Denitrification in foraminifera has ancient origin 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 General 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: Center for Ecology and Evolution in Microbial model Systems-EEMiS, Linnaeus University, Kalmar, 39182, Sweden

* Equally contributing authors

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

Keywords: eukaryotic evolution, Foraminifera, denitrification pathway, nitrogen cycle, microbiome

Abstract

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 origin. However, the prevalence of denitrification genes in foraminifera remains unknown and the missing denitrification pathway components are elusive. Analysing transcriptomes and metagenomes of ten foraminiferal 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 the ability to denitrify for additional foraminiferal species. Additionally, an examination of the foraminiferal microbiota reveals evidence for a stable interaction with *Desulfobacteracea*, which harbour genes that complement the foraminiferal denitrification pathway. Our results provide evidence that foraminiferal denitrification is complemented by the foraminifera-associated microbiome. The interaction of foraminifera with their resident bacteria is at the basis of foraminiferal adaptation to anaerobic environments that manifested in ecological success within oxygen depleted habitats.

Significance Statement

A substantial component of the global nitrogen cycle is the production of biologically inaccessible dinitrogen attributed to anaerobic denitrification by prokaryotes. Recent evidence identified a eukaryote – foraminifera – as new key players in this ‘loss’ of bioavailable nitrogen. The evolution of denitrification in eukaryotes constitutes a rare event and the genetic mechanisms of the denitrification pathway in foraminifera are just starting to be elucidated. We present large-scale sequencing analyses of ten denitrifying foraminiferal species, which reveals the high conservation of the foraminiferal denitrification pathway. We further find evidence for a complementation of denitrification by the foraminiferal microbiome. Together, these findings provide novel insights on the early evolution of a previously overlooked component in the marine nitrogen cycle.

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 reported 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).

103 These findings suggest that genes of the denitrification pathway are widespread in rotaliids,
104 however their distribution remains largely unknown.
105 Foraminifera, like other eukaryotes (19), are populated by bacterial organisms that
106 reside out- and in-side their cell (i.e., test) (20-24). Microscopic observations showed that the
107 bacteria are localized in food vacuoles or in the cytoplasm of the cell, which led to the
108 suggestion that interactions of foraminifera with its microbiota may vary from prey-predator
109 interactions or parasitism to metabolic symbiosis (22). Further studies of the microbiota in
110 clade II rotaliids identified sulphate-reducing and sulphur-oxidizing bacteria that were
111 suggested to participate in Sulphur cycling thus providing carbon and other nutrients to the
112 host (20, 21, 25). Notwithstanding, previous studies of foraminifera-associated bacteria used
113 microscopy observations or sequencing of marker genes (e.g., 16S rRNA) that are limited in
114 their resolution (20-25). It has been suggested that the microbiota of denitrifying foraminifera
115 include denitrifying bacteria that utilize the foraminiferal NO_3^- storage, similarly to observations
116 in other bacteria-protist associations [e.g., gromiids (26), allogromiids (27)]. Indeed,
117 metagenomics of the *Globobulimina* microbiota revealed the presence of a taxonomically
118 diverse species community where several members encode homologs of NapA and NosZ (11).
119 This finding gave rise to the hypothesis that the partial foraminiferal denitrification pathway
120 may be complemented by microbiota functions.
121 Here we investigate the evolutionary history of genes along the denitrification pathway
122 in foraminifera and furthermore examine the functional repertoire of the foraminiferal
123 microbiome. For that purpose, we studied populations of ten rotaliid species known to denitrify
124 (10) (except *Globobulimina pacifica* for which no information is available). Our study supplies
125 insights in the evolution of rotaliids and their microbiota and lays the basis for further research
126 of foraminiferal genome evolution.
127

128 ## Results

129 ### Transcriptomes of ten Peruvian rotaliids

130 For the purpose of our study, we sampled benthic foraminifera in the Peruvian oxygen
131 minimum zone. Individual foraminifers from among ten focal species were manually identified
132 and picked using a stereomicroscope (see sampled species in Figure 1A). Eukaryotic
133 transcriptomes and whole metagenomes were sequenced from the sampled foraminifera with
134 two biological replicates per species (ca. 160 individuals per sample). Our computational
135 analysis includes five additional publicly available transcriptomes of foraminifera (Suppl. Table
136 S1) and the genome data of one monothalamid species (*Reticulomyxa filosa*). The
137 assessment of transcriptome completeness showed that ~90% of the eukaryotic marker

138 proteins are present in the data of the ten newly sequenced species. Furthermore, we
139 examined the purity of the focal species in each transcriptome by testing for redundancy of
140 eukaryotic marker proteins that are expected as single-copy genes. As single-copy genes are
141 expected to exist only once in a given species, their redundancy might indicate for additional
142 eukaryotic species present in our samples. The highest proportion of such bystander species
143 were found in the *B. spissa* (78%) and *B. costata* (50%) transcriptomes (Suppl. Table S1).

144 **The Peruvian rotaliids harbour genes of the denitrification pathway**

145 To explore the denitrification mechanisms in the Peruvian foraminifera, we searched for
146 homologs to denitrification pathway genes. Our results reveal that all sampled species harbor
147 homologous genes to Nrt, NirK and Nor (Figure 1B; note that genes are termed by their
148 corresponding protein symbol). Our search for homologs in the publicly available data yielded
149 a putative homolog of NirK in *Rosalina* sp. and Nrt homologs in all but *R. filosa* (Fig. 1B). The
150 additional species we included here, especially *Ammonia sp.*, *Elphidium margaritaceum* and
151 *Sorites sp.*, are considered to reside mostly in oxygenated habitats, hence they are not
152 expected to be able to perform denitrification. These species lack homologs to the
153 denitrification genes, but encode nitrate transporters (Nrt, Figure 1B). Among this group, only
154 *A. tepida* was so far experimentally tested for denitrification ability and reported as a non-
155 denitrifying species (9). The reconstructed phylogenies indicate that all the examined genes
156 are indeed encoded in the foraminiferal genomes, as all of them group with homologs
157 previously reported to originate from the nuclear genomes of *Globobulimina* spp. (11).
158 Furthermore, the phylogenies of Nrt and NirK, reveal two major sub-clades including most, but
159 not all, of the species sampled (Full and half circles in Fig. 1B; Suppl. Figure S2; Suppl. Table
160 S2). Previously we reported the presence of subclades in the Nrt and NirK phylogenies (11),
161 but could not clarify if those were explained by sequencing errors, gene duplication, or
162 speciation events. The phylogenetic trees presented here show that each subclade is
163 represented in most of the rotaliids sampled in the current study. Consequently, we conclude
164 that ancient gene duplications in the Nrt and NirK gene families led to the evolution of two or
165 more sub-clades that encode for different (paralogous) protein subtypes. Overall, our results
166 demonstrate that the denitrification pathway is highly conserved in the Peruvian rotaliids.
167

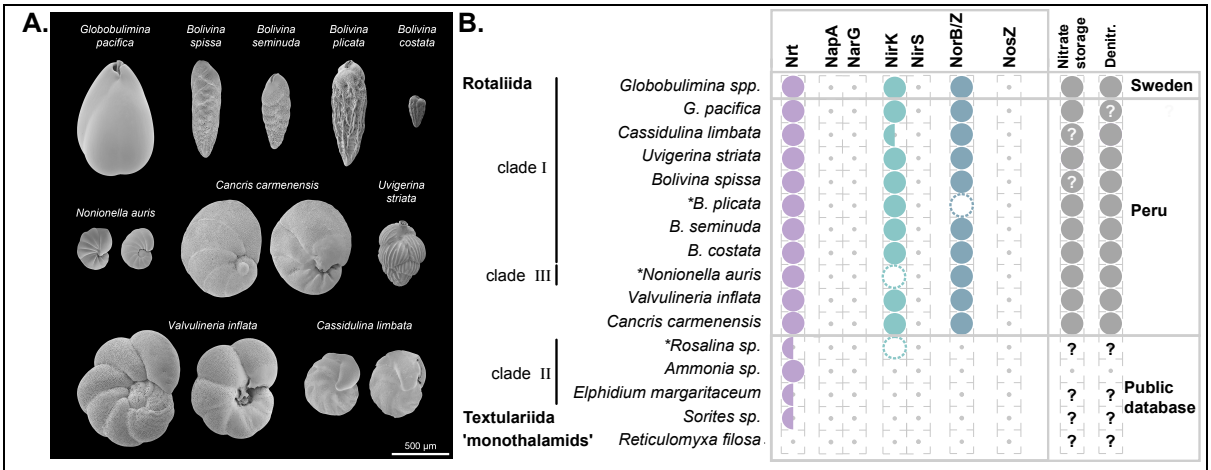


Figure 1. Morphological characteristics of the sampled rotaliids and denitrification gene repertoire. A) Scanning electron micrographs of the sampled species. Species with clearly distinct lateral views are shown from two sides. For more detailed views see Suppl. Figure S1. **B)** Presence/absence of 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 S3; 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 S3). This result is in agreement with previous studies that assumed monothalamids to be an outgroup in foraminiferal 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 of denitrifying species within the order Rotaliida.

185 For the inference of the denitrifying rotaliids last common ancestor (LCA; i.e., the
186 rotaliids root), we reconstructed phylogenetic trees from an extended set of 146 eukaryotic
187 marker protein coding genes considering only members of the Rotaliida (Figure 2; Suppl.
188 Table S4). An examination of the rooted tree topologies revealed that the branch leading to
189 *Ammonia* sp. and *E. margaritaceum* was the most frequently supported root position (14% of
190 single gene trees). This root inference is further supported by an alternative root position at a
191 branch that splits *Ammonia* sp., *E. margaritaceum* and *Rosalina* sp. from the remaining
192 species (10%). Overall, the different alternative root positions support with 35% a Rotaliida
193 LCA within clade II foraminifera (i.e., among *Ammonia sp.*, *E. margaritaceum* & *Rosalina* sp.;
194 Fig. 2). The absence of denitrification genes in clade II members (Fig. 1) further supports the
195 suggestion that the evolution of denitrification in rotaliids occurred by lateral gene transfer from
196 a prokaryotic donor rather than being inherited from the rotaliid LCA (11).
197

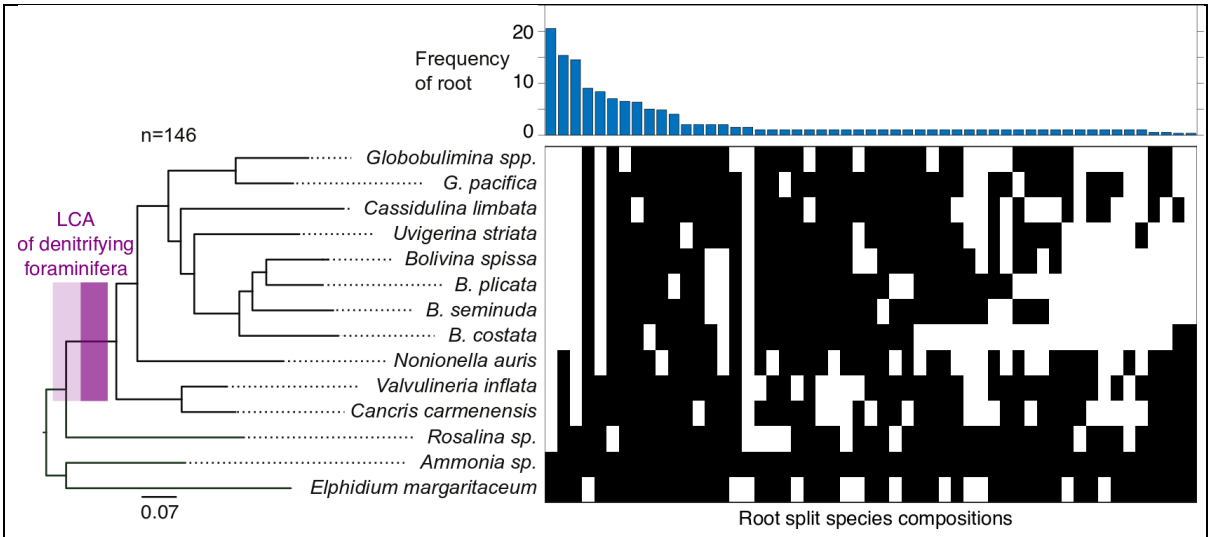


Figure 2. Phylogenetic rooting of rotaliids. A Maximum-likelihood phylogenetic reconstruction of foraminiferal species from 146 eukaryotic protein marker sequences is shown on the left. The branch with the LCA of denitrifying foraminifera is highlighted in purple, where a lighter color includes lower 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.

198

199 **Ancient origin of denitrification in foraminifera**

200 To further study the denitrifying foraminifera LCA, we examined the phylogenetic position of
201 the Peruvian species in the context of a foraminiferal species tree. For that purpose, we
202 extracted sequences of the 18S ribosomal RNA (rRNA) subunit from our transcriptome data
203 and used it in combination with publicly available 18S rRNA sequences to reconstruct a
204 foraminiferal 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 foraminiferal 18S rRNA phylogenies (29, 30).

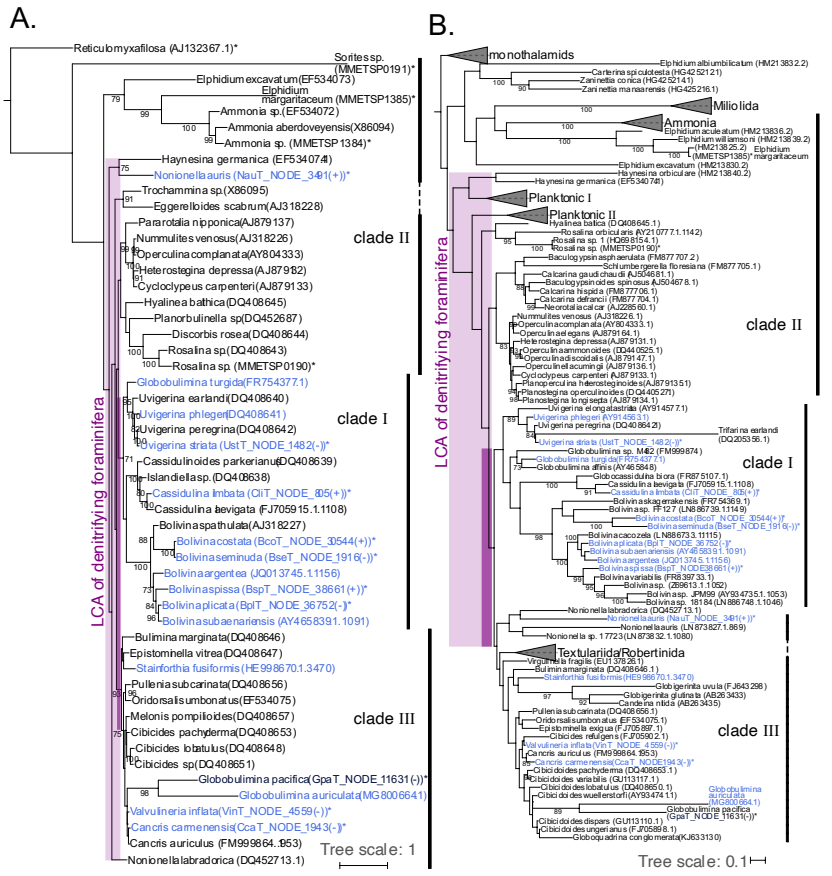


Figure 3. The origin of denitrification in context of the whole foraminiferal group. A) Large-scale phylogenetic representation of Peruvian foraminifera in context of different foraminiferal taxa based on 18S sequences available in public databases. Clades I & III of the order Rotaliida appear are highly 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 1 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 S4.

The inference of the denitrifying foraminiferal LCA enabled us to further reconstruct the origin of denitrification within the foraminiferal species tree. For that purpose, we examined an extended phylogeny with additional foraminiferal groups including planktonic species (Figure 3B; Suppl. Figure S4). While the general taxonomic relationships in the tree were recovered as before (i.e., Figure 3A), many branches had a low statistical support. For example, the rotaliid clades II and III were paraphyletic, in contrast to the phylogeny with less

217 taxa (Figure 3A). Nonetheless, an LCA of clades I and III could be inferred; the previously
218 unassigned genera *Valvulineria* and *Cancris* (9) grouped well with members of clade III in
219 both 18S rRNA phylogenies, hence they can be classified as clade III.
220 Since both clades I and III harbour several denitrifying foraminifera, it is most
221 parsimonious to conclude that LCA of clades I and III was likely a denitrifying organism. Our
222 inference thus suggests the presence of denitrification genes (and hence denitrification
223 capability) in all members of clades I and III including e.g., the genera *Cibicoides* and
224 *Virgulinella*, which were previously not considered as denitrifying species. Indeed,
225 *Cibicidoides wullerstoerfi* has been assumed for a long time to be unable to withstand O_2
226 depletion (31). However, recent studies observed living *Cibicidoides spp.* thriving in
227 environments of < 2 µmol/kg O_2 (32) and fossil specimens in the paleorecord during periods
228 of severe O_2 depletion (33). These recent studies support our prediction that several
229 *Cibicidoides spp.* are able to denitrify.
230

231 **Composition of the foraminiferal microbiome is species-specific**
232 To test for bacterial contribution to the foraminiferal denitrification, we examined the
233 foraminiferal microbiome. For that purpose, we compared the composition of bacterial
234 communities among foraminifera in our sample. In the absence of species-specific interactions,
235 we would expect a strong impact of the environment (i.e., sampling location and water depth)
236 on the microbiome composition. Thus, we compared the microbial community composition
237 between foraminiferal species and their sampling depths (Figure 4A). Our results show that
238 the individual microbiomes are clustered by the foraminiferal species rather than the sampling
239 location, indicating the presence of species-specific bacterial communities in foraminifera. Our
240 results reveal a similar taxonomic composition of bacterial communities associated with
241 *Globobulimina* spp. from Sweden and *G. pacifica* from the Peruvian OMZ (Figure 4A). The
242 similarity in microbiome composition among the *Globobulimina* species indicates the presence
243 of genus-specific microbiome.
244 To identify common key players in foraminiferal microbiota we examined the relative
245 abundance of all bacterial families in their microbiome (Figure 4B; Suppl. Table S6). Our
246 results revealed that the most prevalent bacterial families are Desulfobacteracea followed by
247 Planctomycetaceae. The relative abundance of these two families varied between different
248 species and samples. Furthermore, communities with a high relative abundance of
249 Desulfobacteracea are characterized by a low abundance of Planctomycteaceae and vice
250 versa (r_s=-0.77, p-value < 0.01, using Spearman correlation coefficient and t-test).
251 Desulfobacteracea comprise sulphate-reducing bacteria and several representatives that are
252 able to grow chemoautotrophically (34). Members of this group were previously observed in

253 association with other rotaliids like *Virgulinella fragilis* and members of clade II (20, 25).
254 Planctomycetes are often found in association with macroalgae (35); they are characterized
255 by a diversified metabolism allowing them the of colonization of a wide range of habitats (35).
256

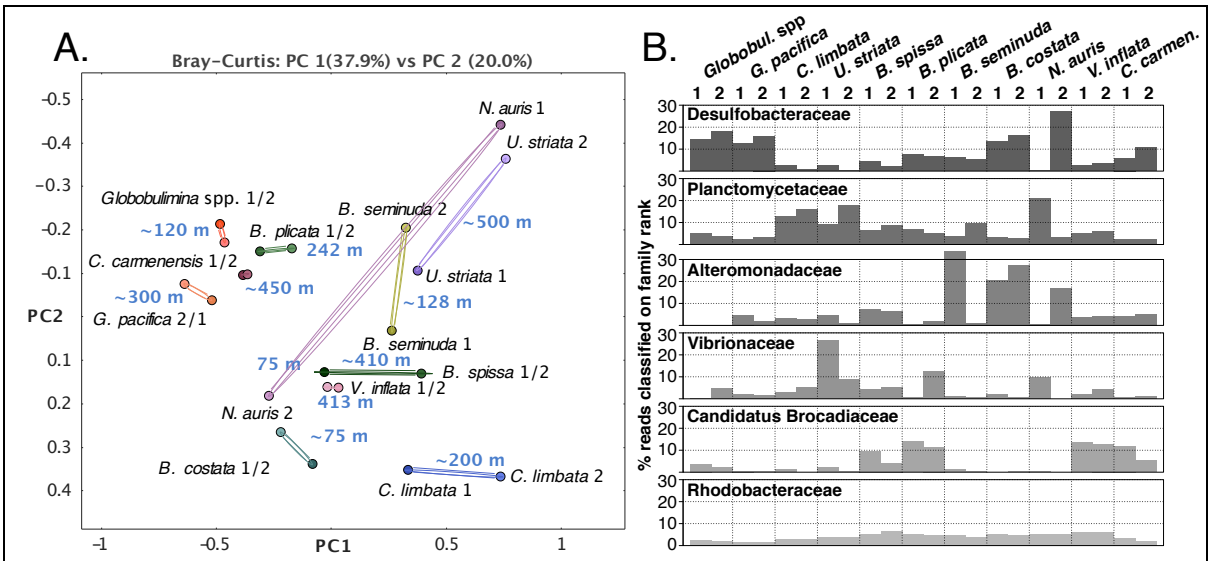


Figure 4. Comparison of microbial communities associated with different foraminifera. A) Principal coordinates analysis (PCoA) clustering of bacterial communities. Each circle represents a bacterial community; duplicates per species are connected by lines. The text next to the circles indicates for species 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.

257

258 The next most abundant families were Alteromonadaceae, Vibrionaceae and
259 `Candidatus Brocadiaceae` that were more diverse in their relative abundances across
260 different species and replicates. Most reported `Candidatus Brocadiaceae` members are
261 autotrophic, obligately anaerobic bacteria performing anammox, which is a dissimilatory
262 pathway where ammonium and nitrite (NO_2^-) are metabolized into N_2 (36). The presence of
263 nitrate storage in foraminifera and NirK suggest that nitrite is readily available inside the
264 foraminiferal test. Vibrionaceae are a diverse group including multiple species that colonize
265 marine organisms either as symbionts (37, 38) or pathogens (39). Alteromonadaceae were
266 isolated from diverse marine environments including eukaryotic microbiota (40). They are
267 considered to be aerobe or facultative anaerobic bacteria that generally lack denitrification
268 capabilities (41). Finally, another abundant family, Rhodobacteracea, is worth mentioning, as
269 it is always found in low, but similar abundance in all samples. Marine Rhodobacteracea
270 species are considered ecological generalists (42-44) and have been reported to colonize
271 marine animals (e.g., fish larvae or sponges) (45, 46). The uniform distribution of

272 Rhodobacteracea in the foraminifera-microbiome suggests that members of this group are
273 permanent residents in the foraminiferal microbiota.
274 Our data thus shows that rotaliids are habitat to bacterial communities whose
275 composition is akin to the microbiota of other marine eukaryotes.

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

278 The metagenomes analysis and classification into bacterial metagenome-assembled
279 genomes (MAGs) resulted in a total of 263 high-quality MAGs (i.e., draft bacterial genomes)
280 of foraminifera-associated bacteria (Suppl. Table S1). The strongest signal for a stable core
281 microbiome in our data was observed in the comparison among the *Globobulimina* species,
282 whose microbiome is characterized by a high frequency of Desulfobacteraceae (Figure 4B).
283 To further explore the association between *Globobulimina* and Desulfobacteraceae we
284 examined the Desulfobacteraceae MAGs in our data. A total of 40 high quality MAGs were
285 obtained from the *G. pacifica* metagenome, including four high quality draft MAGs classified
286 as Desulfobacteracea. These were compared with the 26 previously published MAGs for
287 *Globobulimina* from Sweden, which include two Desulfobacteracea MAGs (11) (Suppl. Table
288 S7). A phylogenetic network of the Desulfobacteracea MAGs reveals two MAGs from the
289 *Globobulimina spp.* and *G. pacifica* metagenomes (Glo_11 & Gpa_30) that appear as sister
290 taxa (Figure 5A). The average nucleotide identity (ANI) between both MAGs was 84%, which
291 is within the range expected for inter-species sequence similarity (e.g., within genera) (47).
292 The common ancestry of these two MAGs suggests that the association between
293 Desulfobacteracea and *Globobulimina* has an ancient origin. Our results thus indicate that the
294 interaction fidelity between *Globobulimina* and Desulfobacteracea is high, similarly to
295 observations in other eukaryote-bacterial symbioses [e.g., oligochaete worms and sulfur
296 bacteria (48) or marine sponges with bacteria of the Poribacteria phylum (49)].
297 A central foundation in the evolution of symbiotic interaction is an exchange of
298 currencies (i.e., resources) between the partners (50). The versatile metabolic capabilities of
299 Desulfobacteraceae and the availability of intracellular NO_3^- storage in foraminifera may
300 suggest that the interaction between Desulfobacteracea and *Globobulimina* is based on
301 nutritional currencies. To further examine possible symbiotic interaction between
302 Desulfobacteracea and *Globobulimina*, we searched for metabolic properties of the
303 Desulfobacteraceae MAGs that could serve as nutritional currencies in the symbiosis. For that
304 purpose, we surveyed the metabolic pathways of representative Desulfobacteracea MAGs
305 with respect to carbon, sulfur and nitrogen metabolism (Figure 5B; Suppl. Table S8). Our
306 results reveal that all MAGs sampled encode the genetic repertoire required in order to
307 perform dissimilatory sulfate reduction as well as the Wood-Ljungdahl pathway. Anaerobic

308 respiration via the dissimilatory sulfate reduction to sulfide is widespread among
309 Desulfobacteraceae. Members of this family that encode the Wood-Ljungdahl pathway are
310 able to oxidize organic compounds to carbon dioxide (51). Alternatively, the Wood-Ljungdahl
311 pathway may function in CO_2 carbon fixation during autotrophic growth.

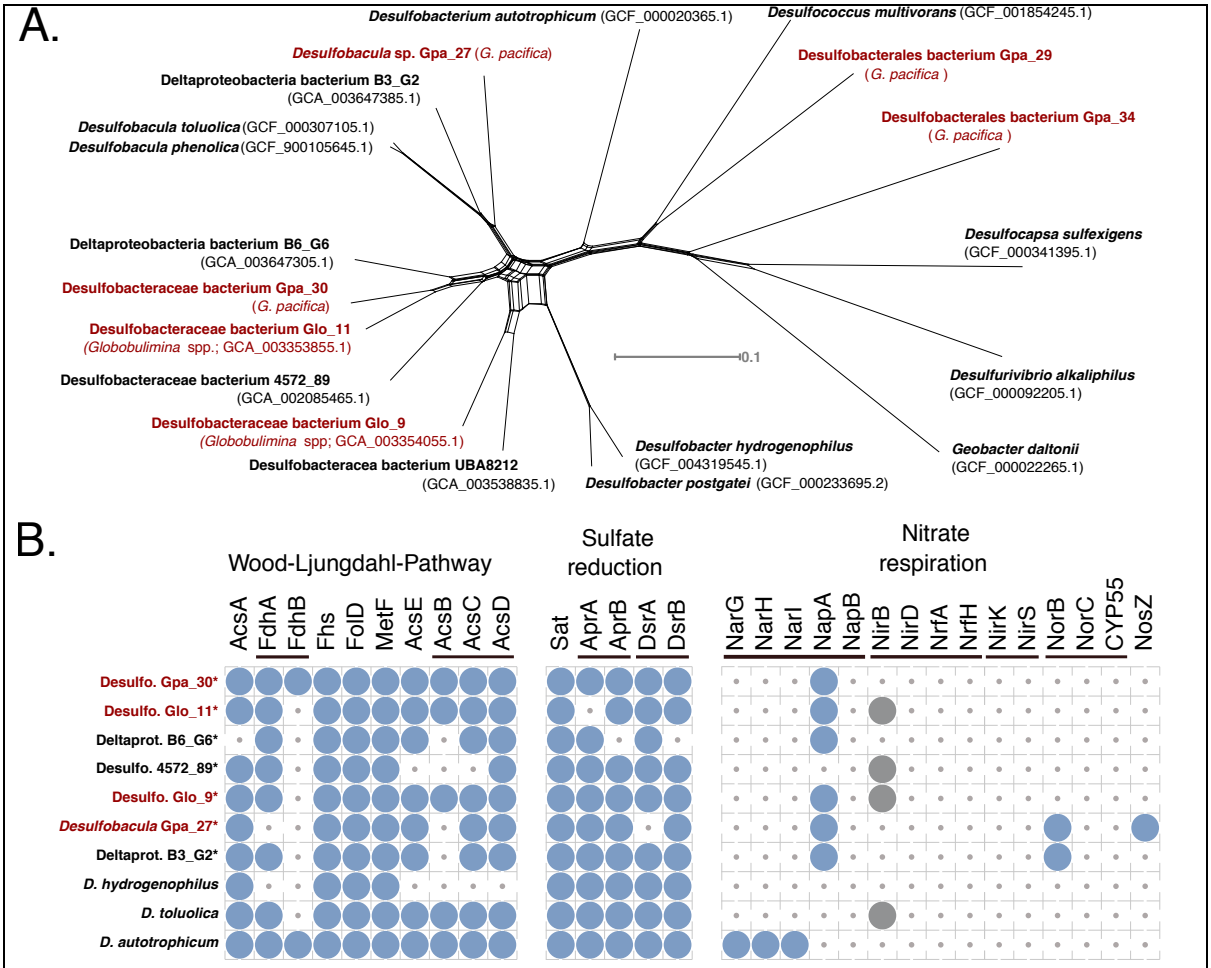


Figure 5. Characteristics of Desulfobacteraceae bacteria associated with Globobulimina. A) Neighbor-net network for Desulfobacteraceae associated with the genus *Globobulimina* from Sweden and Peru (*Globobulimina* spp. and *G. pacifica*, respectively) and additional Deltaproteobacteria. The network 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.

312

313 Genes along the denitrification pathway (or NO_3^- respiration) are mostly absent from
314 all Desulfobacteracea MAGs, except for periplasmic NO_3^- reductases (NapA) homologs that
315 were found in most foraminifera-associated MAGs (for a review in NapA function see ref. 52).
316 All MAGs lack the typical NapB protein forming a complex with NapA. However, *nap* operons

317 without NapB genes have been reported for other members of Deltaproteobacteria, e.g.
318 *Desulfovibrio desulfuricans* (53). Notably, we found that the genome of *Desulfobulbus*
319 *propionicus* (Accession: GCA_000186885.1), a member of the order Desulfobacterales that
320 has been demonstrated to grow based on NO_3^- (54), is lacking a gene for NapB. Hence NO_3^-
321 respiration in Desulfobacterales in the absence of NapB is possible. We found that the NapA
322 sequences of all MAGs (and *D. desulfuricans*) include a signal peptide of the twin-arginine
323 translocation pathway (reviewed in ref. 55) indicating that those proteins could be translocated
324 across cellular membranes. It is tenable to hypothesize that NapA is secreted by the bacteria
325 to *Globobulimina* intra-cellular environment. An alternative scenario is nitrate uptake by the
326 bacterium followed by the reduction to nitrite within the bacterial cellular compartments that
327 are surrounded by membranes. The nitrite is then released from the bacterial cell and can be
328 used by the foraminiferal host. The latter scenario may be more beneficial for the bacterium.
329 Overall, our results suggest that the *Globobulimina*-associated Desulfobacteraceae are able
330 to reduce NO_3^- to NO_2^-, and thus contribute to the foraminiferal denitrification by performing
331 the first reaction in the denitrification pathway for which we found no evidence in the
332 foraminiferal transcriptomes.

333 **Discussion**

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

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

352 MAGs, it is tenable to hypothesize that the reduction of NO_3^- to NO_2^- in foraminifera is
353 performed by resident bacteria.
354 An association between foraminifera and Desulfobacteracea has been previously
355 reported. For example, the presence of a putative Deltaproteobacterium (Desulfobacteracea)
356 was previously described for *Virgulinella fragilis* (20), a foraminiferal species that we here
357 predict to be a denitrifying species (Figure 3B). Furthermore, foraminifera from the genera
358 *Ammonia*, *Elphidium* and *Haynesina* that are not expected to denitrify were found to harbor
359 members of Desulfobacterales bacteria (25). Hence, the interaction between foraminifera and
360 Desulfobacteracea may involve alternative nutritional currencies. Previous studies refer to the
361 role of Desulfobacteracea in sulphur cycling and carbon/nutrient acquisition (20, 21, 25).
362 Especially the carbon fixation capabilities via Wood-Ljungdahl pathway would be highly
363 beneficial for the heterotrophic lifestyle of foraminifera, similarly to other symbioses that
364 involve fixed-carbon as a nutritional currency [e.g., as in deep-sea mussels (57) or sponges
365 (49)].
366 The radiation of early foraminiferal species has been estimated to occur between 690
367 and 1,150 million years ago (mya), while the first primitive rotaliids appeared in the late
368 Permian around 260 mya (58, 59). Most of the rotaliids superfamilies diverged in the mid to
369 late Triassic between approximately 240 and 200 mya (29, 60, 61) and the majority of extant
370 species diverged during the Miocene (23.8 - 5.5 mya) (62). Our phylogenetic analyses suggest
371 that clades I and III diversified from within clade II, while the LCA of clade I probably originated
372 from within clade III (Figures 2 & 3). Although most species reported to denitrify were reported
373 from the Peruvian OMZ and Sweden, representatives of corresponding genera are found all
374 over the world (Suppl. Table S5). Therefore, most extant rotaliids likely diversified from a
375 denitrifying LCA far back in time. The origin of foraminiferal denitrification within or after
376 diversification from members of clade II may coincide with the rapid increase of fossil records
377 at the onset of the late Cretaceous ~100 mya (60, 61). Our results therefore indicate that
378 eukaryotic denitrification by rotaliids emerged late in foraminiferal evolution, possibly during
379 the worldwide Cretaceous Ocean Anoxic Events (63, 64).
380 Considering the high conservation of nitrate transporters (Figure 1) and the
381 observation of nitrate storage in many divergent species (9), it is tenable to speculate that
382 foraminifera had the mechanisms for nitrate import and storage long before they evolved the
383 ability to denitrify. NO_3^- transporters are ancient eukaryotic enzymes that play a role in the
384 NO_3^- assimilation machinery and they are encoded in genomes of photoautotrophs such as
385 plants and saprotrophs like fungi (65, 66). In contrast to bacteria, foraminifera (as most
386 heterotrophs) are unable to perform NO_3^- assimilation on their own. Thus, NO_3^- accumulated
387 by the foraminiferal hosts could have fuelled bacterial NO_3^- metabolism of associated bacteria

388 in exchange for organic compounds. We propose that denitrification by symbiotic bacteria was
389 indeed the ancestral state of denitrification in foraminifera, similarly to gromiids (26), which
390 share a common evolutionary origin with foraminifera (67). The finding of bacterial-like
391 denitrification genes in rotaliids furthermore suggests that foraminifera may have acquired
392 those genes from bacteria (11). Thus, a rare gene acquisition from a foraminifera-associated,
393 denitrifying bacterium could have been at the origin of foraminiferal (eukaryotic) denitrification.
394 Notably, the rotaliids denitrification gene set is incomplete and it varies among the species
395 sampled here (i.e., *N. auris, B.plicata* or even *Rosalina* sp.). Species reported to release N_2O
396 instead of N_2 (9) further support our observation. Thus, the presence of a partial denitrification
397 pathway in rotaliids as well as in their resident Desulfobacteracea may suggest that the
398 acquisition of denitrification ability in foraminifera occurred in multiple stages and is not yet
399 complete.
400

401 ## Methods

402 ### Foraminifera sampling

403 Samples were collected off the Peruvian coast in the 2017 Austral winter (R/V Meteor M137)
404 as described by Glock et al. 2019 (10). Briefly, sediment samples were taken with a video-
405 guided sediment multiple corer (MUC) containing 6 liners along a depth transect at 12°S. The
406 top 1 to 3 cm of the sediment cores were sampled and immediately wet sieved with surface
407 water through staked sieves with mesh size of 2000µm to 63µm to retrieve benthic
408 foraminifera. The foraminifera were rinsed in sterile seawater obtained by filtering core-
409 overlying seawater with a sterile bottle top filtration system (Durapore filter, 0.2 µm) and a
410 vacuum pump within 40 minutes of MUC arrival on deck. Focal species were manually picked
411 using a stereomicroscope. These were characterised morphologically according to the
412 literature (68-71). Up to 160 individuals (classified into focal species) were pooled in one cryo
413 vial (2ml, RNAse free) and flash frozen in liquid nitrogen.

414 ### Microscopy

415 Individual foraminifers were dehydrated in a graded ethanol series (70 %, 80 %, 90 %, 96 %
416 and two times 100 %; 15 min each), air-dried at 20 °C for 12 h in a desiccator and mounted
417 on aluminium stubs (PLANO GmbH) using conductive and adhesive carbon pads (PLANO
418 GmbH). Subsequently, the preparations were sputter-coated with a 10-nm-thick gold-
419 palladium (80/20) layer using a high vacuum sputter coater Leica EM SCD500 (Leica
420 Microsystems GmbH) and visualised with a Hitachi S-4800 field emission scanning electron
421 microscope (Hitachi High-Technologies Corporation) at an acceleration voltage of 3 kV and
422 an emission current of 10 mA applying a combination of the upper detector and the lower

423 detector. The *B. costata* specimen was taken from previous samplings by the General
424 Direction of Research in Oceanography and Climate Change IMARPE (71). The
425 corresponding sample was air-dried at 35–38 °C, sputter-coated with gold and visualised with
426 an FEI Inspect S50 scanning electron microscope (Thermo Fisher Scientific Inc.) at an
427 acceleration voltage of 3 kV.

428 **Nucleic acid extraction, and sequencing**

429 Genomic DNA and total RNA from different biological samples were extracted using either the
430 DNeasy® Plant Mini kit (QIAGEN) for genomics (DNA) or simultaneously purified using
431 InnuPREP DNA/RNA Mini Kit (analytikjena) for transcriptomics (RNA). Foraminiferal cells in
432 each samples (ca. 160 individuals) were disrupted by pestle-crashing on ice after immersion
433 of the containing cryo-vial in liquid nitrogen. Samples for genomics were treated with lysozyme
434 (200µl of 10mg/ml TE) and proteinase K (1mg/100µl). Samples for transcriptomics samples
435 were treated only with lysozyme (6 ul of 20mg/ml TE) before additional crashing and 2-minute
436 incubation was performed. After disruption and initial lysis, manufacturer protocols for the
437 respective nucleic acid extraction were followed with modifications to the elution volumes (2 x
438 50µl for DNA and 2 x 25µl for RNA). Libraries for genomics were produced after DNA
439 fragmentation (Covaris target 400, intensity 5, duty cycle 5% cycles per burst 200, 55sec
440 treatment time) using NEBNext® Ultra II DNA Library Prep Kit for Illumina®. Whereas,
441 transcriptomic libraries were produced using NEBNext® Ultra RNA Library Prep Kit for
442 Illumina® (NEB) with mRNA isolation performed with poly-A mRNA beads. All libraries were
443 produced in duplicates from two different sets of pooled individuals for each species and were
444 produced without protocol interruption. Before sequencing, each library was quantified with a
445 Qubit® fluorometer (Invitrogen by Life TechnologiesTM) and qualified using a TapeStation
446 (Agilent technology). The libraries were sequenced paired-end (2 x 150 bp) on an Illumina®
447 HiSeq 4000 platform.
448

449 **Sequencing datasets of foraminifera**

450 Sequencing resulted in 4.6 billion paired-end reads covering 1.4 tera bases in total. This
451 includes transcriptome and metagenome datasets (BioProject accessions PRJNA494828 &
452 PRJNA503328). Reads were quality-checked by FastQC ver. 0.11.5
453 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc; Aug 2016). Filtering and trimming
454 of reads was performed using Trimmomatic (72) ver. 0.36 (Parameters:
455 ILLUMINACLIP:primers.fa:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:5
456 MINLEN:21; the file 'primers.fa' contained adaptor and contaminant sequences provided by
457 Trimmomatic and FastQC). Processed reads from transcriptomes of the two samples per
458 species were assembled into transcript contigs using SPAdes (73) ver. 3.11.1 ("--rna" option).

459 Protein sequences were translated from transcripts as open reading frames (ORFs) using
460 TransDecoder (74) ver. v5.0.2 (LongOrfs; "-m 30" option). Final protein names consist of the
461 contig ids followed by the sequence positions covered by the CDS and an indicator for the
462 forwards (+) or reverse (-) strand. Transcript abundance of individual transcriptome datasets
463 are referring to Transcripts Per Million (TPM) determined by the Trinity pipeline (74) 2.4.0
464 (Trinity script 'align_and_estimate_abundance.pl') via RSEM (75) ver. 1.2.30 and Bowtie (76)
465 2.1.0 using paired-end reads. Additional raw sequencing reads were obtained from the Marine
466 Microbial Transcriptome Project (MMETSP; SRA accessions: *Rosalina* sp., SRR1296887;
467 *Sorites* sp., SRR1296734; *Ammonia* sp., SRR1300434; *Elphidium margaritaceum*,
468 SRR1300475) and processed as described above. We found that the assembly obtained for
469 *Sorites* sp. contained a high proportion of sequences likely originating from an algal species
470 (Probably *Symbiodinium* sp.), which was removed by applying a binning approach considering
471 only contigs with GC-content ≤38%. Finally, we included data for *Globobulimina* spp. sampled
472 in Sweden from our previous study (Transcriptome data: GloT15) (11) and proteins annotated
473 on the genome assembly of *Reticulomyxa filosa* (NCBI accession: GCA_000512085.1).
474
475 **Species phylogenies and rooting**
476 Transcriptome completeness and heterogeneity was determined by assessing genome
477 completeness via Benchmarking Universal Single-Copy Orthologs (77) (BUSCO v3; lineage
478 'eukaryota') method. Orthologous proteins determined as Complete (or Duplicated) by the
479 BUSCO analysis were merged into protein clusters to study phylogenetic relationships among
480 the species. In case of duplicated BUSCOs in metatranscriptomes, one representative was
481 inferred based on sequence similarity. Therefore, all orthologs retrieved for the same
482 metatranscriptome were compared to all members of the BUSCO protein cluster by global
483 pairwise alignments using needle (EMBOSS tools version 6.6.0) (78). The ortholog with the
484 highest median sequence similarity over all comparisons was picked as the representative
485 sequence. Multiple sequence alignments used in the current study were obtained using
486 MAFFT (79) (Version: 7; parameter: 'linsi') and the phylogenetic trees were reconstructed
487 using IQTREE (80) (Version: 1.5.5 or 1.6.9; default parameters; note that ModelFinder is
488 enabled by default). Phylogenies were reconstructed for the individual BUSCO clusters. Due
489 to their morphological differences from other foraminiferal groups, monothalamids like *R. filosa*
490 have been previously used as an outgroup for phylogenetic studies of foraminifera (29). Since
491 the choice of a distantly related outgroup may lead to erroneous rooted topology due to long
492 branch attraction (81), we further tested the robustness of the root position independent from
493 outgroup species. Root positions were determined using MAD (28) (Parameters: '-bsnn'). The
494 support in each root splits is calculated as the proportion of gene trees where the root position
495 was inferred as such (28). For the overall species trees, multigene phylogenies were

496 reconstructed using IQTREE (parameter: '-spp') considering all protein cluster alignments of
497 either foraminifera or Rotaliida.
498 Reference 18S rRNA sequences used for the large-scale phylogenies (Suppl. Table
499 S5) were obtained from NCBI, MMETSP, SILVA (82), the foram barcoding project
500 (http://forambarcoding.unige.ch/) and the Planktonic Foraminifera Ribosomal Reference
501 database (83) (PFR[2]). These sequences were used as database sequences to identify 18S
502 sequences from Peruvian species transcriptomes based on BLAST (84) (version: 2.2.28+;
503 options: '-task blastn -evalue $1e^{-10}$') searches. Among the hits retrieved we searched for a
504 single representative transcript contig per transcriptome assembly. The transcripts with the
505 highest product of sequence length and sequencing depth (i.e., coverage given by SPAdes)
506 were determined as representative sequences for most of the focal species. However, for *B.*
507 *costata*, *V. inflata*, *G. pacifica* and *N. auris* different representative sequences were chosen
508 by giving sequence information (i.e., longer sequences) a higher weight than sequencing
509 depth. To obtain 18S phylogenies first the reference sequences were aligned with MAFFT and
510 subsequently representatives 18S rRNA sequences of the Peruvian transcriptomes were
511 added (MAFFT options: '--addfragments --adjustdirection') to the alignment followed by the
512 tree reconstruction using IQTREE.
513
514 **Identification of foraminiferal denitrification proteins**
515 To identify homologs to enzymes in the denitrification pathway, we used a similar approach
516 as previously established in Woehle et al. 2018 (11) using the corresponding protein database
517 expanded by the protein sequences identified for *Globobulimina* (11). The search for
518 denitrification enzymes homologs in the transcriptome assemblies was performed with
519 BLASTP (parameter: '-max_target_seqs 1000000 e-value 1e-5'). Protein sequences of hits
520 with query coverage ≥ 40% and sequence identity ≥ 20% were extracted to obtain a first set
521 of homologs. We further applied a cutoff to discard lowly represented transcripts with TPM <
522 2 in at least one of the two replicates sequenced. With the resulting protein set, we reiterated
523 searches in the non-redundant NCBI protein (NR; version May 2018) and the RefSeq 88
524 database. (85) using diamond v0.9.22 applying the '--more-sensitive' option. First best hit
525 sequences per query were obtained and clustered with CD-HIT 4.6 (86) (option: "-c 0.98") to
526 reduce sequence redundancy. All obtained protein sequences of a given enzyme were aligned
527 with MAFFT and phylogenetic trees were reconstructed with IQTREE (parameters: '-bb 1000
528 -alrt 1000'). The trees were rooted using an outgroup, if available, or MAD.
529
530 **Metagenomics processing**
531 For the visualisation of metagenomic composition, trimmed paired-end reads from Peruvian
532 species samples and the *Globobulimina* 'Ambient' samples (11) were subsampled via BBMap

533 (version 36.84; https://sourceforge.net/projects/bbmap/; 'reformat.sh' script; parameters:
534 'samplereadstarget=10000000 addslash=t'). The resulting reads were mapped against the NR
535 database with ac-diamond (87) and classified using MEGAN6 (88) (Version 6.15.0; options '-
536 a2t prot_acc2tax-Nov2018X1.abin -a2eggnog acc2eggnog-Oct2016X.abin -a2seed
537 acc2seed-May2015XX.abin'). MEGAN6 was used for assessment of metagenomics
538 communities and visual representation (See Figure 4) that were classified up to the order rank
539 level with different taxon sets (e.g., all nodes or only the bacterial subtree). Metagenome
540 assemblies were obtained by combining all trimmed reads per focal species using MEGAHIT
541 (89) (Version 1.1.3; default settings). Individual bacterial genomes were obtained using the
542 binning approach implemented with MaxBin2 (90) (Version 2.2.4; parameter: '-
543 min_contig_length 500') using coverage information of the two samples per species of
544 foraminifera. Binning statistics were assessed using CheckM (91) (Version 1.0.11); a
545 threshold for completeness of at least 80% and contamination of maximum 20% was applied
546 to classify bacterial bins as draft genomes. The protein sequences obtained via checkM were
547 used for further classifications. The average nucleotide identity (ANI) was calculated using a
548 perl script obtained from https://github.com/chjp.
549 For the taxonomic assignment of genome bins, we used the diamond tool to find first
550 best hits for each protein of a genome bin against the NR database (option: '-k 10'; *e*-value ≤
551 1e-10) and retrieved the corresponding taxonomic assignments (as we previously described
552 in ref. 11). For each bin, identical taxonomic hierarchies were counted with the genus being
553 the lowest rank considered, sorted accordingly and stepwise searched for the lowest
554 taxonomic rank supporting >50% of protein bin hits starting with the most abundant taxonomy.
555 'Environmental samples' and 'Cellular organisms' were not considered. For the phylogenetic
556 reconstruction of Desulfobacteracea we determined protein families as protein clusters using
557 sequences of Desulfobacteracea bins associated with *Globobulimina* and additional
558 Desulfobacteracea genomes downloaded from NCBI. First, we determined reciprocal best
559 BLAST hit pairs (rBBH; parameter: '-evalue 1e-5') between all the Desulfobacteracea protein-
560 coding sequences (92). Then rBBH pairs were globally aligned with the needle tool and pairs
561 with ≥30% identical amino acids were sorted into clusters using the Markov clustering
562 algorithm (93) (MCL; version 12-135). The 33 resulting protein clusters that contained a single
563 copy for each of the Desulfobacteraceae strains (i.e., universal single-copy clusters) were
564 aligned with MAFFT. The resulting alignments were concatenated and a splits network was
565 reconstructed with SplitsTree (94) (version 4.15.1). Prediction of protein function for bacterial
566 genomes were obtained using the KEGG Automatic Annotation Server (95) (KAAS; BBH
567 method via BLAST and the following species set: hsa, dme, ath, sce, pfa, eco, sty, hin, pae,
568 nme, hpy, rpr, mlo, bsu, sau, lla, spn, cac, mge, mtu, ctr, bbu, syn, aae, mja, afu, pho, ape,
569 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 (96).

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 transcriptome sequencing archive (TSA; See Suppl. Table 1 for accessions) and as whole genome shotgun (WGS) projects (See Suppl. Table 7 for accessions). All other information on accessing data analysed in this study is included in the manuscript or in the supplemental information.

Acknowledgments

We thank Devani Romero Picazo and Anne Kupczok for critical comments on the manuscript. We gratefully acknowledge the scientific party and crew of the R/V Meteor cruise M137 as well as Asmus Petersen and Matthias Türk for their support at sea. Samples from Peru were obtained according to Peruvian access and benefit sharing regulations. We thank the Nagoya officer of Kiel University - Dr. Scarlett Sett - for her support of our research. We further thank Natalia Bernabe Lopez for her assistance in collecting metadata for the 18S sequences. The micrograph of *B. costata* was possible due to the General Direction of Research in Oceanography and Climate Change (IMARPE). Genome sequencing was performed in the Centre for Genome Analysis Kiel that is funded by the German Research Foundation (DFG). The study was supported by the German Research Foundation (DFG) via the SFB 754 on Climate–Biogeochemistry Interactions in the Tropical Ocean, the cluster of excellence The Future Ocean, and the European Research Council (Grant No. 281357 awarded to TD). C.W. was supported by the Kiel Life Science (KLS) Young Scientist Programme and would like to thank the Max Planck-Genome-centre Cologne for their support in data analysis.

Author contributions

C.W., A.-S.R., N.G., T.W., J.W., D.R., J.S. & T.D. designed the research strategy and performed sampling. C.W. analysed the transcriptomes and metagenomes. A.-S.R. performed the experimental lab work. C.H. constructed incubation chambers. D.R. produced the scanning electron micrograph of *B. costata*. J.M. produced all other scanning electron micrographs, which was supported by S.N.G. C.W. and T.D. wrote the manuscript with input from all co-authors.

Declaration of Interests

The authors declare no competing interests.

References:

1. Cox AN (2015) *Allen's Astrophysical Quantities* ed Cox AN (Springer, New York, NY) doi:10.1007/978-1-4612-1186-0.

2. Thamdrup B (2012) New Pathways and Processes in the Global Nitrogen Cycle. *Annu Rev Ecol Evol Syst* 43(1):407–428.

3. Gruber N (2004) The Dynamics of the Marine Nitrogen Cycle and its Influence on Atmospheric CO_2 Variations. *The Ocean Carbon Cycle and Climate* (Springer, Dordrecht, Dordrecht), pp 97–148. 2nd Ed.

4. Gruber N, Galloway JN (2008) An Earth-system perspective of the global nitrogen cycle. *Nature* 451(7176):293–296.

5. Lam P, et al. (2009) Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA* 106(12):4752–4757.

6. Philippot L (2002) Denitrifying genes in bacterial and Archaeal genomes. *Biochim Biophys Acta* 1577(3):355–376.

7. Shoun H, Kim DH, Uchiyama H, Sugiyama J (1992) Denitrification by fungi. *FEMS Microbiology Letters* 73(3):277–281.

8. Risgaard-Petersen N, et al. (2006) Evidence for complete denitrification in a benthic foraminifer. *Nature* 443(7107):93–96.

9. Piña-Ochoa E, et al. (2010) Widespread occurrence of nitrate storage and denitrification among Foraminifera and Gromiida. *Proc Natl Acad Sci USA* 107(3):1148–1153.

10. Glock N, et al. (2019) Metabolic preference of nitrate over oxygen as an electron acceptor in foraminifera from the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA* 116(8):2860–2865.

11. Woehle C, et al. (2018) A Novel Eukaryotic Denitrification Pathway in Foraminifera. *Curr Biol* 28(16):2536–2543.e5.

12. Glock N, et al. (2013) The role of benthic foraminifera in the benthic nitrogen cycle of the Peruvian oxygen minimum zone. *Biogeosciences* 10(7):4767–4783.

13. Schweizer M, Pawlowski J, Kouwenhoven TJ, Guiard J, van der Zwaan B (2008) Molecular phylogeny of Rotaliida (Foraminifera) based on complete small subunit rDNA sequences. *Marine Micropaleontology* 66(3-4):233–246.

14. Murray JW (2006) *Ecology and applications of benthic foraminifera* (Cambridge University Press).

15. LeKieffre C, et al. (2017) Surviving anoxia in marine sediments: The metabolic response of ubiquitous benthic foraminifera (Ammonia tepida). *PLoS One* 12(5):e0177604.

16. Bernhard JM, et al. (2012) Potential importance of physiologically diverse benthic foraminifera in sedimentary nitrate storage and respiration. *J Geophys Res* 117(G3):G03002.

17. Gomaa F, et al. (2021) Multiple integrated metabolic strategies allow foraminiferan protists to thrive in anoxic marine sediments. *Sci Adv* 7(22). doi:10.1126/sciadv.abf1586.

18. Orsi WD, et al. (2020) Anaerobic metabolism of Foraminifera thriving below the seafloor. *ISME J* 14(10):2580–2594.

19. Zilber Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews* 32(5):723–735.

20. Tsuchiya M, et al. (2015) Cytologic and genetic characteristics of endobiotic bacteria and kleptoplasts of *Virgulinella fragilis* (Foraminifera). *J Eukaryotic Microbiology* 62(4):454–469.

21. Nomaki H, et al. (2014) Nitrate uptake by foraminifera and use in conjunction with endobionts under anoxic conditions. *Limnology and Oceanography* 59(6):1879–1888.

22. Bernhard JM, Martin JB, Rathburn AE (2010) Combined carbonate carbon isotopic and cellular ultrastructural studies of individual benthic foraminifera: 2. Toward an understanding of apparent disequilibrium in hydrocarbon seeps. *Paleoceanography* 25(4):PA4206.

23. Bernhard JM, Habura A, Bowser SS (2006) An endobiont‐bearing allogromiid from the Santa Barbara Basin: Implications for the early diversification of foraminifera. *J Geophys Res* 111(G3):399.

24. Bernhard JM (2003) Potential symbionts in bathyal foraminifera. *Science* 299(5608):861–861.

25. Salonen IS, Chronopoulou PM, Bird C, Reichart GJ, Koho KA (2019) Enrichment of intracellular sulphur cycle –associated bacteria in intertidal benthic foraminifera revealed by 16S and aprA gene analysis. *Sci Rep*:1–12.

26. Høgslund S, Cedhagen T, Bowser SS, Risgaard-Petersen N (2017) Sinks and Sources of Intracellular Nitrate in Gromiids. *Front Microbiol* 8:617.

27. Bernhard JM, Edgcomb VP, Casciotti KL, McIlvin MR, Beaudoin DJ (2012) Denitrification likely catalyzed by endobionts in an allogromiid foraminifer. *ISME J* 6(5):951–960.

28. Tria FDK, Landan G, Dagan T (2017) Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol* 1:0193.

29. Pawlowski J, et al. (2003) The evolution of early Foraminifera. *Proc Natl Acad Sci USA* 100(20):11494–11498.

30. Ertan KT, Hemleben V, Hemleben C (2004) Molecular evolution of some selected benthic foraminifera as inferred from sequences of the small subunit ribosomal DNA. *Marine Micropaleontology* 53(3-4):367–388.

31. Mackensen A, Schmiedl G, Harloff J, Giese M (1995) Deep-Sea Foraminifera in the South Atlantic Ocean: Ecology and Assemblage Generation. *Micropaleontology* 41(4):342.

32. Rathburn AE, Willingham J, Ziebis W, Burkett AM, Corliss BH (2018) A New biological proxy for deep-sea paleo-oxygen: Pores of epifaunal benthic foraminifera. *Sci Rep* 8(1):9456–8.

33. Hoogakker BAA, et al. (2018) Glacial expansion of oxygen-depleted seawater in the eastern tropical Pacific. *Nature* 562(7727):410–413.

34. Widdel F (1987) New types of acetate-oxidizing, sulfate-reducing Desulfobacter species, *D. hydrogenophilus* sp. nov., *D. latus* sp. nov., and *D. curvatus* sp. nov. *Arch Microbiol* 148(4):286–291.

35. Lage OM, Bondoso J (2014) Planctomycetes and macroalgae, a striking association. *Front Microbiol* 5(5):267.

36. Oshiki M, Satoh H, Okabe S (2016) Ecology and physiology of anaerobic ammonium oxidizing bacteria. *Environ Microbiol* 18(9):2784–2796.

37. Noguchi T, et al. (1987) *Vibrio alginolyticus*, a tetrodotoxin-producing bacterium, in the intestines of the fish *Fugu vermicularis vermicularis*. *Marine Biology* 94(4):625–630.

38. McFall-Ngai M (2014) Divining the essence of symbiosis: insights from the squid-vibrio model. *PLoS Biol* 12(2):e1001783.

39. Ben-Haim Y, Zicherman-Keren M, Rosenberg E (2003) Temperature-regulated bleaching and lysis of the coral *Pocillopora damicornis* by the novel pathogen *Vibrio coralliilyticus*. *Appl Environ Microbiol* 69(7):4236–4242.

40. López-Pérez M, Rodriguez-Valera F (2014) The Family Alteromonadaceae. *The Prokaryotes: Gammaproteobacteria*, The Prokaryotes: Gammaproteobacteria. eds Rosenberg E, Delong EF, Lory S, Stackebrandt E, Thompson F (Springer Berlin Heidelberg, Berlin, Heidelberg), pp 69–92.

41. Ivanova EP, Flavier S, Christen R (2004) Phylogenetic relationships among marine Alteromonas-like proteobacteria: emended description of the family Alteromonadaceae and proposal of Pseudoalteromonadaceae fam. nov., Colwelliaceae fam. nov., Shewanellaceae fam. nov., Moritellaceae fam. nov., Ferrimonadaceae fam. nov., Idiomarinaceae fam. nov. and Psychromonadaceae fam. nov. *Int J Syst Evol Microbiol* 54(Pt 5):1773–1788.

42. Simon M, et al. (2017) Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J* 11(6):1483–1499.

43. Newton RJ, et al. (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME J* 4(6):784–798.

44. Buchan A, Gonzalez JM, Moran MA (2005) Overview of the marine roseobacter lineage. *Appl Environ Microbiol* 71(10):5665–5677.

45. Hjelm M, et al. (2004) Selection and identification of autochthonous potential probiotic bacteria from turbot larvae (*Scophthalmus maximus*) rearing units. *Syst Appl Microbiol* 27(3):360–371.

46. Webster NS, Negri AP, Munro MMHG, Battershill CN (2004) Diverse microbial communities inhabit Antarctic sponges. *Environ Microbiol* 6(3):288–300.

47. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9(1):5114–8.

48. Dubilier N, et al. (2001) Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* 411(6835):298–302.

49. Kamke J, et al. (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* 7(12):2287–2300.

50. Wein T, et al. (2019) Currency, Exchange, and Inheritance in the Evolution of Symbiosis. *Trends Microbiol* 27(10):836–849.

51. Strittmatter AW, et al. (2009) Genome sequence of *Desulfobacterium autotrophicum* HRM2, a marine sulfate reducer oxidizing organic carbon completely to carbon dioxide. *Environ Microbiol* 11(5):1038–1055.

52. González PJ, Correia C, Moura I, Brondino CD, Moura JJG (2006) Bacterial nitrate reductases: Molecular and biological aspects of nitrate reduction. *J Inorg Biochem* 100(5-6):1015–1023.

53. Dias JM, et al. (1999) Crystal structure of the first dissimilatory nitrate reductase at 1.9 A solved by MAD methods. *Structure* 7(1):65–79.

54. Widdel F, Pfennig N (1982) 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(4):360–365.

55. Berks BC, Ferguson SJ, Moir JW, Richardson DJ (1995) Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. *Biochim Biophys Acta* 1232(3):97–173.

56. Graf JS, et al. (2021) Anaerobic endosymbiont generates energy for ciliate host by denitrification. *Nature* 591(7850):445–450.

57. Duperron S, et al. (2006) 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(8):1441–1447.

58. Maklay M (1964) Novye Rannekamennougolnye Arkhedistsidy. *Min Geol Okhrana Nedr USSR* 1:158–159.

59. Crespin I (1958) Permian Foraminifera of Australia. *Bur Miner Resour Aust Bull* 48:1–207.

60. Tappan H, Loeblich AR (1988) Foraminiferal evolution, diversification, and extinction. *Journal of Paleontology* 62(5):695–714.

61. Loeblich AR, Tappan H (1964) *Treatise on Invertebrate Paleontology. Part C. Protista 2. Sarcodina chiefly "thecamoebians" and Foraminiferida.* (Geological Society of America New York/University of Kansas Lawrence, New York/Lawrence).

62. Kucera M, Schönfeld J (2007) The origin of modern oceanic foraminiferal faunas and Neogene climate change. *The Micropalaeontology Society Special Publications*:409–425.

63. Schlanger SO, Jenkyns HC (1976) Cretaceous oceanic anoxic events: causes and consequences. *Geologie en Mijnbouw* 55(3-4):179–184.

64. Jenkyns HC (1980) Cretaceous anoxic events: from continents to oceans. *Journal of the Geological Society* 137(2):171–188.

65. Forde BG (2000) Nitrate transporters in plants: structure, function and regulation. *Biochim Biophys Acta* 1465(1-2):219–235.

66. Pao SS, Paulsen IT, Saier MH (1998) Major facilitator superfamily. *Microbiol Mol Biol Rev* 62(1):1–34.

67. Sierra R, et al. (2013) Deep relationships of Rhizaria revealed by phylogenomics: A farewell to Haeckel's Radiolaria. *Mol Phylogenet Evol* 67(1):53–59.

68. Figueroa S, Marchant M, Giglio S, Ramírez M (2005) Foraminíferos bentónicos rotalínidos del centro sur de Chile (36°S - 44°S). *Gayana* 69(2):329–363.

69. Erdem Z (2016) 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*.

70. Mallon J (2012) Benthic Foraminifera of the Peruvian and Ecuadorian Continental Margin. *PhD thesis, Kiel University*.

71. Cardich J, et al. (2015) Calcareous benthic foraminifera from the upper central Peruvian margin: control of the assemblage by pore water redox and sedimentary organic matter. *Mar Ecol Prog Ser* 535:63–87.

72. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

73. Nurk S, et al. (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 20(10):714–737.

74. Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494–1512.

75. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

76. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.

77. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.

78. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16(6):276–277.

79. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.

80. Nguyen L-T, Schmidt HA, Haeseler von A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1):268–274.

81. Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27:401–410.

82. Yilmaz P, et al. (2014) The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res* 42(Database issue):D643–8.

83. Morard R, et al. (2015) 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(6):1472–1485.

84. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.

85. Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* 40(Database issue):D130–D135.

86. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.

87. Mai H, et al. (2018) AC-DIAMOND v1: accelerating large-scale DNA-protein alignment. *Bioinformatics* 34(21):3744–3746.

88. Huson DH, et al. (2016) MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol* 12(6):e1004957.

89. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10):1674–1676.

90. Wu Y-W, Simmons BA, Singer SW (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32(4):605–607.

91. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25(7):1043–1055.

92. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278(5338):631–637.

93. Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30(7):1575–1584.

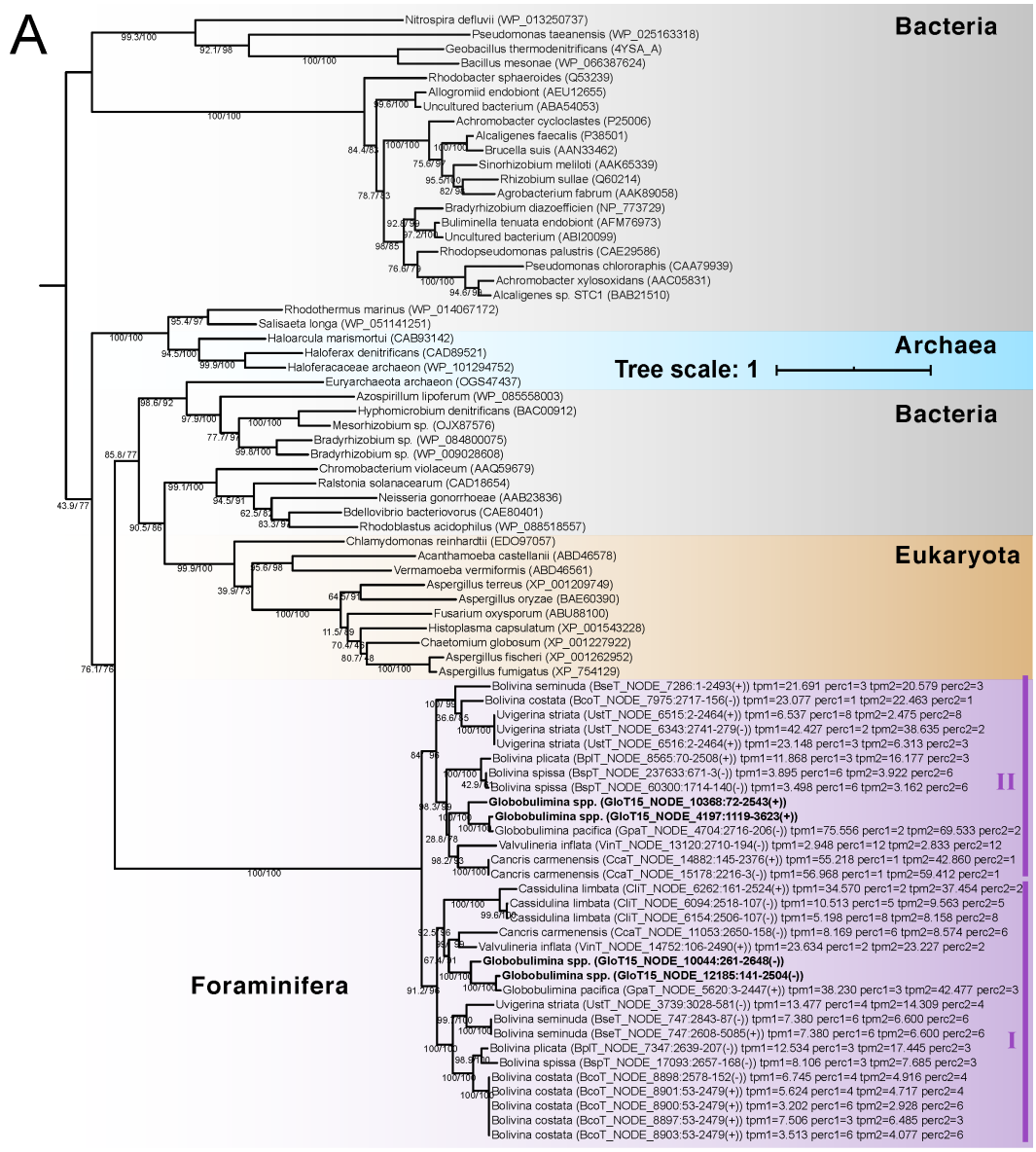
94. Huson DH (2005) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23(2):254–267.

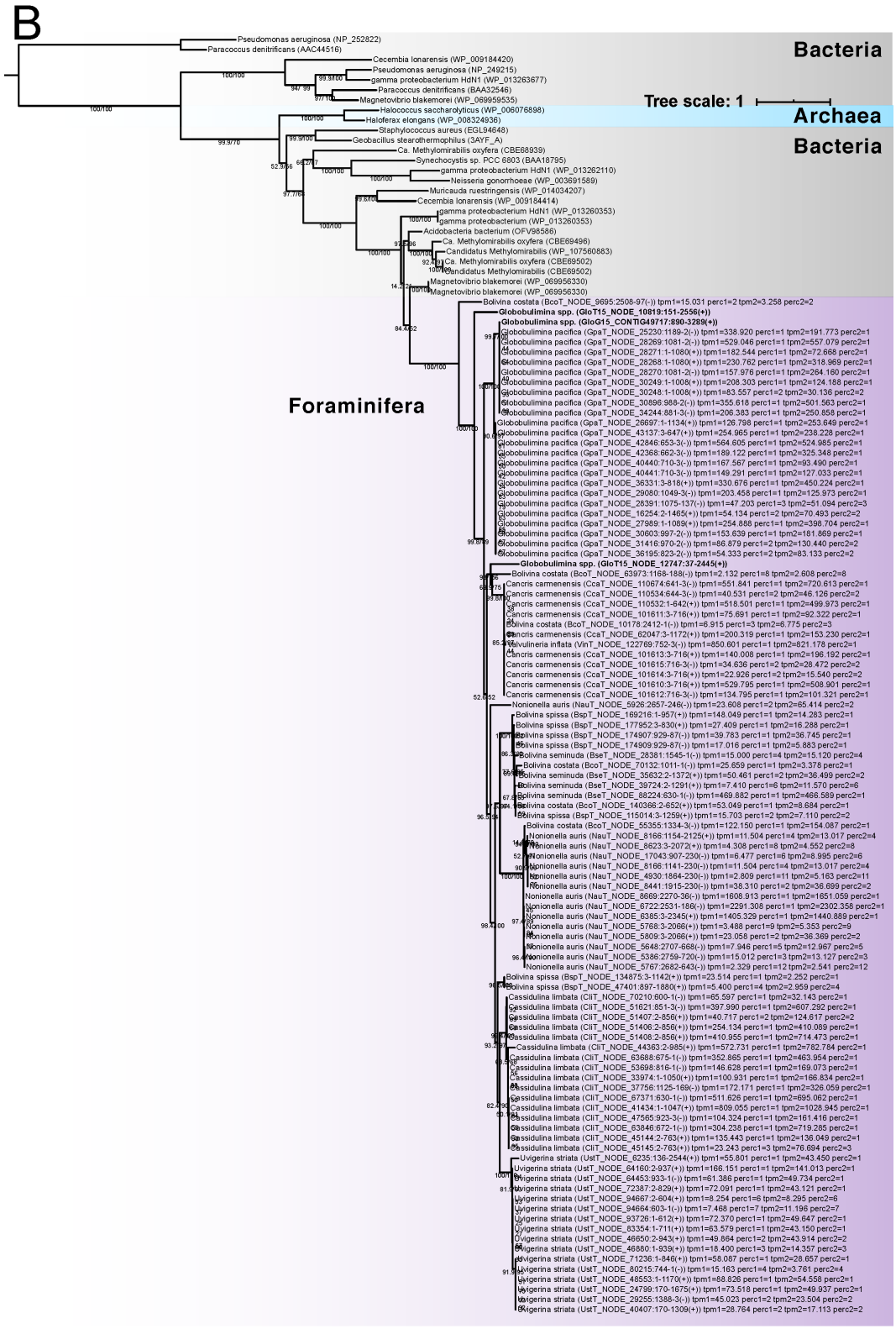
95. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35(Web Server issue):W182–5.

96. Nielsen H (2017) Predicting Secretory Proteins with SignalP. *Methods Mol Biol* 1611(Chapter 6):59–73.

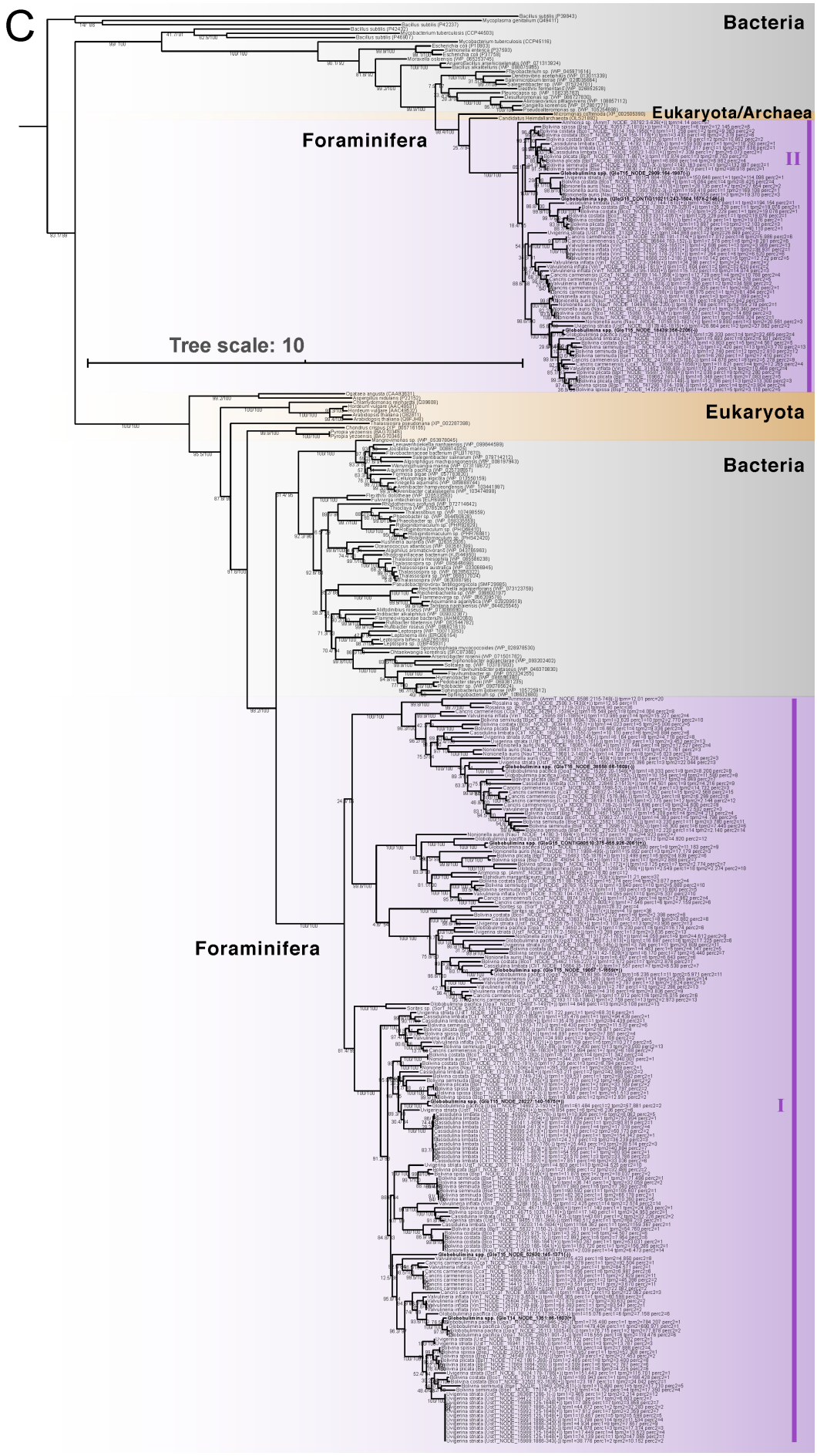
Supplementary material:

Suppl. Figure S1. Scanning electron micrographs of sampled species. Detailed views of the graphs shown in Figure 1A.

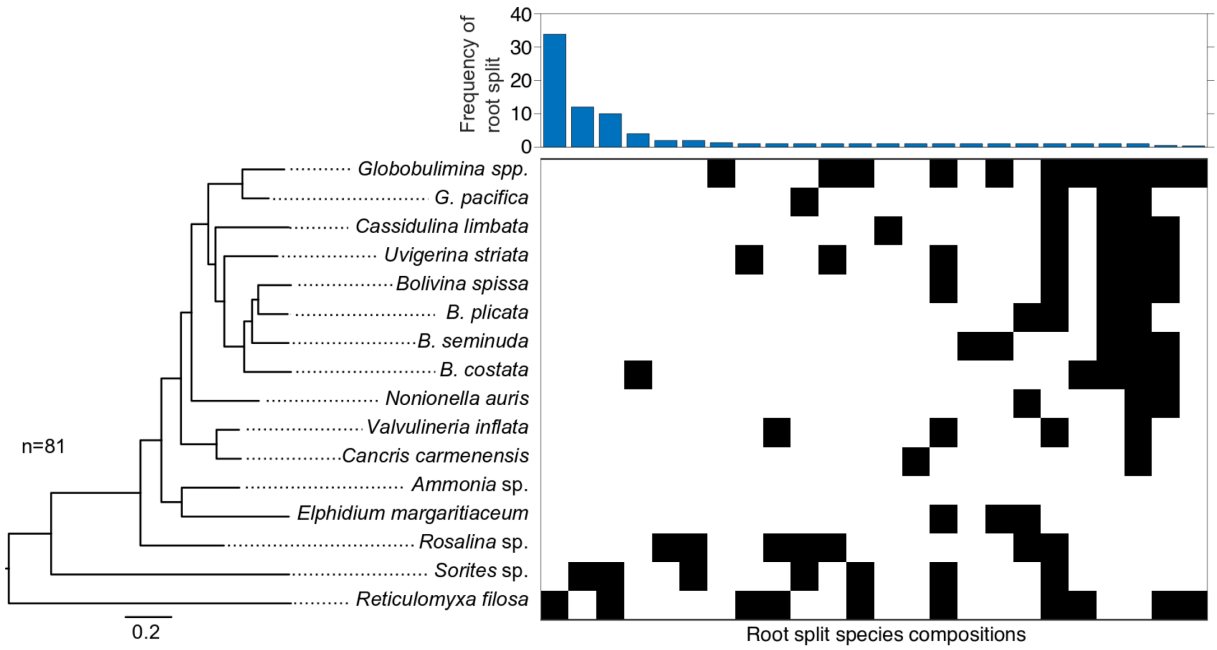



B
Bacteria
Archaea
Bacteria
Foraminifera
Tree scale: 1
Pseudomonas aeruginosa (NP_252822)
Paracoccus denitrificans (AAC44516)
Cecembia lonarensis (WP_009184420)
Pseudomonas aeruginosa (NP_249215)
gamma proteobacterium HdN1 (WP_013263677)
Paracoccus denitrificans (BAA32546)
Magnetovibrio blakemorei (WP_069959535)
Halococcus saccharolyticus (WP_006076898)
Haloferax elongans (WP_008324936)
Staphylococcus aureus (EGL94648)
Geobacillus stearothermophilus (3AYF_A)
Ca. Methylomirabilis oxyfera (CBE68939)
Synechocystis sp. PCC 6803 (BAA18795)
gamma proteobacterium HdN1 (WP_013262110)
Neisseria gonorrhoeae (WP_003691589)
Muricauda ruestringensis (WP_014034207)
Cecembia lonarensis (WP_009184414)
gamma proteobacterium HdN1 (WP_013260353)
gamma proteobacterium (WP_013260353)
Acidobacteria bacterium (OFV98586)
Ca. Methylomirabilis oxyfera (CBE69496)
Candidatus Methylomirabilis (WP_107560883)
Ca. Methylomirabilis oxyfera (CBE69502)
Candidatus Methylomirabilis (CBE69502)
Magnetovibrio blakemorei (WP_069956330)
Magnetovibrio blakemorei (WP_069956330)
Bolivina costata (BcoT_NODE_9695:2508-97(-)) tpm1=15.031 perc1=2 tpm2=3.258 perc2=2
Globobulimina spp. (GloT15_NODE_10819:151-2556(+))
Globobulimina spp. (GloG15_CONTIG49717:890-3289(+))
Globobulimina pacifica (GpaT_NODE_25230:1189-2(-)) tpm1=338.920 perc1=1 tpm2=191.773 perc2=1
Globobulimina pacifica (GpaT_NODE_28269:1081-2(-)) tpm1=529.046 perc1=1 tpm2=557.079 perc2=1
Globobulimina pacifica (GpaT_NODE_28271:1-1080(+)) tpm1=182.544 perc1=1 tpm2=72.668 perc2=1
Globobulimina pacifica (GpaT_NODE_28268:1-1080(+)) tpm1=230.762 perc1=1 tpm2=318.969 perc2=1
Globobulimina pacifica (GpaT_NODE_28270:1081-2(-)) tpm1=157.976 perc1=1 tpm2=264.160 perc2=1
Globobulimina pacifica (GpaT_NODE_30249:1-1008(+)) tpm1=208.303 perc1=1 tpm2=124.188 perc2=1
Globobulimina pacifica (GpaT_NODE_30248:1-1008(+)) tpm1=83.557 perc1=2 tpm2=30.136 perc2=2
Globobulimina pacifica (GpaT_NODE_30896:988-2(-)) tpm1=355.618 perc1=1 tpm2=501.563 perc2=1
Globobulimina pacifica (GpaT_NODE_34244:881-3(-)) tpm1=206.383 perc1=1 tpm2=250.858 perc2=1
Globobulimina pacifica (GpaT_NODE_26697:1-1134(+)) tpm1=126.798 perc1=1 tpm2=253.649 perc2=1
Globobulimina pacifica (GpaT_NODE_43137:3-647(+)) tpm1=254.965 perc1=1 tpm2=238.228 perc2=1
Globobulimina pacifica (GpaT_NODE_42846:653-3(-)) tpm1=564.605 perc1=1 tpm2=524.985 perc2=1
Globobulimina pacifica (GpaT_NODE_42368:662-3(-)) tpm1=189.122 perc1=1 tpm2=325.348 perc2=1
Globobulimina pacifica (GpaT_NODE_40440:710-3(-)) tpm1=167.567 perc1=1 tpm2=93.490 perc2=1
Globobulimina pacifica (GpaT_NODE_40441:710-3(-)) tpm1=149.291 perc1=1 tpm2=127.033 perc2=1
Globobulimina pacifica (GpaT_NODE_36331:3-818(+)) tpm1=330.676 perc1=1 tpm2=450.224 perc2=1
Globobulimina pacifica (GpaT_NODE_29080:1049-3(-)) tpm1=203.458 perc1=1 tpm2=125.973 perc2=1
Globobulimina pacifica (GpaT_NODE_28391:1075-137(-)) tpm1=47.203 perc1=3 tpm2=51.094 perc2=3
Globobulimina pacifica (GpaT_NODE_16254:2-1465(+)) tpm1=54.134 perc1=2 tpm2=70.493 perc2=2
Globobulimina pacifica (GpaT_NODE_27989:1-1089(+)) tpm1=254.888 perc1=1 tpm2=398.704 perc2=1
Globobulimina pacifica (GpaT_NODE_30603:997-2(-)) tpm1=153.639 perc1=1 tpm2=181.869 perc2=1
Globobulimina pacifica (GpaT_NODE_31416:970-2(-)) tpm1=86.879 perc1=2 tpm2=130.440 perc2=2
Globobulimina pacifica (GpaT_NODE_36195:823-2(-)) tpm1=54.333 perc1=2 tpm2=83.133 perc2=2
Globobulimina spp. (GloT15_NODE_12747:37-2445(+))
Bolivina costata (BcoT_NODE_63973:1168-188(-)) tpm1=2.132 perc1=8 tpm2=2.608 perc2=8
Cancris carmenensis (CcaT_NODE_110674:641-3(-)) tpm1=551.841 perc1=1 tpm2=720.613 perc2=1
Cancris carmenensis (CcaT_NODE_110534:644-3(-)) tpm1=40.531 perc1=2 tpm2=46.126 perc2=2
Cancris carmenensis (CcaT_NODE_110532:1-642(+)) tpm1=518.501 perc1=1 tpm2=499.973 perc2=1
Cancris carmenensis (CcaT_NODE_101611:3-716(+)) tpm1=75.691 perc1=1 tpm2=92.322 perc2=1
Bolivina costata (BcoT_NODE_10178:2412-1(-)) tpm1=6.915 perc1=3 tpm2=6.775 perc2=3
Cancris carmenensis (CcaT_NODE_62047:3-1172(+)) tpm1=200.319 perc1=1 tpm2=153.230 perc2=1
Valvulineria inflata (VinT_NODE_122769:752-3(-)) tpm1=850.601 perc1=1 tpm2=821.178 perc2=1
Cancris carmenensis (CcaT_NODE_101613:3-716(+)) tpm1=140.008 perc1=1 tpm2=196.192 perc2=1
Cancris carmenensis (CcaT_NODE_101615:716-3(-)) tpm1=34.636 perc1=2 tpm2=28.472 perc2=2
Cancris carmenensis (CcaT_NODE_101614:3-716(+)) tpm1=22.926 perc1=2 tpm2=15.540 perc2=2
Cancris carmenensis (CcaT_NODE_101610:3-716(+)) tpm1=529.795 perc1=1 tpm2=508.901 perc2=1
Cancris carmenensis (CcaT_NODE_101612:716-3(-)) tpm1=134.795 perc1=1 tpm2=101.321 perc2=1
Nonionella auris (NauT_NODE_5926:2657-246(-)) tpm1=23.608 perc1=2 tpm2=85.414 perc2=2
Bolivina spissa (BspT_NODE_169216:1-957(+)) tpm1=148.049 perc1=1 tpm2=14.283 perc2=1
Bolivina spissa (BspT_NODE_177952:3-830(+)) tpm1=27.409 perc1=1 tpm2=16.288 perc2=1
Bolivina spissa (BspT_NODE_174907:929-87(-)) tpm1=39.783 perc1=1 tpm2=36.745 perc2=1
Bolivina spissa (BspT_NODE_174909:929-87(-)) tpm1=17.016 perc1=1 tpm2=5.883 perc2=1
Bolivina seminuda (BseT_NODE_28381:1545-1(-)) tpm1=15.000 perc1=4 tpm2=15.120 perc2=4
Bolivina costata (BcoT_NODE_70132:1011-1(-)) tpm1=25.659 perc1=1 tpm2=3.378 perc2=1
Bolivina seminuda (BseT_NODE_35632:2-1372(+)) tpm1=50.461 perc1=2 tpm2=36.499 perc2=2
Bolivina seminuda (BseT_NODE_39724:2-1291(+)) tpm1=7.410 perc1=6 tpm2=11.570 perc2=6
Bolivina seminuda (BseT_NODE_88224:630-1(-)) tpm1=469.882 perc1=1 tpm2=466.569 perc2=1
Bolivina costata (BcoT_NODE_140366:2-652(+)) tpm1=53.049 perc1=1 tpm2=8.684 perc2=1
Bolivina spissa (BspT_NODE_115014:3-1259(+)) tpm1=15.703 perc1=2 tpm2=7.110 perc2=2
Bolivina costata (BcoT_NODE_55355:1334-3(-)) tpm1=122.150 perc1=1 tpm2=154.087 perc2=1
Nonionella auris (NauT_NODE_8166:1154-2125(+)) tpm1=11.504 perc1=4 tpm2=13.017 perc2=4
Nonionella auris (NauT_NODE_8623:3-2072(+)) tpm1=4.308 perc1=8 tpm2=4.552 perc2=8
Nonionella auris (NauT_NODE_17043:907-230(-)) tpm1=6.477 perc1=6 tpm2=8.995 perc2=6
Nonionella auris (NauT_NODE_8166:1141-230(-)) tpm1=11.504 perc1=4 tpm2=13.017 perc2=4
Nonionella auris (NauT_NODE_4930:1864-230(-)) tpm1=2.809 perc1=11 tpm2=5.163 perc2=11
Nonionella auris (NauT_NODE_8441:1915-230(-)) tpm1=38.310 perc1=2 tpm2=36.699 perc2=2
Nonionella auris (NauT_NODE_8669:2270-36(-)) tpm1=1608.913 perc1=1 tpm2=1651.059 perc2=1
Nonionella auris (NauT_NODE_6722:2531-186(-)) tpm1=2291.308 perc1=1 tpm2=2302.358 perc2=1
Nonionella auris (NauT_NODE_6385:3-2345(+)) tpm1=1405.329 perc1=1 tpm2=1440.889 perc2=1
Nonionella auris (NauT_NODE_5768:3-2066(+)) tpm1=3.488 perc1=9 tpm2=5.353 perc2=9
Nonionella auris (NauT_NODE_5809:3-2066(+)) tpm1=23.058 perc1=2 tpm2=36.369 perc2=2
Nonionella auris (NauT_NODE_5648:2707-668(-)) tpm1=7.946 perc1=5 tpm2=12.967 perc2=5
Nonionella auris (NauT_NODE_5386:2759-720(-)) tpm1=15.012 perc1=3 tpm2=13.127 perc2=3
Nonionella auris (NauT_NODE_5767:2682-643(-)) tpm1=2.329 perc1=12 tpm2=2.541 perc2=12
Bolivina spissa (BspT_NODE_134875:3-1142(+)) tpm1=23.514 perc1=1 tpm2=2.252 perc2=1
Bolivina spissa (BspT_NODE_47401:897-1880(+)) tpm1=5.400 perc1=4 tpm2=2.959 perc2=4
Cassidulina limbata (CliT_NODE_70210:600-1(-)) tpm1=65.597 perc1=1 tpm2=32.143 perc2=1
Cassidulina limbata (CliT_NODE_51621:851-3(-)) tpm1=397.990 perc1=1 tpm2=607.292 perc2=1
Cassidulina limbata (CliT_NODE_51407:2-856(+)) tpm1=40.717 perc1=2 tpm2=124.617 perc2=2
Cassidulina limbata (CliT_NODE_51406:2-856(+)) tpm1=254.134 perc1=1 tpm2=410.089 perc2=1
Cassidulina limbata (CliT_NODE_51408:2-856(+)) tpm1=410.955 perc1=1 tpm2=714.473 perc2=1
Cassidulina limbata (CliT_NODE_44363:2-985(+)) tpm1=572.731 perc1=1 tpm2=782.784 perc2=1
Cassidulina limbata (CliT_NODE_63688:675-1(-)) tpm1=352.865 perc1=1 tpm2=463.954 perc2=1
Cassidulina limbata (CliT_NODE_53698:816-1(-)) tpm1=146.626 perc1=1 tpm2=169.073 perc2=1
Cassidulina limbata (CliT_NODE_33974:1-1050(+)) tpm1=100.931 perc1=1 tpm2=166.834 perc2=1
Cassidulina limbata (CliT_NODE_37756:1125-169(-)) tpm1=172.171 perc1=1 tpm2=326.059 perc2=1
Cassidulina limbata (CliT_NODE_67371:630-1(-)) tpm1=511.626 perc1=1 tpm2=695.062 perc2=1
Cassidulina limbata (CliT_NODE_41434:1-1047(+)) tpm1=809.055 perc1=1 tpm2=1028.945 perc2=1
Cassidulina limbata (CliT_NODE_47565:923-3(-)) tpm1=104.324 perc1=1 tpm2=181.416 perc2=1
Cassidulina limbata (CliT_NODE_63846:672-1(-)) tpm1=304.238 perc1=1 tpm2=719.285 perc2=1
Cassidulina limbata (CliT_NODE_45144:2-763(+)) tpm1=135.443 perc1=1 tpm2=136.049 perc2=1
Cassidulina limbata (NauT_NODE_45145:2-763(+)) tpm1=23.243 perc1=3 tpm2=76.694 perc2=3
Uvigerina striata (UstT_NODE_6235:136-2544(+)) tpm1=55.801 perc1=1 tpm2=43.450 perc2=1
Uvigerina striata (UstT_NODE_64160:2-937(+)) tpm1=166.151 perc1=1 tpm2=141.013 perc2=1
Uvigerina striata (UstT_NODE_64453:933-1(-)) tpm1=61.386 perc1=1 tpm2=49.734 perc2=1
Uvigerina striata (UstT_NODE_72387:2-829(+)) tpm1=72.091 perc1=1 tpm2=43.121 perc2=1
Uvigerina striata (UstT_NODE_94667:2-604(+)) tpm1=8.254 perc1=6 tpm2=8.295 perc2=6
Uvigerina striata (UstT_NODE_94664:603-1(-)) tpm1=7.468 perc1=7 tpm2=11.196 perc2=7
Uvigerina striata (UstT_NODE_93726:1-612(+)) tpm1=72.370 perc1=1 tpm2=49.647 perc2=1
Uvigerina striata (UstT_NODE_83354:1-711(+)) tpm1=63.579 perc1=1 tpm2=43.150 perc2=1
Uvigerina striata (UstT_NODE_46650:2-943(+)) tpm1=49.864 perc1=2 tpm2=43.914 perc2=2
Uvigerina striata (UstT_NODE_46880:1-939(+)) tpm1=18.400 perc1=3 tpm2=14.357 perc2=3
Uvigerina striata (UstT_NODE_71236:1-846(+)) tpm1=58.087 perc1=1 tpm2=28.657 perc2=1
Uvigerina striata (UstT_NODE_80215:744-1(-)) tpm1=15.163 perc1=4 tpm2=3.761 perc2=4
Uvigerina striata (UstT_NODE_48553:1-1170(+)) tpm1=88.826 perc1=1 tpm2=54.558 perc2=1
Uvigerina striata (UstT_NODE_24799:170-1675(+)) tpm1=73.518 perc1=1 tpm2=49.937 perc2=1
Uvigerina striata (UstT_NODE_29255:1388-3(-)) tpm1=45.023 perc1=2 tpm2=23.504 perc2=2
Uvigerina striata (UstT_NODE_40407:170-1309(+)) tpm1=28.764 perc1=2 tpm2=17.113 perc2=2

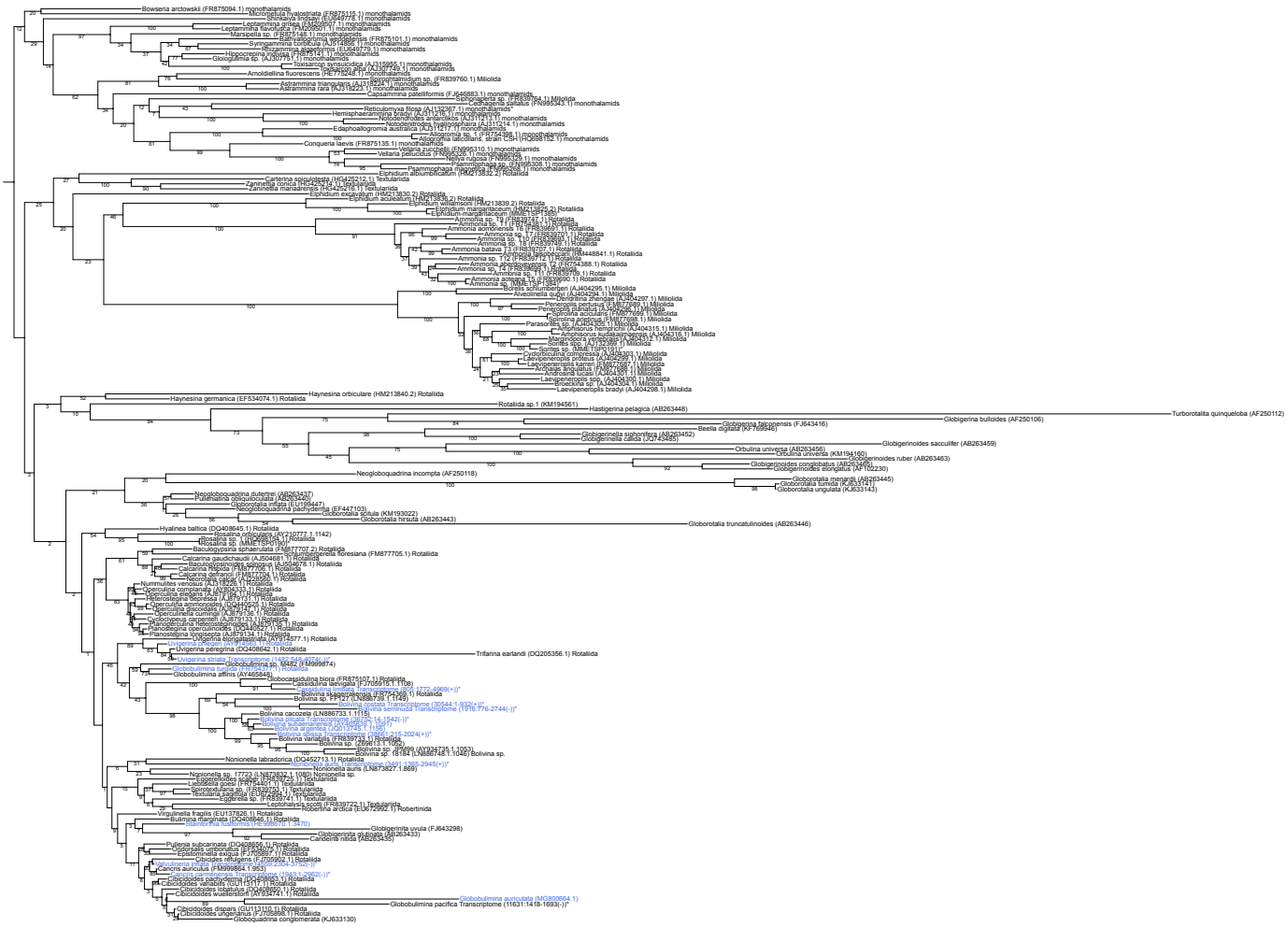
7


C
Bacteria
Eukaryota/Archaea
Foraminifera
II
Tree scale: 10
Eukaryota
Bacteria
Foraminifera
I

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. The background colour coding indicates different taxonomic groups. *Gobobulimina* spp. homologs from Woehle et al. 2018 are shown in bold. Further phylogenetic (sub-)clades, resembling those from Woehle et al. 2018, are highlighted by purple bars. The Nor tree was rooted using two cbb3 oxidases as outgroup, while for the Nrt tree representatives of DHA14 and ACS family of MFS transporters as outgroup. The NirK tree was rooted using the MAD method.



Suppl. Figure S3. 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 S4. 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.