



**HAL**  
open science

## The global genetic diversity of planktonic foraminifera reveals the structure of cryptic speciation in plankton

Raphaël Morard, Kate Darling, Agnes Weiner, Christiane Hassenrück, Chiara Vanni, Tristan Cordier, Nicolas Henry, Mattia Greco, Nele Vollmar, Tamara Milivojevic, et al.

### ► To cite this version:

Raphaël Morard, Kate Darling, Agnes Weiner, Christiane Hassenrück, Chiara Vanni, et al.. The global genetic diversity of planktonic foraminifera reveals the structure of cryptic speciation in plankton. *Biological Reviews*, 2024, 99 (4), pp.1218-1241. 10.1111/brv.13065 . hal-04552961

**HAL Id: hal-04552961**

**<https://univ-lyon1.hal.science/hal-04552961>**

Submitted on 19 Jul 2024





**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# The global genetic diversity of planktonic foraminifera reveals the structure of cryptic speciation in plankton

Raphaël Morard<sup>1,\*</sup> , Kate F. Darling<sup>2,3</sup>, Agnes K. M. Weiner<sup>4</sup>, Christiane Hassenrück<sup>5</sup>, Chiara Vanni<sup>1</sup>, Tristan Cordier<sup>4</sup>, Nicolas Henry<sup>6,7</sup>, Mattia Greco<sup>8</sup> , Nele M. Vollmar<sup>1,4</sup>, Tamara Milivojevic<sup>1,9</sup>, Shirin Nurshan Rahman<sup>1</sup>, Michael Siccha<sup>1</sup> , Julie Meilland<sup>1</sup>, Lukas Jonkers<sup>1</sup>, Frédéric Quillévère<sup>10</sup> , Gilles Escarguel<sup>11</sup>, Christophe J. Douady<sup>11,12</sup>, Thibault de Garidel-Thoron<sup>13</sup>, Colomban de Vargas<sup>6,14</sup> and Michal Kucera<sup>1</sup>

<sup>1</sup>MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, Bremen 28359, Germany

<sup>2</sup>School of GeoSciences, University of Edinburgh, Edinburgh, EH9 3JW, UK

<sup>3</sup>Biological and Environmental Sciences, University of Stirling, Stirling, FK9 4LA, UK

<sup>4</sup>NORCE Climate and Environment, NORCE Norwegian Research Centre AS, Bjerknes Centre for Climate Research, Jahnebakken 5, Bergen 5007, Norway

<sup>5</sup>Biological Oceanography, Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Seestrasse 15, Warnemünde 18119, Germany

<sup>6</sup>CNRS, Sorbonne Université, FR2424, ABiMS, Station Biologique de Roscoff, Roscoff 29680, France

<sup>7</sup>Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE, 3 rue Michel-Ange, Paris 75016, France

<sup>8</sup>Institut de Ciències del Mar, Passeig Marítim de la Barceloneta, Barcelona 37-49, Spain

<sup>9</sup>Max Planck Institute for Marine Microbiology, Bremen, Germany

<sup>10</sup>Univ Lyon, Université Claude Bernard Lyon 1, ENS de Lyon, CNRS, UMR CNRS 5276 LGL-TPE, Villeurbanne F-69622, France

<sup>11</sup>Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, Villeurbanne F-69622, France

<sup>12</sup>Institut Universitaire de France, Paris, France

<sup>13</sup>Aix Marseille Univ, CNRS, IRD, INRAE, CEREGE, Aix-en-Provence, France

<sup>14</sup>Sorbonne Université, CNRS, Station Biologique de Roscoff, AD2M, UMR7144, Place Georges Teissier, Roscoff 29680, France

## ABSTRACT

The nature and extent of diversity in the plankton has fascinated scientists for over a century. Initially, the discovery of many new species in the remarkably uniform and unstructured pelagic environment appeared to challenge the concept of ecological niches. Later, it became obvious that only a fraction of plankton diversity had been formally described, because plankton assemblages are dominated by understudied eukaryotic lineages with small size that lack clearly distinguishable morphological features. The high diversity of the plankton has been confirmed by comprehensive metabarcoding surveys, but interpretation of the underlying molecular taxonomies is hindered by insufficient integration of genetic diversity with morphological taxonomy and ecological observations. Here we use planktonic foraminifera as a study model and reveal the full extent of their genetic diversity and investigate geographical and ecological patterns in their distribution. To this end, we assembled a global data set of ~7600 ribosomal DNA sequences obtained from morphologically characterised individual foraminifera, established a robust molecular taxonomic framework for the observed diversity, and used it to query a global metabarcoding data set covering ~1700 samples with ~2.48 billion reads. This allowed us to extract and assign 1 million reads, enabling characterisation of the structure of the genetic diversity of the group across ~1100 oceanic stations worldwide. Our sampling revealed the existence of, at most, 94 distinct molecular operational taxonomic units (MOTUs) at a level of divergence indicative of biological species. The genetic diversity only

\* Author for correspondence (E-mail: [rmorard@marum.de](mailto:rmorard@marum.de)).

doubles the number of formally described species identified by morphological features. Furthermore, we observed that the allocation of genetic diversity to morphospecies is uneven. Only 16 morphospecies disguise evolutionarily significant genetic diversity, and the proportion of morphospecies that show genetic diversity increases poleward. Finally, we observe that MOTUs have a narrower geographic distribution than morphospecies and that in some cases the MOTUs belonging to the same morphospecies (cryptic species) have different environmental preferences. Overall, our analysis reveals that even in the light of global genetic sampling, planktonic foraminifera diversity is modest and finite. However, the extent and structure of the cryptic diversity reveals that genetic diversification is decoupled from morphological diversification, hinting at different mechanisms acting at different levels of divergence.

**Key words:** protists, planktonic foraminifera, cryptic diversity, molecular taxonomy, automated delimitation, barcoding, metabarcoding, ecology, biogeography, biodiversity.

## CONTENTS

I. Introduction	1219
II. Assembling data sources	1220
(1) Barcoding data set – PFR <sup>2</sup> v.2	1221
(2) Genotyping data	1222
(3) Metabarcoding data set	1222
(4) Metadata standardisation	1222
III. Data standardisation through molecular taxonomy	1223
(1) Molecular nomenclature	1224
(2) Taxonomic assignment	1224
(3) Metabarcoding data set	1227
IV. Diversity patterns	1229
(1) Global diversity	1229
(2) Latitudinal gradient of cryptic diversity	1230
(3) Compartmentalisation of biodiversity	1232
V. Biogeography and ecology	1232
(1) Cryptic diversity patterns of low-latitude morphospecies	1233
(2) Biogeography of mid- to high-latitude morphospecies	1234
VI. Discussion	1235
VII. Conclusions	1237
VIII. Acknowledgements	1238
IX. Data availability statement	1238
X. References	1238
XI. Supporting information	1241

## I. INTRODUCTION

The standardised description of biodiversity started with the proposition of Linnaeus in *Systema Naturae* in 1735 to use a binomial nomenclature system to catalogue life. Preceding the publication of *The Origin of Species* by more than a century, the hierarchical Linnaean taxonomic system implicitly described the major evolutionary events that led to modern biodiversity. The 275 years of taxonomic description that followed the seminal work of Linnaeus led to the description of 1.2 million species that were catalogued in public databases in 2010, but it is estimated that ~86% of species on Earth are still undescribed (Mora *et al.*, 2011). The work carried out by taxonomists throughout the generations relied on morphology-based diagnosis that is particularly time-consuming, since novel species description should always be compared to the existing body of morphological taxonomic knowledge (Blaxter, 2016). Inevitably, morphological diagnosis

cannot capture the full breadth of biological diversity because of morphologically cryptic diversity (Bickford *et al.*, 2007). The establishment of a DNA-based taxonomy to complement the morphology-based diagnosis can assist in a more comprehensive assessment of diversity, and can rely on barcoding data to bridge morphological and molecular diversity (Hebert *et al.*, 2003). Barcoding utilises DNA sequences as a basic unit to facilitate standardisation within a DNA-based taxonomic system. This system relies on specific marker genes that fix sufficient mutations to differentiate even closely related species. Although barcoding is a powerful tool to identify species, it has a limited throughput since the DNA sequences must be sourced from individual specimens (Cristescu, 2014). This limitation was lifted more than a decade ago with the onset of DNA metabarcoding, a procedure that profiles entire communities by sequencing marker genes from the total DNA extracted from environmental samples. This approach can be applied to virtually any

sample source, such as soil (Mahé *et al.*, 2017), the marine water column (de Vargas *et al.*, 2015) or deep-sea sediment (Lejzerowicz *et al.*, 2021).

Metabarcoding emerged as a result of the constant progress of sequencing technologies (Burki, Sandin & Jamy, 2021). The large sequence data sets revealed that a significant share of living communities is still unknown to science, especially in remote environments such as the deep sea (Cordier *et al.*, 2022). However, metabarcoding approaches wholly rely on the completeness and accuracy of taxonomic reference databases to assign environmental sequences to known taxonomic units (Keck, Couton & Altermatt, 2022). Therefore, the full integration of the most recent metabarcoding data sets with the 275 years of classical morphological taxonomy requires the establishment of dedicated reference databases where Linnaean taxonomy can be associated with DNA barcodes (Guillou *et al.*, 2013). The interpretation of metabarcoding data sets is also further complicated by the fact that different marker genes may provide variable taxonomic resolution (Bucklin *et al.*, 2016), and individual marker genes may exhibit different rates of DNA substitution across groups. One particularly challenging aspect is the case of intra-genomic variability, where several versions of the same gene may occur within a single individual which artificially inflates diversity in metabarcoding studies (Sandin, Romac & Not, 2022). These factors make it difficult to harmonise the diversity estimates of metabarcoding studies with classical morphological taxonomy.

Herein we use planktonic foraminifera, a group of calcifying planktonic protists belonging to the Rhizaria clade (Adl *et al.*, 2019), to integrate morphological taxonomy, barcoding and metabarcoding into a single taxonomic system to assess the structure of their diversity on a global scale. The systematic description of this group started with the work of Alcide d'Orbigny in the early 19th century and the two centuries of taxonomic studies that followed led to the formal description of 50 extant morphologically defined species (Brummer & Kucera, 2022). The outstanding preservation of these relatively large calcifying protists (50  $\mu\text{m}$ –1 mm) in marine sediments has resulted in an exceptionally complete marine fossil record, allowing comprehensive reconstruction of the evolutionary history of this group during the Cenozoic (Aze *et al.*, 2011; Fenton *et al.*, 2021). Planktonic foraminifera are palaeoceanographic proxies (Kucera, 2007), i.e. their species composition (Kucera *et al.*, 2005) and geochemistry (Katz *et al.*, 2010) are used to reconstruct the past state of the surface ocean. The global distribution of planktonic foraminifera morphospecies is well documented by surface sediment census counts (Siccha & Kucera, 2017), their seasonal occurrence by sediment traps (Jonkers & Kučera, 2015), and their vertical distribution in the water column by plankton net hauls (Chaabane *et al.*, 2023). This extensive knowledge of modern planktonic foraminifera ecology results from the need to interpret their fossil archive, which relies on a morphology-based taxonomy.

The morphology-based taxonomy of planktonic foraminifera was questioned in the 1990s following the genetic analysis of individual specimens of the morphospecies

*Globigerinella siphonifera*, which revealed the presence of two 'genotypes' or cryptic species in the Caribbean Sea (Huber, Bijma & Darling, 1997). This first report of cryptic diversity in planktonic foraminifera was followed by similar observations for the species *Orbulina universa*, *Globorotalia truncatulinoides*, and again *G. siphonifera*, collected across large geographic gradients (de Vargas *et al.*, 1999, 2001, 2002), and the bipolar morphospecies *Globigerina bulloides*, *Neogloboquadrina pachyderma*, and *Turborotalita quinqueloba* (Darling *et al.*, 2000, 2004). Subsequent studies reported extensive genetic diversity in the investigated morphospecies (Darling, Kucera & Wade, 2007; Ujjié & Lipps, 2009; Aurahs *et al.*, 2009b; Morard *et al.*, 2011; Ujjié *et al.*, 2012; Weiner *et al.*, 2012, 2014) except for the tropical *Trilobatus sacculifer* (André *et al.*, 2013). Inconsistencies in the methodology used to identify cryptic diversity among studies and the resulting profusion of genetic labels produced in these studies make it difficult to assess the true extent of genetic diversity within the group (André *et al.*, 2014; Morard *et al.*, 2016). The first investigation of planktonic foraminiferal biodiversity based on basin-scale DNA metabarcoding (Morard *et al.*, 2018) indicated that the level of genetic diversity was congruent with that estimated from single-cell DNA barcoding. However, to date there is no standardised taxonomic integration of the wealth of barcoding and metabarcoding data available for planktonic foraminifera to enable robust assessment of their global biodiversity. Clarifying the extent of their genetic diversity is crucial to understand the processes acting at different levels of divergence of these organisms that shaped their evolutionary history.

Here we present a comprehensive assessment of planktonic foraminifera diversity. We first curated an extensive observational data set encompassing all available genetic data for these organisms and structured their genetic variability into molecular operational taxonomic units (MOTUs), using the barcodes as a foundation for our taxonomic system. Subsequently, this refined molecular taxonomy served as the key to interpret the diversity of the metabarcoding data set. We then used the resulting global-scale data set to assess the diversity of planktonic foraminifera, and the macroecological patterns at the cryptic diversity level. Finally, we quantified the degree of biogeographic and environmental differentiation between cryptic species and highlight the major patterns of genetic diversity among planktonic foraminifera.

## II. ASSEMBLING DATA SOURCES

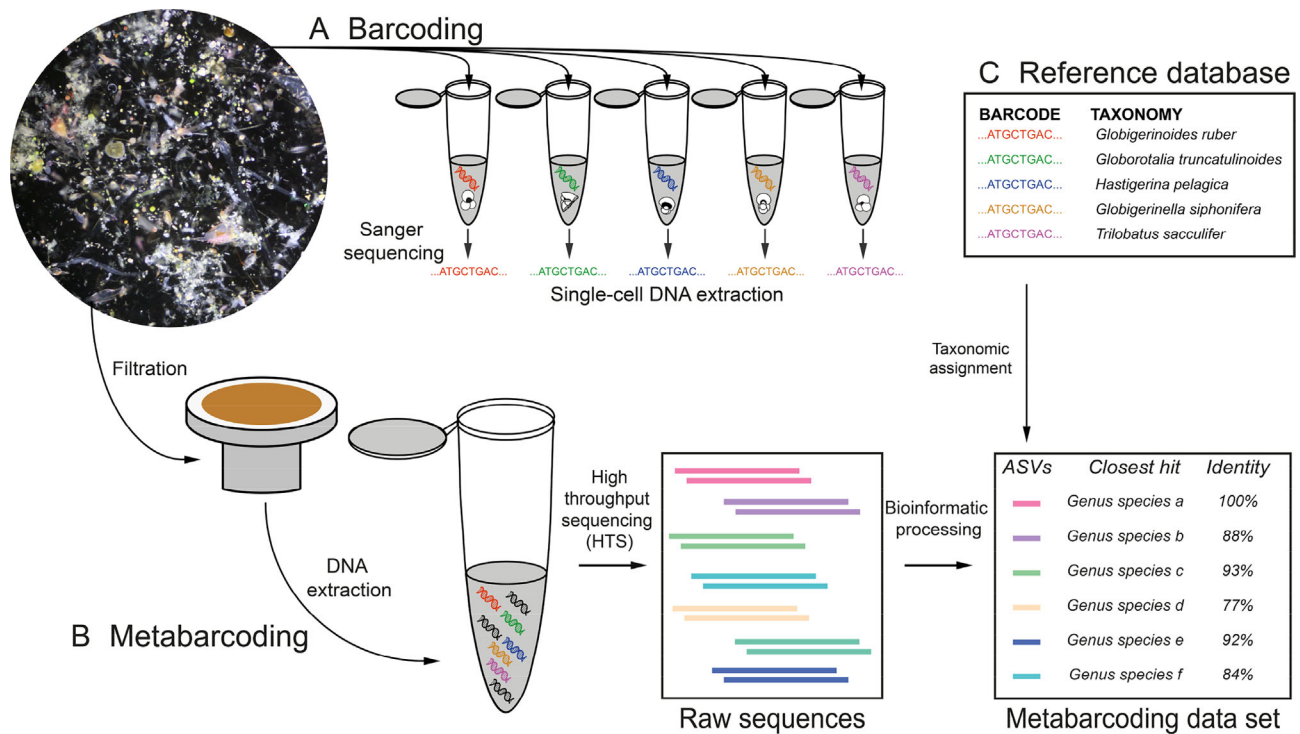
Currently, three distinct data sources are available for studying the genetic diversity of planktonic foraminifera. The first source comprises genetic data obtained from individual specimens that were taxonomically identified based on their morphology (Weiner *et al.*, 2016). These specimens were then Sanger sequenced, mostly targeting the small subunit (SSU) fragment of the ribosomal RNA (rRNA) gene to produce

the ‘barcodes’ (Fig. 1A). The second data source consists of genotyping data, primarily generated using a technique called restriction fragment length polymorphism (RFLP), which was initially used alongside Sanger sequences to explore foraminiferal genetic diversity. This approach helped to reveal the distribution of cryptic species across larger biogeographic areas in a cost-effective manner. The third type of genetic data comes from bulk plankton or sediment samples. In this case, DNA was extracted from the entire community and sequenced using a single taxonomically informative genomic region (Burki *et al.*, 2021). These sequences are termed amplicon sequence variants (ASVs), since they represent the genetic profile of the community rather than individual specimens (Fig. 1B). The taxonomic identification of the sequenced community is indirectly acquired by comparing the ASVs with a reference database constructed from the barcoding data (Fig. 1C). Finally, all the genetic data are associated with information related to the collection of the samples (or metadata), to contextualise their geographic origin, date, depth of collection, and physico-chemical environment. Here, we outline the main methodologies employed to compile the observational data set for the planktonic foraminifera. We provide only essential

methodological information below, and a full technical description as online supporting information in Appendix S1.

### (1) Barcoding data set – PFR<sup>2</sup> v.2

We compiled all 18S rDNA sequences of planktonic foraminifera available to date to update the Planktonic Foraminifera Ribosomal Reference database (PFR<sup>2</sup>; Morard *et al.*, 2015). We gathered all sequences released on National Center for Biotechnology Information after constructing the PFR<sup>2</sup> v.1 and harmonised all metadata linked to the new sequences. Briefly, we linked every sequence to a specimen voucher and morphological taxonomic attribution, documented whether the sequences resulted from direct or clonal sequencing, and compiled the geographic coordinates, depth, and date of collection of each specimen whenever the information was available. Finally, we updated the morphological nomenclature of the barcodes following Brummer & Kucera (2022). The resulting database included 7618 Sanger sequences derived from 5598 specimens of planktonic foraminifera collected globally at 650 discrete localities over nearly three decades (Fig. 2A–C) and covered 41 of the 50 morphospecies of planktonic foraminifera. We provide the PFR<sup>2</sup> v.2 database



**Fig. 1.** Methods used to produce genetic data for foraminifera from a plankton sample. (A) Barcoding. Single specimens are morphologically identified, isolated, and Sanger-sequenced to produce a genetic barcode. The sequences associated with the species name of the specimen are made publicly available in data repositories and later organised in dedicated reference databases to achieve internal taxonomic consistency. Genotyping is similar to barcoding except that the observations are only reported in publications and not public databases. (B) Metabarcoding. The entire sample is filtered, its DNA extracted and a short but informative genetic marker is amplified but the organisms occurring in the sample are never observed. The amplified genomic region is then sequenced with high throughput sequencing (HTS) and the raw sequences are assembled into amplicon sequencing variants (ASVs). (C) The taxonomic identity of the ASVs is then determined by comparing them with the reference database constructed with the barcodes, and their similarity is provided with an identity score, where 100% is a perfect match.

as Table S1 and detail the methodological steps for its construction in Appendix S1.

## (2) Genotyping data

Genotyping data were generated as a cost-effective alternative to sequencing in order to determine the genetic identity of a large number of specimens and are reported as discrete observations in publications. We recovered the genotyping and associated metadata of 966 specimens of *Orbulina universa*, 1105 specimens of *Globorotalia truncatulinoides*, 179 specimens of *Globigerinella siphonifera*, 439 specimens of *Globorotalia inflata*, 855 specimens of *Pulleniatina obliquiloculata*, 600 specimens of *Globigerina bulloides*, and 805 specimens of *Neogloboquadrina pachyderma*. All genotyping data and links to original publications are provided in Table S2.

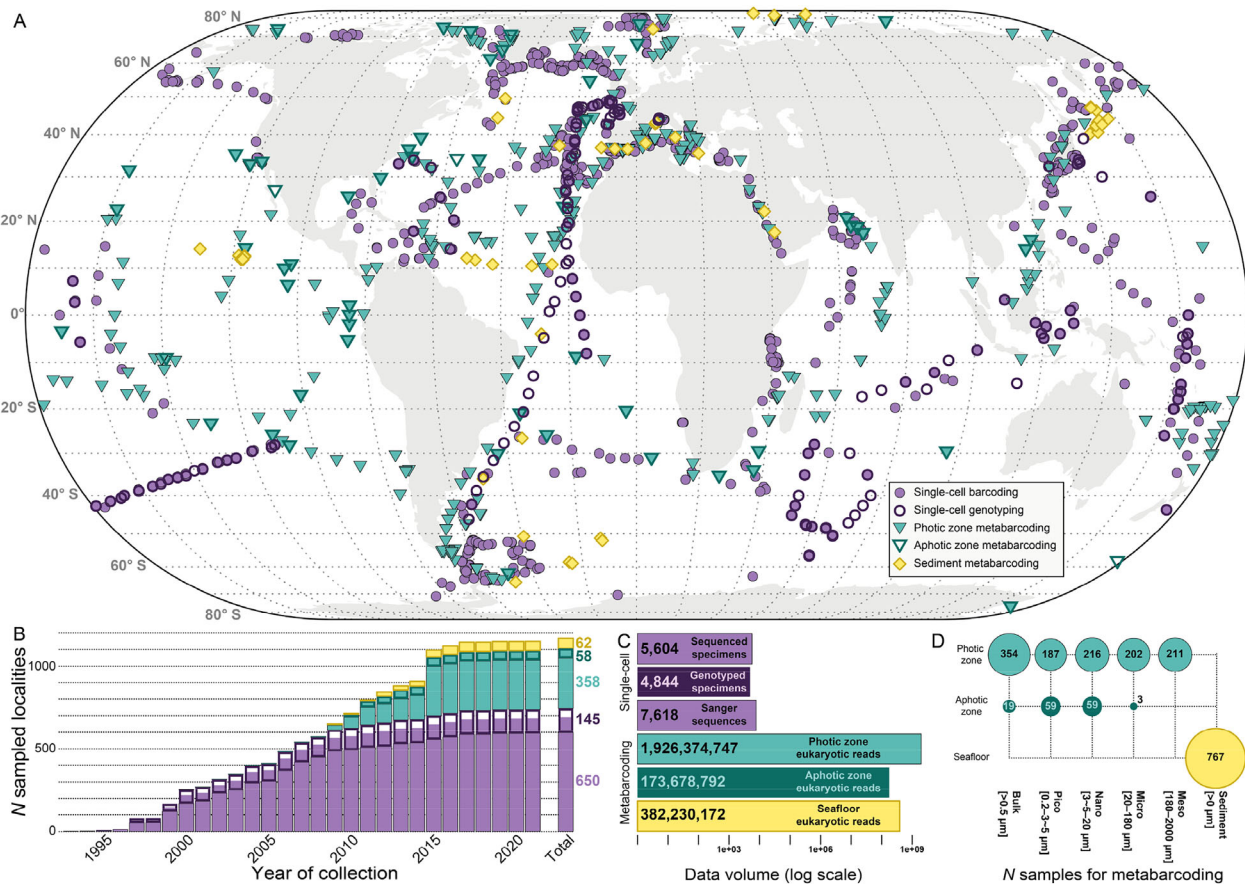
## (3) Metabarcoding data set

We queried the largest metabarcoding data set available to date (Cordier *et al.*, 2022) to retrieve additional foraminifera

genetic data. The data set is a combination of 1716 samples collected at 447 sites with ~2.48 billion reads and ~250,000 ASVs, of the hypervariable V9 region of the 18S rDNA (Fig. 2C). To extract the foraminifera ASVs from the metabarcoding data set, we updated the PR2\_V9 reference database (Henry, de Vargas & Audic, 2019) by integrating the novel sequences of planktonic foraminifera from the PFR<sup>2</sup> v.2 and harmonised the taxonomy of benthic foraminifera sequences. The updated PR2\_V9 database is provided in Table S3. We reassigned the ASVs using the updated PR2\_V9 reference database and initially retained the ASVs with a minimum level of identity of 70% with foraminifera references.

## (4) Metadata standardisation

As a final step of the data set assembly, we harmonised the metadata of the barcoding, genotyping, and metabarcoding data sets to facilitate their combined analysis. This step was necessary because samples collected at the same locations



**Fig. 2.** Sample collections available for exploring the genetic diversity of planktonic foraminifera. (A) World map indicating the location of the 1167 sites sampled for either single-cell barcoding/genotyping or metabarcoding (colour coded by ocean depth). (B) Cumulative number of localities sampled over the years. The overlaps between rectangles with darker borders indicate the simultaneous availability of Sanger sequences and genotyping (purple), and photic and aphotic metabarcoding samples (green) at the same station. (C) Data volume available for single-cell and metabarcoding data sets. (D) Distribution of samples available for metabarcoding data across environments and size fractions from Cordier *et al.* (2022).

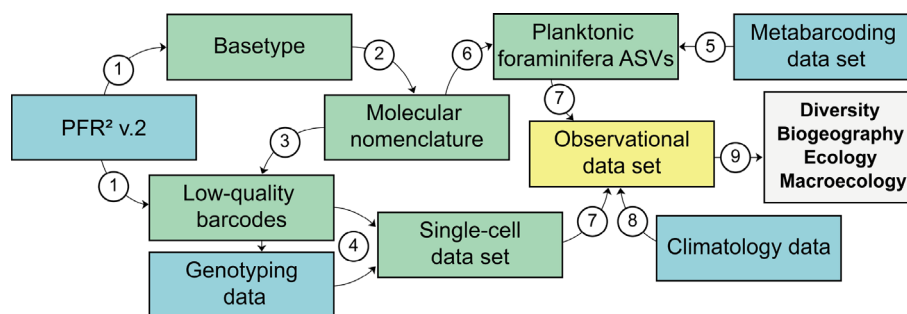
but published in different studies were sometimes associated with slightly different geographic coordinates. We considered all samples that were collected within a radius of 10 km and during the same month as originating from a single locality. Our data set comprises 1167 distinct localities sampled over a timespan of nearly three decades (Fig. 2). Furthermore, we grouped the plankton metabarcoding samples collected between 0 and 200 m depth as ‘photic zone’, those collected below 200 m as ‘aphotic zone’, and those from deep sea sediment samples as ‘seafloor’. We defined five categories for size-fractionated plankton samples: Meso (180–2000  $\mu\text{m}$ ), Micro (20–180  $\mu\text{m}$ ), Nano (3–5–20  $\mu\text{m}$ ), Pico (0.2–3–5  $\mu\text{m}$ ) and Bulk (>0.5  $\mu\text{m}$ ). The sediment samples were not size fractionated. We gathered a suite of climatology data for each sampling site associated with biogeographic classifications based on Spalding *et al.* (2012). All contextual data are provided in Table S4.

### III. DATA STANDARDISATION THROUGH MOLECULAR TAXONOMY

De Queiroz (2007) proposed a comprehensive species concept that defines species as distinct, evolving metapopulations. These metapopulations are interconnected through gene fluxes and undergo changes in morphology, ecology, biogeography, and genetics over time. Understanding the process of evolution, specifically how species develop differentiated traits in morphology, ecology, biogeography, and genetics, is crucial for understanding how diversity emerges and varies in different environments. To quantify this evolutionary process, it is essential to utilise an integrative taxonomic system that combines all available

observations. We designed a hierarchical taxonomic system (Morard *et al.*, 2016) that organises genetic variability using a standardised approach into MOTUs. The fundamental idea behind this hierarchical taxonomic system is to capture the ongoing process of evolution to understand how ‘secondary properties’ of species *sensu* De Queiroz (2007), such as biogeography, ecology or morphology, emerge and are transformed through time.

In this study, our objective is to document the ecology and biogeography of planktonic foraminifera cryptic species, utilising all available evidence up to the present time as detailed above (Fig. 3). The primary aim is to standardise the assessment of genetic variability occurring below the level of morphological species. The most significant challenge is to accommodate the heterogeneity of taxonomic information available in every data source while minimising the potential sources of errors that could lead to taxonomic inflation. To achieve this, we exploit the strength of the barcoding data set, where genetic sequences are linked to individual specimens, thus creating a direct tie between genetic variability and presumed taxonomic identity. Simultaneously, we can also account for intra-genomic variability when a single specimen harbours different rDNA sequences (Weber & Pawlowski, 2014; Greco *et al.*, 2023). However, the barcoding data set coverage of planktonic foraminifera morphospecies is not uniform, and some sequences in the database do not fully cover the entire barcode region. By contrast, the metabarcoding data set offers global coverage and documents the diversity in a sequenced sample comprehensively, rather than being limited to individual specimens. However, metabarcoding sequences are considerably shorter than barcodes [ $\sim 150$  base pairs (bp) *versus*  $\sim 1000$  bp], resulting in reduced taxonomic information. Moreover, the lack of linkage between sequences and singular specimens poses challenges in



**Fig. 3.** Workflow used to organise all data sources within a common taxonomic framework. The publicly available data (blue) are iteratively analysed (green) to construct the observational data set (yellow) that is then used for the formal analysis (grey). (1) The Planktonic Foraminifera Ribosomal Reference database (PFR<sup>2</sup> v.2) is first queried to separate the barcodes that cover entirely the barcode region of the foraminifera and are observed at least three times, referred to as *basetype* (see ‘Molecular taxonomy’ section in Appendix S1 for details), from those that are too short or rarely observed and considered of low quality. (2) The basetypes are used to construct the molecular nomenclature using automated methods of delimitation. (3) Based on the nomenclature, all low-quality barcodes and genotyping data are assigned to the molecular nomenclature and merged into the single-cell data set (4). (5) The planktonic foraminifera amplicon sequence variants (ASVs) are retrieved from the metabarcoding data set, and subsequently analysed (6) to include them in the same taxonomic system as the molecular nomenclature. The single-cell data and ASVs now belonging to the same taxonomic system are merged (7) together with the climatology data (8) and form the observational data set that is analysed in the present study (9).

distinguishing between intra-genomic and inter-specimen genetic variability.

To construct a cohesive and comprehensive taxonomic framework for our genetic data, we developed a stepwise approach centred on our unified molecular nomenclature system. Firstly, we created a molecular taxonomy using high-quality barcoding data, enabling us to assign low-quality barcoding data and genotyping data to the molecular taxonomy to generate the single-cell data set (Steps 1–4 in Fig. 3). Subsequently, we utilised the taxonomic properties from the molecular taxonomy to organise the ASVs within the same taxonomic framework (Steps 5 and 6 in Fig. 3). Next, we merged all genetic data and incorporated climatic data to form our observational data set (Steps 7 and 8 in Fig. 3). This comprehensive data set allowed us to explore the dynamic process of diversification among planktonic foraminifera at the cryptic species level. We outline below the key aspects of this approach (Fig. 3) and provide a fully detailed methodology in Appendix S1.

### (1) Molecular nomenclature

To capture the diversification momentum of planktonic foraminifera, we applied the taxonomic nomenclature system presented in Morard *et al.* (2016) to structure the genetic diversity of planktonic foraminifera into MOTUs. This system organises the genetic variation into three hierarchical levels below the morphospecies level using the standard barcode for foraminifera, located at the 3'-end of the 18S rDNA (Pawlowski & Holzmann, 2014). This fragment of ~1000 bp contains six variable regions including three unique to foraminifera (Pawlowski & Lecroq, 2010) and covers the hyper-variable V9 region used in metabarcoding studies (Cordier *et al.*, 2022), but not the longer V4 fragment. To reduce diversity inflation caused by sequencing errors, we retained only sequences for which the individual hypervariable and conserved regions of the 18S are present at least three times in the database and referred to them as *basetypes*. From the 7618 Sanger sequences derived from 5598 specimens of planktonic foraminifera, we found 1300 sequences matching the criteria, representing 336 distinct basetypes. In addition, we included 17 sequences representative of potential cryptic species described in the literature lacking the desired replication, to avoid excluding potentially genuine variability from our analysis. We used these selected 353 sequences to build the three-level hierarchical nomenclature below the morphospecies level.

We constructed the lowest taxonomical level (MOTUs lvl-3) by grouping together the basetypes co-occurring in one or several individuals into *basegroups*. We obtained 209 basegroups and considered the variability observed within the basegroups to represent the intra-genomic variability and the variability observed among different basegroups to represent the level of inter-population variability.

Planktonic foraminifera are a polyphyletic group (Morard *et al.*, 2022) constituted of three major clades

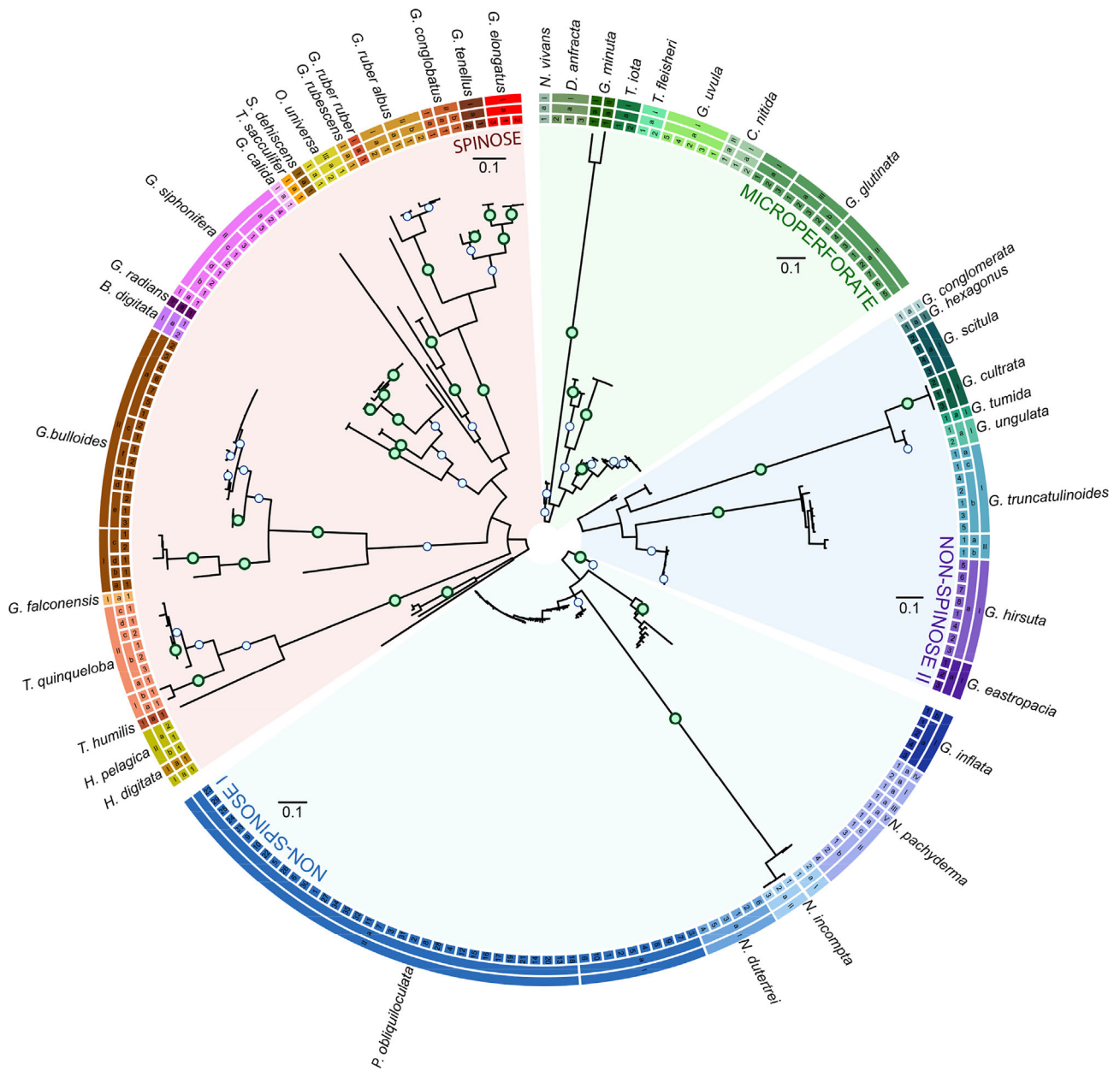
(spinose, non-spinose, and microperforate) and two additional phylogenetically unrelated species: *Neogallitellia vivans* and *Dentigloborotalia anfracta*. We built the molecular nomenclature independently for 80 basetypes of the spinose clade, 206 of the non-spinose clade, and 59 of the microperforate clade. The eight basetypes belonging to *N. vivans* and *D. anfracta* were added to the microperforate alignment because their phylogenetic branch lengths are shorter which is compatible with the evolutionary rate of this clade. We aligned the basetypes of each of the three main clades independently and calculated phylogenetic inferences using maximum likelihood. We submitted the three data sets to two independent automated delimitations methods – Assemble Species by Automatic Partitioning (ASAP; Puillandre, Brouillet & Achaz, 2021) and Poisson Tree Process (PTP; Zhang *et al.*, 2013) – to generate species delimitations (see Figs S1–S3 in Appendix S1), these essentially delimit groups of basetypes into putative (or hypothetical) species. We retained the proposed delimitations as a working hypothesis provided that two basetypes belonging to the same basegroup (and potentially to the same individual) were not attributed to different partitions (oversplitting) and sequences belonging to different morphospecies did not group into the same partition (lumping). We defined the MOTUs lvl-2 as the finest delimitation proposed by ASAP and/or PTP and consider it to represent the level of biological species (or genotype). We defined the MOTUs lvl-1 as the coarsest delimitation proposed by ASAP or PTP and considered it to represent a major disruption in the genetic variability within a given morphospecies organised in monophyletic clusters (lineages) consisting of one or several genotypes. As a result, we obtained a molecular taxonomy where we structured the basetypes belonging to 41 morphospecies into 59 MOTUs lvl-1 and 84 MOTUs lvl-2.

After delineating the MOTUs, we also devised a novel naming scheme or nomenclature. Because there are existing naming schemes for cryptic diversity of morphospecies of planktonic foraminifera in the literature, we maintained as far as possible the continuity between the existing body of knowledge and the present work by preserving the names utilised in the literature. The results of automated classification, resulting molecular nomenclature, and correspondence of all genetic labels found throughout the literature are given in Table S5. The delimitation outcomes and resulting nomenclature are shown in Fig. 4, and the connection with the most commonly used genetic labels in the literature are provided in Table 1. This effort ensures a smooth integration of newly identified MOTUs with previously established taxonomic schemes, facilitating seamless communication and understanding in the scientific community.

### (2) Taxonomic assignment

Only 1,317 sequences of the PFR<sup>2</sup> v.2 database were used to build the molecular taxonomy, and the remaining sequences and genotyping data were assigned to the finest possible level





**Fig. 4.** Composite phylogenetic tree showing the genetic diversity of planktonic foraminifera based on the ~1000 bp 18S rDNA foraminifera barcode only. We used four separate inferences to create a graphical representation of the molecular nomenclature. The non-spinose clade has been split into two parts because of the extreme difference in rates of evolution within this clade. The trees were constructed using only one sequence per basegroup. The trees showing all basetypes and the partitions returned by Poisson Tree Process (PTP) and Assemble Species by Automatic Partitioning (ASAP) are shown as Figs S1–S3. The outer rings indicate the molecular taxonomy based on the automated partitioning, from the outermost to inner rings: MOTUs lvl-1, MOTUs lvl-2 and MOTUs lvl-3, respectively. Each colour represents a single morphological species indicated next to the molecular nomenclature. Each tree support was assessed through 100 pseudoreplicate bootstraps, with maximal branch support indicated by a green circle, and branch support above 90% indicated by a blue circle. MOTU, molecular operational taxonomic unit.

of the nomenclature to maximise the biogeographic coverage. Depending on the taxonomic information available in each sequence, we could assign 4236 sequences to the MOTU lvl-3 molecular taxonomy (including assignments at basetype level), 1851 to the MOTU lvl-2, 1094 to the

MOTU lvl-1 and 437 sequences could not be assigned because they were too short and were not considered further. Finally, we could align the 4844 genotyped specimens to the lvl-2 of our molecular taxonomy using taxonomic equivalence provided in Table S5.

Table 1. Equivalence between existing naming scheme in the literature and the taxonomic scheme built in this work.

Morphospecies	Existing nomenclature presented in	Existing nomenclature	Change in taxonomy	Current nomenclature
<i>Dentigloborotalia anfracta</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
<i>Tenuitellita fleisheri</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
<i>Globigerinita glutinata</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
		IIa	No	IIa
		IIIa	No	IIIa
		IIIb	No	IIIb
<i>Tenuitellita iota</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
<i>Globigerinita minuta</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
		IIa	No	IIa
<i>Candeina nitida</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
	Morard <i>et al.</i> (2019b)	IIa	No	IIa
<i>Globigerinita uwula</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
<i>Neogallitellia vivans</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
<i>Globoquadrina conglomerata</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
<i>Globorotalia cultrata</i>	–	–	–	Ia
<i>Neogloboquadrina dutertrei</i>	Darling <i>et al.</i> (2003)	Ia	Lumping	Ia
		Ib		
		Ic		
<i>Globorotalia eastropacia</i>	–	–	–	Ia
<i>Globorotaloides hexagonus</i>	Morard <i>et al.</i> (2019b)	Ia	No	Ia
<i>Globorotalia hirsuta</i>	–	–	–	Ia
<i>Neogloboquadrina incompta</i>	Darling <i>et al.</i> (2000)	I	No	Ia
		II	No	IIa
<i>Globorotalia inflata</i>	Morard <i>et al.</i> (2011)	I	Nomenclature	Ia
	Morard <i>et al.</i> (2011)	II	Nomenclature	Ib
<i>Pulleniatina obliquiloculata</i>	Ujiié <i>et al.</i> (2012)	Ia	No	Ia
	Ujiié <i>et al.</i> (2012)	IIa	Lumping	IIa
	Ujiié <i>et al.</i> (2012)	IIb		
<i>Neogloboquadrina pachyderma</i>	Darling <i>et al.</i> (2004)	I	No	Ia
	Darling <i>et al.</i> (2007)	II	Nomenclature	IIa
	Darling <i>et al.</i> (2007)	III	Lumping	IIb
	Darling <i>et al.</i> (2007)	V		
	Darling <i>et al.</i> (2007)	VI	Nomenclature	IIc
	Darling <i>et al.</i> (2007)	IV	No	IVa
	Darling <i>et al.</i> (2007)	VII	Nomenclature	IIIa
	Darling <i>et al.</i> (2017)	VIII	Nomenclature	Va
<i>Globorotalia scitula</i>	–	–	–	Ia
<i>Globorotalia truncatulinoidea</i>	de Vargas <i>et al.</i> (2001)	I	Nomenclature	Ia
	de Vargas <i>et al.</i> (2001)	II	Nomenclature	Ib
	Quillévére <i>et al.</i> (2013)	V	Nomenclature	Ic
	de Vargas <i>et al.</i> (2001)	III	Nomenclature	IIa
	de Vargas <i>et al.</i> (2001)	IV	Nomenclature	IIb
<i>Globorotalia tumida</i>	–	–	–	Ia
<i>Globigerina bulloides</i>	Darling <i>et al.</i> (2000, 2003)	Ia	No	Ia
	Kucera & Darling (2002); Darling <i>et al.</i> (2003)	Ib	No	Ib
	Darling <i>et al.</i> (2017)	Ic	No	Ic
	Darling <i>et al.</i> (2017)	Id	No	Id
	Darling <i>et al.</i> (2000)	IIa	No	IIa
	Darling <i>et al.</i> (2000)	IIb	No	IIb
	Darling <i>et al.</i> (2000)	IIc	No	IIc
	Darling <i>et al.</i> (2000)	IIId	No	IIId
	Darling <i>et al.</i> (2007)	IIe	No	IIe
	Darling <i>et al.</i> (2017)	IIIf	No	IIIf
<i>Globigerinella calida</i>	Weiner <i>et al.</i> (2014)	IIIa	Nomenclature	Ia
<i>Globigerinoides conglobatus</i>	Morard <i>et al.</i> (2019a)	Ia1	No	Ia
	Morard <i>et al.</i> (2019a)	Ib1	No	Ib
<i>Sphaeroidinella dehiscens</i>	–	–	–	Ia
<i>Beella digitata</i>	–	–	–	Ia

(Continues on next page)

Table 1. (Cont.)

Morphospecies	Existing nomenclature presented in	Existing nomenclature	Change in taxonomy	Current nomenclature
<i>Hastigerinella digitata</i>	–	–	–	Ia
<i>Globigerinoides elongatus</i>	Morard <i>et al.</i> (2019a)	Ia1	No	Ia
<i>Globigerina falconensis</i>	–	–	–	Ia
<i>Turborotalita humilis</i>	–	–	–	Ia
<i>Hastigerina pelagica</i>	Weiner <i>et al.</i> (2012)	I	No	Ia
	Weiner <i>et al.</i> (2012)	IIa	No	IIa
	Weiner <i>et al.</i> (2012)	IIb	No	IIb
<i>Turborotalita quinqueloba</i>	Darling <i>et al.</i> (2000); Darling & Wade (2008)	Ia	No	Ia
	Darling & Wade (2008)	Ib	No	Ib
	Darling <i>et al.</i> (2000)	IIa	No	IIa
	Darling <i>et al.</i> (2000)	IIb	Lumping	IIb
	Darling <i>et al.</i> (2000)	IIc		
	Darling <i>et al.</i> (2000)	IId	No	IId
	Darling <i>et al.</i> (2000)	IIe	Nomenclature	IIc
	Darling <i>et al.</i> (2017)	Ia	No	Ia
<i>Globigerinella radians</i>	Weiner <i>et al.</i> (2014)	Ia	No	Ia
<i>Globigerinoides ruber albus</i>	Morard <i>et al.</i> (2019a)	Ia	No	Ia
		Ib	Nomenclature	IIb
		Ic	Nomenclature	IIa
<i>Globigerinoides ruber ruber</i>	Morard <i>et al.</i> (2019a)	Ia1	No	Ia
<i>Globoturbotalita rubescens</i>	Morard <i>et al.</i> (2019a)	Ia1	No	Ia
<i>Trilobatus sacculifer</i>	–	–	–	Ia
<i>Globigerinella siphonifera</i>	Weiner <i>et al.</i> (2014)	Ib	Nomenclature	Ia
		IIa	Splitting	IIa
				IIc
				IId
		IIb	No	IIb
<i>Globigerinella radians</i>	Weiner <i>et al.</i> (2015)	<i>G. siphonifera</i> Ia	Nomenclature	Ia
<i>Globigerinella calida</i>	Weiner <i>et al.</i> (2015)	<i>G. siphonifera</i> III	Nomenclature	Ia
<i>Globigerinoides tenellus</i>	Morard <i>et al.</i> (2019a)	Ia1	No	Ia
<i>Orbulina universa</i>	de Vargas <i>et al.</i> (1999)	I	Nomenclature	Ia
		III	Nomenclature	IIIa

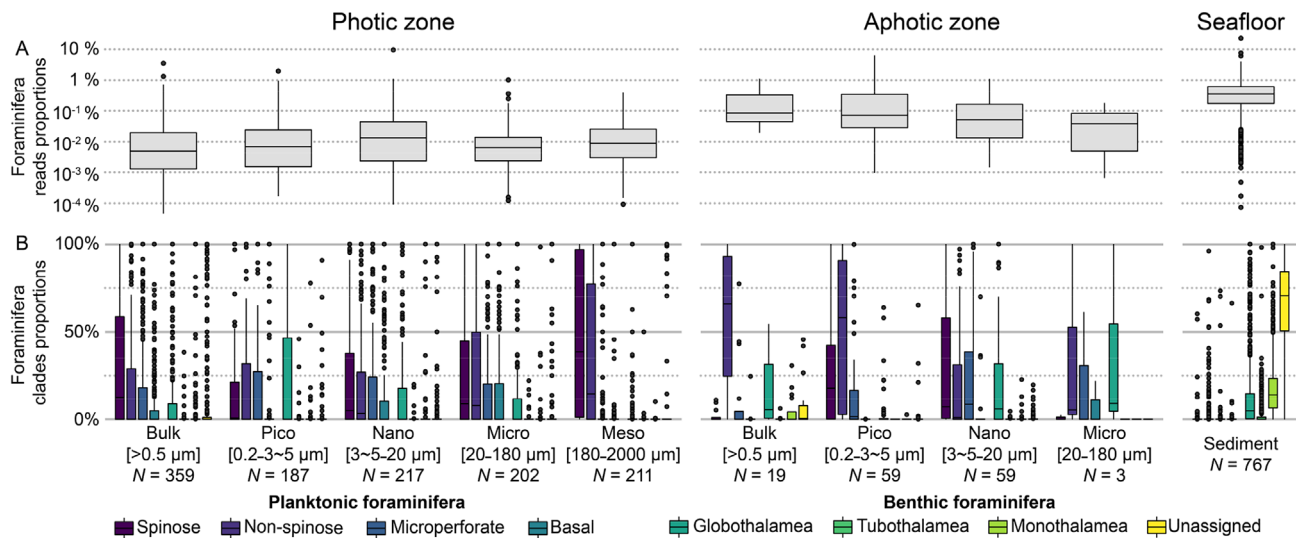
### (3) Metabarcoding data set

The metabarcoding data set represents the richest source of information, but its accurate interpretation relies on our ability to parse reliably the ASV diversity (see Fig. S4A in Appendix S1). In our case, categorising ASVs as either planktonic or benthic foraminifera based solely on their occurrence in the water column or sediments is not possible. This is because planktonic sequences can be found in the benthos due to the sequestration of organic matter from the surface ocean to its base (Morard *et al.*, 2017; Barrenechea Angeles *et al.*, 2020; Cordier *et al.*, 2022), and also because benthic foraminifera of the clade Globothalamea actively disperse into the plankton (Morard *et al.*, 2022). Moreover, there is no unique genetic gap separating benthic and planktonic foraminifera due to the variability in their evolutionary rates (Darling *et al.*, 1997; de Vargas & Pawlowski, 1998).

Consequently, adopting strict fixed thresholds would impede our ability to detect genuine planktonic foraminifera diversity not covered by our reference database, while permissive thresholds would lead to an overestimation of planktonic foraminifera diversity. To address this challenge, we designed a flexible approach and determined a ‘floating’ threshold by determining the distance gap between each of

the planktonic foraminifera morphospecies references and the benthic barcodes in the PR2\_V9 reference database (Fig. S4A in Appendix S1). This allowed us to assign 257 ASVs confidently to planktonic foraminifera out of the initial 13,430 potential foraminifera ASVs.

To ensure that we did not miss any planktonic foraminifera diversity, we assigned the rest of the ASVs to the main benthic clades to visualise their allocation in the three main sampled environments (photic zone, aphotic zone, and seafloor), and size classes. We determined a fixed threshold of 90% to consider an ASV belonging to either Monothalamea, Tubothalamea, or Globothalamea clades (Fig. S4B in Appendix S1) or unassigned if the ASVs were attributed to references with unresolved taxonomic attribution. We then considered all ASVs with an identity of 80–90% as unassigned foraminifera, and those below 80% as ‘other eukaryotes’. As a result, we categorised 1566 ASVs to the Monothalamea, 466 to the Tubothalamea, 1139 to the Globothalamea, and 6051 as unassigned foraminifera (Fig. S4C, D in Appendix S1). On this basis, we could evaluate the contribution of foraminifera to the metabarcoding data set compared to other eukaryotes (Fig. 5A), and the relative contribution of each foraminifera clade in every environment and size



**Fig. 5.** Distribution of foraminifera reads across depths and size fractions in the metabarcoding data set. (A) Proportions of foraminifera reads among all eukaryotes sequenced in the metabarcoding data set. (B) Relative contribution of each foraminifera clade within the foraminifera reads. Boxplots display the interquartile range (box), extrema (whiskers), and outliers (dots).

fraction (Fig. 5B) to ensure we correctly allocated foraminifera diversity using a molecular-taxonomy-only approach.

The results showed that the contribution of foraminifera reads in the metabarcoding data sets varied between 0.001 and 0.1% in the photic zone, between 0.01 and 0.1% in the aphotic zone, and between 0.1 and 1% in the sediment. The planktonic clades are dominant in the photic and aphotic zones and rarely detected in the sediment, whereas the benthic Monothalamea and unclassified foraminifera are abundant in the sediment samples only. The Globothalamea clade is well represented in all environments (except for the largest size fraction in the photic zone) and the Tubothalamea are rare in all environments. The ‘unassigned foraminifera’, which is by far the most diversified category (Fig. S4C in Appendix S1), is abundant only in the sediment, suggesting that the vast majority of the ASVs of this ‘waste-basket’ category belong to benthic foraminifera clades. We are therefore confident that our approach successfully identified the planktonic foraminifera ASVs that constitute the majority of the data volume in the photic and aphotic zones, as they logically should.

Henceforth, we narrow our focus to the 257 ASVs allocated to planktonic foraminifera, accounting for ~1 million reads from the entire metabarcoding data set. These ASVs exhibit varying degrees of similarity to the documented diversity in the reference database, ranging from identical to increasingly dissimilar ones, representing MOTUs that may belong to morphospecies not captured through single-cell sampling. Our goal is to incorporate these ASVs into our molecular taxonomy, making them comparable to other data sources. However, this presents several challenges. Firstly, we must consider that metabarcoding data are not immune to technical biases, implying that not all ASVs may accurately represent genuine biological diversity. Secondly, there are no distinct barcode gaps between successive

MOTUs or even between morphospecies, despite the fact that we established the molecular taxonomy with a consistent methodology across foraminifera clades. Lastly, the taxonomic information available in the V9 hypervariable region is limited and requires cautious interpretation. To address these issues systematically, we have adopted a conservative and stepwise approach, aligned with our methodology for single-cell data processing.

Therefore, from the initial 257 ASVs we discarded 112 ASVs that occurred in a single sample because they could result from potential sequencing artefacts (Flegontova, Lukeš & Horák, 2023). Considering the extensiveness of the metabarcoding data set and the dispersal ability of planktonic foraminifera, we deemed ASVs found in only one sample unlikely to represent genuine diversity. After this filtering, we assigned the remaining 145 ASVs to our molecular nomenclature. To determine the relationships between these ASVs and known molecular taxonomy levels, we utilised the V9 fragment of the basetypes and calculated barcode gaps for various taxonomic levels, including identical, intra-MOTUs lvl-2, inter-MOTUs lvl-2, and inter-MOTUs lvl-1, to consider the potential genetic diversity not covered within morphospecies in our data set. In order to consider potentially missing morphospecies, even distantly related to those in the barcoding data set, we calculated the inter-morphospecies distances from the same genus (when applicable), and inter-morphospecies distances from different genera for each morphospecies. We used patristic distances and identity level (see Figs S5–S7 in Appendix S1) for this purpose.

Out of the 145 ASVs, 72 were found to be identical to existing basetypes, and 35 were in the range of intra-MOTUs lvl-2, representing known MOTUs, which we incorporated into the existing nomenclature. However, 24 ASVs showed marked differences from all basetypes, thus representing genuine diversity not covered by our reference database.

Within this group, two ASVs could represent two novel MOTUs lvl-2: *Neogloboquadrina pachyderma* (named IId) and *Hastigerina pelagica* (named IIc). Additionally, there were 12 ASVs potentially representing novel MOTUs lvl-1, one novel MOTU lvl-1 of *Turborotalita quinqueloba* (named III), one novel MOTU of *Neogloboquadrina pachyderma* (named VI), and one novel MOTU lvl-1 of *Orbulina universa*. For the latter, since our basetype collection only covered two out of three cryptic species described by De Vargas *et al.* (1999), I and III, we assumed the single distant ASV represented the missing MOTU lvl-1 in our data set and named it *Orbulina universa* IIa. Furthermore, we identified 10 ASVs which could belong to potential morphospecies not present in our reference database. As we lacked a direct link with morphology, we refer to them as ‘candidate morphospecies’. Among these ASVs, we found a cluster of four ASVs forming a sister branch to *Turborotalita quinqueloba*, further divided into two subclusters. Since three species are described in the genus *Turborotalita* and two are covered by our barcoding data set (*T. quinqueloba* and *T. humilis*), we considered the four ASVs as candidates for the third morphospecies of the genus: *T. clarkei* composed of two MOTUs lvl-1 (I and II). Additionally, five ASVs related to the *Tenuitellita–Globigerinita* wulfa plexus of the microperforate clade formed two separate subclusters and were too distinct from any basetype to belong to any known barcoded morphospecies. Since the genus *Tenuitella* includes three morphospecies and only two were covered by our barcoding efforts (*T. iota* and *T. fleisheri*), we treated the four closest ASVs as candidates for the third morphospecies, named *T. parkerae*, and organised them into a single MOTU lvl-1. We also considered the remaining single ASV as an unknown microperforate since all morphospecies of this clade were covered by the barcoding effort except for *T. parkerae*. Finally, one ASV was identified as a sister to *Beella digitata* and the genus *Globigerinella*, which we designated as *Globigerinella* sp. since it could not be attributed to any morphotaxon lacking barcoding coverage, as we did for *T. parkerae* and *T. clarkei*.

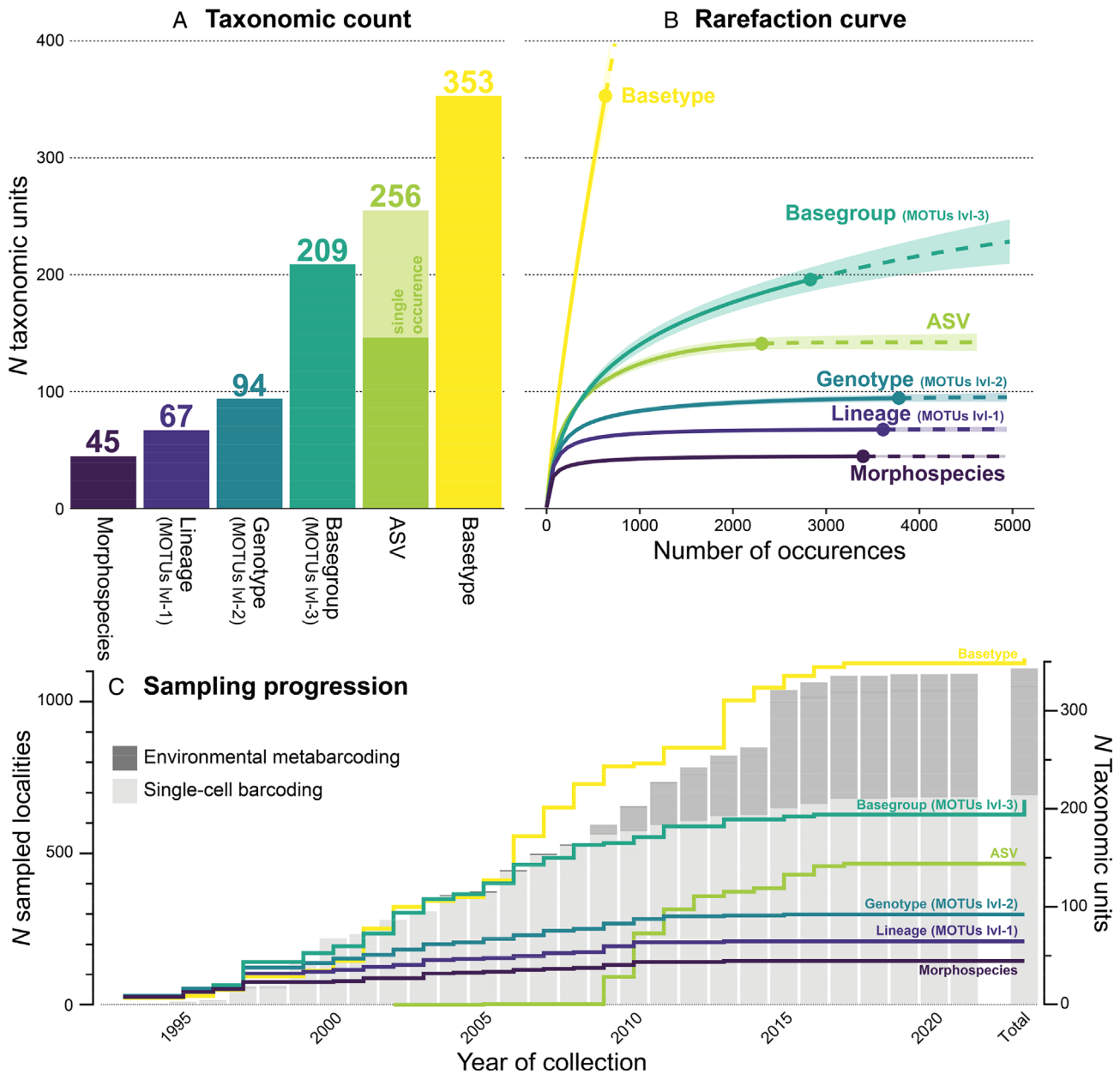
Finally, we identified 14 ASVs that were difficult to align in our molecular nomenclature. This could be (i) because they are either attributed to morphospecies with too few basetypes to interpret the ASVs diversity reliably, as it could either represent intra-genomic variability or novel MOTUs, (ii) because in some species we could not identify a clear barcode gap between successive levels of the nomenclature, or (iii) because in some cases the V9 fragment is identical between closely related morphospecies. To avoid overinterpretation of the metabarcoding data set, we either allocated the ASVs to the lowermost level of the molecular taxonomy or discarded them. As a result, we attributed two ASVs to existing MOTUs lvl-2, three ASVs to existing MOTUs lvl-1, six ASVs to morphospecies level and three ASVs were discarded because of the lack of resolution in the gene marker. We provide the classification of the ASVs in Table S6. Finally, we merged the occurrences of planktonic foraminifera MOTUs of the Sanger, genotyping and metabarcoding data sets into Table S7.

## IV. DIVERSITY PATTERNS

### (1) Global diversity

We enumerated 94 MOTUs lvl-2 and 67 MOTUs lvl-1 from the 45 analysed morphospecies (Fig. 6A). To assess the likelihood that we captured the entire planktonic foraminifera diversity, we ran rarefaction analyses on the entire data set at each taxonomic level of our molecular nomenclature, and also for the basetypes and ASVs that acted as data sources in our approach. The rarefaction analyses indicated that diversity at the level of morphospecies, MOTUs lvl-1 and lvl-2, as well as for the ASVs has been comprehensively sampled (Fig. 6B). These patterns may partly result from our stringent approach to base our molecular taxonomy on the most robust observations of the Sanger data set (albeit with some exceptions) and our exclusion of ASVs occurring in single samples. Unlike *T. clarkei* and *T. parkerae* that were presumably detected despite our lack of barcodes, we could not detect five morphologically defined taxa (*Globorotalia crassaformis*, *Orcadia riedeli*, *Berggrenia pumilio*, *Globigerinella adamsi* and *Globorotaloides oveyi*; Brummer & Kucera, 2022). There are three possibilities that could explain why these morphospecies were not detected in our data set: (i) they were removed because of sequence similarity to benthic foraminifera, (ii) they lacked taxonomic resolution of the metabarcoded region, or (iii) their sequences were discarded as too divergent because of the similarity thresholds applied. Therefore, we cannot claim that the entire diversity of planktonic foraminifera has been captured here, but we are reasonably confident that we quantified the magnitude of their taxonomic richness correctly, at least to MOTU lvl-2. The rarefaction curve did not saturate for the taxonomic levels MOTU lvl-3 and basetypes indicating that they have been largely undersampled. This is due to the pervasive intra-genomic variability, especially within the microperforate and non-spinose clades (Greco *et al.*, 2023), which generates a large number of distinct gene copies, but which does not represent evolutionarily meaningful biological diversity. Our strategy to group co-occurring copies into MOTUs lvl-3 partly mitigates taxonomic inflation, but it might be insufficient in some taxa such as *Pulleniatina obliquiloculata* with 43 MOTUs lvl-3 for only two MOTUs lvl-2.

In addition to the rarefaction analyses, we plotted the pace of discovery of MOTUs over the course of the exploration of the genetic diversity of planktonic foraminifera (Fig. 6C). During the initial phase of exploration based on single-cell barcoding (1993–2010), the discovery of novel biological diversity (MOTUs lvl-1 and lvl-2) kept pace with the sequencing of basetypes. However, this pace of discovery slowed and eventually halted, despite continuous single-cell sequencing efforts. Similarly, when sampling for metabarcoding analyses started in 2009 with the TARA Oceans expedition (Karsenti *et al.*, 2011), it only revealed a modest number of additional MOTUs at lvl-1 and lvl-2. This indicates that metabarcoding mostly re-sampled the diversity that had already been sequenced and described by the



**Fig. 6.** Global diversity of planktonic foraminifera. (A) Number of taxa detected in our global data set for each taxonomic level. ASV, amplicon sequence variant; MOTU, molecular operational taxonomic unit. (B) Rarefaction analysis shows the sampling effort of each taxonomic level against the number of occurrences. The occurrences are observations of MOTUs in independent samples (where several samples can originate from the same locality) with either barcoding or metabarcoding. (C) The bar plots in the background represent the cumulative number of sampled localities during the three decades of exploration of the molecular diversity of planktonic foraminifera. The solid lines represent the cumulative number of taxonomic units sampled through time. We provide a ‘Total’ column on the  $x$ -axis because a few barcode samples did not have information about their collection date.

single-cell approach. Nonetheless, it did uncover diversity in either lightly or non-barcoded morphospecies and expanded the biogeographic coverage of many morphospecies.

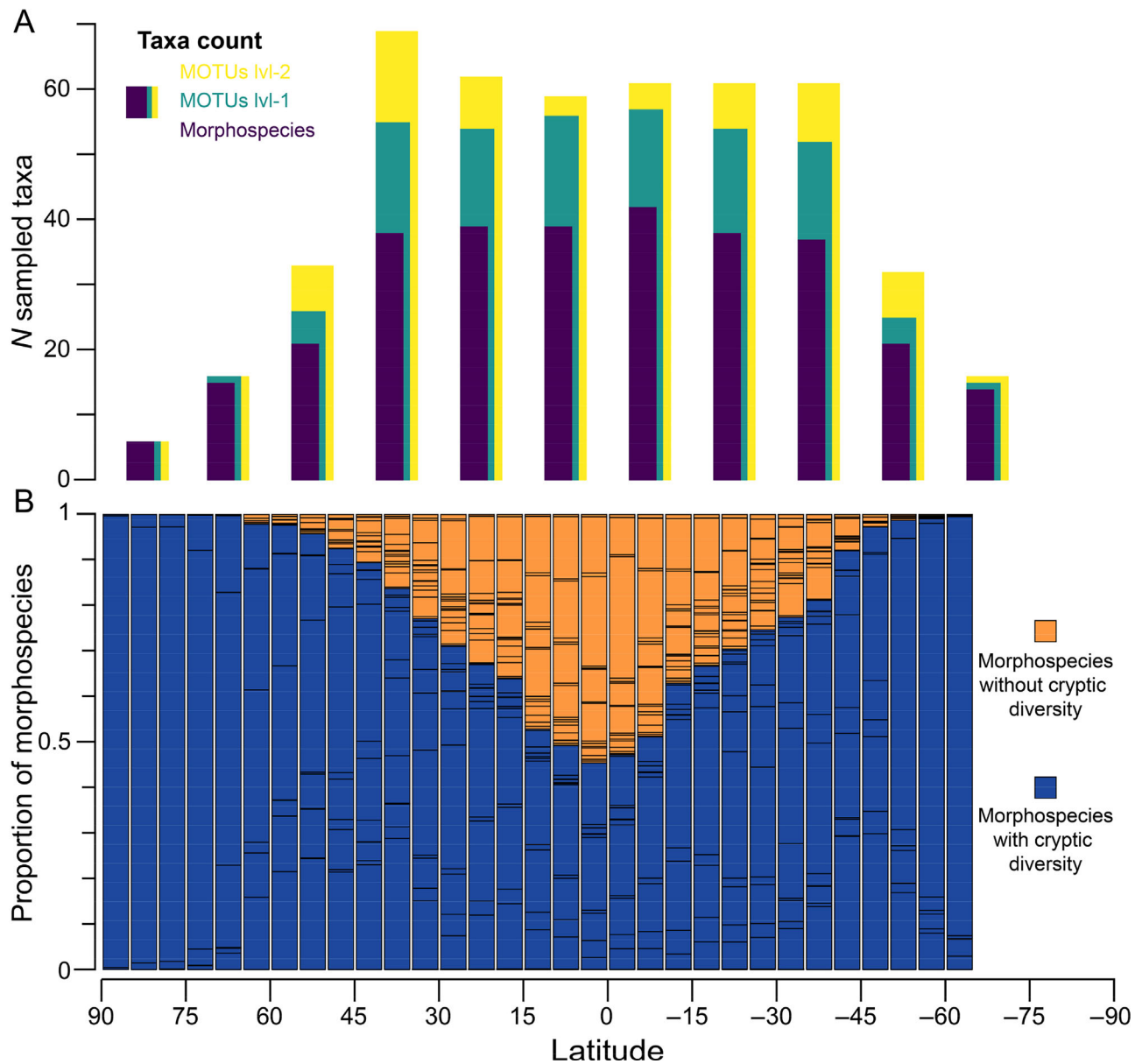
## (2) Latitudinal gradient of cryptic diversity

The latitudinal diversity gradient is a well-known macroecological pattern observed in land and marine ecosystems

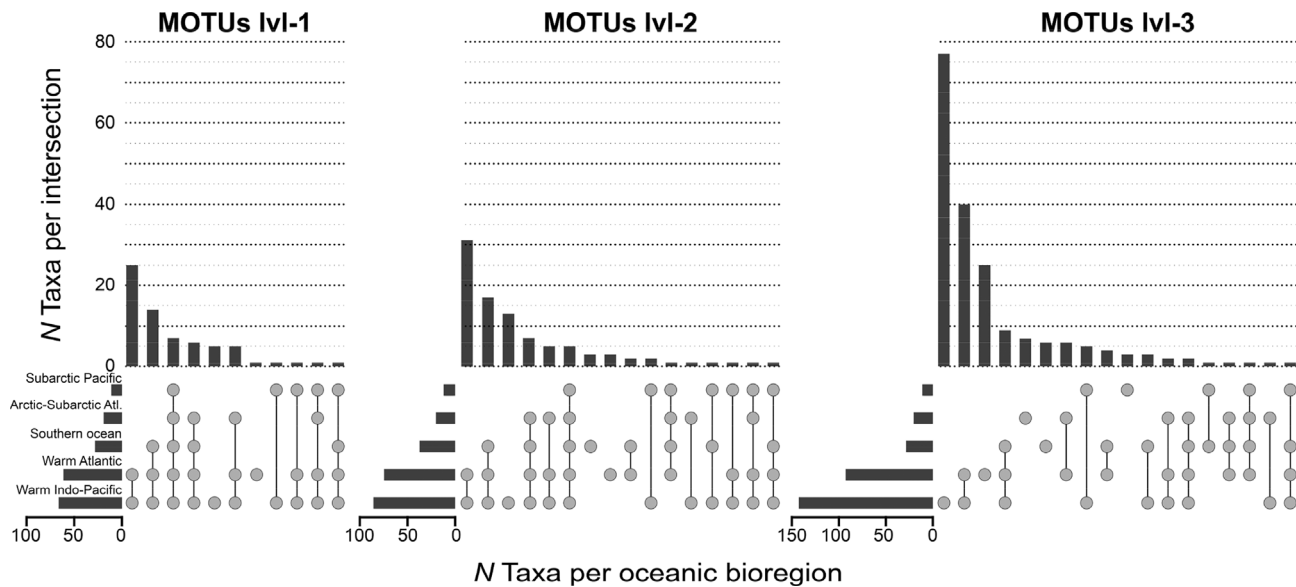
(Hillebrand, 2004), it is characterised by a decline of biodiversity from low to high latitude. The pattern has been studied intensively in planktonic foraminifera thanks to the global coverage of census counts in surface sediment data (Rillo, Woolley & Hillebrand, 2022). The study of the fossil record of planktonic foraminifera showed that this gradient started to emerge in the last 15 million years (Fenton *et al.*, 2023), and the tropical dip in diversity emerged during

the transition from the last glacial maximum to the Holocene from 19 to ~11.5 thousand years ago (Yasuhara *et al.*, 2020; Strack *et al.*, 2022). Our collection also reveals the latitudinal diversity gradient, with a broad maximum in diversity spanning from 45° S to 45° N latitude (Fig. 7A). Genetic diversity shows a steeper poleward gradient in comparison to morphospecies counts. This steeper gradient indicates that cryptic species are more endemic with narrower geographic

ranges than morphospecies as also evidenced in a terrestrial context (Eme *et al.*, 2018). Additionally, we note a marked peak in diversity for MOTUs lvl-2 in the 30–45° latitude range of the Northern Hemisphere. This is due to the presence of endemic MOTUs in the Warm Atlantic and Warm Indo-Pacific, contributing to an overall increase in diversity compared to the Southern Hemisphere, where no physical boundaries separate the oceanic basins (Fig. 8).



**Fig. 7.** (A) Latitudinal gradient of cryptic diversity. The vertical bars represent the number of taxa detected worldwide within 15° latitudinal increments at each of the taxonomic levels. Each morphospecies has a minimum of one MOTU lvl-1 and lvl-2, and different bar thickness is used to facilitate the visualisation of the diversity overlap between the successive levels. (B) Proportion of morphospecies in surface sediments that harbour at least two MOTUs lvl-2 independent of the place of occurrence of these MOTUs lvl-2 (blue), against those where only a single MOTUs lvl-2 has been observed (orange) within 5° latitudinal increments. The proportions were calculated based on the proportion of foraminifera morphospecies provided in the ForCenS database (Siccha & Kucera, 2017). The black horizontal bars represent the limits between individual morphospecies. MOTU, molecular operational taxonomic unit.



**Fig. 8.** UpSet plots showing the compartmentalisation of biodiversity by geography at the three levels of the molecular taxonomy. On each UpSet plot, the left bottom panel shows the counts of molecular operational taxonomic units (MOTUs) in each geographic region, the connected circles show the intersection between the categories, and the top panel the number of MOTUs in the intersections. The partitioning of the geography is based on a simplified biogeographic classification of Spalding *et al.* (2012) where we aggregated the North Atlantic, Mediterranean sea and South Atlantic as ‘Warm Atlantic’ and the Indian Ocean, North Pacific and South Pacific as ‘Warm Indo-Pacific’. The occurrences of MOTUs lvl-2 in each basin is detailed in Figs 9–11.

Next, we evaluated the contribution of morphospecies with cryptic diversity to recent fossil assemblages. The distribution of planktonic foraminifera is richly documented in surface sediments where their shells are preserved and the relative contribution of morphospecies has been documented (Siccha & Kucera, 2017). We calculated the relative contribution of morphospecies harbouring more than one MOTU lvl-2 (thus having cryptic diversity) against those displaying only a single MOTU lvl-2 in surface sediment assemblages (Fig. 7B). We observe that at least 50% of the assemblages are constituted by morphospecies with cryptic diversity in the tropics, and their contribution increases poleward until 60° latitude where almost the whole assemblages are constituted of morphospecies with cryptic diversity.

### (3) Compartmentalisation of biodiversity

Following an analysis of the broader patterns of diversity across latitudinal ranges, we explored how the biodiversity of foraminifera is compartmentalised across different geographical regions (Fig. 8). The regions with warmer waters exhibit the highest levels of diversity, with slightly more MOTUs lvl-1 and lvl-2 identified in the Warm Indo-Pacific region compared to the Warm Atlantic. These two regions also display the highest number of shared MOTUs lvl-1 and lvl-2. However, we identified a disproportionately elevated number of MOTUs level-3 occurring only in the Warm Indo-Pacific that we attribute to the morphospecies *Pulleniatina obliquiloculata* that has a particularly high diversity at this taxonomic level (Fig. 4) and has been extensively

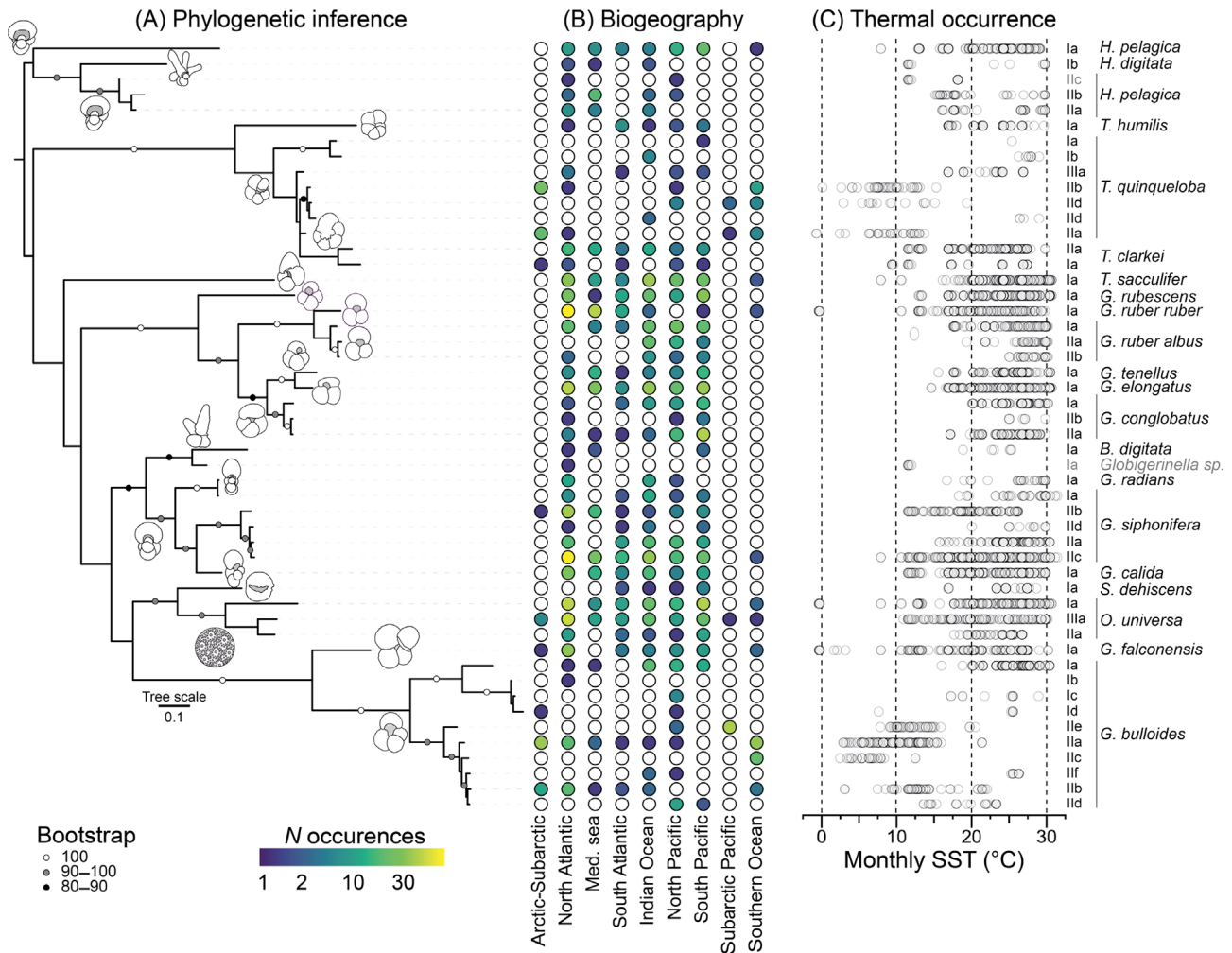
sampled in the Indo-Pacific (Ujiié *et al.*, 2012). The diversity at MOTU lvl-3, especially for *P. obliquiloculata*, may be inflated due to elevated intra-genomic variability, and does not necessarily reflect genuinely genetically distinct populations. However, we note that although the diversity at MOTUs lvl-3 should be interpreted with caution, the shape of the diversity allocation is similar to that at MOTUs lvl-1 and lvl-2. The Arctic–Subarctic Atlantic exhibit lower levels of diversity compared to the colder Southern Ocean water masses, and the latter share more MOTUs at all levels with the warm water environments than with the northern cold waters. We could identify only five MOTUs lvl-2 occurring in all regions, and none at MOTUs lvl-3.

## V. BIOGEOGRAPHY AND ECOLOGY

In this section, we detail the prominent biogeographic trends within cryptic diversity among planktonic foraminifera. Our objective is to elucidate the distinctions in biogeography and environmental preferences among MOTUs belonging to the same morphospecies. We constructed phylogenies for MOTUs lvl-2 for each clade and summarised their biogeographic distributions and thermal preferences (Figs 9–11). We attempted to quantify the ecological differentiation between MOTUs belonging to the same morphospecies (Fig. 12), and generated maps for each MOTU (Appendix S2).

Despite identifying 94 MOTUs lvl-2 within our data set, we found only 16 morphospecies exhibiting multiple





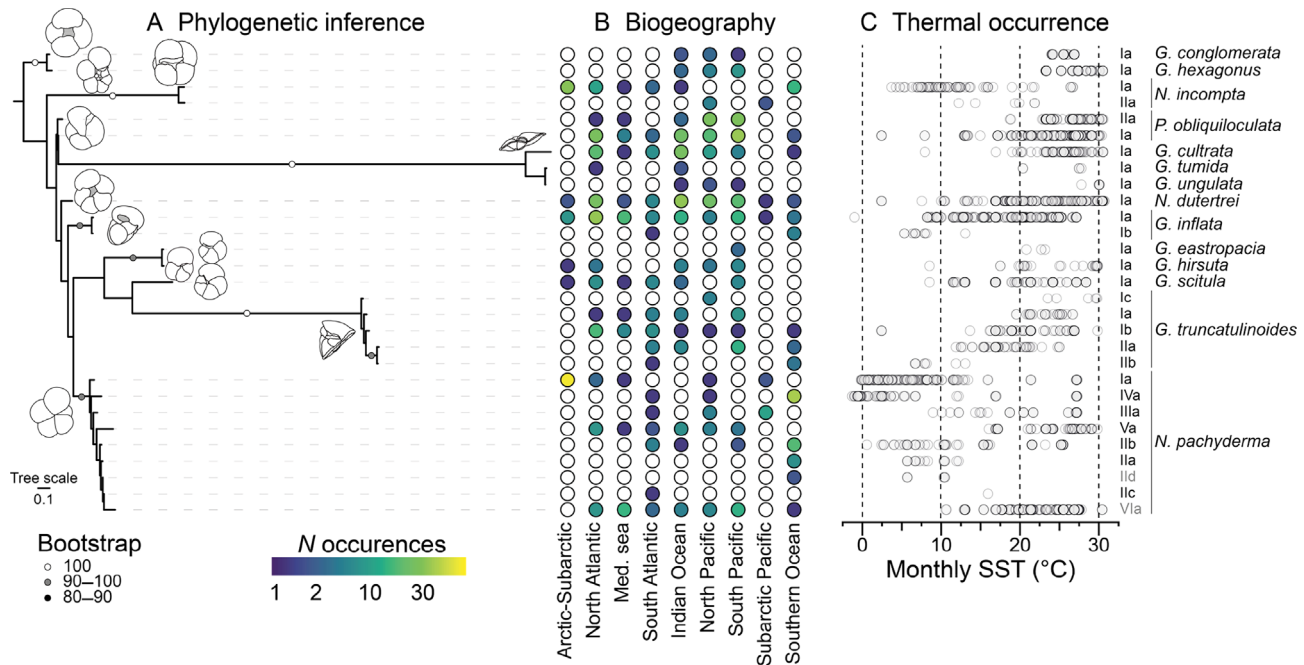
**Fig. 9.** Diversity, biogeography and thermal niches of MOTUs lvl-2 of the spinose clade. MOTU, molecular operational taxonomic unit. (A) Phylogenetic inference of the MOTUs lvl-2 of the spinose clade. A sketch of each morphospecies is provided next to the branches, and the names of the MOTUs are indicated on the right side of the figure. The names of the MOTU or morphospecies is in grey when only an amplicon sequence variable (ASV) is available in our data set. Because we used basetypes and ASV sequences to calculate the inference, we stress that the topology of the tree should not be used to interpret the phylogenetic relationships among genera, and is used here only as a way to display the documented biological diversity. (B) Occurrences of MOTUs in specific biogeographic regions. Open symbols indicate when the given MOTUs have not been detected in a specific biogeographic region. (C) Thermal occurrence of the MOTUs based on the averaged sea surface temperature (SST) of the month and localities at which the MOTUs have been collected. These temperatures do not necessarily represent the thermal optimum of the MOTUs because the depth of collection is not considered here. Note that the colours in B and the density of the symbols in C also reflect sampling effort.

MOTUs at lvl-2. This shows an uneven distribution of cryptic diversity among morphospecies of planktonic foraminifera. Although we acknowledge again that our data set may not encompass all existing MOTUs due to the absence of a few morphospecies, we believe the sampling effort is sufficient to outline general trends in cryptic diversity, represented by the distinction among lvl-2 MOTUs occurring within an individual morphospecies. Since detailing the occurrence patterns of all MOTUs would be too extensive, we identified two main ‘regimes’ of cryptic diversity. The first is for the morphospecies occurring at low latitudes and the second is for those occurring in mid to high latitudes which

we present separately. The most prominent features of each regime are highlighted below.

**(1) Cryptic diversity patterns of low-latitude morphospecies**

Here we focus on morphospecies with multiple MOTUs primarily occurring in tropical and subtropical environments. A significant overlap occurs in the ecological niches of the MOTUs of *G. siphonifera*, *G. glutinata*, *G. conglobatus*, *G. ruber albus*, *O. universa*, *H. pelagica*, and *P. obliquiloculata* (Fig. 12). Only *C. nitida* and *T. clarkei* exhibit differentiated environmental preferences



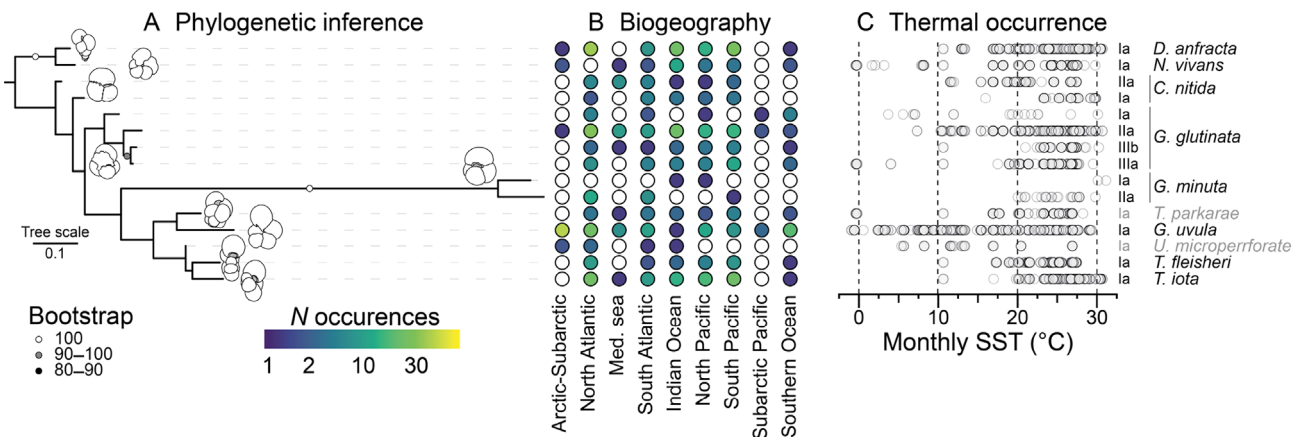
**Fig. 10.** Diversity, biogeography and thermal niches of MOTUs lvl-2 of the non-spinose clade. Other details are as in Fig. 9.

between their MOTUs lvl-2. The prevailing lack of pronounced differentiation in environmental preferences among these MOTUs is mirrored in their geographical distribution. The MOTUs belonging to the tropical and subtropical morphospecies mentioned above occur in warm waters of both the Atlantic and the Indo-Pacific regions. There are some rare exceptions to this pattern, such as *G. siphonifera* IIc, *G. ruber albus* IIa, and *P. obliquiloculata* IIa, which are either scarce or absent in the Atlantic region. Despite these exceptions, our analysis seems to indicate a lack of environmental or geographical divergence among MOTUs belonging to the same morphospecies. This could be caused by our approach that lacks the resolution required locally to unveil subtle ecological differences between MOTUs. For instance, the MOTUs IIa and IIb of *H. pelagica*, occurring at the same location, exhibit a distinct depth distribution (Weiner *et al.*, 2012). However, demonstrating this potential

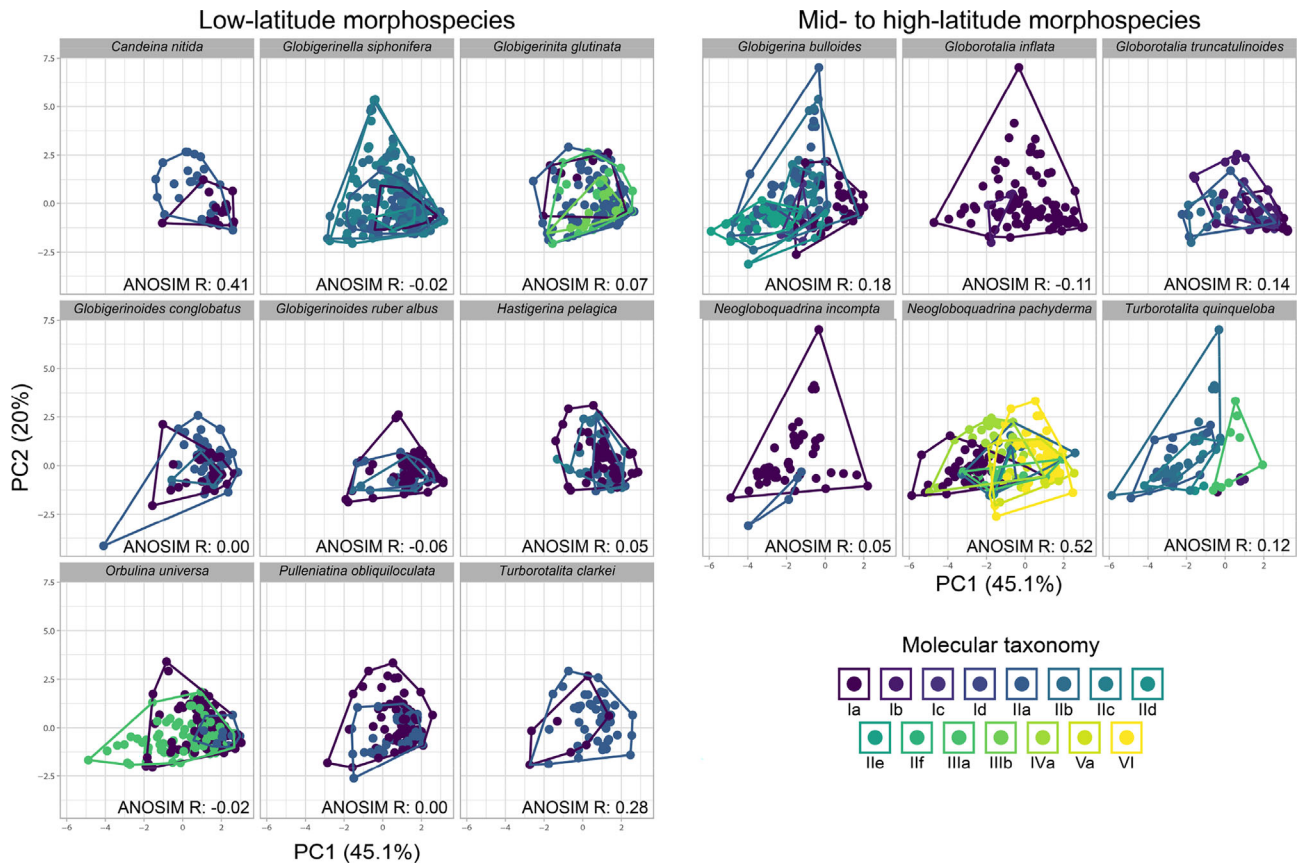
distinction for all morphospecies would require a systematic depth-stratified sampling at all locations, which is rarely available. Our analysis is restricted to the horizontal (biogeographic) distribution of MOTUs and hence we cannot rule out the possibility of environmental differentiation with depth. Nevertheless, it indicates that cryptic diversity in low-latitude morphospecies is rarely accompanied by strong biogeographical or environmental differentiation of the MOTUs in the surface ocean.

## (2) Biogeography of mid- to high-latitude morphospecies

Here we focus on morphospecies with multiple MOTUs primarily occurring in temperate, subpolar and polar environments. We observe a pronounced differentiation in



**Fig. 11.** Diversity, biogeography and thermal niches of MOTUs lvl-2 of the microporiferate clade. Other details are as in Fig. 9.



**Fig. 12.** Environmental differentiation of individual molecular operational taxonomic units (MOTUs) of the 15 morphospecies where cryptic diversity has been identified. We excluded *G. minuta* from the analysis because of an insufficient number of observations. Principal components analysis (PCA) axes were calculated based on climatological data (see Appendix S1). Each dot represents a locality where MOTUs were detected. Positive values on PC1 represent warm and oligotrophic environments as opposed to cold and nutrient-rich environments. Positive values on PC2 represent deeper mixed-layer zones and negative values represent environments with higher photosynthetically active radiation (PAR) values. We provide analysis of similarity (ANOSIM) R values to measure the degree of separation of MOTUs lvl-2 within the morphospecies. A R value of 1 indicates complete environmental separation between MOTUs, a R of zero indicates complete overlap between MOTUs and negative R values indicate larger variability within a MOTU than between MOTUs (Chapman & Underwood, 1999).

environmental preferences between MOTUs of *N. pachyderma*, *G. bulloides*, *G. truncatulinoides*, *T. quinqueloba* and *G. inflata* (Fig. 12). Among mid to high-latitude morphospecies, only *N. incompta* MOTUs do not show signs of environmental differentiation. The differentiation observed for the other morphospecies is caused by MOTUs collected in water below 15 °C such as *G. bulloides* II, *N. pachyderma* I and IV or *T. quinqueloba* II, and MOTUs occurring in water above 15 °C such as *G. bulloides* I, *N. pachyderma* V and VI or *T. quinqueloba* I (Figs 9 and 10). Also, some cold-water MOTUs appear restricted to a single basin such as *G. bulloides* IIe, which occurs in the North Pacific, or *N. pachyderma* IIa and *G. inflata* Ib, which occur in the southern hemisphere only. We also observe bipolar MOTUs for the morphospecies *T. quinqueloba* IIa–IIc, *G. bulloides* IIa and IIb, *N. incompta* Ia and *G. inflata* Ia in temperate waters of both hemispheres (but not necessarily in both the North Pacific and North Atlantic). In contrast to the warm-water morphospecies, cold-water morphospecies show clearly contrasted

biogeography and/or environmental preferences of the MOTUs belonging to the same morphospecies.

## VI. DISCUSSION

The morphological taxonomy of planktonic foraminifera stems from a palaeontological tradition and the vast majority of modern species have been described and typified from fossil material (Brummer & Kucera, 2022). This morphology-based taxonomy has been challenged by the discovery of cryptic diversity within the group (Huber *et al.*, 1997; de Vargas *et al.*, 1999, 2001, 2002; Darling *et al.*, 2000, 2004; Stewart *et al.*, 2001; Bauch *et al.*, 2003; Aurahs *et al.*, 2009b; Morard *et al.*, 2011, 2013; Sears, Darling & Wade, 2012; Ujiie *et al.*, 2012; Weiner *et al.*, 2012). However, in the course of the exploration of cryptic species, detailed morphometric analyses of sequenced specimens showed that part of the

molecular diversity actually matched morphological species concepts. For instance, most right- and left-coiling individuals of specimens identified as *N. pachyderma* constitute distinct and divergent genetic entities, which led to the resurrection of the *N. incompta* morphospecies name for the right-coiling specimens (Darling *et al.*, 2006). Similarly, the ‘genotypes’ IIa and IIb of *G. ruber* (Aurahs *et al.*, 2009b) match the morphological diagnoses of *G. elongatus* (Aurahs *et al.*, 2011) and *G. tenellus* (Morard *et al.*, 2019a), respectively. The genetically hyper-diverse *Globigerinella* genus (Weiner *et al.*, 2014) was also found to be composed of three distinct morphospecies: *G. radians*, *G. calida* and *G. siphonifera* (Weiner *et al.*, 2015). By contrast, the morphologically diverse *T. sacculifer* species, displaying four different morphotypes, showed no sign of genetic diversity (André *et al.*, 2013). Other morphospecies such as *G. bulloides*, *T. quinqueloba* and *N. pachyderma* have an extensive genetic diversity (Darling & Wade, 2008), which is corroborated by the metabarcoding that unveiled novel lineages of *N. pachyderma* (Fig. 10). Therefore, while in some morphospecies the cryptic diversity is limited or absent, in others the phenomenon appears widespread. As a result, assessing the extent of the diversity of the group using single-cell analyses alone would have been difficult to achieve because a robust estimation would have required complete global sampling and Sanger sequencing for every morphospecies. These limitations can in principle be overcome by environmental sequencing but the interpretation of the metabarcoding data set is complicated by the pervasive intra-genomic diversity of foraminifera (Weber & Pawłowski, 2014; Greco *et al.*, 2023), and by the presence of benthic foraminifera DNA in the water column (Morard *et al.*, 2022). To solve this entanglement of ‘diversities’ in our metabarcoding data set, we designed a stepwise approach, starting by parsing the benthic and planktonic foraminifera ASVs and then used our molecular nomenclature to expand its properties on the planktonic foraminifera ASVs. We are confident that our approach captured the planktonic foraminifera diversity from the initial ASV pool because we observed a dominance of planktonic foraminifera clades over benthic clades in foraminifera assemblages of the photic zone (Fig. 5B). In addition, the portion of unclassified foraminifera ASVs, which could include planktonic ASVs, is greatest in sedimentary assemblages, which are dominated by benthic foraminifera. Also, the spinose and non-spinose clades dominate the largest size fraction (180–2000  $\mu\text{m}$ ) in the photic zone, whereas the microperforate clade is more abundant in the smaller size fraction (Fig. 5B), which is consistent with microscopy observations (Brummer, Hemleben & Spindler, 1986). Therefore, the first-order allocation of the ASV diversity matches the expected occurrence patterns of planktonic foraminifera.

Using our combined data sets and harmonised molecular taxonomy, we identified 94 MOTUs, which only doubles the diversity assessment based on shell morphology (Brummer & Kucera, 2022). Sequence-based protist species richness studies tend to inflate diversity (Caron & Hu, 2019), but our estimate integrating all genetic evidence remains rather close to the morphology-based taxonomic count. In addition,

among the 45 morphospecies present in our data set, only 16 exhibited multiple lvl-2 MOTUs. We tested whether the cryptic diversity is distributed randomly among the three clades of planktonic foraminifera with a chi-squared test using as a null hypothesis the case where within each clade all morphospecies have the same number of MOTUs. We found that the test was significant for the spinose ( $\chi^2 = 50.3$ , d.f. = 21,  $P = 0.0003$ ) and non-spinose clade ( $\chi^2 = 32.3$ , d.f. = 13,  $P = 0.002$ ) but not for the microperforate clade ( $\chi^2 = 5.7$ , d.f. = 9,  $P = 0.77$ ). Considering all planktonic foraminifera together, the distribution of cryptic diversity also appears to be non-randomly distributed ( $\chi^2 = 94.9$ , d.f. = 45,  $P = 0.00002$ ). Therefore, we can make the reasonable assumption that cryptic diversity does not arise randomly, otherwise, we would have observed a more equal distribution of the MOTUs among the morphospecies, but we acknowledge that the situation may vary depending on the clade considered. In our data set, we identify a stronger link between cryptic diversity and geography, because we do not observe a systematic environmental differentiation between MOTUs lvl-2 belonging to the same morphospecies (Fig. 12). The environmental separation is minimal for MOTUs lvl-2 of morphospecies inhabiting low-latitude environments such as *G. ruber albus*, *O. universa*, *G. siphonifera*, *G. glutinata*, *G. conglobatus*, *H. pelagica* and *P. obliquiloculata* (Fig. 12), as is their biogeographic separation (Figs 9–11). Their MOTUs lvl-2 occur broadly in the warm Indo-Pacific and Atlantic oceans with the exception of *G. siphonifera* IIc, *G. ruber albus* IIa, *P. obliquiloculata* IIa, which are either absent or rare in the Atlantic. Weiner *et al.* (2014) ascribed the absence of such MOTUs in the Atlantic to incumbency, where the expansion of MOTUs from the Indo-Pacific into the Atlantic would be prevented by incumbent (already established) taxa with similar ecological preferences, or in other words competitive exclusion. This is opposite to the findings of Rillo *et al.* (2019) who did not find evidence for inter-specific competition between extant morphospecies of planktonic foraminifera. While in apparent contradiction, the two findings are not incompatible because competition could act below the morphospecies level, between closely related and co-occurring MOTUs, and fade between co-occurring morphospecies. The situation of the morphospecies inhabiting mid- to high-latitude environments such as *N. pachyderma*, *N. incompta*, *G. bulloides*, *T. quinqueloba*, *G. inflata* and *G. truncatulinooides* seems to be the almost exact opposite of the low-latitude morphospecies. Their MOTUs tend to show different environmental preferences and geographic disconnection (Figs 9–11). This could be because morphological diversity decreases with increasing latitude, hence reducing competitive exclusion and making more niches available for cryptic species. The low-latitude diversity pattern and the mid- to high-latitude cryptic diversity pattern combine to create the observed connectivity structure between oceanic basins, where the warm Indo-Pacific and Atlantic share the highest number of MOTUs, and the north and south subpolar and polar basins share fewer MOTUs (Fig. 8). The poleward increasing contribution of morphospecies with cryptic

diversity in surface sediment assemblages (Fig. 7B) also implies a relationship between cryptic diversity and geography.

The likely factor linked to geography that could influence the degree of cryptic diversity in morphospecies may be the limited ability of planktonic foraminifera to maintain gene flow over large distances. Gene flow is responsible for the genetic homogeneity of populations, and compensates the genetic drift that can lead to the formation of novel biological entities (Slatkin, 1987). In our data set, we consider MOTUs lvl-2 as indicative of biological divergence, in the sense that no gene flow occurs between sister MOTUs lvl-2. Limited ability to maintain gene flow over large distances could explain why the pool of morphospecies without cryptic diversity has a restricted geographical range (Fig. 7B). It would correspond to the extent to which planktonic foraminifera can maintain a gene flow sufficient to sustain the genetic cohesiveness of a morphospecies. The metabarcoding data set unveiled rare but consistent occurrences of MOTUs or morphospecies outside of their core biogeographic range, showing their capability to disperse but not necessarily maintain a strong gene flow. For instance, we observed *N. pachyderma* Ia, which was thought to be restricted to the Arctic and subpolar Atlantic, in the south Atlantic and in the Pacific, while *N. pachyderma* IVa, which was thought to be restricted to polar waters of the Southern hemisphere (Darling *et al.*, 2004), has been observed in the equatorial Pacific. Similarly, the Atlantic and Mediterranean morphosubspecies *Globigerinoides ruber ruber* has a few occurrences in the Indo-Pacific in the metabarcoding data set. These rare occurrences could be ascribed to either technical bias such as tag jumps (Schnell, Bohmann & Gilbert, 2015) or human-assisted dispersal by ballast water (Olenin *et al.*, 2000). However, they could also reflect the presence of naturally dispersed living specimens. It has recently been shown that planktonic foraminifera can undergo asexual reproduction (Kimoto & Tsuchiya, 2006; Davis *et al.*, 2020; Takagi, Kurasawa & Kimoto, 2020; Meilland *et al.*, 2022). This means that dispersed specimens can in principle produce offspring without relying on gamete fusion, which is hardly possible when the population density is low (Weinkauff, Siccha & Weiner, 2022). Asexual reproduction could thus aid the persistence of specimens outside their core range. However, it is unlikely to be the main mechanism to establish new viable populations outside this range. Several bipolar MOTUs, such as *T. quinqueloba* MOTUs IIa–IIc, *G. bulloides* MOTUs IIa and IIb, *N. incompta* MOTU Ia and *G. inflata* MOTU Ia indicate the maximum extension MOTUs lvl-2 can undergo on a north–south axis (Appendix S2). Beyond that range occupied by these MOTUs, they appear to lose cohesiveness and evolve into biogeographically disconnected MOTUs as observed in the high-latitude *N. pachyderma* MOTUs lvl-2. The exception to this pattern is the morphospecies *G. uvula* that occurs globally at high latitudes, yet it is represented only by a single MOTU lvl-2 (Fig. 11). The presumed inability to maintain gene flow over larger distances could cause the fragmentation of morphospecies into MOTUs with differentiated environmental preferences along increasing latitude.

The morphological diversity of planktonic foraminifera (50 morphospecies) is far below those of radiolarians (600–800 morphospecies; Suzuki & Not, 2015), dinoflagellates (~2000 morphospecies; Gomez, 2012) and diatoms that most certainly range in the tens of thousands (Malviya *et al.*, 2016). Since the biological diversity of planktonic foraminifera will most likely not extend far beyond 100 MOTUs, even at the molecular level their diversity is several orders of magnitude lower than that of these other marine protist groups. This low diversity is even more surprising because foraminifera as a whole have conquered all marine habitats from coastal areas (Förderer, Rödder & Langer, 2018), and deep sea (Lecroq *et al.*, 2011), to freshwater environments (Holzmann *et al.*, 2007), and even soils (Lejzerowicz *et al.*, 2010). In our data set, which includes mostly samples from the water column, planktonic foraminifera represent less than 3% of the foraminifera ASV diversity. It remains to be understood why planktonic foraminifera have such low diversity, yet are successful in maintaining a global distribution.

Although we have assembled a global data set, which represents nearly 30 years of collection, single-cell analysis and metabarcoding, the evolutionary and ecological mechanisms sustaining the global biodiversity of planktonic foraminifera remain elusive. In particular, our data set does not permit the resolution of the vertical habitat dimension of cryptic diversity. For instance, different MOTUs of the same morphospecies may co-occur at the same location but occupy different ecological niches in the water column, as shown for *H. pelagica* (Weiner *et al.*, 2012), or co-occur broadly in the same depth range and hence be in direct competition for resources. In addition, we could not examine the connection between MOTUs and abiotic factors or their association with other eukaryotes (Greco, Morard & Kucera, 2021) or prokaryotes (Bird *et al.*, 2018). Current technological advances have opened the way to explore new areas of the foraminifera genome (Macher *et al.*, 2023) and transcriptomics has made multigene phylogenies a reality for foraminifera (Sierra *et al.*, 2022), that could even be applied at the single-cell level (Weiner *et al.*, 2023). These advances could facilitate a robust phylogenetic reconstruction of the deep evolutionary history of planktonic foraminifera that often lacks statistical support and towards population genetic analysis which is problematic to interpret using SSU rDNA because of intra-genomic variability (Greco *et al.*, 2023). While these promising modern analytical approaches offer new perspectives on the molecular biology of planktonic foraminifera, the taxonomic framework established here provides the foundation for this forthcoming era of exploration.

## VII. CONCLUSIONS

(1) In our global survey of the biological diversity of planktonic foraminifera we identified only 94 MOTUs, which is only twice as high as the morphologically based estimates.

- (2) The cryptic diversity is unequally distributed within morphological diversity and occurs in only 16 of the 45 morphospecies that we captured in our data set.
- (3) Mid- to high-latitude morphospecies have the highest degree of cryptic diversity associated with a biogeographic and environmental differentiation between MOTUs.
- (4) Low-latitude morphospecies are less subject to cryptic diversification with weak to no biogeographic and or environmental differentiation between MOTUs.
- (5) Geography is the primary factor along which cryptic diversity is structured.
- (6) Planktonic foraminifera are among the least diversified groups of pelagic protists.

## VIII. ACKNOWLEDGEMENTS

We thank all crew members and scientists for their help in collecting single-cell planktonic foraminifera and samples for metabarcoding during the numerous expeditions that allowed us to assemble a worldwide data set. This work was carried out within the framework of the SCOR/IGBP working group 138: Modern Planktonic Foraminifera and Ocean Changes. M.G. was supported by a Juan de la Cierva-formación 2021 fellowship (EJC2021–047494-I/MCIN/AEI/10.13039/501100011033) from the European Union ‘Next-GenerationEU’/PRTR, F.Q. was supported by the French national program Tellus-INTERRVIE of CNRS-INSU, K. F.D. was supported by the Natural Environment Research Council (NERC) of the United Kingdom (grants NER/J/S/2000/00860 and NE/D009707/1), and R.M., C.V., J. M. and M.K. were supported by the Cluster of Excellence ‘The Ocean Floor—Earth’s Uncharted Interface’ (EXC-2077, Project 390741603) funded through the German Research Foundation (DFG). Finally, we would like to thank the anonymous reviewers for their detailed and constructive criticism and suggestions. Open Access funding enabled and organized by Projekt DEAL.

## IX. DATA AVAILABILITY STATEMENT

The data that support the findings of this study are provided as supporting information and are openly available at Zenodo (10.5281/zenodo.10454474); the R script is provided on GitHub ([https://github.com/Raph-forams/Forams\\_Cryptic\\_Diversity](https://github.com/Raph-forams/Forams_Cryptic_Diversity)).

## X. REFERENCES

References identified with an asterisk (\*) are cited only within the online supporting information.

ADL, S. M., BASS, D., LANE, C. E., LUKEŠ, J., SCHOCH, C. L., SMIRNOV, A., AGATHA, S., BERNEY, C., BROWN, M. W., BURKI, F., CÁRDENAS, P., ČEPIČKA, I., CHISTYAKOVA, L., DEL CAMPO, J., DUNTHORN, M., *ET AL.* (2019).

- Revisions to the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology* **66**, 4–119.
- ANDRÉ, A., QUILLÉVÉRÉ, F., MORARD, R., UJJIÉ, Y., ESCARGUEL, G., DE VARGAS, C., DE GARIDEL-THORON, T. & DOUADY, C. J. (2014). SSU rDNA divergence in planktonic foraminifera: molecular taxonomy and biogeographic implications. *PLoS ONE* **9**, e104641.
- \*ANDRÉ, A., QUILLÉVÉRÉ, F., SCHIEBEL, R., MORARD, R., HOWA, H., MEILLAND, J. & DOUADY, C. J. (2018). Disconnection between genetic and morphological diversity in the planktonic foraminifer *Neogloboquadrina pachyderma* from the Indian sector of the Southern Ocean. *Marine Micropaleontology* **144**, 14–24.
- ANDRÉ, A., WEINER, A., QUILLÉVÉRÉ, F., AURAHNS, R., MORARD, R., DOUADY, C. J., DE GARIDEL-THORON, T., ESCARGUEL, G., DE VARGAS, C. & KUCERA, M. (2013). The cryptic and the apparent reversed: lack of genetic differentiation within the morphologically diverse plexus of the planktonic foraminifer *Globigerinoides sacculifer*. *Paleobiology* **39**, 21–39.
- \*AURAHNS, R., GÖKER, M., GRIMM, G. W., HEMLEBEN, V., HEMLEBEN, C., SCHIEBEL, R. & KUCERA, M. (2009a). Using the multiple analysis approach to reconstruct phylogenetic relationships among planktonic foraminifera from highly divergent and length-polymorphic SSU rDNA sequences. *Bioinformatics and Biology Insights* **3**, 155–177.
- AURAHNS, R., GRIMM, G. W., HEMLEBEN, V., HEMLEBEN, C. & KUCERA, M. (2009b). Geographical distribution of cryptic genetic types in the planktonic foraminifer *Globigerinoides ruber*. *Molecular Ecology* **18**, 1692–1706.
- AURAHNS, R., TREIS, Y., DARLING, K. & KUCERA, M. (2011). A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence. *Marine Micropaleontology* **79**, 1–14.
- AZE, T., EZARD, T. H. G., PURVIS, A., COXALL, H. K., STEWART, D. R. M., WADE, B. S. & PEARSON, P. N. (2011). A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biological Reviews* **86**, 900–927.
- \*BARBERA, P., KOZLOV, A. M., CZECH, L., MOREL, B., DARRIBA, D., FLOURI, T. & STAMATAKIS, A. (2019). EPA-ng: massively parallel evolutionary placement of genetic sequences. *Systematic Biology* **68**, 365–369.
- BARRENECHEA ANGELES, I., LEJZEROWICZ, F., CORDIER, T., SCHEPLITZ, J., KUCERA, M., ARIZTEGUI, D., PAWLOWSKI, J. & MORARD, R. (2020). Planktonic foraminifera eDNA signature deposited on the seafloor remains preserved after burial in marine sediments. *Scientific Reports* **10**, 1–12.
- BAUCH, D., DARLING, K., SIMSTICH, J., BAUCH, H. A. & ERLLENKEUSER, H. (2003). Palaeoceanographic implications of genetic variation in living North Atlantic *Neogloboquadrina pachyderma*. *Nature* **424**, 11–14.
- \*BECKER, R. A., WILKS, A. R., BROWNRIFF, R., MINKA, T. P. & DECKMYN, A. (2021). maps: draw geographical maps. <https://cran.r-project.org/package=maps>.
- BICKFORD, D., LOHMAN, D. J., SODHI, N. S., NG, P. K. L., MEIER, R., WINKER, K., INGRAM, K. K. & DAS, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* **22**, 148–155.
- \*BIRD, C., DARLING, K. F., RUSSELL, A. D., DAVIS, C. V., FEHRENBACHER, J., FREE, A., WYMAN, M. & NGWENYA, B. T. (2017). Cyanobacterial endobionts within a major marine planktonic calcifier (*Globigerina bulloides*, foraminifera) revealed by 16S rRNA metabarcoding. *Biogeosciences* **14**, 901–920.
- BIRD, C., DARLING, K. F., RUSSELL, A. D., FEHRENBACHER, J. S., DAVIS, C. V., FREE, A. & NGWENYA, B. T. (2018). 16S rRNA gene metabarcoding and TEM reveals different ecological strategies within the genus *Neogloboquadrina* (planktonic foraminifer). *PLoS ONE* **13**, e0191653.
- BLAXTER, M. (2016). Imagining Sisyphus happy: DNA barcoding and the unnamed majority. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150329.
- BRUMMER, G. A., HEMLEBEN, C. & SPINDLER, M. (1986). Planktonic foraminiferal ontogeny and new perspectives for micropalaeontology. *Nature* **319**, 50–52.
- BRUMMER, G. A. & KUCERA, M. (2022). Taxonomic review of living planktonic foraminifera. *Journal of Micropalaeontology* **41**, 29–74.
- BUCKLIN, A. N. N., LINDEQUE, P. K., RODRIGUEZ-EZPELETA, N., ALBAINA, A. & LEHTINIEMI, M. (2016). Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *Journal of Plankton Research* **38**, 393–400.
- BURKI, F., SANDIN, M. M. & JAMY, M. (2021). Diversity and ecology of protists revealed by metabarcoding. *Current Biology* **31**, R1267–R1280.
- \*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.
- CARON, D. A. & HU, S. K. (2019). Are we overestimating protistan diversity in nature? *Trends in Microbiology* **27**, 197–205.
- CHAABANE, S., DE GARIDEL-THORON, T., GIRAUD, X., SCHIEBEL, R., BEAUGRAND, G., BRUMMER, G.-J., CASAJUS, N., GRECO, M., GRIGORATOU, M., HOWA, H., JONKERS, L., KUCERA, M., KUROYANAGI, A., MEILLAND, J., MONTEIRO, F., *ET AL.* (2023). The FORCIS database: A global census of planktonic foraminifera from ocean waters. *Scientific Data* **10**, 1–16.
- CHAPMAN, M. G. & UNDERWOOD, A. J. (1999). Ecological patterns in multivariate assemblages: information and interpretation of negative values in ANOSIM tests. *Marine Ecology Progress Series* **180**, 257–265.

- \*CONWAY, J. & GEHLENBORG, N. (2017). UpSetR: a more scalable alternative to Venn and Euler diagrams for visualizing intersecting sets. <https://github.com/hms-dbmi/UpSetR>.
- CORDIER, T., ANGELES, I. B., HENRY, N., LEJZEROWICZ, F., BERNEY, C., MORARD, R., BRANDT, A., CAMBON-BONAVITA, M.-A., GUIDI, L., LOMBARD, F., ARBIZU, P. M., MASSANA, R., OREJAS, C., POULAIN, J., SMITH, C. R., *ET AL.* (2022). Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. *Science Advances* **8**, 1–13.
- CRISTESCU, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution* **29**, 566–571.
- \*DARLING, K. F., KROON, D., WADE, C. M. & LEIGH BROWN, A. J. (1996). Molecular phylogeny of the planktic foraminifera. *The Journal of Foraminiferal Research* **26**, 324–330.
- DARLING, K. F., KUCERA, M., KROON, D. & WADE, C. M. (2006). A resolution for the coiling direction paradox in *Neogloboquadrina pachyderma*. *Paleoceanography* **21**, PA2011.
- DARLING, K. F., KUCERA, M., PUDSEY, C. J. & WADE, C. M. (2004). Molecular evidence links cryptic diversification in polar planktonic protists to quaternary climate dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7657–7662.
- DARLING, K. F., KUCERA, M. & WADE, C. M. (2007). Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5002–5007.
- DARLING, K. F., KUCERA, M., WADE, C. M., VON LANGEN, P. & PAK, D. (2003). Seasonal distribution of genetic types of planktonic foraminifer morphospecies in the Santa Barbara Channel and its paleoceanographic implications. *Paleoceanography* **18**, 1–10.
- DARLING, K. F. & WADE, C. M. (2008). The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. *Marine Micropaleontology* **67**, 216–238.
- DARLING, K. F., WADE, C. M., KROON, D. & BROWN, A. J. L. (1997). Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. *Marine Micropaleontology* **30**, 251–266.
- \*DARLING, K. F., WADE, C. M., KROON, D., BROWN, A. J. L. & BIJMA, J. (1999). The diversity and distribution of modern planktic foraminiferal small subunit ribosomal RNA genotypes and their potential as tracers of present and past circulation. *Paleoceanography* **14**, 3–12.
- DARLING, K. F., WADE, C. M., SICCHA, M., TROMMER, G., SCHULZ, H., ABDOLALIPOUR, S. & KURASAWA, A. (2017). Genetic diversity and ecology of the planktonic foraminifers *Globigerina bulloides*, *Turborotalita quinqueloba* and *Neogloboquadrina pachyderma* off the Oman margin during the late SW monsoon. *Marine Micropaleontology* **137**, 64–77.
- DARLING, K. F., WADE, C. M., STEWART, I. A., KROON, D., DINGLE, R. & BROWN, A. J. (2000). Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature* **405**, 43–47.
- \*DARRIBA, D., POSADA, D., KOZLOV, A. M., STAMATAKIS, A., MOREL, B. & FLOURI, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution* **37**, 291–294.
- DAVIS, C. V., LIVSEY, C. M., PALMER, H. M., HULL, P. M., THOMAS, E., HILL, T. M. & BENITZ-NELSON, C. R. (2020). Extensive morphological variability in asexually produced planktic foraminifera. *Science Advances* **6**, 1–8.
- DE QUEIROZ, K. (2007). Species concepts and species delimitation. *Systematic biology* **56**, 879–886.
- DE VARGAS, C., AUDIC, S., HENRY, N., DECELLE, J., MAHÉ, F., LOGARES, R., LARA, E., BERNEY, C., LE BESCOT, N., PROBERT, I., CARMICHAEL, M., POULAIN, J., ROMAC, S., COLIN, S., AURY, J.-M., *ET AL.* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605.
- DE VARGAS, C., BONZON, M., REES, N. W., PAWLOWSKI, J. & ZANINETTI, L. (2002). A molecular approach to biodiversity and biogeography in the planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Marine Micropaleontology* **45**, 101–116.
- \*DE VARGAS, C., LE BESCOT, N., POLLINA, T., HENRY, N., ROMAC, S., HAËNTJENS, N., CARMICHAEL, M., BERGER, C., LE GUEN, D., DECELLE, J., POULAIN, J., MALPOT, E., BEAUMONT, C., HARDY, M., GUIFFANT, D., *ET AL.* (2022). Plankton planet: a frugal, cooperative measure of aquatic life at the planetary scale. *Frontiers in Marine Science* **9**, 1–17.
- DE VARGAS, C., NORRIS, R., ZANINETTI, L., GIBB, S. W. & PAWLOWSKI, J. (1999). Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2864–2868.
- DE VARGAS, C. & PAWLOWSKI, J. (1998). Molecular versus taxonomic rates of evolution in planktonic foraminifera. *Molecular Phylogenetics and Evolution* **9**, 463–469.
- DE VARGAS, C., RENAUD, S., HILBRECHT, H. & PAWLOWSKI, J. (2001). Pleistocene adaptive radiation in *Globorotalia truncatulinoides*: genetic, morphologic, and environmental evidence. *Paleobiology* **27**, 104–125.
- \*DE VARGAS, C., ZANINETTI, L., HILBRECHT, H. & PAWLOWSKI, J. (1997). Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. *Journal of Molecular Evolution* **45**, 285–294.
- \*DUARTE, C. M. (2015). Seafaring in the 21st century: the Malaspina 2010 circumnavigation expedition. *Limnology and Oceanography Bulletin* **24**, 11–14.
- EME, D., ZAGMAJSTER, M., DELIĆ, T., FIŠER, C., FLOT, J.-F., KONECNY-DUPRÉ, L., PÁLSSON, S., STOCH, F., ZAKŠEK, V., DOUADY, C. J. & MALARD, F. (2018). Do cryptic species matter in macroecology? Sequencing European groundwater crustaceans yields smaller ranges but does not challenge biodiversity determinants. *Ecography* **41**, 424–436.
- \*ERTAN, K. T., HEMLEBEN, V. & HEMLEBEN, C. (2004). Molecular evolution of some selected benthic foraminifera as inferred from sequences of the small subunit ribosomal DNA. *Marine Micropaleontology* **53**, 367–388.
- FENTON, I. S., AZE, T., FARNSWORTH, A., VALDES, P. & SAUPE, E. E. (2023). Origination of the modern-style diversity gradient 15 million years ago. *Nature* **614**, 708–712.
- FENTON, I. S., WOODHOUSE, A., AZE, T., LAZARUS, D., RENAUDIE, J., DUNHILL, A. M., YOUNG, J. R. & SAUPE, E. E. (2021). Triton, a new species-level database of Cenozoic planktonic foraminiferal occurrences. *Scientific Data* **8**, 160.
- FLEGONTOVA, O., LUKEŠ, J. & HORÁK, A. (2023). Intragenomic diversity of the V9 hypervariable domain in eukaryotes has little effect on metabarcoding. *iScience* **26**, 102791.
- FÖRDERER, M., RÖDDER, D. & LANGER, M. R. (2018). Patterns of species richness and the center of diversity in modern Indo-Pacific larger foraminifera. *Scientific Reports* **8**, 8189.
- GOMEZ, F. (2012). A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Systematics and Biodiversity* **10**, 267–275.
- GRECO, M., MORARD, R., DARLING, K. & KUCERA, M. (2023). Macroevolutionary patterns in intragenomic rDNA variability among planktonic foraminifera. *PeerJ* **11**, e15255.
- GRECO, M., MORARD, R. & KUCERA, M. (2021). Single-cell metabarcoding reveals biotic interactions of the Arctic calcifier *Neogloboquadrina pachyderma* with the eukaryotic pelagic community. *Journal of Plankton Research* **43**, 113–125.
- GUILLOU, L., BACHAR, D., AUDIC, S., BASS, D., BERNEY, C., BITTNER, L., BOUTTE, C., BURGAUD, G., DE VARGAS, C., DECELLE, J., DEL CAMPO, J., DOLAN, J. R., DUNTHORN, M., EDVARDSEN, B., HOLZMANN, M., *ET AL.* (2013). The protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* **41**, D597–D604.
- HEBERT, P. D. N., CYWINSKA, A., BALL, S. L. & DEWAARD, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings Biological Sciences/The Royal Society* **270**, 313–321.
- HENRY, N., DE VARGAS, C. & AUDIC, S. (2019). *PR2\_V9, a SSU V9 rDNA reference database with functional annotations*. [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.3768951>
- HILLEBRAND, H. (2004). On the generality of the latitudinal diversity gradient. *American Naturalist* **163**, 192–211.
- HOLZMANN, M., HABURA, A., GILES, H., BOWSER, S. S. & PAWLOWSKI, J. (2007). Freshwater foraminifera revealed by analysis of environmental DNA samples. *The Journal of Eukaryotic Microbiology* **50**, 135–139.
- \*HOLZMANN, M. & PAWLOWSKI, J. (2017). An updated classification of rotalid foraminifera based on ribosomal DNA phylogeny. *Marine Micropaleontology* **132**, 18–34.
- \*HSIEH, T. C., MA, K. H. & CHAO, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution* **7**, 1451–1456.
- HUBER, B. T., BIJMA, J. & DARLING, K. (1997). Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera*. *Paleobiology* **23**, 33–62.
- \*IBARBALZ, F. M., HENRY, N., BRANDAO, M. C., MARTINI, S., BUSSENI, G., BYRNE, H., COELHO, L. P., ENDO, H., GASOL, J. M., GREGORY, A. C., MAHÉ, F., RIGONATO, J., ROYO-LONGH, M., SALAZAR, G., SANZ-SÁEZ, I., *ET AL.* (2019). Global trends in marine plankton diversity across kingdoms of life. *Cell* **179**, 1084–1097.
- JONKERS, L. & KUČERA, M. (2015). Global analysis of seasonality in the shell flux of extant planktonic foraminifera. *Biogeosciences* **12**, 2207–2226.
- KARSENTI, E., ACINAS, S. G., BORK, P., BOWLER, C., DE VARGAS, C., RAES, J., SULLIVAN, M., ARENDT, D., BENZONI, F., CLAVERIE, J.-M., FOLLOWS, M., GORSKY, G., HINGAMP, P., IUDICONE, D., JAILLON, O., *ET AL.* (2011). A holistic approach to marine eco-systems biology. *PLoS Biology* **9**, e1001177.
- \*KASSAMBARA, A. (2016). Factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7. [https://cran.r-project.org/web/packages/factoextra%0A](https://cran.r-project.org/web/packages/factoextra/factoextra%0A).
- \*KATOH, K. & STANDLEY, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.
- KATZ, M. E., CRAMER, B. S., FRANZESE, A., HONISCH, B., MILLER, K. G., ROSENTHAL, Y. & WRIGHT, J. D. (2010). Traditional and emerging geochemical proxies in foraminifera. *The Journal of Foraminiferal Research* **40**, 165–192.
- KECK, F., COUTON, M. & ALTERMATT, F. (2022). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources* **23**, 742–755.
- KIMOTO, K. & TSUCHIYA, M. (2006). The “unusual” reproduction of planktic foraminifera: an asexual reproductive phase of *Neogloboquadrina pachyderma* (Ehrenberg). *Anuário do Instituto de Geociências – UFRJ* **29**, 461.

- KUCERA, M. (2007). Chapter six planktonic foraminifera as tracers of past oceanic environments. In *Developments in Marine Geology* (eds C. Hillaire-Marcel and A. De Vernal), pp. 213–262. GEOTOP-Université du Québec à Montréal, Montréal, Québec, Canada.
- KUCERA, M. & DARLING, K. F. (2002). Cryptic species of planktonic foraminifera: their effect on palaeoceanographic reconstructions. *Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences* **360**, 695–718.
- KUCERA, M., ROSELL-MELÉ, A., SCHNEIDER, R., WELBROECK, C. & WEINELT, M. (2005). Multiproxy approach for the reconstruction of the glacial ocean surface (MARGO). *Quaternary Science Reviews* **24**, 813–819.
- \*KUROYANAGI, A., TSUCHIYA, M., KAWAHATA, H. & KITAZATO, H. (2008). The occurrence of two genotypes of the planktonic foraminifer *Globigerinoides ruber* (white) and paleo-environmental implications. *Marine Micropaleontology* **68**, 236–243.
- LECROQ, B., LEJZEROWICZ, F., BACHAR, D., CHRISTEN, R., ESLING, P., BAERLOCHER, L., OSTERÁS, M., FARINELLI, L. & PAWLOWSKI, J. (2011). Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 13177–13182.
- LEJZEROWICZ, F., GOODAY, A. J., BARRENECHEA ANGELES, I., CORDIER, T., MORARD, R., APOTHÉLOZ-PERRET-GENTIL, L., LINS, L., MENOT, L., BRANDT, A., LEVIN, L. A., MARTINEZ ARBIZU, P., SMITH, C. R. & PAWLOWSKI, J. (2021). Eukaryotic biodiversity and spatial patterns in the Clarion-Clipperton zone and other abyssal regions: insights from sediment DNA and RNA metabarcoding. *Frontiers in Marine Science* **8**, 671033.
- LEJZEROWICZ, F., PAWLOWSKI, J., FRAISSINET-TACHET, L. & MARMEISSE, R. (2010). Molecular evidence for widespread occurrence of foraminifera in soils. *Environmental Microbiology* **12**, 2518–2526.
- \*LETUNIC, I. & BORK, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, 293–296.
- MACHER, J.-N., COOTS, N. L., POH, Y.-P., GIRARD, E. B., LANGERAK, A., MUÑOZ-GÓMEZ, S. A., SINHA, S. D., JIRSOVÁ, D., VOS, R., WISSELS, R., GILE, G. H., RENEMA, W. & WIDEMAN, J. G. (2023). Single-cell genomics reveals the divergent mitochondrial genomes of Retaria (foraminifera and Radiolaria). *mBio* **14**, e00302–e00323.
- MAHÉ, F., DE VARGAS, C., BASS, D., CZECH, L., STAMATAKIS, A., LARA, E., SINGER, D., MAYOR, J., BUNGE, J., SERNAKER, S., SIEMENSMEYER, T., TRAUTMANN, I., ROMAC, S., BERNEY, C., KOZLOV, A., ET AL. (2017). Parasites dominate hyperdiverse soil protist communities in neotropical rainforests. *Nature Ecology and Evolution* **1**, 91.
- MALVIYA, S., SCALCO, E., AUDIC, S., VINCENT, F., VÉLUCHAMY, 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* **113**, E1516–E1525.
- MEILLAND, J., EZAT, M. M., WESTGARD, A., MANNO, C., MORARD, R., SICCHA, M. & KUCERA, M. (2022). Rare but persistent asexual reproduction explains the success of planktonic foraminifera in polar oceans. *Journal of Plankton Research* **45**, 15–32.
- \*MILIVOJEVIĆ, T., RAHMAN, S. N., RAPOSO, D., SICCHA, M., KUCERA, M. & MORARD, R. (2021). High variability in SSU rDNA gene copy number among planktonic foraminifera revealed by single-cell qPCR. *ISME Communications* **1**, 63.
- MORA, C., TITTEENSOR, D. P., ADL, S., SIMPSON, A. G. B. & WORM, B. (2011). How many species are there on Earth and in the ocean? *PLoS biology* **9**, e1001127.
- \*MORARD, R. (2010). For a better characterization of the fossil pelagic record: molecular, biogeographical and ecological diversity of planktonic foraminifera cryptic species. Université Claude Bernard Lyon I.
- MORARD, R., DARLING, K. F., MAHÉ, F., AUDIC, S., UJJIÉ, Y., WEINER, A. K. M., ANDRÉ, A., SEARS, H. A., WADE, C. M., QUILLÉVÉRÉ, F., DOUADY, C. J., ESCARGUEL, G., DE GARIDEL-THORON, T., SICCHA, M., KUCERA, M., ET AL. (2015). PFR<sup>2</sup>: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. *Molecular Ecology Resources* **15**, 1472–1485.
- MORARD, R., ESCARGUEL, G., WEINER, A. K. M., ANDRÉ, A., DOUADY, C. J., WADE, C. M., DARLING, K. F., UJJIÉ, Y., SEARS, H. A., QUILLÉVÉRÉ, F., DE GARIDEL-THORON, T., DE VARGAS, C. & KUCERA, M. (2016). Nomenclature for the nameless: a proposal for an integrative molecular taxonomy of cryptic diversity exemplified by planktonic foraminifera. *Systematic Biology* **65**, 925–940.
- MORARD, R., FÜLLBERG, A., BRUMMER, G. J. A., GRECO, M., JONKERS, L., WIZEMANN, A., WEINER, A. K. M., DARLING, K., SICCHA, M., LEDEVIN, R., KITAZATO, H., DE GARIDEL-THORON, T., DE VARGAS, C. & KUCERA, M. (2019a). Genetic and morphological divergence in the warm-water planktonic foraminifera genus *Globigerinoides*. *PLoS ONE* **14**, 1–30.
- MORARD, R., HASENBRÜCK, C., GRECO, M., FERNANDEZ-GUERRA, A., RIGAUD, S., DOUADY, C. J. & KUCERA, M. (2022). Renewal of planktonic foraminifera diversity after the Cretaceous Paleogene mass extinction by benthic colonizers. *Nature Communications* **13**, 1–9.
- MORARD, R., LEJZEROWICZ, F., DARLING, K. F., LECROQ-BENNET, B., PEDERSEN, M. W., ORLANDO, L., PAWLOWSKI, J., MULITZA, S., DE VARGAS, C. & KUCERA, M. (2017). Planktonic foraminifera-derived environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology. *Biogeosciences* **14**, 2741–2754.
- MORARD, R., MAHÉ, F., ROMAC, S., POULAIN, J., KUCERA, M. & DE VARGAS, C. (2018). Surface ocean metabarcoding confirms limited diversity in planktonic foraminifera but reveals unknown hyper-abundant lineages. *Scientific Reports* **8**, 2539.
- MORARD, R., QUILLÉVÉRÉ, F., DOUADY, C. J., DE VARGAS, C., DE GARIDEL-THORON, T. & ESCARGUEL, G. (2011). Worldwide genotyping in the planktonic foraminifer *Globocornella inflata*: implications for life history and paleoceanography. *PLoS ONE* **6**, e26665.
- MORARD, R., QUILLÉVÉRÉ, F., ESCARGUEL, G., DE GARIDEL-THORON, T., DE VARGAS, C. & KUCERA, M. (2013). Ecological modeling of the temperature dependence of cryptic species of planktonic foraminifera in the southern hemisphere. *Palaeogeography, Palaeoclimatology, Palaeoecology* **391**, 13–33.
- \*MORARD, R., QUILLÉVÉRÉ, F., ESCARGUEL, G., UJJIÉ, Y., DE GARIDEL-THORON, T., NORRIS, R. D. & DE VARGAS, C. (2009). Morphological recognition of cryptic species in the planktonic foraminifer *Orbulina universa*. *Marine Micropaleontology* **71**, 148–165.
- MORARD, R., VOLLMAR, N. M., GRECO, M. & KUCERA, M. (2019b). Unassigned diversity of planktonic foraminifera from environmental sequencing revealed as known but neglected species. *PLoS ONE* **14**, e0213936.
- \*NASA GODDARD SPACE FLIGHT CENTER, OCEAN ECOLOGY LABORATORY, OCEAN BIOLOGY PROCESSING GROUP. MODERATE-RESOLUTION IMAGING SPECTRODIAMETER (MODIS) AQUA SEA SURFACE TEMPERATURE DATA (2014). *Reprocessing*. NASA OBD.AAC, Greenbelt.
- \*OKSANEN, J., KINDT, R., LEGENDRE, P., O'HARA, B., STEVENS, M. H. H., OKSANEN, M. J. & SUGGESTS, M. (2007). The vegan package. *Community Ecology Package* **10**, 631–637.
- OLENIN, S., GOLLASCH, S., JONUŠAS, S. & RIMKUTĖ, I. (2000). En-route investigations of plankton in ballast water on a ship's voyage from the Baltic Sea to the open Atlantic coast of Europe. *International Review of Hydrobiology* **85**, 577–596.
- \*PARADIS, E., BLOMBER, S., BOLKER, B., BROWN, J., CLAUDE, J., HOA, S., DESPER, R., DIDIER, G., DURAND, B., DUTHEIL, J., EWING, R., GASCUEL, O., GUILLERME, T., HEIBL, C., IVES, A., ET AL. (2019). Package 'ape': analyses of phylogenetics and evolution depends R (3.2.0).
- PAWLOWSKI, J. & HOLZMANN, M. (2014). A plea for DNA barcoding of foraminifera. *The Journal of Foraminiferal Research* **44**, 62–67.
- \*PAWLOWSKI, J., HOLZMANN, M. & TYSZKA, J. (2013). New supraordinal classification of foraminifera: molecules meet morphology. *Marine Micropaleontology* **100**, 1–10.
- PAWLOWSKI, J. & LECROQ, B. (2010). Short rDNA barcodes for species identification in foraminifera. *The Journal of Eukaryotic Microbiology* **57**, 197–205.
- PUILLANDRE, N., BROUILLET, S. & ACHAZ, G. (2021). ASAP: assemble species by automatic partitioning. *Molecular Ecology Resources* **21**, 609–620.
- QUILLÉVÉRÉ, F., MORARD, R., ESCARGUEL, G., DOUADY, C. J., UJJIÉ, Y., DE GARIDEL-THORON, T. & DE VARGAS, C. (2013). Global scale same-specimen morpho-genetic analysis of *Truncorotalia truncatulinoides*: A perspective on the morphological species concept in planktonic foraminifera. *Palaeogeography, Palaeoclimatology, Palaeoecology* **391**, 2–12.
- \*R CORE TEAM (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna <https://www.r-project.org/>.
- RILLO, M. C., SUGAWARA, M. T. C., CABELLA, B., JONKERS, L., BARANOWSKI, U. K., KUČERA, M. & EZARD, T. H. G. (2019). On the mismatch in the strength of competition among fossil and modern species of planktonic foraminifera. *Global Ecology and Biogeography* **28**, 1866–1878.
- RILLO, M. C., WOOLLEY, S. & HILLEBRAND, H. (2022). Drivers of global pre-industrial patterns of species turnover in planktonic foraminifera. *Ecography* **2022**, 1–11.
- \*ROGNES, T., FLOURI, T., NICHOLS, B., QUINCE, C. & MAHÉ, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584.
- \*SADEKOV, A. Y., DARLING, K. F., ISHIMURA, T., WADE, C. M., KIMOTO, K., SINGH, A. D., ANAND, P., KROON, D., JUNG, S., GANSEN, G., GANESHAM, R., TSUNOGAI, U. & ELDERFIELD, H. (2016). Geochemical imprints of genotypic variants of *Globigerina bulloides* in the Arabian Sea. *Paleoceanography* **31**, 1440–1452.
- SANDIN, M. M., ROMAC, S. & NOT, F. (2022). Intra-genomic rRNA gene variability of *Nassellaria* and *Spumellaria* (Rhizaria, Radiolaria) assessed by Sanger, MinION and Illumina sequencing. *Environmental Microbiology* **24**, 2979–2993.
- \*SCHMIDTOK, S., JOHNSON, G. C. & LYMAN, J. M. (2013). MIMOC: a global monthly isopycnal upper-ocean climatology with mixed layers. *Journal of Geophysical Research: Oceans* **118**, 1658–1672.
- SCHNELL, I. B., BOHMANN, K. & GILBERT, M. T. P. (2015). Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources* **15**, 1289–1303.
- \*SCHWEIZER, M., PAWLOWSKI, J., KOUWENHOVEN, T. J., GUIARD, J. & VAN DER ZWAAN, B. (2008). Molecular phylogeny of Rotaliida (foraminifera) based on complete small subunit rDNA sequences. *Marine Micropaleontology* **66**, 233–246.
- SEARS, H. A., DARLING, K. F. & WADE, C. M. (2012). Ecological partitioning and diversity in tropical planktonic foraminifera. *BMC Evolutionary Biology* **12**, 1–15.



- SICCHA, M. & KUCERA, M. (2017). ForCenS, a curated database of planktonic foraminifera census counts in marine surface sediment samples. *Scientific Data* **4**, 170109.
- SIERRA, R., MAUFFREY, F., CRUZ, J., HOLZMANN, M., GOODAY, A. J., MAURER-ALCALÁ, X., THAKUR, R., GRECO, M., WEINER, A. K. M., KATZ, L. A. & PAWLOWSKI, J. (2022). Taxon-rich transcriptomics supports higher-level phylogeny and major evolutionary trends in foraminifera. *Molecular Phylogenetics and Evolution* **174**, 1–6.
- SLATKIN, M. (1987). Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792.
- SPALDING, M. D., AGOSTINI, V. N., RICE, J. & GRANT, S. M. (2012). Pelagic provinces of the world: a biogeographic classification of the world's surface pelagic waters. *Ocean & Coastal Management* **60**, 19–30.
- \*SPEZZAFERRI, S., KUCERA, M., PEARSON, P. N., WADE, B. S., RAPPO, S., POOLE, C. R., MORARD, R. & STALDER, C. (2015). Fossil and genetic evidence for the polyphyletic nature of the planktonic foraminifera 'Globigerinoides', and description of the new genus *Tribobatus*. *PLoS ONE* **10**, e0128108.
- STEWART, I. A., DARLING, K. F., KROON, D., WADE, C. M. & TROELSTRA, S. R. (2001). Genotypic variability in subarctic Atlantic planktic foraminifera. *Marine Micropaleontology* **43**, 143–153.
- STRACK, T., JONKERS, L., RILLO, C. M., HILLEBRAND, H. & KUCERA, M. (2022). Plankton response to global warming is characterized by non-uniform shifts in assemblage composition since the last ice age. *Nature Ecology & Evolution* **6**, 1871–1880.
- SUZUKI, N. & NOT, F. (2015). Biology and ecology of Radiolaria. In *Marine Protists: Diversity and Dynamics, 1st Edition*, (eds S. OHTSUKA, T. SUZUKI, T. HORIGUCHI, N. SUZUKI and F. NOT), pp. 179–222. Springer Japan, Tokyo.
- TAKAGI, H., KURASAWA, A. & KIMOTO, K. (2020). Observation of asexual reproduction with symbiont transmission in planktonic foraminifera. *Journal of Plankton Research* **42**, 403–410.
- \*UJHÉ, Y. & ASAMI, T. (2014). Temperature is not responsible for left–right reversal in pelagic unicellular zooplanktons. *Journal of Zoology* **293**, 16–24.
- UJHÉ, Y., ASAMI, T., DE GARIDEL-THORON, T., LIU, H., ISHITANI, Y. & DE VARGAS, C. (2012). Longitudinal differentiation among pelagic populations in a planktic foraminifer. *Ecology and Evolution* **2**, 1725–1737.
- \*UJHÉ, Y., DE GARIDEL-THORON, T., WATANABE, S., WIEBE, P. & DE VARGAS, C. (2010). Coiling dimorphism within a genetic type of the planktonic foraminifer *Globorotalia truncatulinoides*. *Marine Micropaleontology* **77**, 145–153.
- \*UJHÉ, Y. & ISHITANI, Y. (2016). Evolution of a planktonic foraminifer during environmental changes in the tropical oceans. *PLoS ONE* **11**, 1–16.
- \*UJHÉ, Y., KIMOTO, K. & PAWLOWSKI, J. (2008). Molecular evidence for an independent origin of modern triserial planktonic foraminifera from benthic ancestors. *Marine Micropaleontology* **69**, 334–340.
- UJHÉ, Y. & LIPPS, J. H. (2009). Cryptic diversity in planktonic foraminifera in the northwest Pacific ocean. *Journal of Foraminiferal Research* **39**, 145–154.
- \*WADE, C. M., DARLING, K. F., KROON, D. & BROWN, A. J. L. (1996). Early evolutionary origin of the planktonic foraminifera inferred from small subunit rDNA sequence comparisons. *Journal of Molecular Evolution* **43**, 672–677.
- \*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.
- WEBER, A. A.-T. & PAWLOWSKI, J. (2014). Wide occurrence of SSU rDNA intragenomic polymorphism in foraminifera and its implications for molecular species identification. *Protist* **165**, 645–661.
- WEINER, A., AURAHNS, R., KURASAWA, A., KITAZATO, H. & KUCERA, M. (2012). Vertical niche partitioning between cryptic sibling species of a cosmopolitan marine planktonic protist. *Molecular Ecology* **21**, 4063–4073.
- WEINER, A. K. M., MORARD, R., WEINKAUF, M. F., DARLING, K. F., ANDRÉ, A., QUILLÉVÉRÉ, F., UJHÉ, Y., DOUADY, C. J., DE VARGAS, C. & KUCERA, M. (2016). Methodology for single-cell genetic analysis of planktonic foraminifera for studies of protist diversity and evolution. *Frontiers in Marine Science* **3**, 1–15.
- WEINER, A. K. M., SEHEIN, T., COTE-L'HEUREUX, A., SLEITH, R. S., GRECO, M., MALEKSHAHI, C., RYAN-EMBRY, C., OSTRICKER, N. & KATZ, L. A. (2023). Single-cell transcriptomics supports presence of cryptic species and reveals low levels of population genetic diversity in two testate amoebae morphospecies with large population sizes. *Evolution* **77**, 2472–2483.
- WEINER, A. K. M., WEINKAUF, M. F. G., KURASAWA, A., DARLING, K. F. & KUCERA, M. (2015). Genetic and morphometric evidence for parallel evolution of the *Globigerinella calida* morphotype. *Marine Micropaleontology* **114**, 19–35.
- WEINER, A. K. M., WEINKAUF, M. F. G., KURASAWA, A., DARLING, K. F., KUCERA, M. & GRIMM, G. W. (2014). Phylogeography of the tropical planktonic foraminifera lineage *Globigerinella* reveals isolation inconsistent with passive dispersal by ocean currents. *PLoS ONE* **9**, e92148.
- WEINKAUF, M. F. G., SICCHA, M. & WEINER, A. K. M. (2022). Reproduction dynamics of planktonic microbial eukaryotes in the open ocean. *Journal of the Royal Society Interface* **19**, 1–11.
- \*WICKHAM, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
- \*WICKHAM, H., AVERICK, M., BRYAN, J., CHANG, W., MCGOWAN, L., FRANÇOIS, R., GROLEMUND, G., HAYES, A., HENRY, L., HESTER, J., KUHN, M., PEDERSEN, T., MILLER, E., BACHE, S., MÜLLER, K., ET AL. (2019). Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686.
- \*WUYTS, J., VAN DE PEER, Y. & DE WACHTER, R. (2001). Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acids Research* **29**, 5017–5028.
- YASUHARA, M., WEI, C. L., KUCERA, M., COSTELLO, M. J., TITTENSOR, D. P., KIESSLING, W., BONEBRAKE, T. C., TABOR, C. R., FENG, R., BASELGA, A., KRETSCHMER, K., KUSUMOTO, B. & KUBOTA, Y. (2020). Past and future decline of tropical pelagic biodiversity. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 12891–12896.
- ZHANG, J., KAPLI, P., PAVLIDIS, P. & STAMATAKIS, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869–2876.

## XI. SUPPORTING INFORMATION

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

**Appendix S1.** Detailed methods.

**Appendix S2.** Individual maps of occurrence of each sampled morphospecies and constitutive MOTUs in our data set for the three levels of the molecular taxonomy; the numbers in the hexagons represent the number of observations at independent stations within the area.

**Table S1.** Planktonic Foraminifera Ribosomal Reference (PFR<sup>2</sup>) database v.2., with collection details, taxonomic attribution, and Sanger sequences of planktonic foraminifera.

**Table S2.** Genotyping data of the species *Orbulina universa*, *Globorotalia truncatulinoides*, *Globorotalia inflata*, *Globigerina bulloides*, *Pulleniatina obliquiloculata*, *Neogloboquadrina pachyderma* and *Globigerinella siphonifera* with collection details.

**Table S3.** Updated version of the PR2\_V9 reference database used in this study to re-assign the metabarcoding data set of Cordier *et al.* (2022). The description of the fields can be found at: <https://zenodo.org/records/376895>.

**Table S4.** Collection details of samples included in the study consisting of geographic coordinates, date, depth, size fraction associated with biogeographic region and climatology data.

**Table S5.** Molecular nomenclature resulting from the automated delimitation methods, with the correspondence of the newly defined molecular operational taxonomic units (MOTUs) with the labels existing in the literature and the genotyping data also provided.

**Table S6.** Occurrence table of planktonic foraminifera amplicon sequence variants (ASVs) in the metabarcoding data set of Cordier *et al.* (2022) with their sequences and taxonomic assignment to the molecular nomenclature.

**Table S7.** List of the occurrences of molecular operational taxonomic units (MOTUs) at each level of the nomenclature in the data set for individual samples.

(Received 31 August 2023; revised 4 February 2024; accepted 7 February 2024; published online 13 February 2024)