Denitrification in foraminifera has ancient origins and is complemented by associated bacteria

Christian Woehle[1,5*], Alexandra-Sophie Roy[1*], Nicolaas Glock[2,6], Jan Michels[3], Tanita Wein[1,7], Julia Weissenbach[1,8], Dennis Romero[4], Claas Hiebenthal[2], Stanislav N. Gorb[3], Joachim Schönfeld[2], Tal Dagan[1]

[1]Institute of Microbiology, Kiel University, Am Botanischen Garten 11, Kiel, 24118, Germany
[2]GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstrasse, Kiel, 24148, Germany
[3]Zoological Institute, Kiel University, Am Botanischen Garten 1-9, Kiel, 24118, Germany
[4]Dirección General de Investigaciones Oceanográficas y Cambio Climático, Instituto del Mar del Perú, Callao 01, Peru 17
[5]Present address: Miltenyi Biotec B.V. & Co. KG, Friedrich-Ebert-Str. 68, Bergisch Gladbach, 51429, Germany
[6]Present address: Institute for Geology, University of Hamburg, Hamburg, 20146, Germany
[7]Present address: Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel
[8]Present address: Faculty of Biology, Technion – Israel Institute of Technology, Haifa, 3200003, Israel

* Equally contributing authors

Corresponding authors: Christian Woehle woehle@gmail.com (C.W.), Tal Dagan tdagan@ifam.uni-kiel.de (T.D.)

Summary

Benthic foraminifera are unicellular eukaryotes that inhabit sediments of aquatic environments. Several foraminifera of the order Rotaliida are known to store and use nitrate for denitrification, a unique energy metabolism among eukaryotes. The rotaliid *Globobulimina* spp. has been shown to encode an incomplete denitrification pathway of bacterial origins. However, the prevalence of denitrification genes in foraminifera remains unknown and the missing denitrification pathway components are elusive. Analysing transcriptomes and metagenomes of ten foraminifera species from the Peruvian oxygen minimum zone, we show that denitrification genes are highly conserved in foraminifera. We infer of the last common ancestor of denitrifying foraminifera, which enables us to predict further denitrifying species. Additionally, an examination of the foraminifera microbiota reveals evidence for a stable interaction with *Desulfobacteracea*, which harbour genes that complement the foraminifera denitrification pathway. Our results provide evidence that foraminiferal denitrification is complemented by the foraminifera microbiome. The interaction of Foraminifera with their resident bacteria is at the basis of foraminifera adaptation to anaerobic environments that manifested in ecological success within oxygen depleted habitats.

Introduction

Nitrogen is an essential element for life on earth as it forms the basis for the synthesis of nucleotides and amino acids. Nonetheless, while the earth atmosphere is rich in nitrogen gas (up to 78%[1]), this gas is usually inert but can be made accessible for biological processes by nitrogen fixation[2]. Microbial organisms are important players in the global nitrogen cycle as they facilitate the assimilation of nitrogen into bioavailable nitrogen derivatives as well as the dissimilation of nitrogen derivatives into dinitrogen (N_2)[2]. A key dissimilatory pathway is denitrification, where nitrate (NO_3^-) is either partially or completely degraded and the final product – N_2 – is released to the atmosphere (i.e., nitrogen loss)[2]. Marine organisms are considered major contributors to nitrogen loss from the environment with benthic organisms being responsible for about two-thirds of the loss of reactive nitrogen in the ocean[3,4]. Especially oxygen minimum zones (OMZs) are worth mentioning here as they are estimated to be responsible for 20-40% of bioavailable nitrogen removal in the ocean[3,5]. The ability to perform denitrification is abundantly found in eubacteria[6], whereas it is rare among the eukaryotes. Partial or complete denitrification have only been reported for two species of fungi[7] and several foraminifera of the order Rotaliida[8-11]. Denitrifying foraminifera are unicellular eukaryotes commonly found in marine sediments. Studies of foraminifera residing in the Peruvian OMZ

showed that they are found in high densities of up to 600 individuals per cm^2, where they are estimated to contribute 20%-50% to the total benthic nitrate loss in the OMZ[10,12].

Rotaliid foraminifera are traditionally divided into three clades based on their phylogeny[13]. Only species in clades I and III were demonstrated to denitrify, while rotaliids classified in clade II have been shown to lack an intracellular nitrate storage or measurable denitrification activity[9]. Class II rotaliids typically populate environments with poor-nitrate supply, e.g. intertidal to near-shore habitats from the tropics up to boreal bioprovinces[14], hence it is conceivable that they lack the ability to denitrify. One example is *Ammonia tepida* (class II), that can survive episodic oxygen depletion events via dormancy[15]. A study of clade III species sampled from a hypoxic environment (Gulmarfjord, Sweden) – *Globobulimina* spp. – showed that that their genome encodes several genes along the denitrification pathway that are of ancient bacterial origin[11]. These include the copper-containing nitrite reductase (NirK) and nitric oxide reductase (Nor), but not the nitrate reductase (NapA/NarG) and nitrous oxide reductase (NosZ). In addition, the *Globobulimina* genome contains a diverse gene family encoding for nitrate transporters (Nrt)[11], a finding that is consistent with the accumulation of an intracellular NO_3^- storage in denitrifying rotaliids[8-10,16]. More recent studies could validate the presence of those genes in at least two additional species of foraminifera and provide alternative suggestions on the nitrate reductase homolog missing from the earlier model of denitrification in foraminifera[17,18]. The rotaliids ability to respire both oxygen (O_2) and NO_3^- marks them as facultative anaerobes that are able to thrive in both aerobic and anaerobic conditions. Furthermore, while oxygen is generally considered to be preferred over NO_3^- as electron acceptor, rotaliids from the Peruvian OMZ were reported to prefer NO_3^- over O_2[10]. These findings suggest that genes of the denitrification pathway are widespread in rotaliids, however their distribution remains largely unknown.

Foraminifera, like other eukaryotes[19], are habitat to bacterial organisms that reside out- and in-side their cell (i.e., test)[20-24]. Microscopic observations showed that the bacteria are localized in food vacuoles or in the cytoplasm of the cell, which led to the suggestion that interactions of foraminifera with its microbiota may vary from prey-predator interactions or parasitism to metabolic symbiosis[22]. Further studies of the microbiota in clade II rotaliids identified sulphate-reducing and sulphur-oxidizing bacteria that were suggested to participate in Sulphur cycling thus providing carbon and other nutrients to the host[20,21,25]. Notwithstanding, previous studies of foraminifera-associated bacteria used microscopy observations or sequencing of marker genes (e.g., 16S rRNA) that are limited in their resolution[20-25]. It has been suggested that the microbiota of denitrifying foraminifera include denitrifying bacteria that utilize the foraminifera NO_3^- storage, similarly to observations in other bacteria-protist associations (e.g., Gromiids[26], allogromiids[27]). Indeed, metagenomics of the *Globobulimina*

microbiota revealed the presence of a taxonomically diverse species community where several members encode homologs of NapA and NosZ[11]. This finding gave rise to the hypothesis that the partial foraminiferal denitrification pathway may be complemented by microbiota functions.

Here we investigate the evolutionary history of genes along the denitrification pathway in foraminifera and furthermore examine the functional repertoire of the foraminifera microbiome. For that purpose, we studied populations of ten rotaliid species known to denitrify[10] (except *Globobulimina pacifica* for which no information is available). Our study supplies insights in the evolution of rotaliids and their microbiota and lays the basis for further research of foraminifera genome evolution.

Results

Transcriptomes of ten Peruvian rotaliids

Eukaryotic transcriptomes and whole metagenomes were sequenced from populations of foraminifera with two biological replicates per species (ca. 160 individuals per sample; see sampled species in Figure 1A). Our computational analysis includes five additional publicly available transcriptomes of foraminifera (Suppl. Table S1) and the genome data of one monothalamid species (*Reticulomyxa filosa*). The assessment of transcriptome completeness showed that ~90% of the eukaryotic marker proteins are present in the data of the ten newly sequenced species. Furthermore, we examined the purity of the focal species in each transcriptome by testing for an over-representation of eukaryotic marker proteins that are expected as single-copy genes. A high redundancy of single-copy genes was found in the *B. spissa* (78%) and *B. costata* (50%) transcriptomes, suggesting the presence of bystander species in those samples (Suppl. Table S1).

The Peruvian rotaliids harbour genes of the denitrification pathway

To explore the denitrification mechanisms in the Peruvian foraminifera, we searched for homologs to denitrification pathway genes. Our results reveal that all sampled species harbor homologous genes to Nrt, NirK and Nor (Figure 1B; note that genes are termed by their corresponding protein symbol). Our search for homologs in the publicly available data yielded a putative homolog of NirK in *Rosalina* sp. and Nrt homologs in all but *R. filosa* (Fig. 1B). The additional species we included here are considered to reside mostly in oxygenated habitats, hence they are not expected to be able to perform denitrification. Among this group, only *A. tepida* was so far experimentally tested for denitrification ability and reported as a non-denitrifying species[9]. The reconstructed phylogenies of Nrt and NirK reveal two major sub-clades where each clade includes most of the rotaliids in our study (Suppl. Figure S2; Suppl.

Table S2). This result indicates that Nrt and NirK underwent an ancient gene duplication where the two sub-clades encode for different subtypes of the proteins. Our results thus demonstrate that the denitrification pathway is highly conserved in the Peruvian rotaliids.

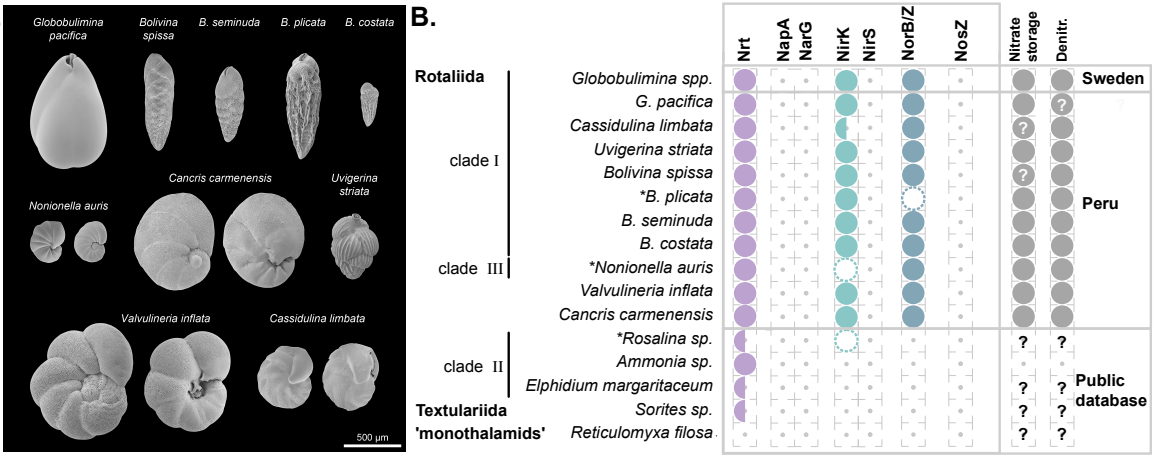


Figure 1. Morphological characteristics of the sampled Rotaliids and denitrification gene repertoire. A) Scanning electron micrographs of the sampled species. Species with clearly distinct lateral views are shown from two sides. **B)** Presence/absence of homologs of denitrification proteins identified for different foraminifera. Half circles indicate species where only a single, of multiple subtypes, was found. Dashed circles illustrate homologs discarded due to low coverage and asterisks highlight corresponding species**.** Evidence for NO_3^- storage and/or denitrification activity is illustrated by closed circles in the last column. Question marks denote missing information. Foraminifera sampled in this study are highlighted by the location 'Peru'. Protein symbols: Nrt: nitrate/nitrite transporter; NapA: periplasmic nitrate reductase; NarG: membrane-bound nitrate reductase; NirK: copper-containing nitrite reductase; NirS: cd_1-containing nitrite reductase; Nor: nitric oxide reductase; NosZ: nitrous oxide reductase. Note that the identification of homologous genes is based not only on sequence similarity but also on transcript abundance in order to exclude bystander species in the data. The additional data filtration stage affected our findings for *Rosalina* sp., *B. plicata* and *N. auris* (Suppl. Table S2 & S3), where the presence of at least one the crucial homologs (i.e, NirK or Nor) remained in the *B. plicata* and *N. auris* metatranscriptomes.

Foraminiferal denitrification evolved in the Rotaliida ancestor

The presence of NirK and Nor homologs the tested rotaliids suggests that those genes may have an ancient origin in this order. To study the origin of denitrification in foraminifera, we reconstructed the phylogenetic relations among species in our data using 81 eukaryotic marker proteins that have homologs in all species analyzed here (Suppl. Figure S1; Suppl. Table S4). Here, we applied a phylogenomics approach where the root position is inferred from all marker genes independent of a single species tree[28] (see methods). Our results show that the best-supported root position (42% of the gene trees) was at the branch leading to *R. filosa* (Suppl. Figure S1). This result is in agreement with previous studies that assumed monothalamids to be an outgroup in foraminifera phylogenies[29]. The second most frequent root position was found at the branch leading to the miliolid *Sorites* sp. (15% of single gene trees), which was followed closely by a root position on the branch leading to *Rosalina* sp.

(12%). Taken together, the inferences of rooted topologies show that the Peruvian rotaliids, together with *Globobulimina* spp. form a monophyletic group. Thus, our results reveal a shared origin of denitrification in foraminifera and hence the evolution denitrifying foraminifera from within the order Rotaliida.

For the inference of the denitrifying rotaliids last common ancestor (LCA; i.e., the rotaliids root), we reconstructed phylogenetic trees from an extended set of 146 eukaryotic marker proteins considering only members of the Rotaliida (Figure 2; Suppl. Table S4). An examination of the rooted tree topologies revealed that the branch leading to *Ammonia* sp. and *E. margaritaceum* was the most frequently supported root position (14% of single gene trees). This root inference is further supported by an alternative root position at a branch that splits *Ammonia* sp., *E. margaritaceum* and *Rosalina* sp. from the remaining species (10%). Overall, the different alternative root positions support with 35% a Rotaliida LCA within clade II foraminifera (i.e., among *Ammonia sp.*, *E. margaritaceum & Rosalina* sp.; Fig. 2). The absence of denitrification genes in clade II members (Fig. 1) further supports the suggestion that the evolution of denitrification in rotaliids occurred by gene gain[11].

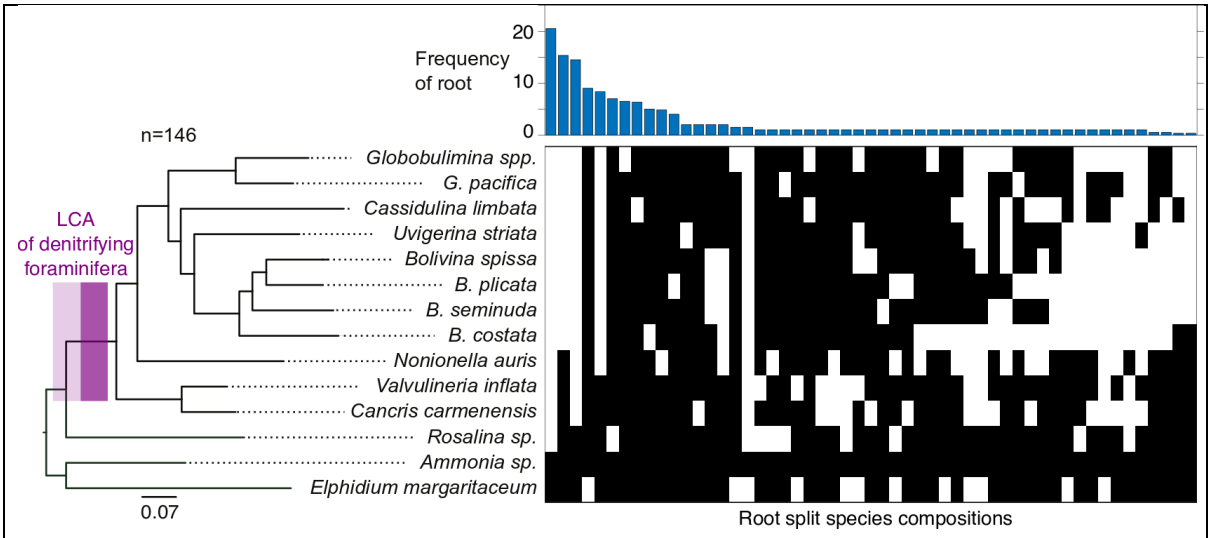


Figure 2. Phylogenetic rooting of rotaliids. A Maximum-likelihood phylogenetic reconstruction of foraminifera species from 146 eukaryotic protein marker sequences is shown on the left. The branch with the LCA of denitrifying foraminifera is highlighted in purple, where a lighter color includes lower bound for the origin of denitrification. Nonparametric bootstrap support is 1000/1000 at all the branches. A split representation of root splits determined via the MAD approach for the 146 single gene trees is shown on the right. Each column represents a putative root branch reported as split of two groups (black and white boxes) indicating for the species found on either side of the branch. The bar graph reports the single gene tree count supporting the corresponding root split.

Ancient origin of denitrification in foraminifera

To further study the denitrifying foraminifera LCA, we examined the phylogenetic position of the Peruvian species in the context of a foraminifera species tree. For that purpose, we

extracted sequences of the 18S ribosomal RNA (rRNA) subunit from our transcriptome data and used it in combination with publicly available 18S rRNA sequences to reconstruct a foraminifera species phylogeny (Figure 3A; Suppl. Table S5). The resulting rooted topology was mostly in agreement with the rotaliids phylogeny, where the focal species are found next to their closest relatives. Most species grouped well with their previously defined clades with the exception of *Nonionella* and *Globobulimina* representatives. Furthermore, several taxa were not monophyletic, as has been previously observed in foraminifera 18S rRNA phylogenies[29,30].

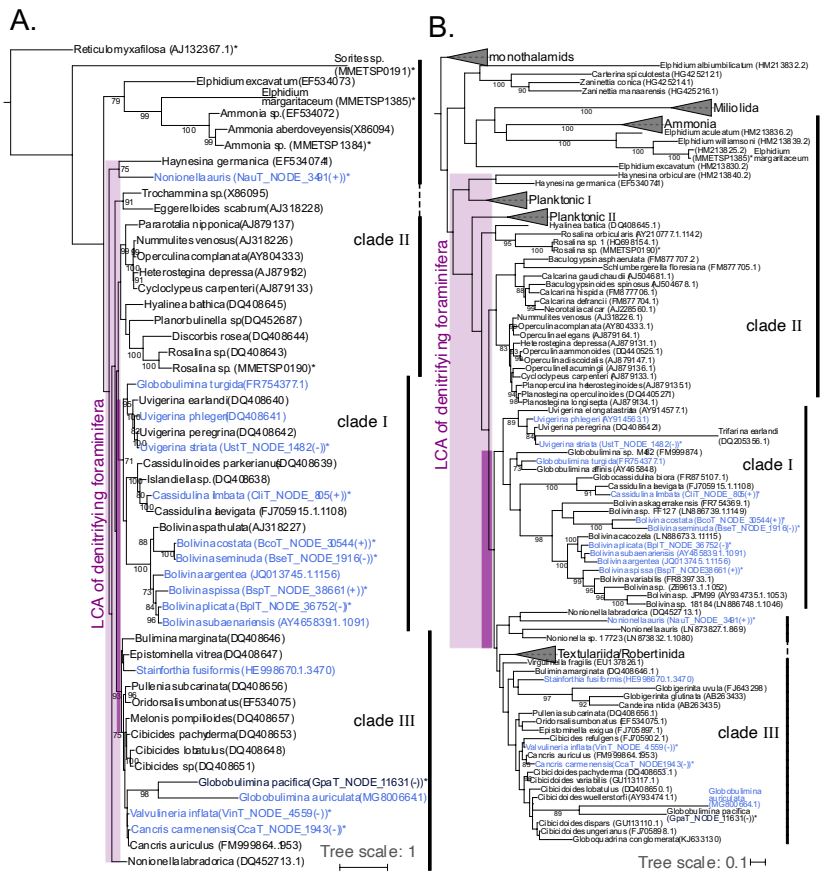


Figure 3. The origin of denitrification in context of the whole foraminifera group. A) Large-scale phylogenetic representation of Peruvian foraminifera in context of different foraminiferal taxa based on 18S sequences available in public databases. Clades I & III of the order Rotaliida appear are highly supported by bootstrap values. Questionable branching of *Nonionella* outside class III is not characteristic to data of the current study as exemplified for *N. labradorica* that as well does not group directly with clade III. *Globobulimina pacifica* forms a clade with *G. auriculata*, but within clade III. However, we consider the branching of these two species as uncertain due to the long branches and the low bootstrap support, which we did not observe for the marker protein phylogeny (Figure 2). Lower (light purple; including *N. labradorica*) and higher (dark purple) boundaries for the origin of foraminiferal denitrification are highlighted by boxes. **B)** Extended phylogenetic representation of foraminifera including planktonic species. Lower (light purple) and higher (dark purple) boundaries for the origin of foraminiferal denitrification are highlighted by boxes. Planktonic I & III designate two to three distinct clusters comprising planktonic foraminifera. Species experimentally shown to denitrify are highlighted in blue. The only exception is *Stainforthia fusiformis*, where denitrification activity has been shown for an unspecified species of the same genus. Species also considered in Figure 2 or Suppl. Figure S3 are highlighted by asterisks. Several species labels contain as well the contig id and orientation of 18S sequences in the corresponding transcriptome assemblies. Bootstrap support values (>=70) with 1000 replicates are shown at the branches. The trees were rooted by the clade of monothalamids containing *R. filosa*. A detailed phylogeny is presented in Supplementary Figure S2.

The inference of the denitrifying foraminifera LCA enabled us to further reconstruct the origin of denitrification within the foraminifera. For that purpose, we examined an extended phylogeny with additional foraminifera groups including planktonic species (Figure 3B; Suppl.

Figure S3). While the general taxonomic relationships in the tree were recovered as before (i.e., Fig 3A), many branches had a low bootstrap support. For example, the rotaliid clades II and III are paraphyletic, in contrast to the phylogeny with less taxa (Figure 3A). Nonetheless, an LCA of clades I and III can be recovered. The previously unassigned genera *Valvulineria* and *Cancris*[9] grouped well with members of clade III in both 18S rRNA phylogenies, hence they can be classified as clade III.

Since both clades I and III harbour several denitrifying foraminifera, it is most parsimonious to conclude that LCA of clades I and III was likely a denitrifying organism. Our inference thus suggests the presence of denitrification genes (and hence the capability) in all members of clades I and III including e.g., the genera *Cibicoides* and *Virgulinella*, which were previously not considered as denitrifying. Indeed, *Cibicidoides wullerstoerfi* has been assumed for a long time to be unable to withstand O_2 depletion[31]. However, recent studies observed living *Cibicidoides spp.* thriving in environments of < 2 µmol/kg O_2[32] and fossil specimens in the paleorecord during periods of severe O_2 depletion[33]. These recent studies support our prediction that several *Cibicidoides spp.* are able to denitrify.

Composition of the foraminifera microbiome is species-specific

To test for bacterial contribution to the foraminifera denitrification, we examined the foraminifera microbiome. For that purpose, we compared the composition of bacterial communities among foraminifera in our sample. In the absence of species-specific interactions, we would expect a strong impact of the environment (i.e., sampling location and water depth) on the microbiome composition. Thus, we compared the microbial community composition between foraminifera species and their sampling depths (Figure 4A). Our results show that the individual microbiomes are clustered by the foraminifera species rather than the sampling location, indicating the presence of species-specific bacterial communities in foraminifera. Our results reveal a similar taxonomic composition of bacterial communities associated with *Globobulimina* spp. from Sweden and *G. pacifica* from the Peruvian OMZ (Figure 4A). The similarity in microbiome composition among the *Globobulimina* species indicates the presence of genus-specific microbiome.

To identify common key players in foraminifera microbiota we examined the relative abundance of all bacterial families in their microbiome (Figure 4B; Suppl. Table S6). Our results revealed that the most prevalent bacterial families are Desulfobacteracea followed by Planctomycetaceae. The relative abundance of these two families varied between different species and samples. Furthermore, communities with a high relative abundance of Desulfobacteracea are characterized by a low abundance of Planctomycteaceae and vice versa (r_s=-0.77, p-value < 0.01, using Spearman correlation coefficient and t-test).

Desulfobacteracea comprise sulphate-reducing bacteria and several representatives that are able to grow chemoautotrophically[34]. Members of this group were previously observed in association with other rotaliids like *Virgulinella fragilis* and members of clade II[20,25]. Planctomycetes are often found in association with macroalgae[35]; they are characterized by a diversified metabolism allowing them the of colonization of a wide range of habitats[35].

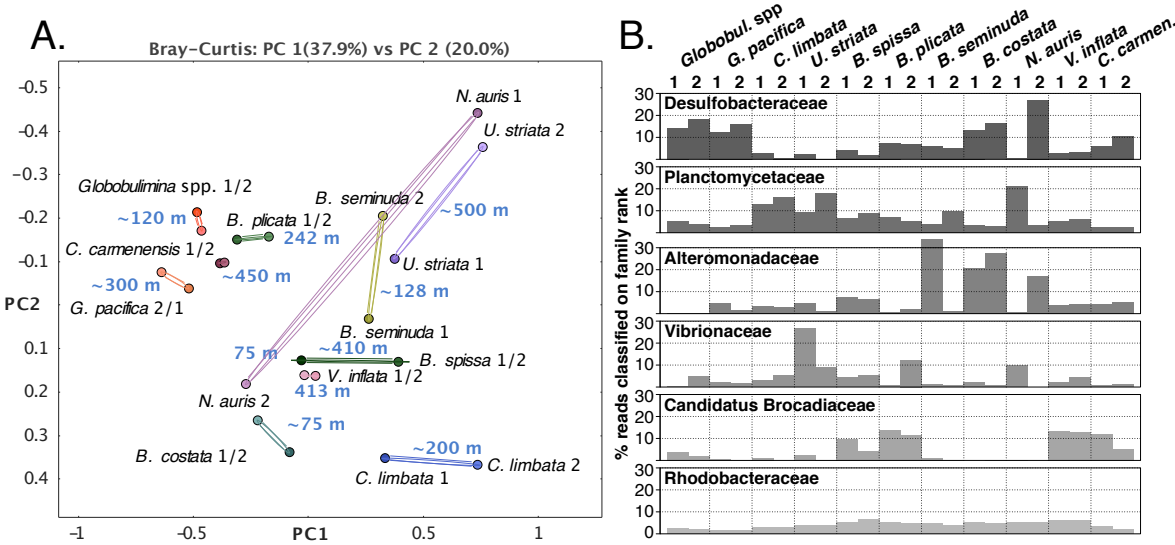


Figure 4. Comparison of microbial communities associated with different Foraminifera. A) Principal coordinates analysis (PCoA) clustering of bacterial communities. Each circle represents a bacterial community; duplicates per species are connected by lines. The text next to the circles indicates for species name, replicate number or sampling depth. The clustering of most species-associated communities indicates for species-specificity. Grouping by higher taxonomy was only observed for *Globobulimina* spp. from Sweden and *Globobulimina pacifica* from the Peruvian oxygen minimum zone (OMZ). **B)** Bar graphs of the six most abundant bacterial families found in the foraminiferal bacterial communities beginning at the top. Each bar represents the proportion of reads assigned to the individual taxon relative to the total number of reads in a sample classified on family rank or lower. Note the prevalence of Desulfobacteraceae shared between the Globobulimina samples, while none of the other bacterial families showed high prevalence in the genus.

The next most abundant families were Alteromonadaceae, Vibrionaceae and `Candidatus Brocadiaceae` that were more diverse in their relative abundances across different species and replicates. Most reported `Candidatus Brocadiaceae` members are autotrophic, obligately anaerobic bacteria performing anammox, which is a dissimilatory pathway where ammonium and nitrite (NO_2^-) are metabolized into N_2[36]. The presence of nitrate storage in foraminifera and NirK suggest that nitrite is readily available inside the foraminifera test. Vibrionaceae are a diverse group including multiple species that colonize marine organisms either as symbionts[37,38] or pathogens[39]. Alteromonadaceae were isolated from diverse marine environments including eukaryotic microbiota[40]. They are considered to be aerobe or facultative anaerobe bacteria and generally lack denitrification capabilities[41]. Finally, another abundant family, Rhodobacteracea, is worth mentioning, as it is always found in low, but similar abundance in all samples. Marine Rhodobacteracea species are considered

ecological generalists[42-44] and have been reported to colonize marine animals (e.g., fish larvae or sponges[45,46]). The uniform distribution of Rhodobacteracea in the foraminifera-microbiome suggests that members of this group are permanent residents in the foraminifera microbiota.

Our data thus shows that rotaliids are habitat to bacterial communities whose composition is akin to the microbiota of other marine eukaryotes.

An ancient interaction between Desulfobacteraceae and *Globobulimina* hints to metabolic dependency

The metagenomes analysis and classification into bacterial metagenome-assembled genomes (MAGs) resulted in a total of 263 high-quality MAGs (i.e., draft bacterial genomes) of foraminifera-associated bacteria (Suppl. Table S1). The strongest signal for a stable core microbiome in our data was observed in the comparison among the *Globobulimina* species, whose microbiome is characterized by a high frequency of Desulfobacteraceae (Figure 4B). To further explore the association between *Globobulimina* and Desulfobacteraceae we examined the Desulfobacteraceae MAGs in our data. A total of 40 high quality MAGs were obtained from the *G. pacifica* metagenome, including four high quality draft MAGs classified as Desulfobacteracea. These were compared with the 26 previously published MAGs for *Globobulimina* from Sweden, which include two Desulfobacteracea MAGs[11] (Suppl. Table S7). A phylogenetic network of the Desulfobacteracea MAGs reveals two MAGs from the *Globobulimina spp.* and *G. pacifica* metagenomes (Glo_11 & Gpa_30) that appear as sister taxa. The average nucleotide identity (ANI) between both MAGs was 84%, which is within the range expected for inter-species sequence similarity (e.g., within genera[47]). The common ancestry of these two MAGs suggests that the association between Desulfobacteracea and *Globobulimina* has an ancient origin. Our results thus indicate that the interaction fidelity between *Globobulimina* and Desulfobacteracea is high, similarly to observations in other eukaryote-bacterial symbioses (e.g., oligochaete worms and sulfur bacteria[48] or marine sponges with bacteria of the Poribacteria phylum[49]).

A central foundation in the evolution of symbiotic interaction is an exchange of currencies (i.e., resources) between the partners[50]. The versatile metabolic capabilities of Desulfobacteraceae and the availability of intracellular NO_3^- storage in foraminifera may suggest that the interaction between Desulfobacteracea and *Globobulimina* is based on nutritional currencies. To further examine possible symbiotic interaction between Desulfobacteracea and *Globobulimina*, we searched for metabolic properties of the Desulfobacteraceae MAGs that could serve as nutritional currencies in the symbiosis. For that purpose, we surveyed the metabolic pathways of representative Desulfobacteracea MAGs with respect to carbon, sulfur and nitrogen metabolism (Figure 5B; Suppl. Table S8). Our results reveal that all MAGs sampled encode the genetic repertoire required in order to

perform dissimilatory sulfate reduction as well as the Wood-Ljungdahl pathway. Anaerobic respiration via the dissimilatory sulfate reduction to sulfide is widespread among Desulfobacteraceae. Members of this family that encode the Wood-Ljungdahl pathway are able to oxidize organic compounds to carbon dioxide[51]. Alternatively, the Wood-Ljungdahl pathway may function in CO_2 carbon fixation during autotrophic growth.

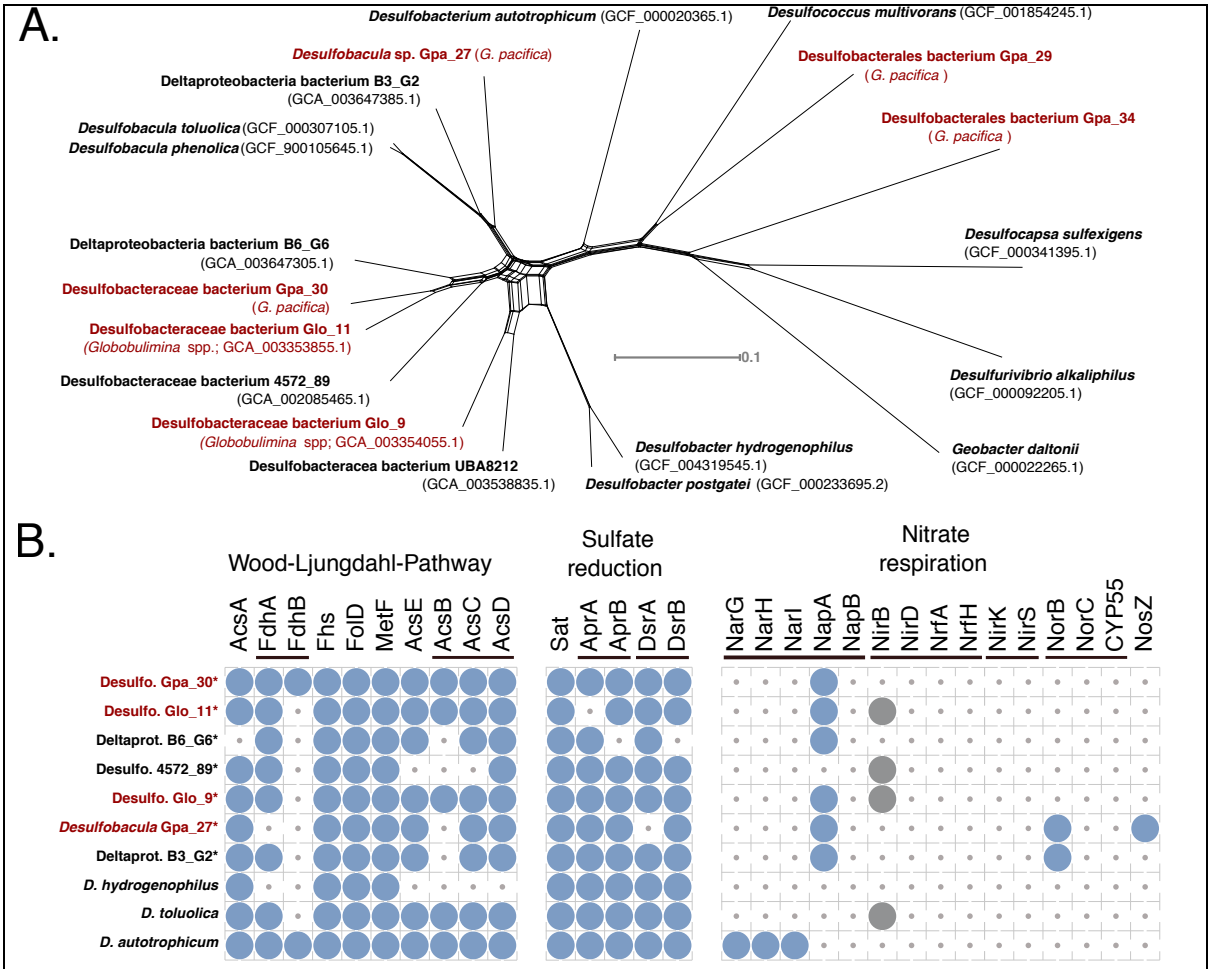


Figure 5. Characteristics of Desulfobacteraceae bacteria associated with Globobulimina. A) Neighbor-net network for Desulfobacteraceae associated with the genus *Globobulimina* from Sweden and Peru (*Globobulimina* spp. and *G. pacifica*, respectively) and additional Deltaproteobacteria. The network is based on a concatenated alignment of 33 single copy orthologous protein sequences found in all isolates shown. Red fonts highlight isolates found in association with members of the genus *Globobulimina.* The monophyly of the MAG Gpa_30 and MAG Glo_11 suggests a common ancestry. Note that the closest relatives to that pair are two uncultivated MAGs (Deltaproteobacteria bacterium B6_G6 & Desulfobacteraceae bacterium 4572_89) sampled from sediments in deep-sea hydrothermal vent sites at the Gulf of California. **B)** Presence of predicted proteins encoding metabolic pathways in representative members of Desulfobacteraceae. Three pathways are shown: Wood-Ljungdahl pathway, (dissimilatory) sulphate reduction & nitrate respiration. Asterisks highlight genomes in draft state. Circles indicate for the presence of homologs identified by the KEGG KAAS server. The thin bars link enzymes that participate in the same reaction step either by being an alternative enzyme or being part of one protein complex. The grey circles indicate for cases that were discarded following a manual inspection.

Genes along the denitrification pathway (or NO_3^- respiration) are mostly absent from all Desulfobacteracea MAGs, except for periplasmic NO_3^- reductases (NapA) homologs that were found in most foraminifera-associated MAGs (for a review in NapA function see [52]). All

MAGs lack the typical NapB protein forming a complex with NapA. However, *nap* operons without NapB genes have been reported for other members of Deltaproteobacteria, e.g. *Desulfovibrio desulfuricans*[53]. Notably, we found that the genome of *Desulfobulbus propionicus* (Accession: GCA_000186885.1), a member of the order Desulfobacterales that has been demonstrated to grow based on NO_3^-[54], is lacking a gene for NapB. Hence NO_3^- respiration in Desulfobacterales in the absence of NapB is possible. We further hypothesize that NapA may be secreted by the bacteria to *Globobulimina* intra-cellular environment. Indeed, we found that the NapA sequences of all MAGs (and *D. desulfuricans*) include signal peptides of the twin-arginine translocation pathway (reviewed in [55]) indicating that those proteins can be translocated across cellular membranes. Our results suggest that the *Globobulimina*-associated Desulfobacteraceae are able to reduce NO_3^- to NO_2^-, and thus contribute to the foraminiferal denitrification by performing the first reaction in the denitrification pathway for which we found no evidence in the foraminifera transcriptomes.

Discussion

Our results demonstrate that foraminifera are habitat to bacterial communities that may play a role in their ability to thrive in oxygen-depleted habitats. A recent study demonstrates the endosymbiotic contribution to denitrification within a ciliate host[56]. However, previous studies of denitrification in foraminifera argued against the possibility of bacterial contribution to foraminiferal denitrification[8,11]. Notably, most species in the foraminifera microbiota are considered strict anaerobes, hence, when exposed to oxygen the foraminifera may lose their associated microbiota. Metagenomics sequencing of samples frozen directly after sampling is thus an important source of information on the composition and function of the foraminifera microbiome.

Our results indicate that Desulfobacteracea members of the foraminifera microbiota can utilize the NO_3^- storage accumulated by their host. Whether the NO_3^- reduction by bacteria is beneficial to the foraminifera remains, however, an open question. The bacteria may use the NO_3^- it for their own respiratory processes or to build up organic compounds via the assimilation of NO_3^-. For example, it has been suggested that a proportion of NO_3^- taken up by the foraminifera *Ammonia beccari* was used for amino acid synthesis, probably by resident bacteria[21]. Previously we speculated about a foraminiferal sulfite reductase homolog performing the conversion of NO_3^-[11]. Yet, considering the absence of foraminifera homologs to NO_3^- reductase and the presence of NapA in all sampled Desulfobacteraceae MAGs, it is tenable to hypothesize that the reduction of NO_3^- to NO_2^- in foraminifera is performed by resident bacteria.

An association between foraminifera and Desulfobacteracea has been previously reported. For example, the presence of a putative Deltaproteobacterium (Desulfobacteracea) was previously described for *Virgulinella fragilis*[20], a foraminifera that we here predict to be a denitrifying species (Figure 3B). Furthermore, foraminifera from the genera *Ammonia*, *Elphidium* and *Haynesina* that are not expected to denitrify were found to harbor members of Desulfobacterales bacteria[25]. Hence, the interaction between foraminifera and Desulfobacteracea may involve alternative nutritional currencies. Previous studies refer to the role of Desulfobacteracea in sulphur cycling and carbon/nutrient acquisition[20,21,25]. Especially the carbon fixation capabilities via Wood-Ljungdahl pathway would be highly beneficial for the heterotrophic lifestyle of foraminifera, similarly to other symbioses that involve fixed-carbon as a nutritional currency (e.g., as in deep-sea mussels[57] or sponges[49]).

The radiation of early foraminifera species has been estimated to occur between 690 and 1,150 million years ago (mya), while the first primitive rotaliids appeared in the late Permian around 260 mya[58,59]. Most of the rotaliids superfamilies diverged in the mid to late Triassic between approximately 240 and 200 mya[29,60,61] and the majority of extant species diverged during the Miocene (23.8 - 5.5 mya[62]). Our phylogenetic analyses suggest that clades I and III diversified from within clade II, while the LCA of clade I probably originated from within clade III (Figures 2 & 3). Although most species reported to denitrify were reported from the Peruvian OMZ and Sweden, representatives of corresponding genera are found all over the world (Suppl. Table S5). Therefore, most extant rotaliids likely diversified from a denitrifying LCA far back in time. The origin of foraminiferal denitrification within or after diversification from members of clade II may coincide with the rapid increase of fossil records at the onset of the late Cretaceous ~100 mya[60,61]. Our results therefore indicate that eukaryotic denitrification by rotaliids emerged late in foraminiferal evolution, possibly during the worldwide Cretaceous Ocean Anoxic Events[63,64].

Considering the high conservation of nitrate transporters (Figure 1) and the observation of nitrate storage in many divergent species[9], it is tenable to speculate that foraminifera had the mechanisms for nitrate import and storage long before they evolved the ability to denitrify. NO_3^- transporters are ancient eukaryotic enzymes that play a role in the NO_3^- assimilation machinery and they are encoded in genomes of photoautotrophs such as plants and saprotrophs like fungi[65,66]. In contrast to bacteria, foraminifera (as most heterotrophs) are unable to perform NO_3^- assimilation on their own. Thus, NO_3^- accumulated by the foraminiferal hosts could have fuelled bacterial NO_3^- metabolism of associated bacteria in exchange for organic compounds. We propose that denitrification by symbiotic bacteria was indeed the ancestral state of denitrification in foraminifera, similarly to gromiids[26], which share a common evolutionary origin with foraminifera[67]. The finding of bacterial-like denitrification

354 genes in rotaliids furthermore suggests that foraminifera may have acquired those genes from
355 bacteria[11]. Thus, a rare gene acquisition from a foraminifera-associated, denitrifying bacterium
356 could have been at the origins of foraminiferal (eukaryotic) denitrification. Notably, the rotaliids
357 denitrification gene set is incomplete and it varies among the species sampled here (i.e., *N.*
358 *auris, B.plicata* or even *Rosalina* sp.). Species reported to release N_2O instead of N_2[9] further
359 support our observation. Thus, the presence of a partial denitrification pathway in rotaliids as
360 well as in their resident Desulfobacteracea may suggest that the acquisition of denitrification
361 ability in foraminifera occurred in multiple stages and is not yet complete.
362

363 ## Methods

364 ### Foraminifera sampling

365 Samples were collected off the Peruvian coast in the 2017 Austral winter (R/V Meteor M137)
366 as described by Glock et al. 2019[10]. Briefly, sediment samples were taken with a video-guided
367 sediment multiple corer (MUC) containing 6 liners along a depth transect at 12°S. The top 1
368 to 3 cm of the sediment cores were sampled and immediately wet sieved with surface water
369 through staked sieves with mesh size of 2000µm to 63µm to retrieve benthic foraminifera. The
370 foraminifera were rinsed in sterile seawater obtained by filtering core-overlying seawater with
371 a sterile bottle top filtration system (Durapore filter, 0.2 µm) and a vacuum pump within 40
372 minutes of MUC arrival on deck. Focal species were manually picked using a
373 stereomicroscope. These were characterised morphologically according to the literature[68-70].
374 Up to 160 individuals (classified into focal species) were pooled in one cryo vial (2ml, RNAse
375 free) and flash frozen in liquid nitrogen.

376 ### Microscopy

377 Individual foraminifera were dehydrated in a graded ethanol series (70 %, 80 %, 90 %, 96 %
378 and two times 100 %; 15 min each), air-dried for 12 h in a desiccator and mounted on
379 aluminium stubs (PLANO GmbH) using conductive and adhesive carbon pads (PLANO
380 GmbH). Subsequently, the preparations were sputter-coated with a 10-nm-thick gold-
381 palladium (80/20) layer using a high vacuum sputter coater Leica EM SCD500 (Leica
382 Microsystems GmbH) and visualised with a Hitachi S-4800 field emission scanning electron
383 microscope (Hitachi High-Technologies Corporation) at an acceleration voltages of 3 kV and
384 an emission current of 10 mA applying a combination of the upper detector and the lower
385 detector.

Nucleic acid extraction, and sequencing

Genomic DNA and total RNA from different biological samples were extracted using either the DNeasy® Plant Mini kit (QIAGEN) for genomics (DNA) or simultaneously purified using InnuPREP DNA/RNA Mini Kit (analytikjena) for transcriptomics (RNA). Foraminifera cells in each samples (ca. 160 individuals) were disrupted by pestle-crashing on ice after immersion of the containing cryo-vial in liquid nitrogen. Samples for genomics were treated with lysozyme (200µl of 10mg/ml TE) and proteinase K (1mg/100µl). Samples for transcriptomics samples were treated only with lysozyme (6 ul of 20mg/ml TE) before additional crashing and 2-minute incubation was performed. After disruption and initial lysis, manufacturer protocols for the respective nucleic acid extraction were followed with modifications to the elution volumes (2 x 50µl for DNA and 2 x 25µl for RNA). Libraries for genomics were produced after DNA fragmentation (Covaris target 400, intensity 5, duty cycle 5% cycles per burst 200, 55sec treatment time) using NEBNext® Ultra II DNA Library Prep Kit for Illumina®. Whereas, transcriptomic libraries were produced using NEBNext® Ultra RNA Library Prep Kit for Illumina® (NEB) with mRNA isolation performed with poly-A mRNA beads. All libraries were produced in duplicates from two different sets of pooled individuals for each species and were produced without protocol interruption. Before sequencing, each library was quantified with a Qubit® fluorometer (Invitrogen by Life TechnologiesTM) and qualified using a TapeStation (Agilent technology). The libraries were sequenced paired-end (2 x 150 bp) on an Illumina® HiSeq 4000 platform.

Sequencing datasets of foraminifera

Sequencing resulted in 4.6 billion paired-end reads covering 1.4 tera bases in total. This includes transcriptome and metagenome datasets (BioProject accessions PRJNA494828 & PRJNA503328). Reads were quality-checked by FastQC ver. 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc; Aug 2016). Filtering and trimming of reads was performed using Trimmomatic[71] ver. 0.36 (Parameters: ILLUMINACLIP:primers.fa:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:5 MINLEN:21; the file 'primers.fa' contained adaptor and contaminant sequences provided by Trimmomatic and FastQC). Processed reads from transcriptomes of the two samples per species were assembled into transcript contigs using SPAdes[72] ver. 3.11.1 ("--rna" option). Protein sequences were translated from transcripts as the longest open reading frame (ORF) using TransDecoder[73] ver. v5.0.2 ("-m 30" option). Final protein names consist of the contig ids followed by the sequence positions covered by the CDS and an indicator for the forwards (+) or reverse (-) strand. Transcript abundance of individual transcriptome datasets are referring to Transcripts Per Million (TPM) determined by the Trinity pipeline[73] 2.4.0 (Trinity script 'align_and_estimate_abundance.pl') via RSEM[74] ver. 1.2.30 and Bowtie[75] 2.1.0 using

423 paired-end reads. Additional raw sequencing reads were obtained from the Marine Microbial
424 Transcriptome Project (MMETSP; SRA accessions: *Rosalina* sp., SRR1296887; *Sorites* sp.,
425 SRR1296734; *Ammonia* sp., SRR1300434; *Elphidium margaritaceum*, SRR1300475) and
426 processed as described above. We found that the assembly obtained for *Sorites* sp. contained
427 a high proportion of sequences likely originating from an algal species (Probably
428 *Symbiodinium* sp.), which was removed by applying a binning approach considering only
429 contigs with GC-content ≤38%. Finally, we included data for *Globobulimina* spp. sampled in
430 Sweden from our previous study (Transcriptome data: GloT15[11]) and proteins annotated on
431 the genome assembly of *Reticulomyxa filosa* (NCBI accession: GCA_000512085.1).
432
433 **Species phylogenies and rooting**
434 Transcriptome completeness and heterogeneity was determined by assessing genome
435 completeness via Benchmarking Universal Single-Copy Orthologs[76] (BUSCO v3; lineage
436 'eukaryota') method. Orthologous proteins determined as Complete (or Duplicated) by the
437 BUSCO analysis were merged into protein clusters to study phylogenetic relationships among
438 the species. In case of duplicated BUSCOs in metatranscriptomes, one representative was
439 inferred based on sequence similarity. Therefore, all orthologs retrieved for the same
440 metatranscriptome were compared to all members of the BUSCO protein cluster by global
441 pairwise alignments using needle (EMBOSS tools version 6.6.0[77]). The ortholog with the
442 highest median sequence similarity over all comparisons was picked as the representative
443 sequence. Multiple sequence alignments used in the current study were obtained using
444 MAFFT[78] (Version: 7; parameter: 'linsi') and the phylogenetic trees were reconstructed using
445 IQTREE[79] (Version: 1.5.5 or 1.6.9; default parameters; note that ModelFinder is enabled by
446 default). Phylogenies were reconstructed for the individual BUSCO clusters. Due to their
447 morphological differences from other foraminifera groups, monothalamids like *R. filosa* have
448 been previously used as an outgroup for phylogenetic studies of foraminifera[29]. Since the
449 choice of a distantly related outgroup may lead to erroneous rooted topology due to long
450 branch attraction[80], we further tested the robustness of the root position independent from
451 outgroup species. Root positions were determined using MAD[28] (Parameters: '-bsnn'). The
452 support in each root splits is calculated as the proportion of gene trees where the root position
453 was inferred as such[28]. For the overall species trees, multigene phylogenies were
454 reconstructed using IQTREE (parameter: '-spp') considering all protein cluster alignments of
455 either foraminifera or Rotaliida.
456 Reference 18S rRNA sequences used for the large-scale phylogenies (Suppl. Table
457 S5) were obtained from NCBI, MMETSP, SILVA[81], the foram barcoding project
458 (http://forambarcoding.unige.ch/) and the Planktonic Foraminifera Ribosomal Reference
459 database[82] (PFR[2]). These sequences were used as database sequences to identify 18S

sequences from Peruvian species transcriptomes based on BLAST[83] (version: 2.2.28+; options: '-task blastn -evalue 1e^{-10}') searches. Among the hits retrieved we searched for a single representative transcript contig per transcriptome assembly. The transcripts with the highest product of sequence length and sequencing depth (i.e., coverage given by SPAdes) were determined as representative sequences for most of the focal species. However, for *B. costata*, *V. inflata*, *G. pacifica* and *N. auris* different representative sequences were chosen by giving sequence information (i.e., longer sequences) a higher weight than sequencing depth. To obtain 18S phylogenies first the reference sequences were aligned with MAFFT and subsequently representatives 18S rRNA sequences of the Peruvian transcriptomes were added (MAFFT options: '--addfragments --adjustdirection') to the alignment followed by the tree reconstruction using IQTREE.

Identification of foraminiferal denitrification proteins

To identify homologs to enzymes in the denitrification pathway, we used a similar approach as previously established in Woehle et al. 2018[11] using the corresponding protein database expanded by the protein sequences identified for *Globobulimina*[11]. The search for denitrification enzymes homologs in the transcriptome assemblies was performed with BLASTP (parameter: '-max_target_seqs 1000000 e-value 1e-5'). Protein sequences of hits with query coverage ≥ 40% and sequence identity ≥ 20% were extracted to obtain a first set of homologs. We further applied a cutoff to discard lowly represented transcripts with TPM < 2 in at least one of the two replicates sequenced. With the resulting protein set, we reiterated searches in the non-redundant NCBI protein (NR; version May 2018) and the RefSeq 88 database[84] using diamond v0.9.22 applying the '--more-sensitive' option. First best hit sequences per query were obtained and clustered with CD-HIT 4.6[85] (option: "-c 0.98") to reduce sequence redundancy. All obtained protein sequences of a given enzyme were aligned with MAFFT and phylogenetic trees were reconstructed with IQTREE (parameters: '-bb 1000 -alrt 1000'). The trees were rooted using an outgroup, if available, or MAD.

Metagenomics processing

For the visualisation of metagenomic composition, trimmed paired-end reads from Peruvian species samples and the *Globobulimina* 'Ambient' samples[11] were subsampled via BBMap (version 36.84; https://sourceforge.net/projects/bbmap/; 'reformat.sh' script; parameters: 'samplereadstarget=10000000 addslash=t'). The resulting reads were mapped against the NR database with ac-diamond[86]) and classified using MEGAN6[87] (Version 6.15.0; options '-a2t prot_acc2tax-Nov2018X1.abin -a2eggnog acc2eggnog-Oct2016X.abin -a2seed acc2seed-May2015XX.abin'). MEGAN6 was used for assessment of metagenomics communities and visual representation (See Figure 4) that were classified up to the order rank level with different

taxon sets (e.g., all nodes or only the bacterial subtree). Metagenome assemblies were obtained by combining all trimmed reads per focal species using MEGAHIT[88] (Version 1.1.3; default settings). Individual bacterial genomes were obtained using the binning approach implemented with MaxBin2[89] (Version 2.2.4; parameter: '-min_contig_length 500') using coverage information of the two samples per species of foraminifera. Binning statistics were assessed using CheckM[90] (Version 1.0.11); a threshold for completeness of at least 80% and contamination of maximum 20% was applied to classify bacterial bins as draft genomes. The protein sequences obtained via checkM were used for further classifications. The average nucleotide identity (ANI) was calculated using a perl script obtained from https://github.com/chjp.

For the taxonomic assignment of genome bins, we used the diamond tool to find first best hits for each protein of a genome bin against the NR database (option: '-k 10'; *e*-value ≤ 1e-10) and retrieved the corresponding taxonomic assignments (as we previously described[11]). For each bin, identical taxonomic hierarchies were counted with the genus being the lowest rank considered, sorted accordingly and stepwise searched for the lowest taxonomic rank supporting >50% of protein bin hits starting with the most abundant taxonomy. 'Environmental samples' and 'Cellular organisms' were not considered. For the phylogenetic reconstruction of Desulfobacteracea we determined protein families as protein clusters using sequences of Desulfobacteracea bins associated with *Globobulimina* and additional Desulfobacteracea genomes downloaded from NCBI. First, we determined reciprocal best BLAST hit pairs (rBBH; parameter: '-evalue 1e-5') between all the Desulfobacteracea protein-coding sequences[91]. Then rBBH pairs were globally aligned with the needle tool and pairs with ≥30% identical amino acids were sorted into clusters using the Markov clustering algorithm[92] (MCL; version 12-135). The 33 resulting protein clusters that contained a single copy for each of the Desulfobacteraceae strains (i.e., universal single-copy clusters) were aligned with MAFFT. The resulting alignments were concatenated and a splits network was reconstructed with SplitsTree[93] (version 4.15.1). Prediction of protein function for bacterial genomes were obtained using the KEGG Automatic Annotation Server[94] (KAAS; BBH method via BLAST and the following species set: hsa, dme, ath, sce, pfa, eco, sty, hin, pae, nme, hpy, rpr, mlo, bsu, sau, lla, spn, cac, mge, mtu, ctr, bbu, syn, aae, mja, afu, pho, ape, geo, dvu, dat, dpr, dol, dal, dak, dps, drt, dba, dao, dbr). Global pairwise identities for NapA protein sequences were inferred with needle. Signal peptides were predicted using the SignalP-5.0 webservice for gram-negative bacteria[95].

Data availability

Sequencing reads are deposited at the single read archive (SRA) accessions SRR8144071 to SRR8144090 and SRR7971179 to SRR7971198. The assemblies are available at the

534 transcriptome sequencing archive (TSA; See Suppl. Table 1 for accessions) and as whole
535 genome shotgun (WGS) projects (See Suppl. Table 7 for accessions). All other information
536 on accessing data analysed in this study is included in the manuscript or in the supplemental
537 information.
538
539 **Acknowledgments**
540 We thank Devani Romero Picazo and Anne Kupczok for critical comments on the manuscript.
541 We gratefully acknowledge the scientific party and crew of the R/V Meteor cruise M137 as
542 well as Asmus Petersen and Matthias Türk for their support at sea. Samples from Peru were
543 obtained according to Peruvian access and benefit sharing regulations. We thank the Nagoya
544 officer of Kiel University - Dr. Scarlett Sett - for her support of our research. We further thank
545 Natalia Bernabe Lopez for her assistance in collecting metadata for the 18S sequences.
546 Genome sequencing was performed in the Centre for Genome Analysis Kiel that is funded by
547 the German Research Foundation (DFG). The study was supported by the Deutsche
548 Forschungsgemeinschaft (DFG) via the SFB 754 on Climate–Biogeochemistry Interactions in
549 the Tropical Ocean, the cluster of excellence The Future Ocean, and the European Research
550 Council (Grant No. 281357 awarded to TD). C.W. was supported by the Kiel Life Science (KLS)
551 Young Scientist Programme and would like to thank the Max Planck-Genome-centre Cologne
552 for their support in data analysis.

553 **Author contributions**

554 C.W., A.-S.R., N.G., T.W., J.W., D.R., J.S. & T.D. designed the research strategy and
555 performed sampling. C.W. analysed the transcriptomes and metagenomes. A.-S.R. performed
556 the experimental lab work. C.H. constructed incubation chambers. J.M. and S.N.G. produced
557 micrographs. C.W. and T.D. wrote the manuscript with input from all co-authors.
558

559 **Declaration of Interests**

560 The authors declare no competing interests.

561

562 **References:**

563 1. Cox, A. N. *Allen's Astrophysical Quantities*. (Springer, 2015). doi:10.1007/978-1-4612-
564 1186-0
565 2. Thamdrup, B. New Pathways and Processes in the Global Nitrogen Cycle. *Annu. Rev.*
566 *Ecol. Evol. Syst.* **43,** 407–428 (2012).

3. Gruber, N. in *The Ocean Carbon Cycle and Climate* **35,** 97–148 (Springer, Dordrecht, 2004).
4. Gruber, N. & Galloway, J. N. An Earth-system perspective of the global nitrogen cycle. *Nature* **451,** 293–296 (2008).
5. Lam, P. *et al.* Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA* **106,** 4752–4757 (2009).
6. Philippot, L. Denitrifying genes in bacterial and Archaeal genomes. *Biochim. Biophys. Acta* **1577,** 355–376 (2002).
7. Shoun, H., Kim, D. H., Uchiyama, H. & Sugiyama, J. Denitrification by fungi. *FEMS Microbiology Letters* **73,** 277–281 (1992).
8. Risgaard-Petersen, N. *et al.* Evidence for complete denitrification in a benthic foraminifer. *Nature* **443,** 93–96 (2006).
9. Piña-Ochoa, E. *et al.* Widespread occurrence of nitrate storage and denitrification among Foraminifera and Gromiida. *Proc Natl Acad Sci USA* **107,** 1148–1153 (2010).
10. Glock, N. *et al.* Metabolic preference of nitrate over oxygen as an electron acceptor in foraminifera from the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA* **116,** 2860–2865 (2019).
11. Woehle, C. *et al.* A Novel Eukaryotic Denitrification Pathway in Foraminifera. *Curr Biol* **28,** 2536–2543.e5 (2018).
12. Glock, N. *et al.* The role of benthic foraminifera in the benthic nitrogen cycle of the Peruvian oxygen minimum zone. *Biogeosciences* **10,** 4767–4783 (2013).
13. Schweizer, M., Pawlowski, J., Kouwenhoven, T. J., Guiard, J. & van der Zwaan, B. Molecular phylogeny of Rotaliida (Foraminifera) based on complete small subunit rDNA sequences. *Marine Micropaleontology* **66,** 233–246 (2008).
14. Murray, J. W. *Ecology and applications of benthic foraminifera*. (Cambridge University Press, 2006).
15. LeKieffre, C. *et al.* Surviving anoxia in marine sediments: The metabolic response of ubiquitous benthic foraminifera (Ammonia tepida). *PLoS One* **12,** e0177604 (2017).
16. Bernhard, J. M. *et al.* Potential importance of physiologically diverse benthic foraminifera in sedimentary nitrate storage and respiration. *J Geophys Res* **117,** G03002 (2012).
17. Gomaa, F. *et al.* Multiple integrated metabolic strategies allow foraminiferan protists to thrive in anoxic marine sediments. *Sci Adv* **7,** (2021).
18. Orsi, W. D. *et al.* Anaerobic metabolism of Foraminifera thriving below the seafloor. *ISME J* **14,** 2580–2594 (2020).
19. Zilber Rosenberg, I. & Rosenberg, E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews* **32,** 723–735 (2008).
20. Tsuchiya, M. *et al.* Cytologic and genetic characteristics of endobiotic bacteria and kleptoplasts of *Virgulinella fragilis* (Foraminifera). *J Eukaryotic Microbiology* **62,** 454–469 (2015).
21. Nomaki, H. *et al.* Nitrate uptake by foraminifera and use in conjunction with endobionts under anoxic conditions. *Limnology and Oceanography* **59,** 1879–1888 (2014).
22. Bernhard, J. M., Martin, J. B. & Rathburn, A. E. Combined carbonate carbon isotopic and cellular ultrastructural studies of individual benthic foraminifera: 2. Toward an understanding of apparent disequilibrium in hydrocarbon seeps. *Paleoceanography* **25,** PA4206 (2010).
23. Bernhard, J. M., Habura, A. & Bowser, S. S. An endobiont-bearing allogromiid from the Santa Barbara Basin: Implications for the early diversification of foraminifera. *J Geophys Res* **111,** 399 (2006).
24. Bernhard, J. M. Potential symbionts in bathyal foraminifera. *Science* **299,** 861–861 (2003).
25. Salonen, I. S., Chronopoulou, P. M., Bird, C., Reichart, G. J. & Koho, K. A. Enrichment of intracellular sulphur cycle –associated bacteria in intertidal benthic

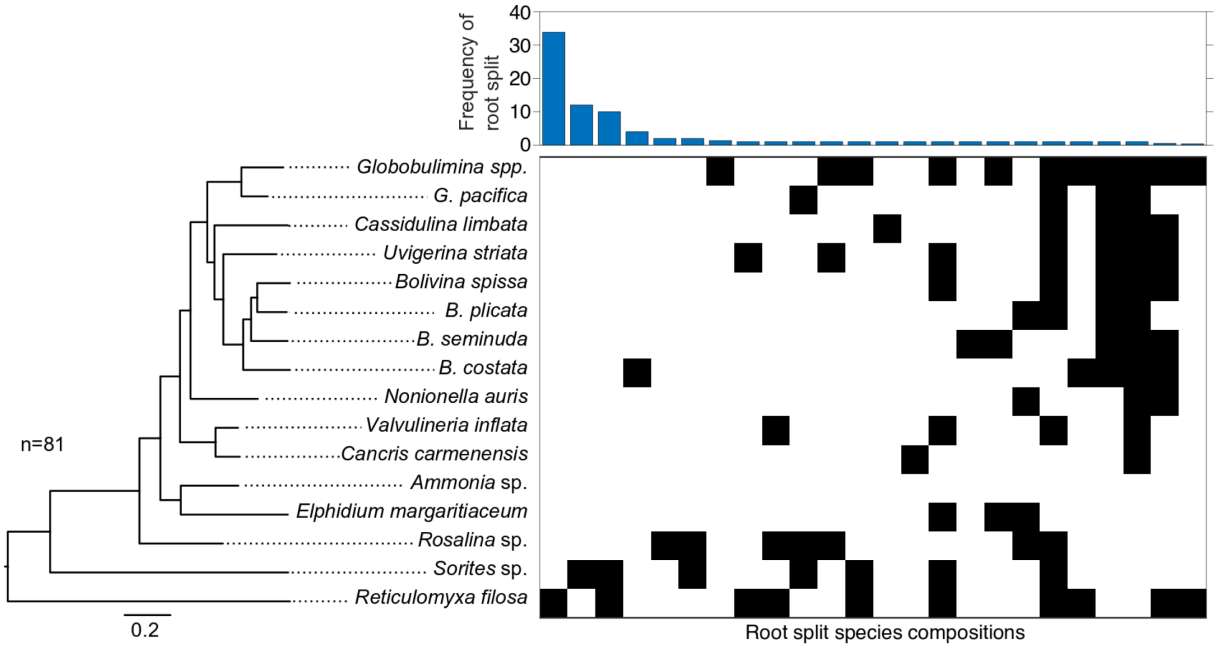
622 foraminifera revealed by 16S and aprA gene analysis. *Sci Rep* 1–12 (2019).
623 doi:10.1038/s41598-019-48166-5
624 26. Høgslund, S., Cedhagen, T., Bowser, S. S. & Risgaard-Petersen, N. Sinks and
625 Sources of Intracellular Nitrate in Gromiids. *Front Microbiol* **8,** 617 (2017).
626 27. Bernhard, J. M., Edgcomb, V. P., Casciotti, K. L., McIlvin, M. R. & Beaudoin, D. J.
627 Denitrification likely catalyzed by endobionts in an allogromiid foraminifer. *ISME J* **6,**
628 951–960 (2012).
629 28. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor
630 deviation. *Nat Ecol Evol* **1,** 0193 (2017).
631 29. Pawlowski, J. *et al.* The evolution of early Foraminifera. *Proc Natl Acad Sci USA* **100,**
632 11494–11498 (2003).
633 30. Ertan, K. T., Hemleben, V. & Hemleben, C. Molecular evolution of some selected
634 benthic foraminifera as inferred from sequences of the small subunit ribosomal DNA.
635 *Marine Micropaleontology* **53,** 367–388 (2004).
636 31. Mackensen, A., Schmiedl, G., Harloff, J. & Giese, M. Deep-Sea Foraminifera in the
637 South Atlantic Ocean: Ecology and Assemblage Generation. *Micropaleontology* **41,**
638 342 (1995).
639 32. Rathburn, A. E., Willingham, J., Ziebis, W., Burkett, A. M. & Corliss, B. H. A New
640 biological proxy for deep-sea paleo-oxygen: Pores of epifaunal benthic foraminifera.
641 *Sci Rep* **8,** 9456–8 (2018).
642 33. Hoogakker, B. A. A. *et al.* Glacial expansion of oxygen-depleted seawater in the
643 eastern tropical Pacific. *Nature* **562,** 410–413 (2018).
644 34. Widdel, F. New types of acetate-oxidizing, sulfate-reducing Desulfobacter species, *D.*
645 *hydrogenophilus* sp. nov., *D. latus* sp. nov., and *D. curvatus* sp. nov. *Arch. Microbiol.*
646 **148,** 286–291 (1987).
647 35. Lage, O. M. & Bondoso, J. Planctomycetes and macroalgae, a striking association.
648 *Front Microbiol* **5,** 267 (2014).
649 36. Oshiki, M., Satoh, H. & Okabe, S. Ecology and physiology of anaerobic ammonium
650 oxidizing bacteria. *Environ. Microbiol.* **18,** 2784–2796 (2016).
651 37. Noguchi, T. *et al. Vibrio alginolyticus*, a tetrodotoxin-producing bacterium, in the
652 intestines of the fish *Fugu vermicularis vermicularis*. *Marine Biology* **94,** 625–630
653 (1987).
654 38. McFall-Ngai, M. Divining the essence of symbiosis: insights from the squid-vibrio
655 model. *PLoS Biol.* **12,** e1001783 (2014).
656 39. Ben-Haim, Y., Zicherman-Keren, M. & Rosenberg, E. Temperature-regulated
657 bleaching and lysis of the coral *Pocillopora damicornis* by the novel pathogen *Vibrio*
658 *coralliilyticus*. *Appl. Environ. Microbiol.* **69,** 4236–4242 (2003).
659 40. López-Pérez, M. & Rodriguez-Valera, F. in *The Prokaryotes: Gammaproteobacteria*
660 (eds. Rosenberg, E., Delong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 69–92
661 (Springer Berlin Heidelberg, 2014).
662 41. Ivanova, E. P., Flavier, S. & Christen, R. Phylogenetic relationships among marine
663 Alteromonas-like proteobacteria: emended description of the family
664 Alteromonadaceae and proposal of Pseudoalteromonadaceae fam. nov.,
665 Colwelliaceae fam. nov., Shewanellaceae fam. nov., Moritellaceae fam. nov.,
666 Ferrimonadaceae fam. nov., Idiomarinaceae fam. nov. and Psychromonadaceae fam.
667 nov. *Int J Syst Evol Microbiol* **54,** 1773–1788 (2004).
668 42. Simon, M. *et al.* Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation
669 to marine and non-marine habitats. *ISME J* **11,** 1483–1499 (2017).
670 43. Newton, R. J. *et al.* Genome characteristics of a generalist marine bacterial lineage.
671 *ISME J* **4,** 784–798 (2010).
672 44. Buchan, A., Gonzalez, J. M. & Moran, M. A. Overview of the marine roseobacter
673 lineage. *Appl. Environ. Microbiol.* **71,** 5665–5677 (2005).
674 45. Hjelm, M. *et al.* Selection and identification of autochthonous potential probiotic
675 bacteria from turbot larvae (*Scophthalmus maximus*) rearing units. *Syst. Appl.*
676 *Microbiol.* **27,** 360–371 (2004).

46. Webster, N. S., Negri, A. P., Munro, M. M. H. G. & Battershill, C. N. Diverse microbial communities inhabit Antarctic sponges. *Environ. Microbiol.* **6,** 288–300 (2004).
47. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9,** 5114–8 (2018).
48. Dubilier, N. *et al.* Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* **411,** 298–302 (2001).
49. Kamke, J. *et al.* Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* **7,** 2287–2300 (2013).
50. Wein, T. *et al.* Currency, Exchange, and Inheritance in the Evolution of Symbiosis. *Trends Microbiol.* **27,** 836–849 (2019).
51. Strittmatter, A. W. *et al.* Genome sequence of *Desulfobacterium autotrophicum* HRM2, a marine sulfate reducer oxidizing organic carbon completely to carbon dioxide. *Environ. Microbiol.* **11,** 1038–1055 (2009).
52. González, P. J., Correia, C., Moura, I., Brondino, C. D. & Moura, J. J. G. Bacterial nitrate reductases: Molecular and biological aspects of nitrate reduction. *J. Inorg. Biochem.* **100,** 1015–1023 (2006).
53. Dias, J. M. *et al.* Crystal structure of the first dissimilatory nitrate reductase at 1.9 A solved by MAD methods. *Structure* **7,** 65–79 (1999).
54. Widdel, F. & Pfennig, N. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids II. Incomplete oxidation of propionate by *Desulfobulbus propionicus* gen. nov., sp. nov. *Arch. Microbiol.* **131,** 360–365 (1982).
55. Berks, B. C., Ferguson, S. J., Moir, J. W. & Richardson, D. J. Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. *Biochim. Biophys. Acta* **1232,** 97–173 (1995).
56. Graf, J. S. *et al.* Anaerobic endosymbiont generates energy for ciliate host by denitrification. *Nature* **591,** 445–450 (2021).
57. Duperron, S. *et al.* A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environ. Microbiol.* **8,** 1441–1447 (2006).
58. Maklay, M. Novye Rannekamennougolnye Arkhedistsidy. *Min. Geol. Okhrana. Nedr. USSR* **1,** 158–159 (1964).
59. Crespin, I. Permian Foraminifera of Australia. *Bur. Miner. Resour. Aust. Bull.* **48,** 1–207 (1958).
60. Tappan, H. & Loeblich, A. R. Foraminiferal evolution, diversification, and extinction. *Journal of Paleontology* **62,** 695–714 (1988).
61. Loeblich, A. R. & Tappan, H. *Treatise on Invertebrate Paleontology. Part C. Protista 2. Sarcodina chiefly 'thecamoebians' and Foraminiferida.* (Geological Society of America New York/University of Kansas Lawrence, 1964).
62. Kucera, M. & Schönfeld, J. The origin of modern oceanic foraminiferal faunas and Neogene climate change. *The Micropalaeontology Society Special Publications* 409–425 (2007).
63. Schlanger, S. O. & Jenkyns, H. C. Cretaceous oceanic anoxic events: causes and consequences. *Geologie en Mijnbouw* **55,** 179–184 (1976).
64. Jenkyns, H. C. Cretaceous anoxic events: from continents to oceans. *Journal of the Geological Society* **137,** 171–188 (1980).
65. Forde, B. G. Nitrate transporters in plants: structure, function and regulation. *Biochim. Biophys. Acta* **1465,** 219–235 (2000).
66. Pao, S. S., Paulsen, I. T. & Saier, M. H. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62,** 1–34 (1998).
67. Sierra, R. *et al.* Deep relationships of Rhizaria revealed by phylogenomics: A farewell to Haeckel's Radiolaria. *Mol Phylogenet Evol* **67,** 53–59 (2013).
68. Figueroa, S., Marchant, M., Giglio, S. & Ramírez, M. Foraminíferos bentónicos rotalínidos del centro sur de Chile (36°S - 44°S). *Gayana* **69,** 329–363 (2005).

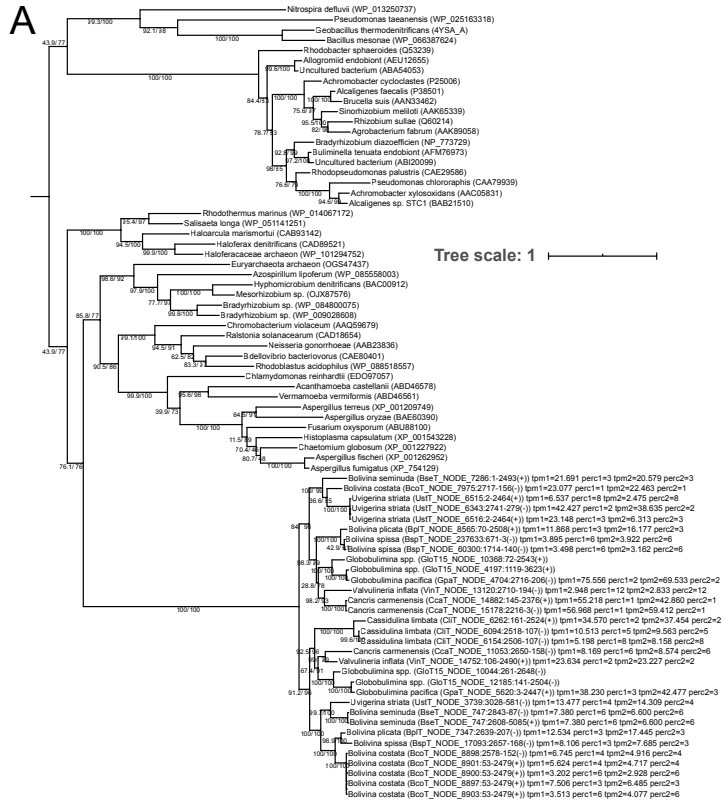
69. Erdem, Z. Reconstruction of past bottom water conditions of the Peruvian Oxygen Minimum Zone (OMZ) for the last 22,000 years and the benthic foraminiferal response to (de)oxygenation. *PhD thesis, Kiel University* (2016).
70. Mallon, J. Benthic Foraminifera of the Peruvian and Ecuadorian Continental Margin. *PhD thesis, Kiel University* (2012).
71. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
72. Nurk, S. *et al.* Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* **20,** 714–737 (2013).
73. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8,** 1494–1512 (2013).
74. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
75. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359 (2012).
76. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).
77. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16,** 276–277 (2000).
78. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30,** 772–780 (2013).
79. Nguyen, L.-T., Schmidt, H. A., Haeseler, von, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32,** 268–274 (2015).
80. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27,** 401–410 (1978).
81. Yilmaz, P. *et al.* The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res* **42,** D643–8 (2014).
82. 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. *Molecular Ecology Resources* **15,** 1472–1485 (2015).
83. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25,** 3389–3402 (1997).
84. Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* **40,** D130–D135 (2012).
85. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22,** 1658–1659 (2006).
86. Mai, H. *et al.* AC-DIAMOND v1: accelerating large-scale DNA-protein alignment. *Bioinformatics* **34,** 3744–3746 (2018).
87. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol* **12,** e1004957 (2016).
88. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31,** 1674–1676 (2015).
89. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32,** 605–607 (2016).
90. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25,** 1043–1055 (2015).

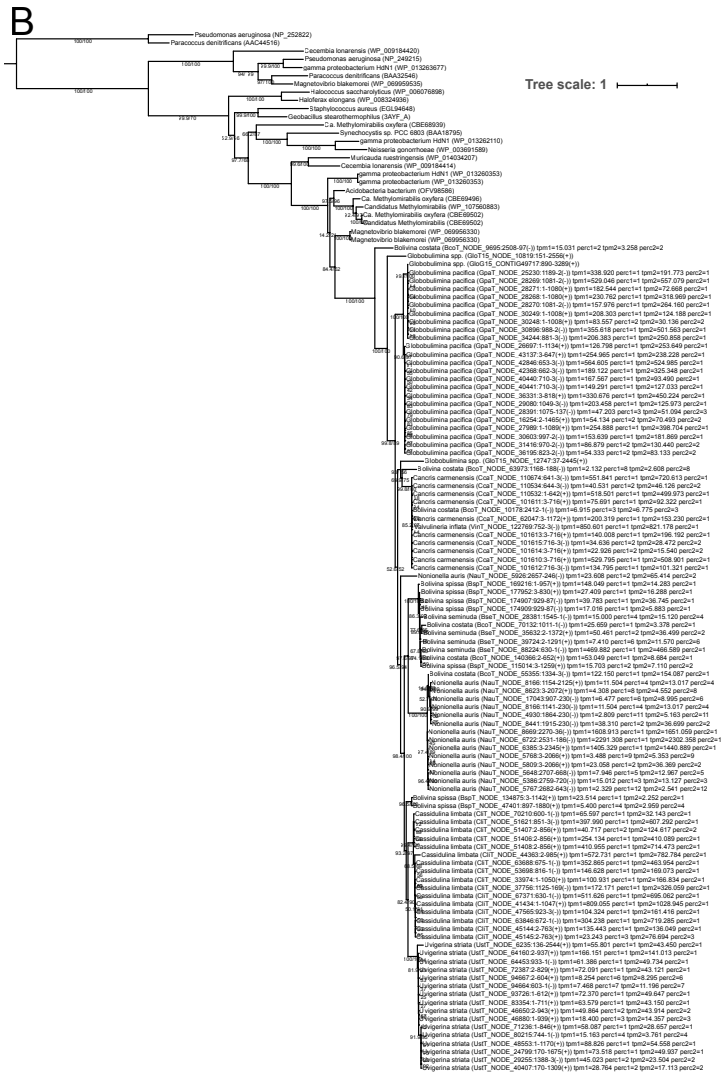
91. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278,** 631–637 (1997).
92. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30,** 1575–1584 (2002).
93. Huson, D. H. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23,** 254–267 (2005).
94. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35,** W182–5 (2007).
95. Nielsen, H. Predicting Secretory Proteins with SignalP. *Methods Mol. Biol.* **1611,** 59–73 (2017).

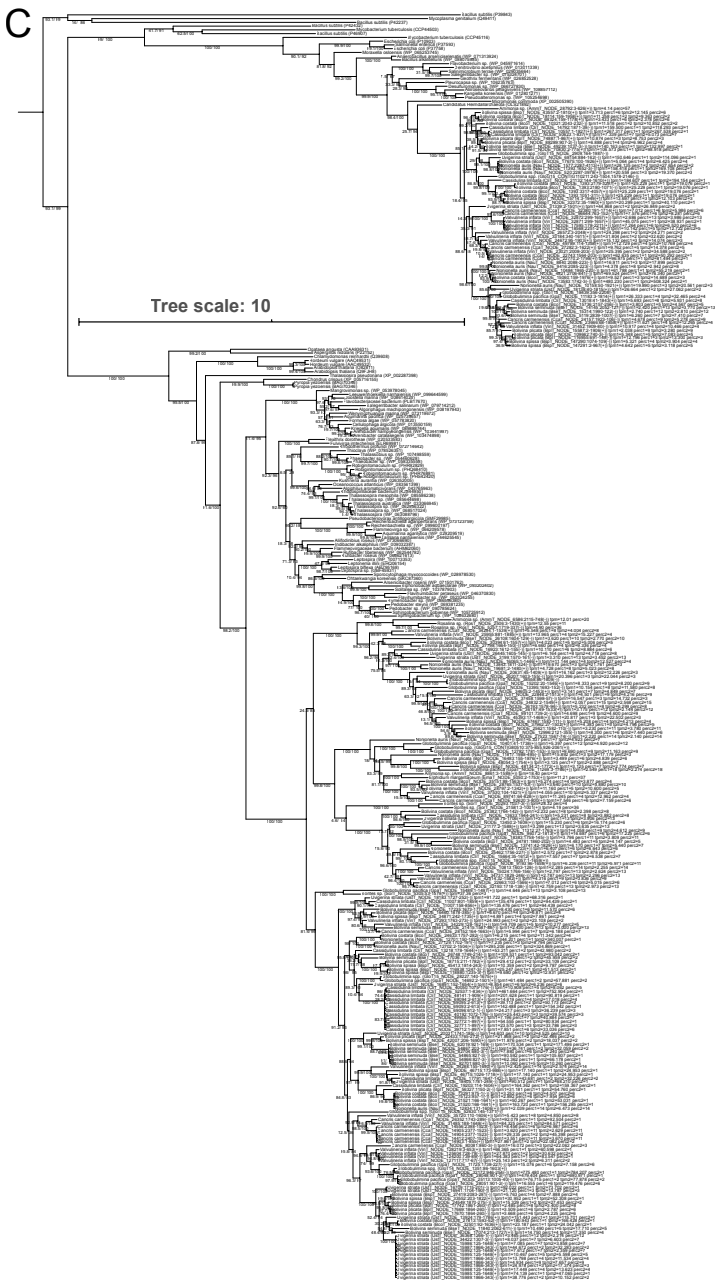
Supplementary material



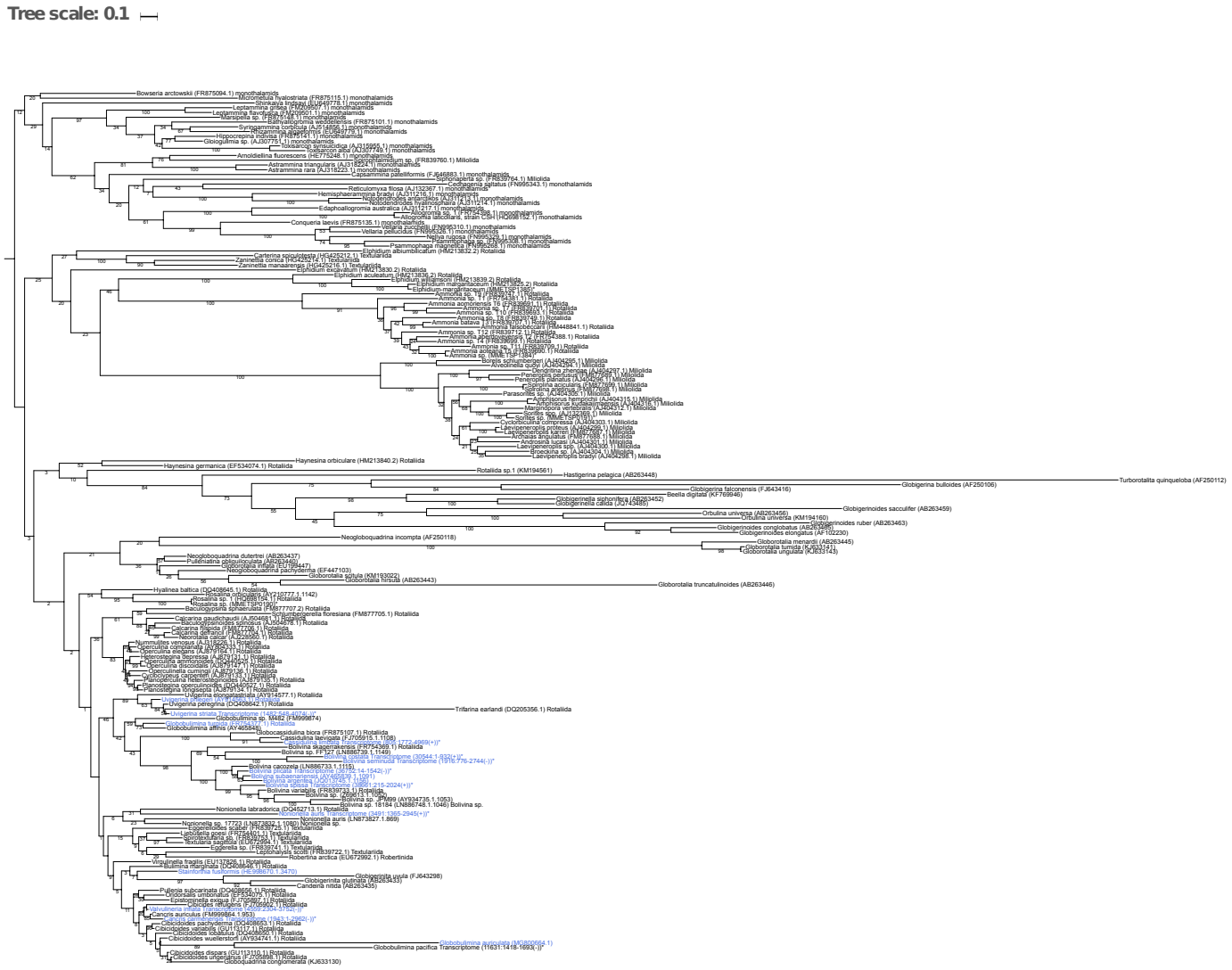
Suppl. Figure S1. Foraminifera rooting based on single-protein trees. A Maximum-likelihood phylogenetic reconstruction of foraminifera species based on 81 eukaryotic protein marker sequences is shown on the left. Parametric bootstrap support is 1000/1000 at all the branches. Right of the phylogeny a split representation of root splits determined via the MAD approach for the 81 single gene trees is shown. Each column represents a root. Root branches are reported as separation of two groups (black and white boxes) indicating for the species found on either side of the root split. The bar graph (top) reports the single gene tree count supporting the corresponding root splits. These were ranked by frequency.







Suppl. Figure S2. Complete phylogenies of NirK, Nor and Nrt homologs. Phylogenetics trees of the foraminiferal denitrification proteins A) NirK B) Nor and C) Nrt that survived the cutoff together with homologs from public databases. Transcriptome labels contain transcription values of the two replicates (tpm1 & tpm2) as well as their percentile expression rank within corresponding transcriptomes. The labels contain as well the transcript id, region and orientation for individual open reading frames used. Numbers of the branches represent statistical support by ultrafast bootstrap and SH-like approximate likelihood ratio test results (following the '/') each with 1000 replicates. Branch support values of zero are omitted. If only one value is shown it represents the SH-like approximate likelihood ratio test.



Suppl. Figure S3. Uncollapsed 18S phylogeny of foraminifera. Species shown to denitrify experimentally are highlighted in blue. The only exception is *Stainforthia fusiformis*, where denitrification activity has been shown for an unspecified species of the same genus. Species also shown in Figure 2 or Figure S1 are highlighted by asterisks. Bootstrap support values with 1000 replicates are shown at the branches. The trees were rooted by the clade of monothalamids containing *R. filosa*. The phylogeny corresponds to the one shown in Figure 3.

Suppl. Table S1. Sequencing statistics for metagenomics & transcriptomics.

Supp. Table S2. List of foraminiferal homologs of denitrification proteins remaining after applying cutoffs.

Supp. Table S3. List of foraminiferal homologs of denitrification proteins removed by cutoffs.

Suppl. Table S4. List of marker proteins used for phylogenetic reconstructions.

Suppl. Table S5. 18S sequences including metadata for phylogenetic reconstructions.

Suppl. Table S6. Proportions of family rank classification of foraminifera-associated microbial communities.

Suppl. Table S7. High quality genomic bins of foraminifera-associated species communities.

Suppl. Table S8. Metabolic classification of Desulfobacteracea genomes.