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 Morard1,*, Kate F. Darling2,3, Agnes K. M. Weiner4, Christiane Hassenrück5, Chiara Vanni1, Tristan Cordier1, Nicolas Henry6,7, Mattia Greco8, Nele M. Vollma1,4, Tamara Milivojevic1,9, Shirin Nurshan Rahman1, Michael Sicchaa1, Julie Meillard1, Lukas Jonkers1, Frédéric Quillévéré10, Gilles Escarguel11, Christophe J. Douady11,12, Thibault de Garidel-Thoron13, Colomban de Vargas6,14 and Michal Kucera1

1MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, Bremen 28359, Germany
2School of GeoSciences, University of Edinburgh, Edinburgh, EH9 3JW, UK
3Biological and Environmental Sciences, University of Stirling, Stirling, FK9 4LA, UK
4NORCE Climate and Environment, NORCE Norwegian Research Centre AS, Bjerkes Centre for Climate Research, Jahnebakken 5, Bergen 5007, Norway
5Biological Oceanography, Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Seestrasse 15, Warnemünde 18119, Germany
6CNRS, Sorbonne Université, FR2424, ABiMS, Station Biologique de Roscoff, Roscoff 29680, France
7Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE, 3 rue Michel-Ange, Paris 75016, France
8Institut de Géosciences del Mar, Passeig Marítim de la Barceloneta, Barcelona 37-49, Spain
9Max Planck Institute for Marine Microbiology, Bremen, Germany
10Univ Lyon, Université Claude Bernard Lyon 1, ENS de Lyon, CNRS, UMR CNRS 5276 LGL-TPE, Villeurbanne F-69622, France
11Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTP, UMR 5023 LEHNA, Villeurbanne F-69622, France
12Institut Universitaire de France, Paris, France
13Aix Marseille Univ, CNRS, IRD, INRAE, CEREGE, Aix-en-Provence, France
14Sorbonne 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: r morard@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* 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 & Altermann, 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 (Brumman & Kucera, 2022). The outstanding preservation of these relatively large calcifying protists (50 μm–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 & Kucera, 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, Neogloborubinella pachyderma, and Turborotalita quinqueloba (Darling et al., 2000, 2004). Subsequent studies reported extensive genetic diversity in the investigated morphospecies (Darling, Kucera & Wade, 2007; Ujiié & Lipps, 2009; Aurahs et al., 2009b; Morard et al., 2011; Ujiié 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 – PFR2 v.2

We compiled all 18S rDNA sequences of planktonic foraminifera available to date to update the Planktonic Foraminifera Ribosomal Reference database (PFR2; Morard et al., 2015). We gathered all sequences released on National Center for Biotechnology Information after constructing the PFR2 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 PFR2 v.2 database...
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 *Pulmo inquilina 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 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...
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 (~150 base pairs (bp) versus ~1000 bp), resulting in reduced taxonomic information. Moreover, the lack of linkage between sequences and singular specimens poses challenges in
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 hypervariable 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 vivax* 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. vivax* 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 base-group (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 PFR² v.2 database were used to build the molecular taxonomy, and the remaining sequences and genotyping data were assigned to the finest possible level
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

**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.

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.

<table>
<thead>
<tr>
<th>Morphospecies</th>
<th>Existing nomenclature presented in</th>
<th>Existing nomenclature</th>
<th>Change in taxonomy</th>
<th>Current nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentigrabolaria anfracta</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Tenuitellita fleischeri</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinita glutinata</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IIa</td>
<td>No</td>
<td>IIa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IIIa</td>
<td>No</td>
<td>IIIa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IIb</td>
<td>No</td>
<td>IIb</td>
</tr>
<tr>
<td>Tenuitellita iota</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinita minuta</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>IIa</td>
<td>No</td>
<td>IIa</td>
</tr>
<tr>
<td>Candea nitida</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinita uvula</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Neogallitella vivans</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globoquadrina conglobulata</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globoquadrina duriatensis</td>
<td>Darling et al. (2003)</td>
<td>Ia</td>
<td>Lumping</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ib</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>lc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Globoquadrina estropacia</td>
<td></td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
<tr>
<td>Globoquadrina hexagonas</td>
<td>Morard et al. (2019)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globoquadrina hirsuta</td>
<td></td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
<tr>
<td>Neogloboquadrina inompta</td>
<td>Darling et al. (2000)</td>
<td>I</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globorotalite inflata</td>
<td>Morard et al. (2011)</td>
<td>I</td>
<td>Nomenclature</td>
<td>Ia</td>
</tr>
<tr>
<td>Morard et al. (2011)</td>
<td>I</td>
<td>II</td>
<td>Nomenclature</td>
<td>Ib</td>
</tr>
<tr>
<td>Pulmonatina obliquiloculata</td>
<td>Ujiié et al. (2012)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td>Ujiié et al. (2012)</td>
<td>IIa</td>
<td>Lumping</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td>Ujiié et al. (2012)</td>
<td>IIb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neogloboquadrina pachyderma</td>
<td>Darling et al. (2004)</td>
<td>I</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2007)</td>
<td>II</td>
<td>Nomenclature</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2007)</td>
<td>III</td>
<td>Lumping</td>
<td>Ib</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2007)</td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2007)</td>
<td>VI</td>
<td>Nomenclature</td>
<td>Ic</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2007)</td>
<td>IV</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2007)</td>
<td>VII</td>
<td>Nomenclature</td>
<td>IIIa</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2017)</td>
<td>VIII</td>
<td>Nomenclature</td>
<td>Va</td>
</tr>
<tr>
<td>Globoquadrina scitula</td>
<td></td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
<tr>
<td>Globoquadrina truncatulinoides</td>
<td>de Vargas et al. (2001)</td>
<td>I</td>
<td>Nomenclature</td>
<td>Ia</td>
</tr>
<tr>
<td>de Vargas et al. (2001)</td>
<td>I</td>
<td>II</td>
<td>Nomenclature</td>
<td>Ib</td>
</tr>
<tr>
<td>Quillévévé et al. (2013)</td>
<td>V</td>
<td>III</td>
<td>Nomenclature</td>
<td>Ia</td>
</tr>
<tr>
<td>de Vargas et al. (2001)</td>
<td>III</td>
<td>IV</td>
<td>Nomenclature</td>
<td>Ic</td>
</tr>
<tr>
<td>Globoquadrina tumida</td>
<td></td>
<td>–</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globoquarina bulboides</td>
<td>Darling et al. (2000, 2003)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td></td>
<td>Kucera &amp; Darling (2002); Darling et al. (2003)</td>
<td>Ib</td>
<td>No</td>
<td>Ib</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2017)</td>
<td>Ic</td>
<td>No</td>
<td>Ic</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2017)</td>
<td>Id</td>
<td>No</td>
<td>Id</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2000)</td>
<td>IIa</td>
<td>No</td>
<td>IIa</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2000)</td>
<td>IIb</td>
<td>No</td>
<td>IIb</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2000)</td>
<td>IIc</td>
<td>No</td>
<td>IIc</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2000)</td>
<td>IId</td>
<td>No</td>
<td>IId</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2007)</td>
<td>IIe</td>
<td>No</td>
<td>IIe</td>
</tr>
<tr>
<td></td>
<td>Darling et al. (2017)</td>
<td>IIf</td>
<td>No</td>
<td>IIf</td>
</tr>
<tr>
<td>Globigerinella calida</td>
<td>Weiner et al. (2014)</td>
<td>IIIa</td>
<td>Nomenclature</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinoides conglobulatus</td>
<td>Morard et al. (2013a)</td>
<td>Ia1</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Morard et al. (2013a)</td>
<td>Ia1</td>
<td>Ib1</td>
<td>No</td>
<td>Ib</td>
</tr>
<tr>
<td>Sphaerosidinella dehiscens</td>
<td></td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
<tr>
<td>Beella digitata</td>
<td></td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
</tbody>
</table>

(Continues on next page)
Table 1. (Cont.)

<table>
<thead>
<tr>
<th>Morphospecies</th>
<th>Existing nomenclature presented in</th>
<th>Existing nomenclature</th>
<th>Change in taxonomy</th>
<th>Current nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hastigerinella digitata</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinoides elongatus</td>
<td>Morard et al. (2019a)</td>
<td>Ia1</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinoides falconensis</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
<tr>
<td>Turborotalita humiles</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
<tr>
<td>Hastigerina pelagica</td>
<td>Weiner et al. (2012)</td>
<td>I</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Turborotalita quinquelonga</td>
<td>Darling et al. (2000); Darling &amp; Wade (2000)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Turborotalita quinquelonga</td>
<td>Darling &amp; Wade (2000)</td>
<td>Ib</td>
<td>No</td>
<td>Ib</td>
</tr>
<tr>
<td>Turborotalita quinquelonga</td>
<td>Darling et al. (2000)</td>
<td>Ila</td>
<td>Lumping</td>
<td>Ia</td>
</tr>
<tr>
<td>Turborotalita quinquelonga</td>
<td>Darling et al. (2000)</td>
<td>IIC</td>
<td>No</td>
<td>IIC</td>
</tr>
<tr>
<td>Turborotalita quinquelonga</td>
<td>Darling et al. (2017)</td>
<td>IId</td>
<td>Nomenclature</td>
<td>IId</td>
</tr>
<tr>
<td>Globigerinella radians</td>
<td>Weiner et al. (2014)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinoides ruber albus</td>
<td>Morard et al. (2019a)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinoides ruber ruber</td>
<td>Morard et al. (2019a)</td>
<td>Ia1</td>
<td>Nomenclature</td>
<td>Ia1</td>
</tr>
<tr>
<td>Globigerinoides ruber ruber</td>
<td>Morard et al. (2019a)</td>
<td>Ia</td>
<td>Nomenclature</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinoides ruber ruber</td>
<td>Morard et al. (2019a)</td>
<td>Ia</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Trilobatus sacculifer</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinella siphonifera</td>
<td>Weiner et al. (2014)</td>
<td>Ib</td>
<td>Nomenclature</td>
<td>Ib</td>
</tr>
<tr>
<td>Globigerinella siphonifera</td>
<td>Weiner et al. (2015)</td>
<td>Ia1</td>
<td>Splitting</td>
<td>Ia1</td>
</tr>
<tr>
<td>Globigerinella radicans</td>
<td>Weiner et al. (2015)</td>
<td>G. siphonifera Ia</td>
<td>Nomenclature</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinella calida</td>
<td>Weiner et al. (2015)</td>
<td>G. siphonifera III</td>
<td>Nomenclature</td>
<td>Ia</td>
</tr>
<tr>
<td>Globigerinoides tenella</td>
<td>Morard et al. (2019a)</td>
<td>Ia1</td>
<td>No</td>
<td>Ia</td>
</tr>
<tr>
<td>Orbulina universa</td>
<td>de Vargas et al. (1999)</td>
<td>III</td>
<td>Nomenclature</td>
<td>III</td>
</tr>
</tbody>
</table>

(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; Barrenchea 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
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 metbarcoding 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 metbarcoding 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 metbarcoding 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, Lakeš & Horák, 2023). Considering the extensiveness of the metbarcoding 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 intermorphospecies 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 new novel MOTUs lvl-1, one novel MOTU lvl-1 of Turborotalita quinqueloba (named I), 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 Tentaculitida–Globigerinita wuallis 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 Tentaculitina includes three morphospecies and only two were covered by our barcoding efforts (T. iota and T. flexilis), 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 Beilla 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 cressafornis, Orcadia riedeli, Berggrenia pumilio, Globigerinella adamsi and Globorotaloides ovata; 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 barcoded 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 Pahuminia 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
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

---

**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.
the transition from the last glacial maximum to the Holocene from 19 to \( \sim 11.5 \) 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 MOTU 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.
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°/C14 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 (Ujié 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
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.
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* IId, *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. 10. Diversity, biogeography and thermal niches of MOTUs lvl-2 of the non-spinose clade. Other details are as in Fig. 9.](image)

![Fig. 11. Diversity, biogeography and thermal niches of MOTUs lvl-2 of the microperforate clade. Other details are as in Fig. 9.](image)
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* Ia and *G. inflata* Ib, which occur in the southern hemisphere only. We also observe bipolar MOTUs for the morphospecies *T. quinqueloba* Ia–Ic, *G. bulloides* Ia 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 (Brunner & Kucera, 2022). This morphological-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; Secars, 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’ Ia and Ib of *G. ruber* (Aurahs et al., 2009b) match the morphological diagnoses of *G. elongatus* (Aurahs et al., 2011) and *G. tenella* (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. radiana*, *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 (Andrè 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 & Pawlowski, 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 µm) 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 (χ² = 50.3, d.f. = 21, *P* = 0.0003) and non-spinose clade (χ² = 32.3, d.f. = 13, *P* = 0.002) but not for the microperforate clade (χ² = 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 (χ² = 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. conglobata*, *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* IId, *G. ruber albus* Ia, *P. obliquiloculata* Ia, 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. truncatulinoides* 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-latitudes 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
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 multi-gene 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.
Revisions to the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology* 66, 4–119.


Chapter six planktonic foraminifera as tracers of past oceanic environments.

For a better characterization of the fossil pelagic record:

Molecular evidence for widespread occurrence of foraminifera in soils.

Surface ocean metabarcoding conforms limited diversity in planktonic foraminifera but reveals unknown hyper-abundant lineages.

High variability in SSU rDNA gene copy number among planktonic foraminifera revealed by single-cell qPCR.

Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation.

Molecular phylogenetics of cryptic species in the planktonic foraminifer Globorotalia inflata: implications for life history and paleoceanography.

Unassigned diversity of planktonic foraminifera from environmental sequencing revealed as known but neglected species. PLoS ONE 14, e0213956.


The Journal of Eukaryotic Microbiology 57, 197–203.


The Journal of Foraminiferal Research 41, 15–32.

The open Atlantic coast of Europe.


The Journal of Foraminiferal Research 45, 15–32.

Biodiversity and spatial patterns in the Clarion-Clipperton zone and other abyssal regions: insights from sediment DNA and RNA metabarcoding. PLoS ONE 16, e26665.


A perspective on the diversity of planktonic foraminifera from environmental sequencing revealed as known but neglected species. PLoS ONE 14, e0213956.

Multiproxy approach for the reconstruction of the glacial ocean surface conditions in the open Atlantic coast of Europe.

Developments in Marine Geology 87, 67–91.


Foraminifera: molecules meet morphology.

Ecological partitioning and biogeographic diversity in the world’s ocean. Proceedings of the National Academy of Science 113, E1516–E1525.

This study presents the first demonstration of a multiplex metabarcoding approach on the entire planktonic foraminifera species diversity. Université Claude Bernard Lyon I.


The Journal of Foraminiferal Research 45, 15–32.


The Journal of Foraminiferal Research 43, 55–70.


Research Paper 8, 2539.

BMC Evolutionary Biology 12, 1–5.
**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 thethree 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) 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*, *Pallénula obliquiloculata*, *Neogloboquadrina pachyderma* and *Globigerinoides 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.

---

**Reference:**


