., & Soininen, J. (2015). Testing the theory of Island biogeography for microorganisms—Patterns for spring diatoms. *Aquatic Microbial Ecology*, *75*, 239–250.

Tipton, L., Zahn, G. L., Darcy, J. L., Amend, A. S., & Hynson, N. A. (2022). Hawaiian fungal amplicon sequence variants reveal otherwise hidden biogeography. *Microbial Ecology*, *83*, 48–57.

Trobajo, R., Clavero, E., Chepurnov, V. A., Sabbe, K., Mann, D. G., Ishihara, S., & Cox, E. J. (2009). Morphological, genetic and mating diversity within the widespread bioindicator *Nitzschia palea* (Bacillariophyceae). *Phycologia*, *48*, 443–459.

Trobajo, R., Mann, D. G., Clavero, E., Evans, K. M., Vanormelingen, P., & McGregor, R. C. (2010). The use of partial *cox* 1, *rbc*L and LSU rDNA sequences for phylogenetics and species identification within the *Nitzschia palea* species complex (Bacillariophyceae). *European Journal of Phycology*, *45*, 413–425.

Vasselon, V., Rimet, F., Tapolczai, K., & Bouchez, A. (2017). Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte Island, France). *Ecological Indicators*, *82*, 1–12.

Verleyen, E., van de Vijver, B., Tytgat, B., Pinseel, E., Hodgson, D. A., Kopalová, K., Chown, S. L., van Ranst, E., Imura, S., Kudoh, S., van Nieuwenhuyze, W., ANTDIAT consortium, Sabbe, K., & Vyverman, W. (2021). Diatoms define a novel freshwater biogeography of the Antarctic. *Ecography*, *44*, 548–560.

Verleyen, E., Vyverman, W., Sterken, M., Hodgson, D. A., de Wever, A., Juggins, S., van de Vijver, B., Jones, V. J., Vanormelingen, P., Roberts, D., Flower, R., Kilroy, C., Souffreau, C., & Sabbe, K. (2009). The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos*, *118*, 1239–1249.

Vyverman, W., Verleyen, E., Sabbe, K., Vanhoutte, K., Sterken, M., Hodgson, D. A., Mann, D. G., Juggins, S., Vijver, B. V., Jones, V., Flower, R., Roberts, D., Chepurnov, V. A., Kilroy, C., Vanormelingen, P., & Wever, A. D. (2007). Historical processes constrain patterns in global diatom diversity. *Ecology*, *88*, 1924–1931.

Wallace, A. R. (1876). *The geographical distribution of animals*. Harper and Brothers, Publishers.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*, 5261–5267.

Wickham, H. (2016). *ggplot2*. Springer International Publishing.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.