Current Biology

A Novel Eukaryotic Denitrification Pathway in Foraminifera

Highlights

- Large-scale genome and transcriptome analysis of foraminifera

- A novel eukaryotic denitrification pathway is encoded in the *Globobulimina* genome

- *Globobulimina* denitrification enzymes are an ancient acquisition from prokaryotes

Authors

Christian Woehle,
Alexandra-Sophie Roy,
Nicolaas Glock, ..., Jan Michels,
Joachim Schönfeld, Tal Dagan

Correspondence

cwoehle@ifam.uni-kiel.de (C.W.),
sroy@ifam.uni-kiel.de (A.-S.R.)

In Brief

Woehle, Roy, et al. report a novel eukaryotic denitrification pathway in foraminiferal genomes. The enzymes nitrite reductase (NirK) and nitric oxide reductase (Nor) are encoded in *Globobulimina.* A phylogenetic analysis provides insights into the genetic capacity for denitrification in foraminifera and its evolutionary origin in prokaryotes.

CellPress

A Novel Eukaryotic Denitrification Pathway in Foraminifera

Christian Woehle,[1,6,7,*] Alexandra-Sophie Roy,[1,6,*] Nicolaas Glock,[2] Tanita Wein,[1] Julia Weissenbach,[1,5]
Philip Rosenstiel,[3] Claas Hiebenthal,[2] Jan Michels,[4] Joachim Schönfeld,[2] and 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]Institute of Clinical Molecular Biology, Kiel University, Am Botanischen Garten 11, Kiel 24118, Germany
[4]Institute of Zoology, Kiel University, Am Botanischen Garten 1–9, Kiel 24118, Germany
[5]Present address: Faculty of Biology, Technion – Israel Institute of Technology, Haifa 3200003, Israel
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: cwoehle@ifam.uni-kiel.de (C.W.), sroy@ifam.uni-kiel.de (A.-S.R.)
https://doi.org/10.1016/j.cub.2018.06.027

SUMMARY

Benthic foraminifera are unicellular eukaryotes inhabiting sediments of aquatic environments. Several species were shown to store and use nitrate for complete denitrification, a unique energy metabolism among eukaryotes. The population of benthic foraminifera reaches high densities in oxygen-depleted marine habitats, where they play a key role in the marine nitrogen cycle. However, the mechanisms of denitrification in foraminifera are still unknown, and the possibility of a contribution of associated bacteria is debated. Here, we present evidence for a novel eukaryotic denitrification pathway that is encoded in foraminiferal genomes. Large-scale genome and transcriptomes analyses reveal the presence of a denitrification pathway in foraminifera species of the genus *Globobulimina*. This includes the enzymes nitrite reductase (NirK) and nitric oxide reductase (Nor) as well as a wide range of nitrate transporters (Nrt). A phylogenetic reconstruction of the enzymes' evolutionary history uncovers evidence for an ancient acquisition of the foraminiferal denitrification pathway from prokaryotes. We propose a model for denitrification in foraminifera, where a common electron transport chain is used for anaerobic and aerobic respiration. The evolution of hybrid respiration in foraminifera likely contributed to their ecological success, which is well documented in palaeontological records since the Cambrian period.

INTRODUCTION

Production of biologically inaccessible dinitrogen (N_2) is attributed to anaerobic oxidation of ammonium (NH_4^+) and to the anaerobic respiration of nitrate (NO_3^-) to N_2, named denitrification [1]. These processes in the oceans are considered of major importance in the global nitrogen cycle [2]. Indeed, oxygen-depleted environments that are densely populated by denitrifying organisms constitute major sinks for bioavailable nitrogen species [1]. Although the production of N_2 by denitrification is widespread among prokaryotes [3, 4], in eukaryotes, it has been reported only in foraminifera [5].

Foraminifera are known to colonize a wide range of marine habitats where abiotic factors, especially fluctuation or depletion of oxygen availability, are key to species diversity and success [6]. Several benthic species of the order Rotaliida show denitrification activity [5, 7]. Recent studies predicted that denitrifying foraminifera contribute up to 100% of total benthic denitrification in the Peruvian oxygen minimum zone [8], where foraminifera reach abundances of >500 individuals per cm^2 [8, 9]. Furthermore, foraminifera were shown to have a NO_3^- storage that is suggested to accumulate in intracellular vacuoles [10] and sustains denitrification activity for months [11].

Several foraminifera species (e.g., *Buliminella tenuata* and *Virgulinella fragilis*) were shown to harbor intracellular bacteria [12–15], and it has been suggested that these bacteria perform the denitrification reported in these species [14, 16]. In contrast, the denitrifying foraminifera *Globobulimina turgida* [17] (previously determined as *G. pseudospinescens*) [18] harbors intracellular bacteria in a low abundance such that the denitrification rates measured for this species cannot be accounted for a substantial bacterial contribution [5]. A eukaryotic denitrification pathway has previously been described in fungi [19, 20]. However, the fungal denitrification is incomplete where the end product is nitrous oxide (N_2O) rather than N_2. So far, only prokaryotes are known to encode the genetic repertoire to perform complete denitrification.

The commonly known denitrification pathway includes four enzymatic steps that catalyze the reactions $NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$. Dissimilatory reduction of NO_3^- to NO_2^- is facilitated by periplasmic or membrane-bound nitrate reductase (NapA or NarG, respectively). The second denitrification step is catalyzed by cd_1-containing (NirS) or copper-containing (NirK) nitrite reductase. Although NirS is exclusively found in prokaryotes, NirK homologs are found in a few protists and fungi [21]. Further reduction of NO to the greenhouse gas N_2O is catalyzed in prokaryotes by nitric oxide reductase (Nor), and an alternative enzyme (P450Nor) is documented in fungi. The last step of denitrification in prokaryotes is catalyzed by nitrous oxide reductase (NosZ). Therefore, all denitrifying organisms share a similar gene
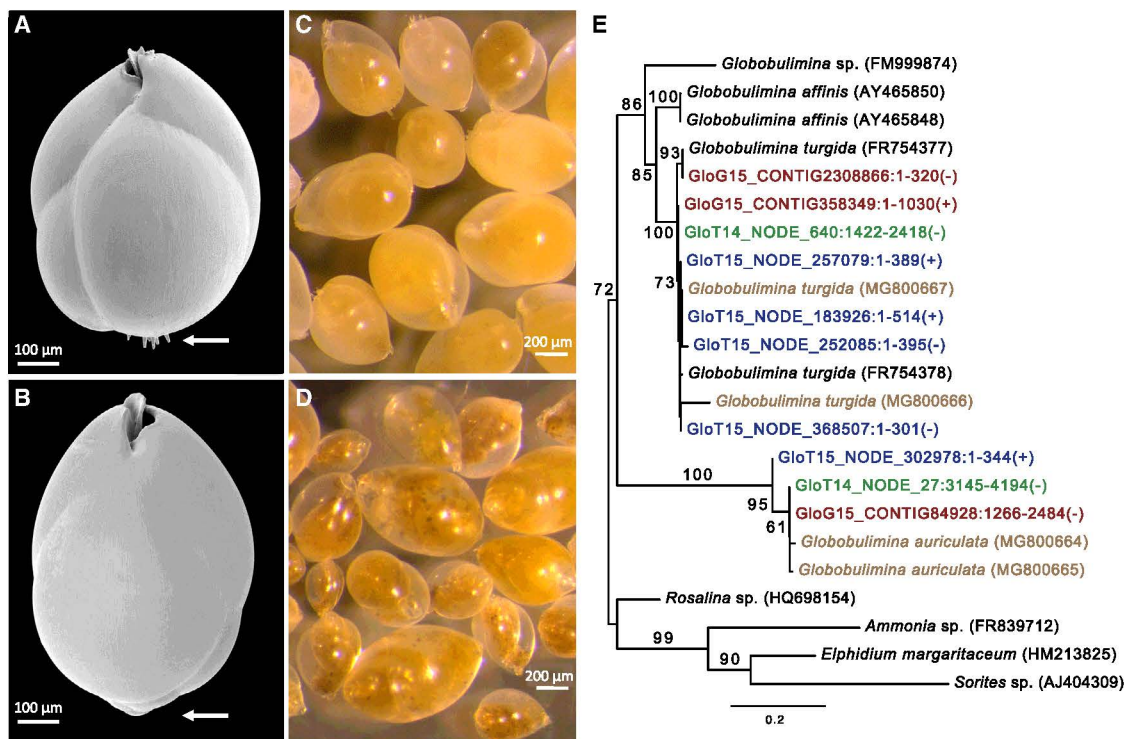
Figure 1. *Globobulimina turgida* and *G. auriculata*

(A) Scanning electron micrograph of a *G. turgida* test. This species is characterized by spines at the proloculus (arrow) and a serrated tooth plate margin.

(B) Scanning electron micrograph of a *G. auriculata* test demonstrating the species' lack of spines (arrow).

(C and D) Stereo micrographs showing numerous living *G. turgida* (C) and *G. auriculata* (D) individuals, highlighting the differences in cytoplasm color and test transparency.

(E) A species phylogenetic tree based on 18S rRNA gene sequences. GloG15 and GloT14/GloT15 prefixes correspond to genomics and transcriptomics data from our study, respectively (in red and green/blue). 18S rRNA gene sequences from specific individuals are shown in brown. Identifiers of sequences derived from assemblies of the current study include the contig identifier, sequence position, and strand orientation. The tree was rooted with 18S rRNA gene sequences from other foraminifera species as outgroup. Bootstrap support values (>50) are given at the branches, and the scale bar refers to substitutions per site. The tree supports the taxonomic classification of the species in our study and demonstrates a clear divergence of the two species into distinct clades. See also Tables S1 and S3.

repertoire of the denitrification pathway. Nonetheless, the denitrification gene set in foraminifera remains unknown. Here, we investigate the genetic repertoire of the denitrification pathway in foraminifera. We analyzed the whole genomes and transcriptomes of two *Globobulimina* species from the Gullmar Fjord (Sweden). We find eukaryotic genes encoding for the denitrification enzymes in foraminifera and investigate their evolutionary origin.

RESULTS

Homologs of Denitrification Enzymes in *Globobulimina* spp.

Foraminifera were obtained from sediment sampled in the Gullmar Fjord, Sweden. A total of 3,360 viable individuals of *Globobulimina turgida* and *Globobulimina auriculata* [17] were manually picked based on their morphological characteristics (Figure 1). Denitrification rates for both species were calculated from N_2O production measurements following protocols previously established for foraminifera [5, 7, 22]. The denitrification rate measured for *G. turgida* ranges between 28 and 1,712

pmol individual^{-1} day^{-1} (Table S1) and are within the range of previously measured rates for this species [5, 11]. The rate calculated for *G. auriculata* ranges between 29 and 124 pmol individual^{-1} day^{-1} (Table S1).

The genome and the eukaryotic transcriptome were sequenced from batches of pooled individuals of both species. Genomic sequences were sorted into genomic bins based on coverage and tetra-nucleotide frequencies. Six bins containing substantial genomic information of *Globobulimina*, as validated by the transcriptome information, were classified as *Globobulimina* draft genome. The draft genome includes 48,370 contigs covering a total length of ~70 Mb. The number of duplicates per gene shows that the *Globobulimina* draft genome is enriched for a single species; hence, the species heterogeneity is low (see STAR Methods). Additional 26 bins were classified as bacterial taxa, and their coverage was sufficient to assemble draft genomes. The remaining contigs (~2,260,000 contigs) were classified as unassigned bins. The presence of genes involved in the denitrification pathway in *Globobulimina* was tested by sequence similarity search using prokaryotic and eukaryotic query sequences. This revealed *Globobulimina* homologs of
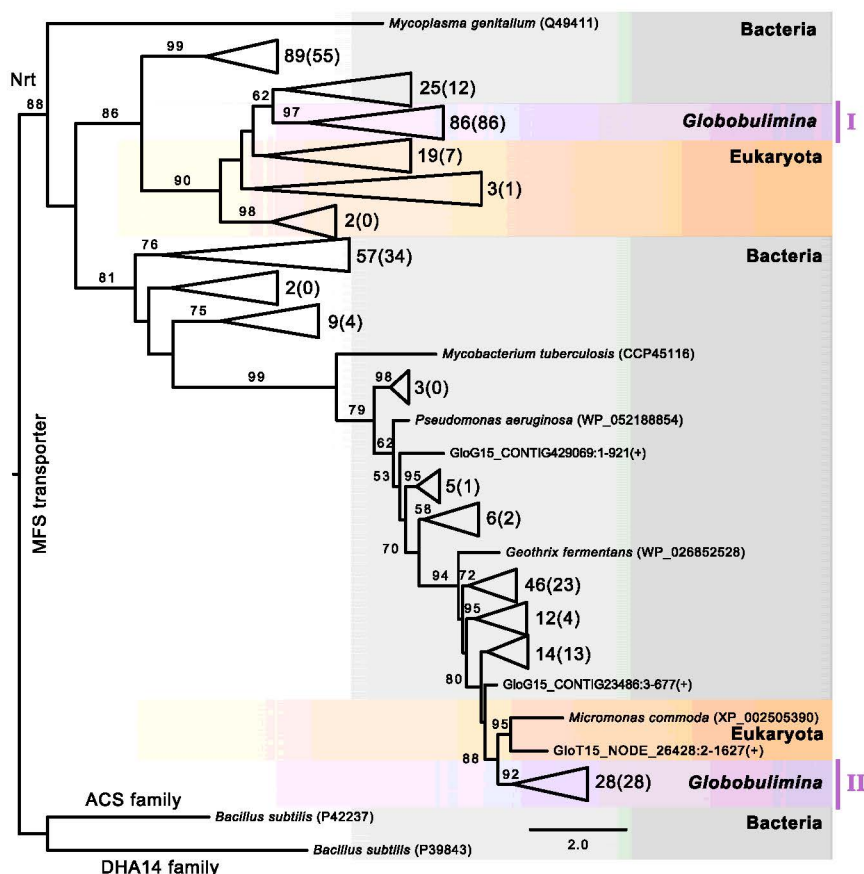
Figure 2. Phylogeny of the *Globobulimina* Nitrate/Nitrite Transporter

Color coding indicates taxonomic classification where *Globobulimina* clades are highlighted in purple. The tree was rooted with homologs of the DHA14 and ACS protein family as outgroup. The number of sequences included in collapsed branches is shown next to the corresponding triangles (the number of sequences derived from the current study is indicated in brackets). Additional sequences are from the genome (GloG15) and transcriptome assembly (GloT15). Bootstrap support values using 100 replicates (>50) are denoted at the branches, and the scale bar refers to substitutions per site (see complete phylogeny in Data S2A). See also Tables S2, S4, and S5 and Data S1 and S2.

clade I and II indicate that NO_3^- transport is an important property of *Globobulimina*.

Our analysis identified homologs of NirK in *Globobulimina*, which contain cupredoxin domains and conserved copper binding sites that are typical for that enzyme, supporting their functional annotation as a nitrite reductase (Figure 3A; Data S1B). Notably, the *Globobulimina* homologs contain introns that are flanked by the canonical eukaryotic splice sites [23] (Figure 3B; Data S1B). A phylogenetic reconstruction of NirK homologs reveals two highly supported sister clades (Figure 3C), including homologs having two (clade I) or three (clade II) introns (Figure 3B). Exon sequences of the NirK-encoding gene from both clades were validated experimentally (Figures S1A and S1B). The phylogenetic tree further shows that the *Globobulimina* NirK homologs are neighbors to a prokaryotic clade (Figure 3C). To further validate the prokaryotic neighborhood of the *Globobulimina* NirK, we tested the likelihood of alternative tree topologies. A eukaryotic neighborhood of the *Globobulimina* clade could be rejected, albeit at a marginal confidence level (p value 0.043; using approximate unbiased test). This shows that the *Globobulimina* clade is deeply branching in the tree rather than grouping with a specific NirK subclade. Consequently, we conclude that the NirK origin in *Globobulimina* is ancient. Indeed, multiple topologies of deeper branching positions could not be statistically rejected (Figure 3C). Further search for the alternative nitrite reductase NirS reveals the presence of homologs, which are found in genomes of associated bacteria in our dataset, but not in the *Globobulimina* transcriptome or draft genome (Data S2C).
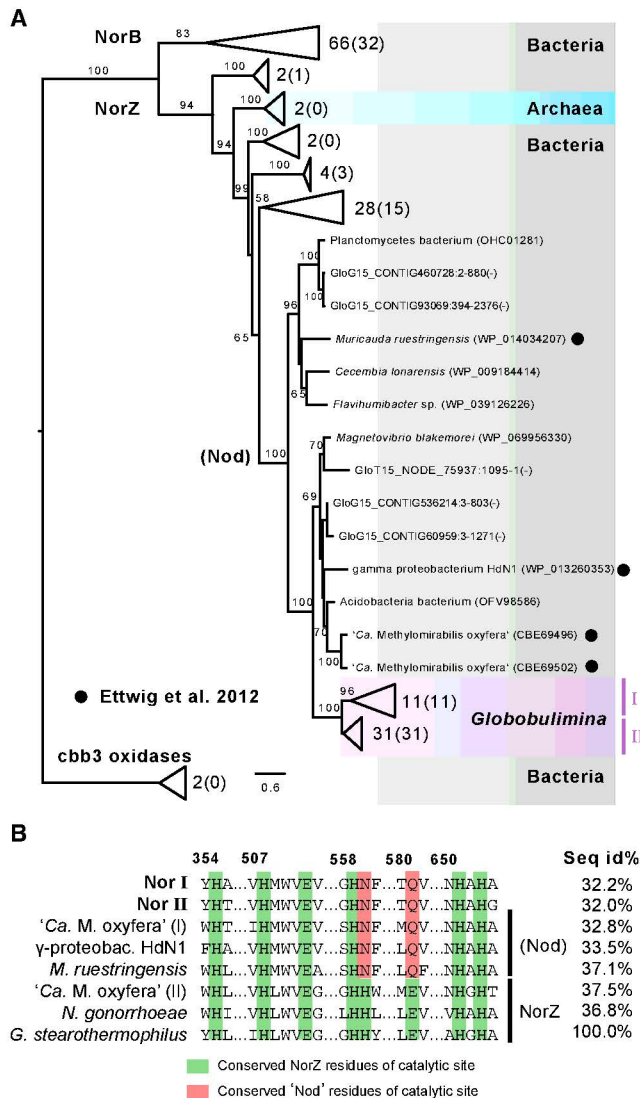
A search for Nor homologs revealed that the *Globobulimina* genome encodes the prokaryotic Nor whose functional annotation is supported by the presence of a conserved cytochrome oxidase I domain and a conserved catalytic site (Data S1C). One gene encoding a Nor homolog is located adjacent to a eukaryote-specific gene, and the Nor-encoding gene sequence could be validated experimentally (Figure S1C). These findings provide further

nitrate transporters (Nrt) and two enzymes in the denitrification pathway: copper-containing nitrite reductase (NirK) and nitric oxide reductase (Nor).

Evidence for Functional Conservation of *Globobulimina* spp. Nrt, NirK, and Nor Protein Sequences

The utilization of NO_3^- in denitrification suggests that foraminifera harbor mechanisms to transport NO_3^-. The search for Nrt homologs in the *Globobulimina* draft genome revealed multiple protein coding sequences related to the NarK protein superfamily (Data S1A). A phylogenetic analysis of Nrt protein sequences and identified homologs show two distinct *Globobulimina* clades that likely correspond to two diverged protein families (clades I and II; Figure 2). Within each clade, we observe a multitude of diversified homologs (Data S2A). Clade I clusters with bacterial encoded Nrt, which would suggest an origin of the *Globobulimina* Nrt that is independent from other eukaryotes. To validate the bacterial neighborhood, we tested an alternative topology where we constrained a monophyletic clade, including the eukaryotic homologs and *Globobulimina* clade I. This revealed that the likelihood of eukaryotic monophyly is not significantly different from the bacterial neighborhood position (p value 0.352; using approximate unbiased test). The second *Globobulimina* clade (clade II) has a highly supported eukaryotic nearest neighbor (*Micromonas commoda*); from this, we conclude that the Nrt is of eukaryotic ancestry. This and a massive diversification of homologs in

Figure 3. Sequence and Phylogeny of the *Globobulimina* Nitrite Reductase

(A) Sequence conservation of copper ligand binding sites in *Globobulimina* nitrite reductase (NirK). Copper ligand binding sites are highlighted in green. Type 1 or 2 copper binding sites are indicated. Column numbers refer to the corresponding sequence positions in the *Geobacillus thermodenitrificans* sequence. "Seq id%" refers to the local alignment sequence identity compared to *Geobacillus thermodenitrificans* (100%). For details and the complete alignment, see Data S1B.

(B) Gene structure in *Globobulimina nirK* clades based on genomic contigs. Dashed lines indicate regions that are supported by transcriptome assembly but were not covered by genomic contigs. Although the automatic annotation of both genomic representatives of *Globobulimina* NirK was predicted to encode complete N termini, homology to predicted transcriptome contigs indicates an incomplete gene start.

(C) Collapsed phylogenetic tree of *Globobulimina* NirK. Color coding indicates taxonomic classification, where two *Globobulimina* clades (clade I and II) are highlighted and form a monophyletic clade. The number of sequences represented by collapsed branches is shown next to the corresponding triangles. (The number of sequences derived from the current study is indicated in brackets.) Alternative tested topologies of the *Globobulimina* clades are indicated by circled numbers (① Grouping of the *Globobulimina* clade with other eukaryotes in contrast to all remaining sequences in the tree is shown. ② The *Globobulimina* clade groups independently from NirK class II clade. ③ Grouping of the *Globobulimina* clade with a eukaryote-containing subclade of NirK class II is shown). For more details and the complete phylogeny, see Data S2B. The tree was rooted with the minimal ancestor deviation (MAD) method. Bootstrap support values using 100 replicates (>50) are denoted at the branches. The scale bar refers to substitutions per site.

See also Figure S1; Tables S2–S5; and Data S1 and S2.

support for the conclusion that Nor is indeed encoded in the *Globobulimina* genome. A phylogenetic reconstruction shows that the *Globobulimina* Nor homologs are monophyletic and nested within the bacterial NorZ clade (Figure 4A). Notably, the *Globobulimina* clade is neighbor to a clade containing two "*Candidatus* Methylomirabilis oxyfera" Nor homologs. Denitrification in "*Ca.* M. oxyfera" was reported to lack the intermediary N_2O, where N_2 and O_2 are released [24]. Consequently, it has been proposed that the *Ca.* M. oxyfera Nor homolog is, in fact, a nitric oxide dismutase (Nod) that catalyzes the enzymatic reaction $2NO \rightarrow N_2 + O_2$ [25]. The amino acid residues of the *Globobulimina* Nor enzyme catalytic site are similar to Nod, where two of the eight sites are identical (Figure 4B). However, because we measured N_2O production in both foraminifera species in our study, we can conclude that the *Globobulimina* Nor function is not Nod-like.

Missing Enzymatic Reactions and Bacterial Contribution

Our search for known denitrification genes did not detect *Globobulimina* homologs to enzymes catalyzing the first and last steps of denitrification. Homologs of the dissimilatory nitrate reductases (NapA and NarG) as well as NosZ were found in the associated bacteria draft genomes (Figure S2; Data S2E–S2G). However, the possibility of bacterial contribution to the

Globobulimina denitrification has been previously discarded due to their low abundance [5]. Considering the presence of nitrate storage in *Globobulimina* [5], it is likely that nitrate reduction occurs within the cell. Therefore, if *Globobulimina* encodes a dissimilatory nitrate reductase, it has to be a yet uncharacterized gene. Notably, we found homologs of the eukaryotic assimilatory nitrate reductase (Nr) in the *Globobulimina* draft genome. This enzyme has been shown to catalyze denitrification in the fungus *Cylindrocarpon tonkinense* under specific conditions [26]. The phylogenetic reconstruction and assessment of functional domains reveal that the *Globobulimina* Nr homologs are more closely related to sulfite oxidases than to Nr; therefore, they are better described as sulfite oxidases (Data S2H; Figure S3). Previous studies of sulfite oxidases in animals showed that these enzymes can acquire nitrate-reducing activity following a replacement of only two amino acid residues (e.g., in human and chicken) [27]. Thus, it is possible that the *Globobulimina* Nr homolog has an assimilatory nitrate reductase function.

To further validate the presence of denitrification enzymes in foraminifera, we searched the public databases for homologs to the *Globobulimina* enzymes. Foraminifera sequencing data are scarce in public databases, yet we could find several representatives (Table S2). Only one highly fragmented draft genome of the distant species *Reticulomyxa filosa* [28] was found in

Figure 4. Phylogeny and Conserved Amino Acid Residues of the *Globobulimina* Nitric Oxide Reductase

(A) Color coding indicates for taxonomic classification. Two *Globobulimina* clades (clade I and II) are monophyletic and highlighted in purple color. The number of sequences represented by collapsed branches is shown next to the corresponding triangles (the number of sequences derived from the current study is indicated in brackets). The tree was rooted with two cbb3 oxidases as outgroup. Nod candidates as in Ettwig et al. [24, 25] are highlighted by black circles. Uncollapsed sequences represent protein predictions from the genome (GloG15) and transcriptome assembly (GloT15). Bootstrap support values using 100 replicates are denoted at the branches (>50). The scale bar refers to substitutions per site. For more details and the complete phylogeny, see Data S2D.

(B) Sequence conservation of catalytic sites. Amino acid residues of the catalytic site are highlighted in green and red. Two different sequences found in "*Candidatus* Methylomirabilis oxyfera" are indicated by "(I)" and "(II)." "Seq id%" refers to the local alignment sequence identity compared to *Geobacillus thermodenitrificans* (100%), and column numbers refer to the corresponding sequence positions in *G. thermodenitrificans*. Details and the complete alignment are shown in Data S1C.

See also Figure S1; Tables S2–S5; and Data S1 and S2.

databases, and no genomes or transcriptomes of species reported to denitrify have been published. Homologs to NirK were found in the rotaliids *Brizalina* sp. and *Rosalina* sp., and Nor homologs were found in *Brizalina* only. Homologs of Nrt were found in the intertidal rotaliids *Elphidium margaritaceum* and *Ammonia* spp., the miliolid foraminifera *Sorites* sp., as well as *Rosalina* sp., *Nonionellina* sp., and *Bulimina marginata*. Most of those homologous sequences are short gene fragments; nonetheless, their phylogenetic reconstruction shows a robust clustering with *Globobulimina* (Data S2I–S2K).

DISCUSSION

Our results demonstrate that the main denitrification enzymes are encoded in the genome of *Globobulimina*, revealing a so far undescribed eukaryotic denitrification pathway. The pathway origin is independent of known eukaryotic enzymes and has a prokaryotic ancestry. Most of the eukaryotic genes of prokaryotic ancestry were shown to have been acquired by endosymbiotic gene transfer (EGT) from the mitochondrion ancestor at the origins of eukaryotes [29, 30]. This process is still ongoing [31], and the mechanisms involved are beginning to be unraveled [32]. Eukaryotic gene acquisition from prokaryotic donors by lateral gene transfer (LGT) (as opposed to EGT) is frequently reported; however, these are often hampered by signatures of bacterial contamination, and therefore, their interpretation is often complicated [33, 34]. We note that, in our study, we controlled for possible prokaryotic contamination by analyzing the genome and transcriptome of *Globobulimina* in parallel. Furthermore, gene origin in our pipeline was determined by genomic binning approach, genomic context, and phylogenetics.

The phylogeny of Nrt supports an ancient origin of the nitrate metabolism in foraminifera. The foraminiferal NirK phylogeny and the absence of eukaryotic homologs to the Nor found in foraminifera further indicate that the origin of the foraminiferal denitrification pathway is independent from the known fungal pathway. The fungal NirK gene origin is likely an EGT [21], and indeed, it is clustering with other eukaryotic taxa in our phylogenetic analysis as expected for genes of endosymbiotic origin (Figure 3C; Data S2B). The *Globobulimina* NirK and Nor are deep branching in all phylogenies, and the enzymes' monophyly in foraminifera is supported by the presence of homologs in other rotaliids (Data S2I–S2K). These indicate an ancient origin of denitrification enzymes in foraminifera by LGT from a prokaryotic donor.

Our findings demonstrate that denitrification is performed by foraminifera rather than associated bacteria. Based on our results, we propose a denitrification model for foraminifera that draws upon known denitrification properties of fungi (Figure 5). Nitrate transport is a necessary step preceding denitrification that can be facilitated by the Nrt. The Nrt function must not be limited to transport into the cell; it can also be integrated into NO_3^- storage vacuoles and NO_3^- transport into the mitochondria. In certain foraminifera species (family Bolivinidae), mitochondria were observed to cluster near the tests' pores in oxygen-depleted environments [35, 36]. Consequently, denitrification can be performed inside the mitochondrion, as previously reported for fungi [37] and suggested for the foraminifera *Bolivina spissa* [38]. We hypothesize that enzymatic components of foraminifera localize similarly to their fungal and bacterial
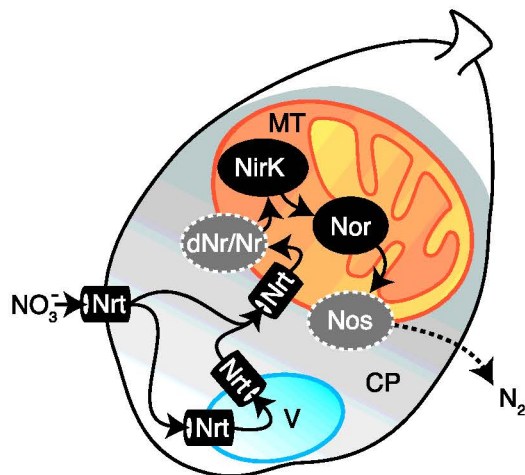
Figure 5. A Model for Denitrification in Foraminifera

A schematic representation of a foraminiferal cell containing cytoplasm, a vacuole, and a mitochondrion is shown. Arrows indicate the suggested flow of nitrogen compounds in foraminiferal denitrification from NO_3^- to N_2. A dashed arrow indicates for the diffusion of gaseous N_2 to the outside of the cell. Hypothesized enzymatic reactions for the first and the last step of denitrification are highlighted by dark gray color and surrounded by a dashed line. CP, cytoplasm; dNr, dissimilatory nitrate reductase; MT, mitochondrion; NirK, copper-containing nitrite reductase; Nor, nitric oxide reductase; Nos, nitrous oxide reductase; Nr, eukaryotic nitrate reductase; Nrt, nitrate/nitrite transporter.

Related to Figures S2 and S3.

homologs. Therefore, NirK can be localized in the mitochondrial intermembrane space and Nor in the mitochondrial inner membrane [37, 39]. Our study is lacking evidence for foraminiferal genes performing the first and last denitrification reactions. The *Globobulimina* sulfite oxidases homologous to Nr (Figure S3) could, hypothetically, catalyze the nitrate reduction step. However, it is also possible that so far uncharacterized or unrecognized proteins are catalyzing the missing denitrification reactions. Considering a tight association of denitrifying enzymes with the respiratory chain, the putative dissimilatory nitrate reductase (dNr) and nitrous oxide reductase (Nos) enzymes are likely localized inside the mitochondrion.

The proposed foraminiferal model suggests sharing of a common electron transport chain between aerobic respiration and denitrification, permitting the use of both electron acceptors in parallel without the need of assembling new protein complexes as reported in the fungus *Fusarium oxysporum* [37]. We speculate that this ability lends foraminifera a substantial ecological advantage when exposed to hypoxia or in response to fluctuating oxygen levels and explains their success in populating a wide range of marine habitats.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Sites description and sampling
- METHOD DETAILS
  - Nucleic acid isolation and sequencing
  - Validation of gene sequences
  - Taxonomy
  - Foraminiferal denitrification rate
  - Microscopic visualizations of tests and living individuals of *Globobulimina* spp.
  - Processing of sequencing data
  - Gene identification and phylogenies
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
  - Data availability

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, five tables, and two supplemental data files and can be found with this article online at https://doi.org/10.1016/j.cub.2018.06.027.

AUTHOR CONTRIBUTIONS

A.-S.R., J.W., T.W., C.W., J.S., N.G., and T.D. designed the research strategy and performed the sampling. A.-S.R. carried out the experimental lab work. C.W. performed the bioinformatics analysis. C.H. built the incubation systems. N.G. performed the denitrification rate measurements. J.M. and A.-S.R. created the scanning electron and stereo micrographs. P.R. sequenced the genomes and transcriptomes. All authors interpreted the results and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Thamdrup, B. (2012). New pathways and processes in the global nitrogen cycle. Annu. Rev. Ecol. Evol. Syst. *43*, 407–428.

2. Canfield, D.E., Glazer, A.N., and Falkowski, P.G. (2010). The evolution and future of Earth's nitrogen cycle. Science *330*, 192–196.

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

4. Zumft, W.G. (1997). Cell biology and molecular basis of denitrification. Microbiol. Mol. Biol. Rev. *61*, 533–616.

5. Risgaard-Petersen, N., Langezaal, A.M., Ingvardsen, S., Schmid, M.C., Jetten, M.S.M., Op den Camp, H.J.M., Derksen, J.W.M., Piña-Ochoa, E., Eriksson, S.P., Nielsen, L.P., et al. (2006). Evidence for complete denitrification in a benthic foraminifer. Nature *443*, 93–96.

6. Murray, J.W. (2006). Ecology and Applications of Benthic Foraminifera (Cambridge University Press).

7. Piña-Ochoa, E., Høgslund, S., Geslin, E., Cedhagen, T., Revsbech, N.P., Nielsen, L.P., Schweizer, M., Jorissen, F., Rysgaard, S., and Risgaard-Petersen, N. (2010). Widespread occurrence of nitrate storage and denitrification among Foraminifera and Gromiida. Proc. Natl. Acad. Sci. USA *107*, 1148–1153.

8. Glock, N., Schönfeld, J., Eisenhauer, A., Hensen, C., Mallon, J., and Sommer, S. (2013). The role of benthic foraminifera in the benthic nitrogen cycle of the Peruvian oxygen minimum zone. Biogeosciences *10*, 4767–4783.

9. Gooday, A.J., Bernhard, J.M., Levin, L.A., and Suhr, S.B. (2000). Foraminifera in the Arabian Sea oxygen minimum zone and other oxygen-deficient settings: taxonomic composition, diversity, and relation to metazoan faunas. Deep Sea Res. Part II Top. Stud. Oceanogr. *47*, 25–54.

10. Bernhard, J.M., Edgcomb, V.P., Casciotti, K.L., McIlvin, M.R., and Beaudoin, D.J. (2012). Denitrification likely catalyzed by endobionts in an allogromiid foraminifer. ISME J. *6*, 951–960.

11. Piña-Ochoa, E., Koho, K.A., Geslin, E., and Risgaard-Petersen, N. (2010). Survival and life strategy of the foraminiferan *Globobulimina turgida* through nitrate storage and denitrification. Mar. Ecol. Prog. Ser. *417*, 39–49.

12. Nomaki, H., Chikaraishi, Y., Tsuchiya, M., Toyofuku, T., Ohkouchi, N., Uematsu, K., Tame, A., and Kitazato, H. (2014). Nitrate uptake by foraminifera and use in conjunction with endobionts under anoxic conditions. Limnol. Oceanogr. *59*, 1879–1888.

13. Bernhard, J.M., Martin, J.B., and Rathburn, A.E. (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*, PA4206.

14. Bernhard, J.M., Habura, A., and Bowser, S.S. (2006). An endobiont-bearing allogromiid from the Santa Barbara Basin: Implications for the early diversification of foraminifera. J. Geophys. Res. *111*, G03002.

15. Bernhard, J.M. (2003). Potential symbionts in bathyal foraminifera. Science *299*, 861.

16. Bernhard, J.M., Casciotti, K.L., McIlvin, M.R., Beaudoin, D.J., Visscher, P.T., and Edgcomb, V.P. (2012). Potential importance of physiologically diverse benthic foraminifera in sedimentary nitrate storage and respiration. J. Geophys. Res. *117*, G03002.

17. Bailey, J.W. (1851). Microscopical examination of soundings, made by the United States Coast Survey off the Atlantic Coast of the United States. Smithsonian Inst. Contr. Knowledge *2*, 12.

18. Emiliani, C. (1949). Studio micropaleontologico di una serie calabriana. Riv. Ital. Paleontol. Stratigr. *54*, 75–77.

19. Shoun, H., Fushinobu, S., Jiang, L., Kim, S.-W., and Wakagi, T. (2012). Fungal denitrification and nitric oxide reductase cytochrome P450nor. Philos. Trans. R. Soc. Lond. B Biol. Sci. *367*, 1186–1194.

20. Shoun, H., Kim, D.H., Uchiyama, H., and Sugiyama, J. (1992). Denitrification by fungi. FEMS Microbiol. Lett. *73*, 277–281.

21. Kim, S.-W., Fushinobu, S., Zhou, S., Wakagi, T., and Shoun, H. (2009). Eukaryotic nirK genes encoding copper-containing nitrite reductase: originating from the protomitochondrion? Appl. Environ. Microbiol. *75*, 2652–2658.

22. Høgslund, S., Revsbech, N.P., Cedhagen, T., Nielsen, L.P., and Gallardo, V.A. (2008). Denitrification, nitrate turnover, and aerobic respiration by benthic foraminiferans in the oxygen minimum zone off Chile. J. Exp. Mar. Biol. Ecol. *359*, 85–91.

23. Thanaraj, T.A., and Clark, F. (2001). Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res. *29*, 2581–2593.

24. Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M.M., Schreiber, F., Dutilh, B.E., Zedelius, J., de Beer, D., et al. (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. Nature *464*, 543–548.

25. Ettwig, K.F., Speth, D.R., Reimann, J., Wu, M.L., Jetten, M.S.M., and Keltjens, J.T. (2012). Bacterial oxygen production in the dark. Front. Microbiol. *3*, 273.

26. Watsuji, T.-O., Takaya, N., Nakamura, A., and Shoun, H. (2003). Denitrification of nitrate by the fungus *Cylindrocarpon tonkinense*. Biosci. Biotechnol. Biochem. *67*, 1115–1120.

27. Qiu, J.A., Wilson, H.L., and Rajagopalan, K.V. (2012). Structure-based alteration of substrate specificity and catalytic activity of sulfite oxidase from sulfite oxidation to nitrate reduction. Biochemistry *51*, 1134–1147.

28. Glöckner, G., Hülsmann, N., Schleicher, M., Noegel, A.A., Eichinger, L., Gallinger, C., Pawlowski, J., Sierra, R., Euteneuer, U., Pillet, L., et al. (2014). The genome of the foraminiferan *Reticulomyxa filosa*. Curr. Biol. *24*, 11–18.

29. Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F.L., Lockhart, P.J., Bryant, D., Hazkani-Covo, E., McInerney, J.O., Landan, G., and Martin, W.F. (2015). Endosymbiotic origin and differential loss of eukaryotic genes. Nature *524*, 427–432.

30. Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. *5*, 123–135.

31. Ju, Y.S., Tubio, J.M.C., Mifsud, W., Fu, B., Davies, H.R., Ramakrishna, M., Li, Y., Yates, L., Gundem, G., Tarpey, P.S., et al.; ICGC Prostate Cancer Working Group; ICGC Bone Cancer Working Group; ICGC Breast Cancer Working Group (2015). Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells. Genome Res. *25*, 814–824.

32. Diner, R.E., Noddings, C.M., Lian, N.C., Kang, A.K., McQuaid, J.B., Jablanovic, J., Espinoza, J.L., Nguyen, N.A., Anzelmatti, M.A., Jr., Jansson, J., et al. (2017). Diatom centromeres suggest a mechanism for nuclear DNA acquisition. Proc. Natl. Acad. Sci. USA *114*, E6015–E6024.

33. Ku, C., and Martin, W.F. (2016). A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70 % rule. BMC Biol. *14*, 89.

34. Koutsovoulos, G., Kumar, S., Laetsch, D.R., Stevens, L., Daub, J., Conlon, C., Maroon, H., Thomas, F., Aboobaker, A.A., and Blaxter, M. (2016). No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. Proc. Natl. Acad. Sci. USA *113*, 5053–5058.

35. Leutenegger, S., and Hansen, H.J. (1979). Ultrastructural and radiotracer studies of pore function in foraminifera. Mar. Biol. *54*, 11–16.

36. Bernhard, J.M., Goldstein, S.T., and Bowser, S.S. (2010). An ectobiont-bearing foraminiferan, Bolivina pacifica, that inhabits microxic pore waters: cell-biological and paleoceanographic insights. Environ. Microbiol. *12*, 2107–2119.

37. Takaya, N., Kuwazaki, S., Adachi, Y., Suzuki, S., Kikuchi, T., Nakamura, H., Shiro, Y., and Shoun, H. (2003). Hybrid respiration in the denitrifying mitochondria of *Fusarium oxysporum*. J. Biochem. *133*, 461–465.

38. Glock, N., Eisenhauer, A., Milker, Y., Liebetrau, V., Schoenfeld, J., Mallon, J., Sommer, S., and Hensen, C. (2011). Environmental influences on the pore-density in tests of *Bolivina spissa* (Cushman). J. Foraminiferal Res. *41*, 22–32.

39. Berks, B.C., Ferguson, S.J., Moir, J.W., and Richardson, D.J. (1995). Enzymes and associated electron transport systems that catalyse the

respiratory reduction of nitrogen oxides and oxyanions. Biochim. Biophys. Acta 1232, 97–173.

40. Kester, D.R., Duedall, I.W., Connors, D.N., and Pytkowicz, R.M. (1967). Preparation of artificial seawater. Limnol. Oceanogr. 12, 176–179.

41. Pawlowski, J., and Lecroq, B. (2010). Short rDNA barcodes for species identification in foraminifera. J. Eukaryot. Microbiol. 57, 197–205.

42. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

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

44. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

45. Smith, M.S., Firestone, M.K., and Tiedje, J.M. (1978). The acetylene inhibition method for short-term measurement of soil denitrification and its evaluation using nitrogen131. Soil Sci. Soc. Am. J. 42, 611–615.

46. Andersen, K., Kjær, T., and Revsbech, N.P. (2001). An oxygen insensitive microsensor for nitrous oxide. Sens. Actuators B 81, 42–48.

47. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

48. Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J. Comput. Biol. 20, 714–737.

49. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.

50. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics 12, 323.

51. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

52. Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428.

53. Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607.

54. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055.

55. Hyatt, D., LoCascio, P.F., Hauser, L.J., and Uberbacher, E.C. (2012). Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics 28, 2223–2230.

56. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360.

57. Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32, 767–769.

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

59. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.

60. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

61. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31, 365–370.

62. Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 40, D130–D135.

63. Suzuki, S., Kakuta, M., Ishida, T., and Akiyama, Y. (2015). Faster sequence homology searches by clustering subsequences. Bioinformatics 31, 1183–1190.

64. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26, 1641–1650.

65. Tria, F.D.K., Landan, G., and Dagan, T. (2017). Phylogenetic rooting using minimal ancestor deviation. Nat. Ecol. Evol. 1, 0193.

66. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

67. Matsumoto, Y., Tosha, T., Pisliakov, A.V., Hino, T., Sugimoto, H., Nagano, S., Sugita, Y., and Shiro, Y. (2012). Crystal structure of quinol-dependent nitric oxide reductase from Geobacillus stearothermophilus. Nat. Struct. Mol. Biol. 19, 238–245.

68. Adman, E.T., Godden, J.W., and Turley, S. (1995). The structure of copper-nitrite reductase from Achromobacter cycloclastes at five pH values, with NO_2^- bound and with type II copper depleted. J. Biol. Chem. 270, 27458–27474.

69. Marchler-Bauer, A., and Bryant, S.H. (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32, W327–W331.

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

71. Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12, e1001889.

72. Huang, X., and Madan, A. (1999). CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Foraminifera	This paper	Not applicable
Deposited Data		
Sequencing data generated in the project	This paper	Bioproject (RRID:SCR_004801) accession: PRJNA415343
Raw sequencing reads	This paper	NCBI (RRID:SCR_006472): SRR6202052 – SRR6202078
Globobulimina draft genome assembly (GloG15)	This paper	NCBI: PIVH00000000
Globobulimina transcriptome assemblies (GloT14 & GloT15)	This paper	NCBI: GGCE00000000 & GGCD00000000
Globobulimina-associated bacterial genome assemblies & unbinned contigs	This paper	NCBI: PIVI00000000– PIWH00000000, PJEL00000000
18S rRNA sequences	This paper	NCBI: MG800664–MG800667
Non-redundant protein database (NR), January 2017	ftp://ftp.ncbi.nlm.nih.gov/blast/db/	Not found
RefSeq 79	ftp://ftp.ncbi.nlm.nih.gov/refseq/release/complete/	RRID:SCR_003496
Foraminifera NCBI genomic data	https://www.ncbi.nlm.nih.gov/	NCBI: GCA_000512085.1, GCA_000211355.2
Foraminifera NCBI transcriptome reads	https://www.ncbi.nlm.nih.gov/	NCBI: SRR2003403, SRR2003397, SRR2003388, SRR2003283,
Foraminifera MMETSP (Marine Microbial Eukaryote Transcriptome Sequencing Project) transcriptome data	http://datacommons.cyverse.org/browse/iplant/home/shared/imicrobe/projects/104	MMETSP: MMETSP0190, MMETSP0191, MMETSP1384, MMETSP1385
Experimental Models: Organisms/Strains		
Globobulimina turgida	This paper	Not applicable
Globobulimina auriculata	This paper	Not applicable
Oligonucleotides		
18S (see Table S3)	This paper	Not applicable
nirK clade I (see Table S3)	This paper	Not applicable
nirK clade II (see Table S3)	This paper	Not applicable
Nor clade I (see Table S3)	This paper	Not applicable
Software and Algorithms		
FastQC ver 0.11.5	http://www.bioinformatics.babraham.ac.uk/projects/fastqc	RRID:SCR_014583
NCBI	http://www.ncbi.nlm.nih.gov	RRID:SCR_006472
BLASTP	http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome	RRID:SCR_001010
MAFFT	http://mafft.cbrc.jp/alignment/server/	RRID:SCR_011811
PhyML ver. 20131022	http://www.atgc-montpellier.fr/phyml/	RRID:SCR_014629
Trimmomatic	http://www.usadellab.org/cms/index.php?page=trimmomatic	RRID:SCR_011848
SPAdes ver. 3.9.1	http://bioinf.spbau.ru/spades/	RRID:SCR_000131
TransDecoder ver. 3.0.1	https://github.com/TransDecoder/TransDecoder/releases/tag/v3.0.1	Not found
RSEM ver. 1.2.30	http://deweylab.biostat.wisc.edu/rsem/	RRID:SCR_013027

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Trinity ver. 2.4.0	https://github.com/trinityrnaseq/trinityrnaseq/wiki	RRID:SCR_013048
Bowtie ver. 2.1.0	http://bowtie-bio.sourceforge.net/bowtie2	Not found
IDBA-UD ver. 1.1.1	http://i.cs.hku.hk/~alse/hkubrg/projects/idba_ud/	RRID:SCR_011912
MaxBin ver. 2.2	https://sourceforge.net/projects/maxbin2/	Not found
CheckM ver. 1.0.5	http://ecogenomics.github.io/CheckM/	Not found
MetaProdigal ver. 2.6.2	https://github.com/hyattpd/Prodigal	RRID:SCR_011936
HISAT2 ver. 2.0.5	http://ccb.jhu.edu/software/hisat2/index.shtml	RRID:SCR_015530
BRAKER ver. 1.9.	http://exon.gatech.edu/braker1.html	Not found
CD-HIT ver. 4.6	http://weizhong-lab.ucsd.edu/cd-hit/	RRID:SCR_007105
DIAMOND tool 0.7.11.60	http://www.nitrc.org/projects/diamond/	RRID:SCR_009457
Swiss-Prot	http://www.uniprot.org/	RRID:SCR_002380
GHOSTZ 1.0.0	http://www.bi.cs.titech.ac.jp/ghostz/	Not found
FastTree 2.1.7	http://www.microbesonline.org/fasttree/	RRID:SCR_015501
CONSEL ver. 1.20.	http://stat.sys.i.kyoto-u.ac.jp/prog/consel/	Not found
EMBOSS	http://www.ebi.ac.uk/Tools/emboss/cpgplot/indexhtml	RRID:SCR_007254
CAP3	http://seq.cs.iastate.edu/cap3.html	RRID:SCR_007250

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Christian Woehle (cwoehle@ifam.uni-kiel.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sites description and sampling

Living foraminifera (*Globobulimina turgida* and *Globobulimina auriculata*) were sampled during two consecutive expeditions (2014 & 2015) to the Gullmar Fjord, Sweden. Sediment samples were obtained from the Alsbäck Deep (58°19.38'N, 11°32.74'E) at a depth of ca. 117 m using a 4-tube sediment interface corer (Mini Muc K/MT410). The top 3 cm of the sediments were sampled and wet sieved directly upon return to the station. Foraminifera were individually picked from the 125 to 2000 μm size fraction and cleaned in sterile artificial seawater (ASW) with 30 psu and a nitrate concentration of 20 μmol l^{-1} (see below, modified from Kester et al. [40]). Batches of foraminifera were frozen in liquid nitrogen (ambient condition) or incubated in culturing vessels filled with sterile ASW (see next paragraph) and flash frozen after ca. 45 h (Figure S4). The culturing vessels were incorporated to one of two culturing systems assembled to reproduce different Alsbäck Deep environmental conditions. One system was dedicated to natural oxygen concentration of the Alsbäck Deep (≈125 μmol/l) during summer whereas the second system was completely drawn down of oxygen (< 10 μmol/l). Each system was filled with sterile ASW and sparged to oxygen concentration of interests with N$_2$, CO$_2$ and O$_2$ pre-mixed gas (AGA Gas A; gas entries indicated by red arrows) before and during the length of the experiments. Frozen samples were stored at −80°C.

The sterile ASW used for rinsing and incubations was based on the natural salinity and nitrate concentration based on CTD and nutrients measurements. The water was obtained by dissolving the following salts in 3/5 of the end volume. Here we present weights and molecular weight (MW) for a 1 l end volume ASW. A total of 23.38 g of NaCL (MW = 58.44 g/mol), 4.93 g MGSO$_4$·7H$_2$O (MW = 246.48 g/mol), 1.11 g CaCl$_2$·2H$_2$O (MW = 147.02 g/mol), 0.2 g KBr (MW = 119.01 g/mol), 0.75 g KCL (MW = 74.56 g/mol), 4.06 g MgCL$_2$·6H20 (MW = 203.3 g/mol), and 1 mL H$_3$BO$_3$ (of 61.83 g/mol) were added to double-distilled-water. After the addition of the previous salts 0.023 mL of NaNO$_3$ was added, the pH adjusted to 8 and the media autoclaved. It was important to dissolve each salt individually before the addition of a new one to decrease chance of salt precipitation. The nutrients necessary to produce the adequate ASW were autoclave separately in 10 mL glass vials. In our case 1 mL of NaH$_2$PO$_4$·H$_2$O, 1 mL Na$_2$SiO$_3$·9H$_2$O, and 1 mL of trace metal mix (see below) containing 100 μL selenium (H$_2$SeO$_3$, MW = 128,97) were added to the medium after sterilization. The carbonate system was also added after sterilization of the medium by sterile filtering (0.22 um) 0.17 g of NaHCO$_3$ dissolved in 30 mL double-distilled-water and 500 μL of the vitamins directly to the medium.

The trace metal mix used for the ASW consists of 11,65 μM FeCl$_3$·6H$_2$O (MW = 270.3 g/mol), 11.71 μM of Na$_2$EDTA·2H$_2$O (MW = 372.24 g/mol), 0.039 μM of CuSO$_4$·5H$_2$O (MW = 249.7 g/mol), 0.026 μM of Na$_2$MoO4·2H$_2$O (MW = 241.9 g/mol), 0.077 μM of ZnSO$_4$·7H$_2$O (MW = 287.54 g/mol), 0.042 μM of CoCl$_2$·6H$_2$O (MW = 237.93 g/mol), and 0.91 μM of MnCl$_2$·4H$_2$O

(MW = 197.9 g/mol). Vitamins were made from 0.002 μM of Biotin (MW = 244.32 g/mol), 0.0004 μM of B12 (MW = 1355.38 g/mol) and 0.30 μM of Thiamine-HCL (MW = 337.28 g/mol). The vitamins mix needed to be sterile filtered and kept in the dark.

METHOD DETAILS

Nucleic acid isolation and sequencing
Foraminifera lysis and disruption (including lysozyme buffer) were done via immersion in liquid nitrogen followed by pestle-crashing. Genomic DNA and total RNA were simultaneously isolated using the AllPrep DNA/RNA Micro Kit (QIAGEN) or innuPREP DNA/RNA Mini Kit (AnalytikJena). Transcriptome libraries of two consecutives years (2 independent biological replicates) were synthesized from cDNA and enrichment for eukaryotic transcripts. Libraries were prepared with TotalScript RNA-Seq kit (Epicenter®) with Oligo (dT) primers or NEBNext® Ultra RNA Library Prep Kit for Illumina® (NEB) with mRNA isolation performed with poly-A mRNA beads. Transcriptomics paired end libraries (2x 100 bp and 2x 125 bp) were sequenced on an Illumina HiSeq 2000 platform. Genomic libraries were prepared with the NEBNext® Ultra II DNA Library Prep Kit for Illumina® and were directly used for whole genome shotgun sequencing on an Illumina HiSeq 2000 platform (2x 100 bp). All samples were quantified and qualified using a Qubit® fluorometer (Invitrogen by Life Technologies™) and a Bioanalyzer & TapeStation (Agilent technology). Sequence data were deposited in NCBI (BioSamples SAMN07821823, SAMN07821824).

Validation of gene sequences
The gene sequence of the genes encoding NirK and Nor was confirmed by amplifying exon fragments using RT-PCR with cDNA synthesized with Oligo (dT) primers (NEB) as template. The reactions were performed with iTaq Universal SYBR® on the Bio-Rad CFX connect Real-Time System (Bio-Rad Laboratories). The reaction was composed of 10 μL iTaq Universal SYBR® master mix (Bio-Rad), 0.5 μL of 10 μM forward and reverse primers each (Table S3), 8 μL PCR-H_2O and 1 μL of poly-A cDNA template. RT-PCR cycling conditions were 3 min at 95°C (once), 39 cycles of 0:10 min at 95°C, 0:30 min at 60°C, followed by a melt curve from 58°C to 98°C (increment 0.2°C 0:50). Duplicates of one biological sample per gene and no template control (NTC) were run for each primer pair.

Taxonomy
Six foraminifera-specific hypervariability 18S rRNA gene loci were amplified as described by Pawlowski and Lecroq 2010 [41] and used to separate closely related species. The foraminiferal 18S rRNA gene sequence was amplified from 25–44 individuals using Phusion polymerase (NEB) and the primers 14F1 and B (Table S3). This yielded a fragment of about 1200 bp for the *Globobulimina* species. The PCR products were evaluated by electrophoresis, and positive samples were purified (GeneJet Gel extraction and DNA Cleanup Micro Kit; Thermo Scientific) for Sanger sequencing. Duplicates are shown in the 18S phylogenetic tree (Figure 1E). Additional 18S rRNA gene homologs were detected in NCBI public databases and the genomes and transcriptomes presented in the current study by BLAST [42]. All 18S rRNA gene sequences were aligned using MAFFT [43] 7.123b ('linsi' option), and a phylogeny was reconstructed using PhyML [44] ver. 20131022 ('-b 100 -m HKY85'; Figure 1E).

Foraminiferal denitrification rate
The foraminiferal denitrification rates were calculated from linear steady state gradients of N_2O in microtubes with 3–5 individuals (at least quadruplicates measurements per species). Nitrous oxide reductase was inhibited with acetylene, thereby altering the complete denitrification by making N_2O the final product [45]. The steady state diffusion fluxes in the tubes corresponding to the respiration rates were calculated by Fick's first law of diffusion: Equation 1: $J = -D \cdot dC/dx$ where J = flux; dC/dx = the measured concentration gradient; D = free diffusion coefficient of N_2O. All measurements were performed in a cooling room at a constant temperature of 9°C. However, since the temperature could not be continuously recorded inside the microchamber medium, sporadic temperature measurements were taken within a water bath next to the incubation chamber (average temperature 9.2 ± 0.1°C; N = 19). The production of N_2O was measured with a N_2O microsensor [46] (Table S1), following methods previously established for foraminifera [5, 7, 11, 22]. N_2O production microprofiles of ambient foraminifera were performed with freshly picked foraminifera directly after wet sieving and rinsing twice with nitrate free ASW (red sea salt) prior to the transfer to a microchamber.

Microscopic visualizations of tests and living individuals of *Globobulimina* spp.
Specimens of the two *Globobulimina* species were removed from sediment samples, washed with sterile artificial seawater, dehydrated in a graded ethanol series (70%, 80%, 90%, 96% and two times 100%; 15 min each), air-dried for 12 h in a desiccator and mounted on aluminum stubs (PLANO GmbH) using conductive and adhesive carbon pads (PLANO GmbH). Subsequently, the preparations were sputter-coated with a 10-nm-thick gold-palladium (80/20) layer using a high vacuum sputter coater Leica EM SCD500 (Leica Microsystems GmbH) and visualized with a Hitachi S-4800 field emission scanning electron microscope (Hitachi High-Technologies Corporation) at an acceleration voltage of 3 kV and an emission current of 10 mA applying a combination of the upper detector and the lower detector. Stereo micrographs of living specimens were created with a ZEISS SteREO Discovery.V8dissecting microscope mounted with a ZEISS colour camera.

Processing of sequencing data

Sequencing resulted in total in 1.3 billion paired-end reads. This includes transcriptome datasets of two consecutive years (SRA accessions: SRR6202056, SRR6202059–SRR6202078) as well as genome data of *Globobulimina* (SRA accessions: SRR6202052–SRR6202055, SRR6202057–SRR6202058*)*. Reads were quality-checked by FastQC ver. 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc; Aug 2016). Filtering and trimming of the reads was performed using Trimmomatic [47] ver. 0.36 (Parameters: ILLUMINACLIP:primers.fa:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:5 MINLEN:21; the file 'primers.fa' contains adaptor and contaminant sequences provided by Trimmomatic and FastQC). Processed paired-end reads of the transcriptome samples were assembled into contigs using SPAdes [48] ver. 3.9.1 ('–rna' option), which yielded 161,222 (termed 'GloT14', NCBI accession: GGCE00000000) and 906,588 (termed 'GloT15', NCBI accession: GGCD00000000) transcripts. The contigs shorter than 200 nucleotides and contamination identified by NCBI Transcriptome Shotgun Assembly (TSA) submission pipeline were excluded. Protein sequences were translated from transcripts as the longest open reading frame (ORF) using TransDecoder [49] ver. 3.0.1 ('-m 30' option). Transcript abundances of individual transcriptome datasets was determined as Transcripts Per Million (TPM) by the Trinity pipeline [49] 2.4.0 (Trinity script 'align_and_estimate_abundance.pl') via RSEM [50] ver. 1.2.30 and Bowtie [51] 2.1.0 using paired-end reads. Paired-end reads of *Globobulimina* spp. genome datasets were assembled using IDBA-UD [52] ver. 1.1.1 (parameters:–pre_correction–mink 20–maxk 120). Resulting contig sequences (termed 'GloG15') were classified into 241 genomic bins by MaxBin [53] ver. 2.2 (parameter: -min_contig_length 500) providing reads of individual sequencing samples separately. We assessed the quality of prokaryotic genomic bins, their completeness, coverage and first prediction of protein sequences on binned contigs with checkM [54] ver. 1.0.5. Thresholds of completeness $\geq$ 80% and contamination $\leq$ 15% were applied to classify bacterial bins as draft genomes (26 in total; Figure S2). The protein sequences obtained by checkM were combined with those predictions for unclassified contigs using MetaProdigal [55] ver. 2.6.2, and additional protein sequences were predicted based on genome mapping of transcriptome paired-end reads using HISAT2 [56] ver. 2.0.5 ('–non-deterministic' option) and BRAKER [57] ver. 1.9. Similar sequences were clustered with CD-HIT [58] ver. 4.6 ('-c 0.98' option) in order to reduce redundancy of the complete protein catalog. The final protein names consist of the contig IDs followed by the sequence positions covered by the CDS and an indicator for the forward (+) or reverse (-) strand. In case of multiple exons, individual regions are separated by comma. For a high resolution taxonomic classification of prokaryotic genomic bins, DIAMOND [59] tool 0.7.11.60 ('-k 10' option) similarity searches of proteins predicted via checkM were applied. The first best hit (by score) per protein sequence to the NCBI non-redundant database (NR; January 2017; *e*-value $\leq$ 1e-10) was obtained, and the corresponding NCBI taxonomy assignment was further used. Identical protein taxonomy assignments per genomic bin were counted and sorted accordingly. Beginning with the most abundant taxonomy, the lowest taxonomic rank was searched that was supported by > 50% of bin protein hits and accepted as taxonomic classification of the corresponding draft genome. Taxonomic assignments containing 'environmental samples' and the rank 'Cellular organisms' were not considered. Contigs that were not classified as bacterial or *Globobulimina* draft genome represent the unclassified metagenome associated with *Globobulimina* (NCBI accession: PJEL00000000). Six genomic bins (No. 179, 190–194) were classified as the *Globobulimina* draft genome (NCBI accession: PIVH00000000). All of them were annotated as eukaryotic, based on the previously described approach. Additionally, each of these bins covers more than 5% of transcriptome read pairs mapped to binned contigs, and their coverage over different samples is highly correlated (Pairwise Pearson correlations $\geq$ 0.99, FDR adjusted p values < 10^{-7}). Overall, 93.3% of transcriptome read pairs mapped to the *Globobulimina* draft genome. The corresponding assembly consists of 48,370 contigs with a GC-content of 34.25% and a N50 of 1,797 nucleotides. The assessment of genome completeness by Benchmarking Universal Single-Copy Orthologs [60] (BUSCO v3; lineage 'eukaryota') method was done using all 132,080 protein predictions for *Globobulimina* genome contigs and recovered 63.7% of BUSCOs completely, which shows that more than half of the genome is covered. We analyzed pooled samples of two species of the genus *Globobulimina* (*G. turgida* & *G. auriculata*) and therefore expected our sequencing results to contain sequences derived from both genomes. As BUSCOs represent universal single-copy genes, the presence of multiple BUSCOs of the same type indicates multiple organisms in the analysis (i.e., species heterogeneity within the dataset). In the described BUSCO analysis, 44.1% of recovered BUSCOs were inferred as multi-copy orthologs. However, this still includes duplicated predictions for the same gene locus that resulted by differing predictions based on prodigal or BRAKER. Focusing only on the 27,238 protein sequences predicted via BRAKER reduced the species heterogeneity such that only 7.12% of complete BUSCOs were found multiple times. From this, we conclude that the *Globobulimina* draft genome is enriched for a single species, either *G. turgida* or *G. auriculata*. The BUSCO analysis of the transcriptome assemblies revealed a higher level of heterogeneity where 37.5% (GloT14) or 86% (GloT15) of the complete orthologs represented multiple copies.

Gene identification and phylogenies

To identify homologs of enzymes in the denitrification pathway, we collected query protein sequences of known enzymes from Swiss-Prot [61] and the literature (Table S4). The search for denitrification enzymes homologs in our (and external) datasets was performed with BLASTP [42] applying an *e*-value $\leq$ 1e-5 threshold. Protein sequences of hits with query coverage $\geq$ 40% and sequence identity $\geq$ 20% were extracted to obtain a first set of homologs. With this preliminary set, we reiterated the search in NR and RefSeq 79 database [62] using GHOSTZ [63] 1.0.0 applying a threshold of *e*-value $\leq$ 1e-5. Hit sequences were obtained and clustered with CD-HIT 4.6 (option: '-c 0.98') to reduce sequence redundancy. Protein sequences were aligned with MAFFT ('linsi' option), and phylogenetic trees were reconstructed with FastTree [64] 2.1.7. Putative paralogs were excluded as following: i) the trees were rooted using outgroups (See Table S4), if available, and/or the Minimal Ancestor Deviation (MAD) method [65]; ii) the resulting root separated the tree into two main clades. Only one of these clades, containing orthologs, was retained, which was defined by sequence

annotation and sequence representation of proteins from literature (Table S4). Further sequences that exhibit long branches or increase redundancy of the phylogenetic information were as well excluded. For a complete list of homologous sequences that were excluded from the final phylogenies see Table S5. The refined sequence set was used for the reconstruction of phylogenetic trees using PhyML ('-m LG -b 100' options). The resulting trees were rooted using outgroups or MAD. Clades of protein sequences encoded by the *Globobulimina* spp. were defined by representation in the *Globobulimina* genome data and transcriptome assemblies. Statistical support of alternative topologies was estimated using CONSEL [66] ver. 1.20. Homologs for comparison of functional domains and conserved amino acid residues of copper binding and catalytic sites were chosen based on available literature [25, 67, 68] (Table S4). For each enzyme, sequences were aligned using MAFFT ('linsi' option). The protein domains (shown in Data S1 and Figure S3) were determined using the CD-Search webserver at NCBI [69] (Aug–Nov 2017), and regions spanned by hits to specific protein domains were extracted. Pairwise local sequence identities were obtained using 'water' tool in the EMBOSS [70] package.

In the search for homologs of *Globobulimina* NirK, Nor and Nrt in further foraminifera species and related taxa, *Globobulimina* protein sequences were used as queries for BLAST (with a threshold of e-value ≤ 1e-5). Datasets in use include NCBI genome data (*Reticulomyxa filosa*, GCA_000512085.1; *Astrammina rara*, GCA_000211355.2), NCBI transcriptomic reads (*Nonionella* sp., SRR2003403; *Bulimina marginata*, SRR2003397; *Brizalina* sp., SRR2003388; *Ammonia* sp., SRR2003283) and MMETSP [71] transcriptomes (*Rosalina* sp., MMETSP0190; *Sorites* sp., MMETSP0191; *Ammonia* sp., MMETSP1384; *Elphidium margaritaceum*, MMETSP1385). Nrt homologs with high sequence identity (≥50%) were identified in protein sequences of MMETSP datasets and in sequencing reads of NCBI transcriptomes. Protein sequences of hits were used without modification, and the obtained sequencing reads were assembled into contigs using CAP3 [72]. Final Nrt protein sequences were determined by predicting longest ORF using TransDecoder on contigs and remaining unassembled reads. Hits for *Globobulimina* NirK and Nor with high sequence identity (≥50%) in the same dataset were only observed for *Brizalina* sp. sequencing reads. Further 24 reads of *Rosalina* were covered by hits to NirK using the same cut-off. Read sequences of individual hits to NirK and Nor were obtained and individually assembled using CAP3. *Rosalina* sp. associated reads were short, and the parameters recommended for assembly of short sequences ('-i 30 -j 31 -o 18 -s 300' options) were applied. Resulting contigs were translated into the reading frame that exhibited most sequence similarity to corresponding proteins. The sequence similarity of *Brizalina* sp. sequences to NirK was limited to a short region on one read (accession: SRR2003388.68811). The homologous region was extracted and translated. If represented, stop codons were replaced by 'X's. For protein sequences of gene fragments obtained see Table S2. Predicted protein sequences that did not show sequence similarity to corresponding query protein sequences were discarded. Protein sequences obtained in this way were added to the alignments of the denitrification proteins described before (MAFFT; '–addfragments' option) for subsequent phylogenetic reconstruction via PhyML (Data S2I–S2K).

QUANTIFICATION AND STATISTICAL ANALYSIS

The standard deviation (σ_{sd}), standard of the mean (σ_{SEM}) and the number of individuals (N) for the denitrification measurements are presented in Table S1. An Approximate Unbiased (AU) test was used to compare optimal phylogenetic trees to alternative topologies without application of further testing procedure. Alternative topologies were inferred by constraining individual clades for tree reconstruction and their tree likelihoods were compared to the unconstrained phylogeny using AU-test via CONSEL (See Figure 2, eukaryotic monophyly of clade I: *p-value* 0.352; See Figure 3, Grouping of the *Globobulimina* clade with other eukaryotes: *p-value* 0.043, The *Globobulimina* clade groups independently from NirK class II clade: *p-value* 0.373, Grouping of the *Globobulimina* clade with a eukaryote-containing subclade of NirK class II: *p-value* 0.581). An α of 0.05 was applied to determine a significant difference of the alternative topology in contrast to the optimal tree.

DATA AND SOFTWARE AVAILABILITY

Data availability
Sequencing reads are available from the single read archive (SRA) accessions SRR6202052 to SRR6202078. Transcriptome assemblies were deposited at the transcriptome sequencing archive (TSA) accessions GGCE00000000 and GGCD00000000. The genome sequencing assembly is available at NCBI with the accessions PIVH00000000 to PIWH00000000 representing draft genomes and PJEL00000000 the unassigned contigs. Individually amplified 18S rRNA gene sequences of *G. turgida* and *G. auriculata* were submitted to GenBank (accessions: MG800664 to MG800667). All other information on accessing data analyzed in this study is included in the manuscript or in the supplemental information. All additional datasets generated during the current study are available from the corresponding authors upon request.