

Renewal of planktonic foraminifera diversity after the Cretaceous Paleogene mass extinction by benthic colonizers

Received: 9 February 2022

Accepted: 8 November 2022

Published online: 21 November 2022



Raphaël Morard ¹✉, Christiane Hassenrück ^{1,2}, Mattia Greco ³, Antonio Fernandez-Guerra ⁴, Sylvain Rigaud⁵, Christophe J. Douady^{6,7} & Michal Kucera ¹

The biotic crisis following the end-Cretaceous asteroid impact resulted in a dramatic renewal of pelagic biodiversity. Considering the severe and immediate effect of the asteroid impact on the pelagic environment, it is remarkable that some of the most affected pelagic groups, like the planktonic foraminifera, survived at all. Here we queried a surface ocean metabarcoding dataset to show that calcareous benthic foraminifera of the clade Globobulimina are able to disperse actively in the plankton, and we show using molecular clock phylogeny that the modern planktonic clades originated from different benthic ancestors that colonized the plankton after the end-Cretaceous crisis. We conclude that the diversity of planktonic foraminifera has been the result of a constant leakage of benthic foraminifera diversity into the plankton, continuously refueling the planktonic niche, and challenge the classical interpretation of the fossil record that suggests that Mesozoic planktonic foraminifera gave rise to the modern communities.

Despite the pivotal role of the fossil record of planktonic foraminifera in revealing past climates and studying plankton evolution, the origin of the group remains elusive. Conflicting evidence exists between paleontological and molecular studies on how the planktonic foraminifera emerged and diversified. The earliest record of planktonic foraminifera dates back to the early Jurassic, where their appearance is thought to have been the response to widespread oceanic anoxia¹. The group became diverse and abundant in the early Cretaceous and in most paleontological phylogenies^{2,3}, the main extant clades are traced back to that time. However, molecular genetic data imply that the extant foraminifera in the plankton may be the result of repeated invasions from the benthos. This is supported by the independent colonization of the planktonic niche by the triserial *Gallitellia* during the Miocene⁴ and the ongoing transition into the plankton observed in the biserial *Bolivina*^{5,6}. Understanding the origin of the planktonic

foraminifera is of key importance because their stratigraphic record is used to study the interplay between diversity and climate⁷ and past biological crises⁸. So far, these studies assume that the fossil record of planktonic foraminifera represents the waxing and waning of diversity generated by speciation and extinction within long-ranging clades that were able to survive past environmental crises⁷. If planktonic foraminifera represent different clades bestowed with different life traits and if the pelagic niche has been repopulated repeatedly from the benthos, the interpretation of the biotic response of planktonic foraminifera to environmental upheavals throughout the Meso- and Cenozoic would have to be fundamentally reassessed.

The fossil record of the extant foraminifera is largely represented by agglutinated and calcareous forms belonging to the clades Globobulimina, Tubobulimina, and Lagenida⁹, but environmental surveys show that globally the diversity in the group is dominated by the naked

¹MARUM Center for Marine Environmental Sciences, University of Bremen, Leobener Strasse, 28359 Bremen, Germany. ²Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Seestrasse 15, 18119 Rostock-Warnemünde, Germany. ³Institute of Oceanology, Polish Academy of Sciences, Sopot, Poland. ⁴Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, Copenhagen 1350K, Denmark. ⁵52 chemin de Claret, F-05700 Serres, France. ⁶Université Lyon, Université Claude Bernard Lyon 1, CNRS, UMR 5023, ENTPE, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, Villeurbanne, France. ⁷Institut Universitaire de France, 103 Boulevard Saint-Michel, 75005 Paris, France. ✉e-mail: rmorard@marum.de

Monothalamea¹⁰ that likely emerged during the Precambrian¹¹. One surprising aspect of the evolution of foraminifera is that many key innovations, such as biomineralization, emerged repeatedly and independently in the group (Fig. 1). Yet, despite the existence of the group since the Precambrian¹¹ and of biomineralization at least since the upper Devonian¹², foraminifera colonized the plankton only in the Jurassic and all of their planktonic representatives appear to belong to one clade of the Globothalamea, the Rotaliida. Could it be that unlike many other innovations in the group, the transition into the plankton only occurred once? With the emergence of large metabarcoding surveys¹³ and improved coverage of barcode references¹⁴, the occurrence and identity of foraminifera in the plankton can now be studied from environmental DNA sequences, which would reveal which lineages, irrespective of size and presence of shells, dwell in the plankton. Here, we show that the benthic foraminifera of the clade Globothalamea can disperse actively in the plankton and form independent holoplanktonic clades that renewed the diversity of the group in the pelagic realm after the end-Cretaceous biological crisis.

Results and discussion

Living benthos in the plankton

To evaluate the diversity of foraminifera in the plankton, we re-analyzed the comprehensive catalog of eukaryotic diversity of the TARA Ocean dataset¹³. We queried the environmental dataset against a reference database updated with ~2000 additional foraminifera reference sequences with curated taxonomy that include 40 of the 45 morphological species of holoplanktonic foraminifera and all main clades of benthic foraminifera, albeit with Lagenida underrepresented. We retrieved 1094 Molecular Operational Taxonomic Units (MOTUs) with at least 90% identity with the updated reference database that accounted for 1,157,287 sequences and occurred in 1002 of the

1048 samples of the TARA Ocean dataset (Supplementary Fig. 1). From those MOTUs, 346 (32%) were attributed to known holoplanktonic clades, representing 83% of the sequences (Fig. 2A). The remaining 748 MOTUs (68%) were attributed to foraminifera clades that are only known from the benthos, with the largest part belonging to the Globothalamea (416 MOTUs).

Finding DNA of benthic foraminifera in pelagic samples is not surprising. Some benthic microorganisms will always be found in the plankton because of the passive entrainment of sediment particles during storms or similar events. However, such inadvertent inhabitants of the plankton would rapidly decrease in abundance away from the source of their advection. Indeed, the likelihood of observing MOTUs of each holoplanktonic clade in our analysis does not decrease with distance from the coast, whereas the benthic Tubothalamea and “Monothalamea” OTUs always become rarer away from the coast (Fig. 2C). Nonetheless, the “benthic” Globothalamea contain some MOTUs that also persist further off-shore and whose occurrence in the plankton thus cannot be explained by passive entrainment.

To unravel which of the Globothalamea MOTUs appeared plankton-like, we carried out the analysis of their occurrence as a function of distance from shore at the level of single MOTUs (Fig. 2D). This is a simplified model because other environmental parameters are likely to control the presence or absence of individual MOTUs, but it allowed us to differentiate MOTUs that may be passively transported compared to those capable to persist in the plankton. Indeed, the vast majority of MOTUs assigned to the holoplanktonic clades showed an increasing or constant probability of occurrence with greater distance to the coast. Inversely, the MOTUs belonging to Monothalamea and Tubothalamea mostly displayed decreasing probability of occurrence with a greater distance to the coast. Based on this observation, we used the holoplanktonic MOTUs on the one hand, and the Monothalamea and

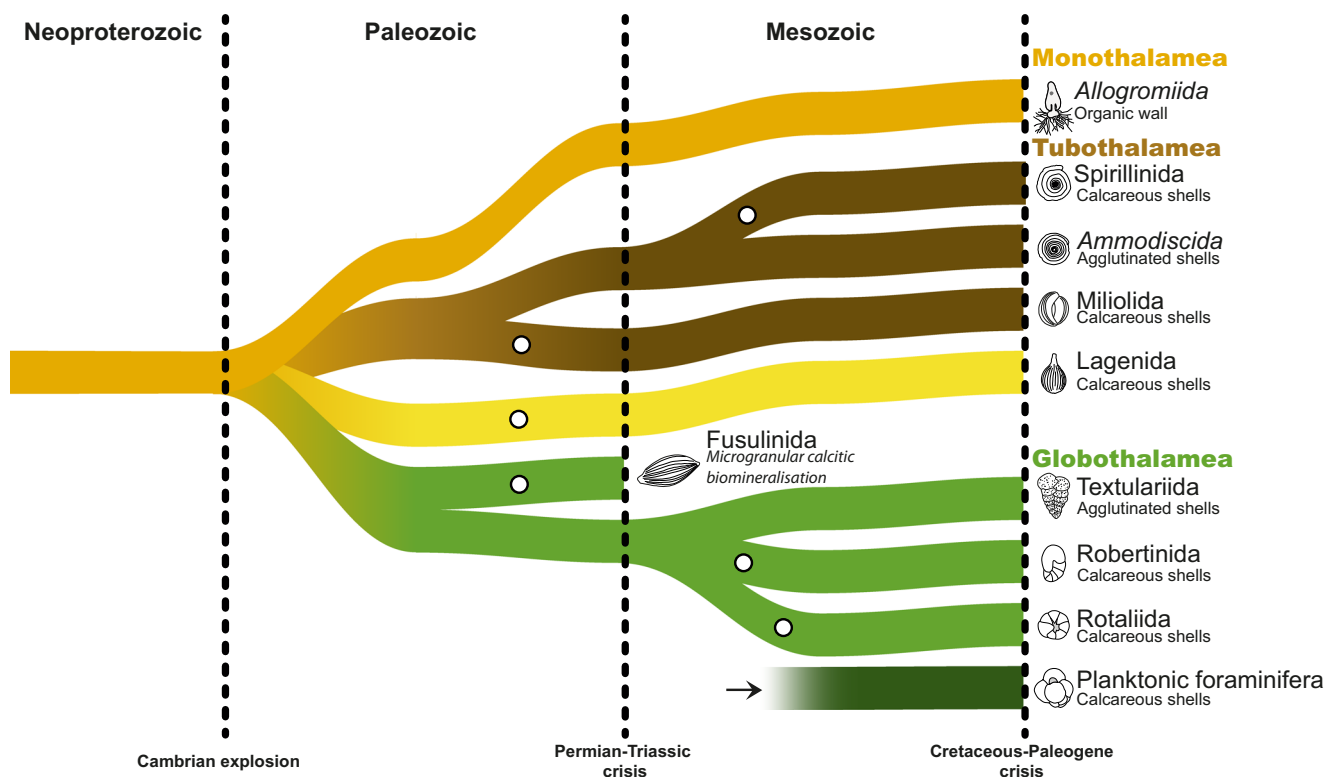


Fig. 1 | Simplified evolutionary history of foraminifera. The scheme depicts the diversification of foraminifera from their emergence in the Precambrian until the end of the Mesozoic with a simplified stratigraphy. The polytomies reflect uncertainties in the evolutionary relationships between the depicted lineages. The Globothalamea and Tubothalamea have a class rank and the status of the

Monothalamea is uncertain. The relationship of Lagenida to other foraminifera lineages is unclear but it is a distinct lineage from the three other main clades. The open circles approximately indicate when completely biomineralized shells appeared in each lineage. The figure is based on refs. 9, 11, 57, 58.

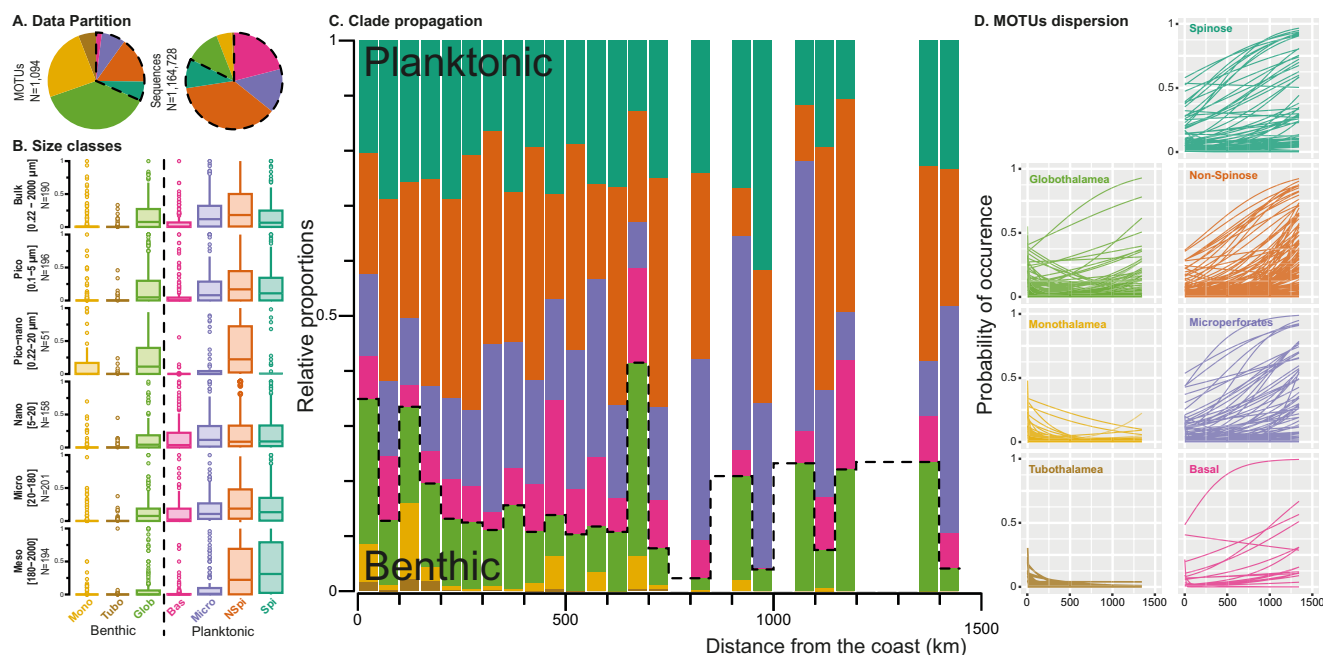


Fig. 2 | Dispersion of benthic foraminifera in the plankton. A Relative contribution of benthic and planktonic clades to MOTUs and sequences in the TARA Ocean dataset. **B** Sequence proportions of the main foraminifera clades across size fractions of the TARA Ocean dataset (Boxplots display the interquartile range (box), extrema (whiskers), and outliers (dots)). **C** Average sequence proportions of

planktonic and benthic clades with distance from the coast in the TARA Ocean dataset. **D** Probability of occurrence of individual MOTUs of planktonic clades as a function of distance to coast based on their absence/presence in the TARA Ocean dataset. The data used to produce the figure are provided in Supplementary Data 2.

Tubothalamea on the other hand, to train two complementary supervised Random Forest models to classify the Globothalamea MOTUs as either “planktonic” or “benthic” according to (i) their presence/absence patterns and (ii) probability of occurrence (Fig. 3A). The Random Forest models performed with an accuracy of 94.4% (i) and 96.9% (ii) associated with an F1 score of 0.8 (i) and 0.9 (ii), respectively, and were in agreement in 71% of the predictions, representing 46 Globothalamea MOTUs that were considered as “putative planktonic”, i.e., showing a pattern of occurrence in the plankton, which is consistent with the pattern of occurrence of planktonic taxa.

To determine the taxonomic affiliation of the Globothalamea MOTUs found in the plankton, we produced a backbone phylogeny of Globothalamea that encompassed all documented families of the rotaliids¹⁵ and mapped the MOTUs using a phylogenetic placement approach (Fig. 3B). This analysis revealed that the ability to persist in the plankton is widespread among the Globothalamea and should therefore be considered a synapomorphy of the clade. The occurrence of the “putative planktonic” Globothalamea across all size classes in the TARA dataset implies (Fig. 2B) that the ability to persist in the plankton occurs among adult individuals and is not limited to gametes or juveniles. With the exception of the tychopelagic *Bolivina*⁵, actual specimens (or their shells) of the “putative planktonic” Globothalamea have only rarely been reported from the plankton¹⁶. This is consistent with the small share of these MOTUs among the sequences and suggests that the ability to persist in the plankton may be a dispersal strategy rather than a persistent lifestyle.

A widespread ability to disperse in the plankton, being able to remain buoyant and feed in the plankton⁵ provides an obvious stepping stone on the transition from the benthos into the plankton¹⁷. However, this innovation was apparently in itself not sufficient to develop a full holoplanktonic lifestyle. Since among the Globothalamea, only lineages with biomineralized shells have completed the full transition to the planktonic lifestyle, the second stepping stone on the transition into the plankton may have been the ability to secrete mineralized shells. The reason why biomineralisation would be the key

to holoplanktonic lifestyle among the Globothalamea may be simple: we note that the non-biomineralising Globothalamea (textulariids) build their shells by agglutinating sediment particles and such particles are not present in the plankton, possibly preventing holoplanktonic lifestyle in the absence of complete biomineralization.

Origin of the extant holoplanktonic foraminifera clades

The evolutionary history of planktonic foraminifera is typically presented as a narrative of a Mesozoic origin and a history of extinction and radiation events leading to their modern diversity^{3,8}. This narrative is supported by the observed continuous occupation of the planktonic niche by planktonic foraminifera since their emergence in the Jurassic. It implies that the transition into the plankton occurred once, or was rare afterward, and that the main clades survived all environmental and biotic crises at least since the early Cretaceous where the fossil record is resolved, including the Cretaceous–Paleogene mass extinction⁸. Considering that the two main prerequisites for conquering the plankton, the ability to disperse in the plankton and the ability to produce biomineralized shells, were widespread among the Globothalamea, it would be surprising that this transition did not occur repeatedly.

The extant planktonic foraminifera have long been considered to represent descendants of Cretaceous lineages, with the two most diverse clades (the macroperforate Globigerinidae and Globorotaliidae) originating from among the rare survivors of the K/Pg crisis². However, there exists no independent evidence for a common origin of the extant holoplanktonic clades, not even for a common origin of the two macroperforate clades. Having barcoded the entire diversity of the extant planktonic foraminifera using their SSU rRNA gene, we can now ask whether or not their molecular phylogeny supports a shared common ancestry hypothesis. To this end, we constructed a reference phylogeny using representative sequences of major benthic families of the Globothalamea, adding sequences representing all holoplanktonic foraminifera clades, but excluding long-branch taxa (see “Methods”).

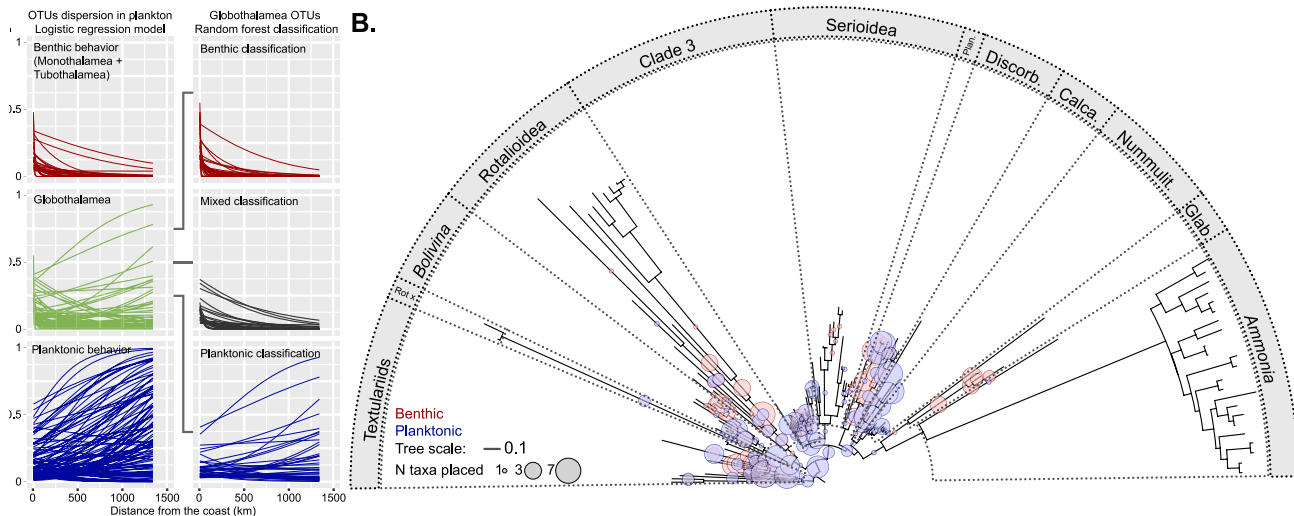


Fig. 3 | Identification of “putative planktonic” Globothalamea. A Random Forest classification of the Globothalamea MOTUs as either planktonic or benthic. **B** Phylogenetic placement of the MOTUs on the Globothalamea tree. The main clades of benthic foraminifera are delimited on the tree. The size of the circles represent the number of potential MOTUs affiliated with a branch. Note that the

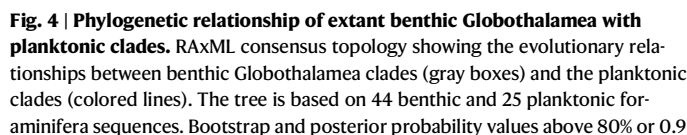
genera *Ammonia* and *Bolivina* are branching out of their home clades (Serioidea and Rotalioidea). Rot X = Rotaliida X, Plan. = Planorbuloidea, Calc. = Calcarinoidea, Nummulit. = Nummulitoidea, Glab. = Glabratelloidea. The data used to produce the figure are provided in Supplementary Data 2 for panel A and Supplementary Data 3 for panel B.

The phylogenetic inferences (Fig. 4) revealed topologies with all Globothalamea families resolved and all three main holoplanktonic clades being monophyletic (Supplementary Fig. 2). The holoplanktonic clades, however, branch in different parts of the tree, implying at least five independent origins of the extant holoplanktonic foraminifera. This topology is consistent with earlier molecular phylogenies¹⁸, but since the bootstrap support is not sufficient to robustly resolve their phylogenetic relationships, we subsequently formally tested two alternative hypotheses for the origin of the extant planktonic foraminifera. A Swofford Olsen Waddell Hillis test implemented in SOWHAT¹⁹ confirmed that it is unlikely that all the extant planktonic foraminifera are monophyletic ($P = 0.001$), or that even the two macroperforate lineages are monophyletic ($P = 0.003$). A similar result is obtained using the Approximately Unbiased (AU) test²⁰, which rejected the monophyly of all clades ($P < 0.05$ irrespective of internal topology, Supplementary Fig. 3) and provided no support for monophyly of the two macroperforate clades ($P = 0.15$ against $P = 0.91$ for the topology in Fig. 4). This would explain why the two main clades of planktonic foraminifera that diversified after the K-Pg boundary have stark morphological differences (with and without spines), which was difficult to reconcile with a common ancestry. Clearly, a complete or partial monophyletic origin of the holoplanktonic clades is not compatible with the molecular dataset and the extant diversity of planktonic foraminifera is therefore more likely the result of multiple independent invasions of the plankton from different benthic ancestors.

Next, we can use phylogenetic inference to estimate the time of divergence of each of the extant planktonic clades from their nearest extant benthic relative. Because the foraminifera has different rates of evolution²¹, we used a relaxed clock model and applied multiple dates to calibrate the molecular tree. Since the topologies of the dated trees are not supported, we cannot assume that the nodes between the sister benthic and planktonic clades represent their closest, thus youngest, last common ancestry. However, since there is no evidence for any of the known fossil or extant clades of planktonic foraminifera to have returned back to the benthos, the last common ancestor of an extant planktonic lineage and any relative from among the extant benthic foraminifera must have been benthic and therefore the age of

the inferred divergence is informative, irrespective of the support for the topology. In fact, it is important to stress that the resulting divergence age estimates provide the maximum ages of a transition into the plankton in each clade. This is because the earliest representative of a lineage leading to the planktonic clade could have still been benthic and the transition into the plankton may have occurred later. Here we retain the divergence age from the nearest benthic ancestor as a conservative estimate of the benthic-planktonic transition and use these conservative estimates to discuss the compatibility between fossil and molecular evidence concerning the evolution of the modern planktonic foraminifera (see Fig. 4 and “Methods” for details). The results reveal that the mean benthic divergence age estimates for the four main clades all cluster around the K/Pg boundary, and that the most recently diverged *D. anfracta* and *G. vivans* very likely colonized the plankton later, during the last 30 Ma of the Cenozoic (Fig. 5). The time-calibrated phylogenies based on the maximum likelihood and Bayesian inferences returned essentially the same results (Supplementary Data 5). The large uncertainties on the divergence age estimates reflect the heterogeneity in substitution rates, and are a common feature for time-calibrated trees even for inferences based on phylogenomic datasets that have perfect branch support and include sequences from hundreds of genes²². However, even with these large uncertainties, the divergence age estimates for all four main extant clades are incompatible with their origin from the Cretaceous planktonic foraminifera.

If the extant clades originated among the Cretaceous planktonic foraminifera, their divergence from nearest benthic ancestors must have occurred before the planktonic lineages emerged in the fossil record. Since the fossil record of Cretaceous planktonic foraminifera is well documented³, and it is well known that these lineages have been planktonic throughout their existence²³, their divergence from the benthos must have occurred in the Early Cretaceous at the latest. Specifically, the macroperforate Cenozoic planktonic foraminifera are considered to have originated from two species of *Hedbergella*, which appear to have survived the K/Pg crisis²⁴. However, the *Hedbergella* lineage can be traced in the fossil record to 140 Ma, implying a divergence age from the benthos, which is older than any of the molecular clock estimates for the benthic divergence of the extant taxa, even



5

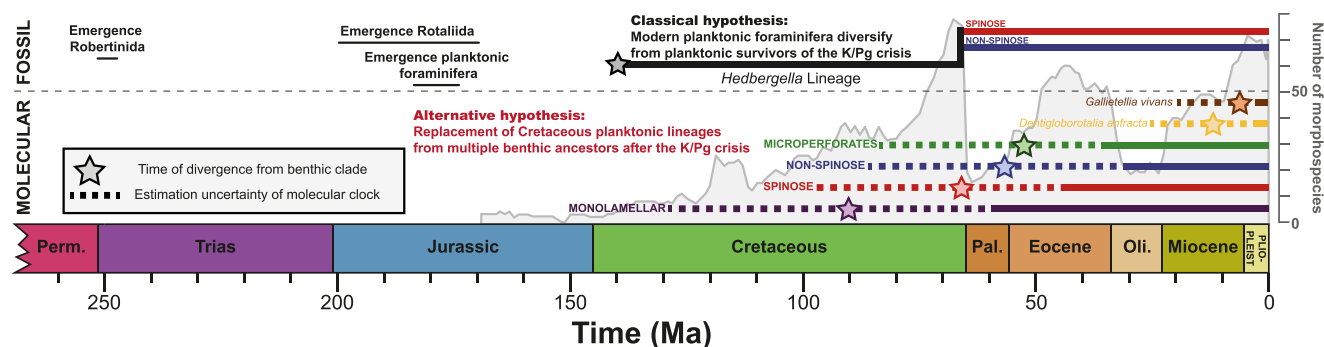


Fig. 5 | Emergence of modern planktonic foraminifera. The emergence of the modern clades of planktonic foraminifera in relation to the evolution of their diversity through the Mesozoic and Cenozoic following the classical hypothesis based on fossil record data and our alternative hypothesis based on molecular data. The gray curve represents the number of planktonic species observed in the fossil record (from ref. 8). The time of divergence between the extant clades and their

nearest benthic relative in the maximum likelihood molecular phylogeny is represented by a star. It represents the earliest possible date of transition into the plankton. Dashed lines represent the uncertainty of the time of divergence. Time range of evolutionary events in the Mesozoic are provided above the graph. The molecular clocked phylogenies are provided as Supplementary Data 5.

sequences labeled as Foraminifera in the PR²_V9 database and replaced them with 3390 foraminifera barcode sequences with updated taxonomy following the 10 rank system of the PR² database¹⁴. The database then included 2044 planktonic foraminifera sequences from the PFR² database³³ and sequences published in recent work^{34,35}, that are organized in 4 clades (Spinose, Non-Spinose, Microperforates, Basal) and cover 40 of the ~45 species described by morphotaxonomy. The benthic foraminifera sequences are classified in the three major clades Monothalamea, Tubothalamea, and Globothalamea after ref. 9 and the taxonomy of the Globothalamea has been further updated after ref. 15. The updated V9 reference for foraminifera includes 496 unique taxonomic paths and is provided as Supplementary Data 1.

We re-classified the TARA Ocean V9 MOTUs using blastn³⁶ v. 2.7.1 against the updated reference database. We retrieved all sequences with a percentage of identity of 90% or more to the reference for further downstream analysis. Based on the taxonomic affiliation returned by BLAST, we classified each MOTUs into four categories for planktonic foraminifera: “Spinose”, “Non-Spinose”, “Microperforates” and “Basal”, and three categories of Benthic foraminifera: “Tubothalamea”, “Monothalamea” and “Globothalamea”. The results of the assignment are shown in Supplementary Fig. 1 and the occurrence of foraminifera MOTUs with the updated assignment is available in Supplementary Data 2.

Dispersion of benthic OTUs in the plankton

Based on the updated taxonomic assignment, we calculated the portion of the diversity (percentage of MOTUs) and the volume of the data attributed to each clade (Fig. 2A), their distribution in each size class of the TARA Ocean dataset (Fig. 2B), the average proportion of each clade against the distance to the coast (Fig. 2C), and the probability of occurrence of individual MOTUs based on their presence/absence at each station using logistic regression (Generalized Linear Model implemented in R v. 4.0.2³⁷; Fig. 2D). To predict the lifestyle of the Globothalamea OTUs as either planktonic or benthic, we pursued a dual approach. We applied random forest models using (1) the coefficients of the logistic regression models, implemented in the package randomForest³⁸, and (2) the presence/absence MOTU table directly, implemented in the mlr3 package³⁹. Only MOTUs occurring in at least five stations were considered for this approach. The random forest models were trained and validated with 112 MOTUs affiliated with holoplanktonic foraminifera lineages (“Planktonic”) and 32 MOTUs affiliated with the monothalamids and Tubothalamea clades (“Benthic”), to predict the ecology of 126 Globothalamea MOTUs. To

reduce the risk of false positives, we only considered the Globothalamea MOTUs that were predicted as putative planktonic and benthic by both approaches for further analyses. The results of the classification approaches are provided as part of Supplementary Data 2 and the analysis code is provided on Github (<https://github.com/chassenr/ForamsOrigin>) and Zenodo (<https://doi.org/10.5281/zenodo.7274980>).

Phylogenetic placement of Globothalamea MOTUs

To compare the phylogenetic affiliation of the Globothalamea foraminifera MOTUs classified with either a “planktonic” or “benthic” behavior, we relied on a phylogenetic placement approach. We constructed an alignment that included 137 non-redundant SSU rRNA gene sequences representative of Globothalamea foraminifera diversity that covered the V9 region. We automatically aligned the sequences with MAFFT v.7⁴⁰, chose the best model of evolution according to Modeltest-ng⁴¹ and inferred the topology using RAXML-ng⁴² with 100 rapid bootstraps using the model TVM+I+G4. The returned topology was consistent with the results of ref. 15 that indicated the respective monophyly of the families within the Globothalamea. Then, we separately aligned the MOTUs classified as “Benthic” or “Planktonic” to the backbone alignment using the –add option of the online version of MAFFT. We then used phylogenetic placement EPA-ng⁴³ with default options and ITOL⁴⁴ to display the results (Fig. 3). The results of the automated alignments, model selection, phylogenetic inference, and EPA-ng placement are provided as Supplementary Data 3.

Phylogenetic context of the emergence of planktonic foraminifera

We established a phylogenetic framework to contextualize the emergence of modern planktonic foraminifera. We selected 44 representative sequences of the recognized seven superfamilies of Rotaliida and the “Clade 3” as describe in ref. 15 complemented with textulariid sequences to root the phylogeny, the aim being to reconstruct all evolutionary splits leading to the present day diversity of the Rotaliida. We thus included 11 sequences belonging to the spinose and monolamellar planktonic foraminifera clades, seven sequences of the non-spinose clade, five sequences of the microperforate clade and two sequences of planktonic species with uncertain phylogenetic affiliations labeled as “Basal”. We purposely excluded species with long branches to avoid artifacts and not disrupt the phylogenetic inference (*G. minuta*, *N. incompta*, *G. truncatulinoides*, *G. menardii*, *G. unguolata*, *G. tumida*). We used the partial fragment of ~1000 bp located at 5′-end of the 18S because most of the planktonic foraminifera sequences are covering this

fragment only. We automatically aligned the 69 sequences with the phylogeny aware alignment method PRANK⁴⁵ that is better suited than MAFFT because of the higher heterogeneity between the sequences. The model GTR + I + G4 was selected with Modeltest-ng and the Maximum Likelihood topology was inferred using RAXML-ng with 1000 rapid bootstrapping pseudo-replicates. The Bayesian inference was performed with MrBayes v.3.2.7⁴⁶ and consisted of two simultaneous chains run for 10,000,000 generations that converged with an average standard deviation of split frequencies of 0.01, and 40,000 trees there were sampled of which 10,000 were discarded as burn-in. The results of the automated alignments, model selection, and phylogenetic inferences are provided as Supplementary Data 4. The RAXML majority rule is shown as Fig. 4 with the bootstrap values and posterior probability provided next to the branches and the topologies with branch lengths are shown on Supplementary Fig. 2.

Because the inferences with all planktonic clades returned a topology with each planktonic foraminifera clade being monophyletic, hence suggesting a polyphyletic origin, but with poor branch support (Fig. 4), we tested for the two hypotheses for the origin of planktonic foraminifera suggested by the fossil record. The first hypothesis postulates a single origin of the three main clades of planktonic foraminifera, consistent with the hypothesis of a single emergence in the early Jurassic¹. We tested two topologies for this hypothesis, one where the microperforate clade would be at the base of the planktonic clade, and one topology where the spinose would be at the base. The second hypothesis assumes a common origin of the Spinose and Non-Spinose clade as descending from the *Hedbergella* genus that survived the KT crisis as suggested by ref. 2. We manually constructed the corresponding phylogenetic trees and used the Swofford Olsen Waddell Hillis test implemented in SOWHAT¹⁹ and the Approximately Unbiased test²⁰ implemented in IQ-TREE²⁷ to assess if these constructed topologies were potentially congruent with the dataset. The constructed topologies and the results of the SOWHAT and AU test are provided as Supplementary Fig. S3.

To estimate the time of divergence of the planktonic clades from their benthic ancestors, we applied a molecular clock estimation to the Maximum Likelihood and Bayesian topologies. Molecular clocked phylogenies aim to translate molecular distances into absolute times of divergence between the branches and rates of evolution, which requires the establishment of prior probability of distributions on parameters such as fossil calibration and branching model⁴⁸. Bayesian molecular clock dating uses statistical distributions to characterize uncertainties in model parameters which is translated into large confidence intervals in the time divergence estimation between branches, even when using genome wide datasets²². The uncertainties in posterior time of divergence estimates are influenced by fossil calibration that is crucial to detect variation of the rate of evolution between species or clades. Because foraminifera display high heterogeneity in their rate of evolution, we provided the tree with multiple calibrations. Although the Globobulimina appeared during the Paleozoic¹¹, the fully calcified extant taxa of the group occurred in the Mesozoic after the end-Permian crisis that caused a drastic loss of foraminifera diversity and the disappearance of the largest benthic foraminifera species⁴⁹. Right after the crisis during the Olenekian (251.2–247.2 Ma) emerged the first robertinids characterized by arago-agglutinated or mixed arago-agglutinated-secreted (with aragonite) walls^{30,31} from which the robertinids developed fully mineralized tests. The emergence of the rotaliids, with fully calcitic tests is not clear. The traditional view is that the rotaliids are directly descended from the robertinids and their earliest occurrences were in the lower Jurassic (199.3–170.3 Ma) with the first observation of buliminids⁵⁰. An alternative hypothesis is that the rotaliids evolved directly from the textulariids and

developed calcitic mineralization independently from the robertinids⁵¹. We chose to be conservative and considered a divergence between the textulariids and the rotaliids during Olenekian (251.2–247.2 Ma), which is compatible with either a direct divergence from the textulariids or the robertinids and places this split at the earliest possible date. Second, we added a constraint within the benthic Rotaliida using the divergence between the clades Nummulitoidea (planispiral) and Calcarinoidea (trochospiral). Their separation from each other and from early Rotaliidae took place in the Cretaceous⁵². The first illustrated occurrence of the calcarinoid genus *Pararotalia* being in the Cenomanian (93.9–100.5 Ma)⁵³, the separation of the two clades is necessarily older. To not artificially constrain the tree toward younger date estimates, we retained a potential split between these two clades in the lower Cretaceous–Cenomanian (145–93.9).

Finally and to calibrate the planktonic foraminifera, we deliberately chose split dates in the recent history of the clade because of their higher reliability and to not constrain the most internal nodes close to the divergence from benthic clades. For the Spinose clade, we used the divergence between *G. rubescens* and the genus *Globigerinoides* at 23.8 Ma⁷, the divergence between *Beella digitata* and *Globigerinella* at 11.08 Ma⁵⁴, and the divergence between *Turborotalita humulis* and *Turborotalita quinqueloba* at 12.9 Ma⁷. For the Non-Spinose clade, we used the divergence time between *Globorotalia hirsuta* and *Globorotalia scitula* at 8.2 Ma, and the divergence between *Globorotalia inflata* and *Neoglobobulimina dutertrei* at 18.8 Ma⁷. Last, for the microperforate clade, we used the divergence time between *Candeina nitida* and *Globigerinita glutinata* at 9.9 Ma and the first appearance datum of *Globigerinita uvula* at 29 Ma⁵⁴.

To calculate the molecular clock phylogeny, we used a relaxed clock model implemented in BEAST v.1.8.4⁵⁵ and model parameters were set using BEAUTi v.1.8.4⁵⁵. The distribution of the fixed node age prior was considered normal and the speciation rate was assumed constant under the Yule-Process. The GTR model was selected as a substitution model. A custom R script was used to derive a rooted and fully bifurcated tree from the RAXML and MrBayes returned topologies that were then used as starting tree. Markov-Chain- Monte Carlo (MCMC) analyses were conducted for 10,000,000 generations for the RAXML topology and 15,000,000 generation for the Bayesian topology, with a burn-in of 1000 generations and saving each 1000th generation. The maximum clade credibility tree with median node heights was calculated in TREEAnnotator from the BEAST package, with a burn-in of 100 trees and a posterior probability limit of 0. The resulting trees were then visualized in FigTree v. 1.3.1⁵⁶ and provided as Supplementary Data 5, and the date of divergence between the planktonic clades and their nearest neighbor in the RAXML phylogeny is shown in Fig. 5.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used in the manuscript are publically available on Zenodo (<https://doi.org/10.5281/zenodo.3768509>, <https://zenodo.org/record/3768951>) and PANGAEA (<https://doi.pangaea.de/10.1594/PANGAEA.875577>), the relevant intermediary files are provided as part of the Supplementary Data files. Source data for Fig. 2, consisting of occurrences of foraminifera MOTUs in Tara Ocean samples and relevant metadata of the samples, are provided in Supplementary Data 2.

Code availability

The code used to run the random forest analyses is provided on Github <https://github.com/chassennr/ForamsOrigin> and Zenodo (<https://doi.org/10.5281/zenodo.7274980>).

References

- Hart, M. B. et al. The search for the origin of the planktic foraminifera. *J. Geol. Soc. Lond.* **160**, 341–343 (2003).
- Aze, T. et al. A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biol. Rev.* **86**, 900–927 (2011).
- Gradstein, F., Waskowska, A. & Glinskikh, L. The first 40 million years of planktonic foraminifera. *Geosci* **11**, 1–25 (2021).
- Ujiie, Y., Kimoto, K. & Pawlowski, J. Molecular evidence for an independent origin of modern triserial planktonic foraminifera from benthic ancestors. *Mar. Micropaleontol.* **69**, 334–340 (2008).
- Darling, K. F. et al. Surviving mass extinction by bridging the benthic/planktic divide. *Proc. Natl Acad. Sci. USA* **106**, 12629–33 (2009).
- Kucera, M. et al. Caught in the act: anatomy of an ongoing benthic–planktonic transition in a marine protist. *J. Plankton Res.* **39**, 436–449 (2017).
- Ezard, T. H. G., Aze, T., Pearson, P. N. & Purvis, A. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. *Science* **332**, 349–352 (2011).
- Lowery, C. M., Bown, P. R., Fraass, A. J. & Hull, P. M. Ecological response of plankton to environmental change: thresholds for extinction. *Annu. Rev. Earth Planet. Sci.* **48**, 403–429 (2020).
- Pawlowski, J., Holzmann, M. & Tyszk, J. New supraordinal classification of foraminifera: molecules meet morphology. *Mar. Micropaleontol.* **100**, 1–10 (2013).
- Lecroq, B. et al. Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proc. Natl Acad. Sci. USA* **108**, 13177–13182 (2011).
- Pawlowski, J. et al. The evolution of early foraminifera. *Proc. Natl Acad. Sci. USA* **100**, 11494–8 (2003).
- Vachard, D. Macroevolution and biostratigraphy of paleozoic foraminifers. in *Stratigraphy and Timescales* (Ed. Montenari, M.) Vol. 1, 257–323 (Academic Press, 2016).
- Ibarbalz, F. M. et al. Global trends in marine plankton diversity across kingdoms of life. *Cell* **179**, 1084–1097.e21 (2019).
- Guillou, L. et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–D604 (2013).
- Holzmann, M. & Pawlowski, J. An updated classification of rotaliid foraminifera based on ribosomal DNA phylogeny. *Mar. Micropaleontol.* **132**, 18–34 (2017).
- John, A. W. G. The regular occurrence of Reophax Scottie Chaster, a benthic foraminiferan, in plankton samples from the North Sea. *J. Micropaleontol.* **6**, 61–63 (1987).
- Kucera, M. et al. Caught in the act: anatomy of an ongoing benthic–planktonic transition in a marine protist. *J. Plankton Res.* **39**, 436–449 (2017).
- Darling, K. F., Wade, C. M., Kroon, D. & Brown, A. J. L. Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. *Mar. Micropaleontol.* **30**, 251–266 (1997).
- Church, S. H., Ryan, J. F. & Dunn, C. W. Automation and evaluation of the SOWHAT test with SOWHAT. *Syst. Biol.* **64**, 1048–1058 (2015).
- Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).
- Pawlowski, J. et al. Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. *Mol. Biol. Evol.* **14**, 498–505 (1997).
- Peijnenburg, K. T. C. A. et al. The origin and diversification of pteropods precede past perturbations in the Earth's carbon cycle. *Proc. Natl Acad. Sci. USA* **117**, 25609–25617 (2020).
- O'Brien, C. L. et al. Cretaceous sea-surface temperature evolution: constraints from TEX86 and planktonic foraminiferal oxygen isotopes. *Earth-Sci. Rev.* **172**, 224–247 (2017).
- Olsson, R. K., Berggren, W. A., Hemleben, C. & Huber, B. T. Atlas of Paleocene planktonic foraminifera. *Smithson. Contrib. Paleobiol.* 1–252 <https://doi.org/10.5479/si.00810266.85.1> (1999).
- Arenillas, I. & Arz, J. A. Benthic origin and earliest evolution of the first planktonic foraminifera after the Cretaceous/Palaeogene boundary mass extinction. *Hist. Biol.* **29**, 25–42 (2017).
- Huber, B. T., Petrizzo, M. R. & MacLeod, K. G. Planktonic foraminiferal endemism at southern high latitudes following the terminal cretaceous extinction. *J. Foraminifer. Res.* **50**, 382–402 (2020).
- Arenillas, I., Arz, J. A. & Gilabert, V. An updated suprageneric classification of planktic foraminifera after growing evidence of multiple benthic–planktic transitions. *Spanish J. Palaeontol.* <https://doi.org/10.7203/sjp.22189> (2022).
- Culver, S. J. Benthic foraminifera across the Cretaceous–Tertiary (K–T) boundary: a review. *Mar. Micropaleontol.* **47**, 177–226 (2003).
- Widmark, J. G. V. & Malmgren, B. A. Benthic foraminiferal changes across the Cretaceous/Tertiary boundary in the deep sea; DSDP sites 525, 527, and 465. *J. Foraminifer. Res.* **22**, 81–113 (1992).
- Rigaud, S., Martini, R. & Vachard, D. Early evolution and new classification of the order Robertinida (foraminifera). *J. Foraminifer. Res.* **45**, 3–28 (2015).
- Rigaud, S., Granier, B. & Masse, J. P. Aragonitic foraminifers: an unsuspected wall diversity. *J. Syst. Palaeontol.* **19**, 461–488 (2021).
- Hull, P. M. et al. On impact and volcanism across the Cretaceous–Paleogene boundary. *Science* **367**, 266–272 (2020).
- Morard, R. et al. PFR²: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. *Mol. Ecol. Resour.* **15**, 1472–1485 (2015).
- Morard, R. et al. Genetic and morphological divergence in the warm-water planktonic foraminifera genus Globigerinoides. *PLoS ONE* **14**, 1–30 (2019).
- Morard, R., Vollmar, N. M., Greco, M. & Kucera, M. Unassigned diversity of planktonic foraminifera from environmental sequencing revealed as known but neglected species. *PLoS ONE* **14**, e0213936 (2019).
- Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinforma.* **10**, 1–9 (2009).
- R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2014).
- Liaw, A. & Wiener, M. Classification and Regression by random Forest. *R. N.* **2**, 18–22 (2002).
- Lang, M. et al. mlr3: a modern object-oriented machine learning framework in R. *J. Open Source Softw.* **4**, 1903 (2019).
- Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- Darriba, D. et al. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).
- Kozlov, A. M. et al. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
- Barbera, P. et al. EPA-ng: massively parallel evolutionary placement of genetic sequences. *Syst. Biol.* **68**, 365–369 (2019).
- Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, 293–296 (2021).
- Löytynoja, A. & Goldman, N. WebPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinform.* **11**, 1–7 (2010).

46. Ronquist, F. et al. MrBayes 3. 2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
47. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
48. Dos Reis, M., Donoghue, P. C. J. & Yang, Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* **17**, 71–80 (2016).
49. Song, H., Tong, J. & Chen, Z. Q. Evolutionary dynamics of the Permian-Triassic foraminifer size: Evidence for Lilliput effect in the end-Permian mass extinction and its aftermath. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **308**, 98–110 (2011).
50. Copestake, P. & Johnson, B. Lower Jurassic Foraminifera from the Llanbedr (Mochras Farm) Borehole, North Wales, UK. *Monogr. Palaeontogr. Soc.* **167**, 1–403 (2013).
51. Rigaud, S. & Blau, J. New Robertinid Foraminifers from the Early Jurassic of Adnet, Austria and Their Evolutionary Importance. *Acta Palaeontol. Pol.* **61**, 721–734 (2016).
52. Boudagher-fadel, M. K. *Evolution and Geological Significance of Larger Benthic Foraminifera. Evolution and Geological Significance of Larger Benthic Foraminifera* (UCL Press, 2018).
53. Piuze, A. & Meister, C. Cenomanian rotaliids (Foraminiferida) from Oman and Morocco. *Swiss J. Palaeontol.* **132**, 81–97 (2013).
54. Kucera, M. & Schönfeld, J. The origin of modern oceanic foraminiferal faunas and Neogene climate change. in *Deep-Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies*. (ed. The Micropalaeontological Society, S. P.) 409–425 (The Geological Society, 2007).
55. Drummond, A. J. & Suchard, M. A. Bayesian random local clocks, or one rate to rule them all. *BMC Biol.* **8**, 114 (2010).
56. Rambaut, A. FigTree version 1.3.1. <http://tree.bio.ed.ac.uk> (2009).
57. Groussin, M., Pawłowski, J. & Yang, Z. Bayesian relaxed clock estimation of divergence times in foraminifera. *Mol. Phylogenet. Evol.* **61**, 157–166 (2011).
58. Loeblich Jr, A. R. & Tappan, H. *Foraminiferal Genera and Their Classification* (Springer, 1988).

Acknowledgements

The project was supported by the BMBF-funded German network for bioinformatics infrastructure (de.NBI) [project ForamsOrigin] through the provision of computational resources and user support, and by the Cluster of Excellence “The Ocean Floor—Earth’s Uncharted Interface” (EXC-2077, Project 390741603) funded by the German Research Foundation (DFG).

Author contributions

R.M. and M.K. designed the study; C.H. and R.M. performed bioinformatic and statistical analyses; A.F.G. administrated the computational resources; R.M., A.F.G., C.D., and M.G. performed the phylogenetic analyses; R.M., M.K., and S.R. discussed the paleontological interpretation. R.M. and M.K. wrote the manuscript, which was edited and approved by all authors.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-34794-5>.

Correspondence and requests for materials should be addressed to Raphaël Morard.

Peer review information *Nature Communications* thanks Tracy Aze, Brian Huber and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022