Supplementary Information

for

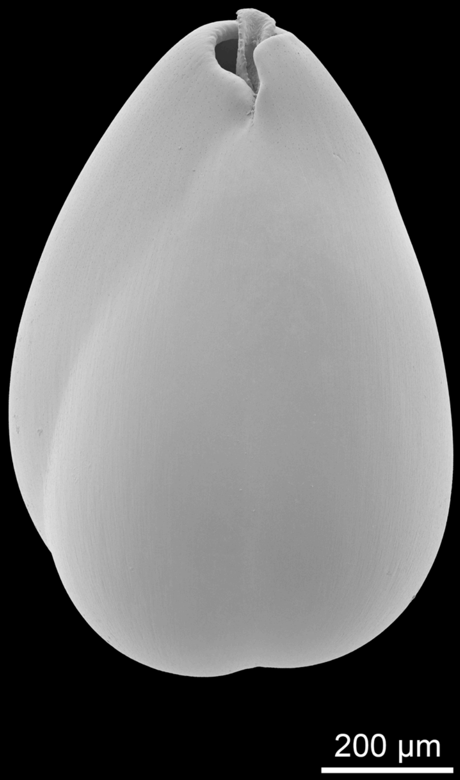
Denitrification in foraminifera has an ancient origin and is complemented by associated bacteria

Christian Woehle, Alexandra-Sophie Roy, Nicolaas Glock, Jan Michels, Tanita Wein, Julia Weissenbach, Dennis Romero, Claas Hiebenthal, Stanislav N. Gorb, Joachim Schönfeld, Tal Dagan

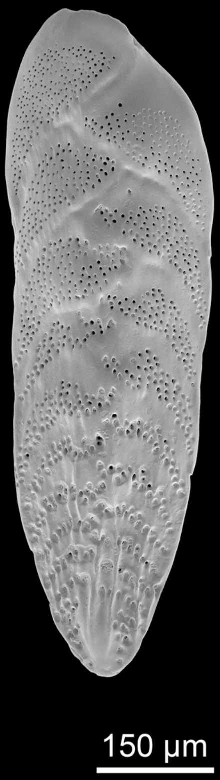
Supplementary Figures S1 – S4

Titles: Dataset 1 – 8

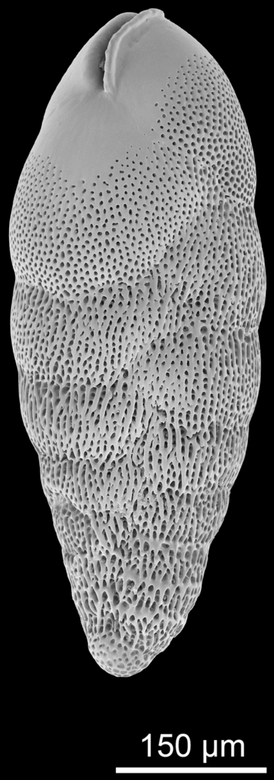
Globobulimina pacifica



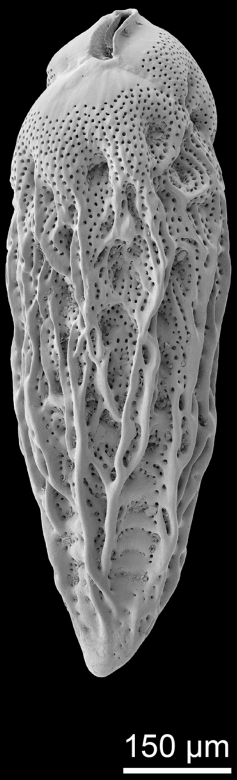
Bolivina spissa



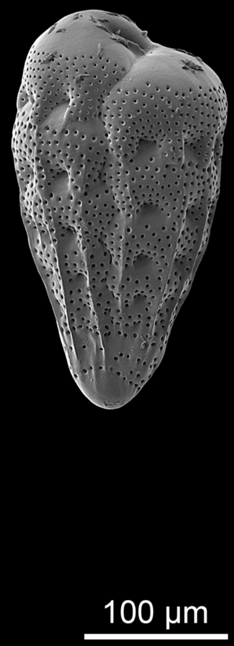
Bolivina seminuda



Bolivina plicata

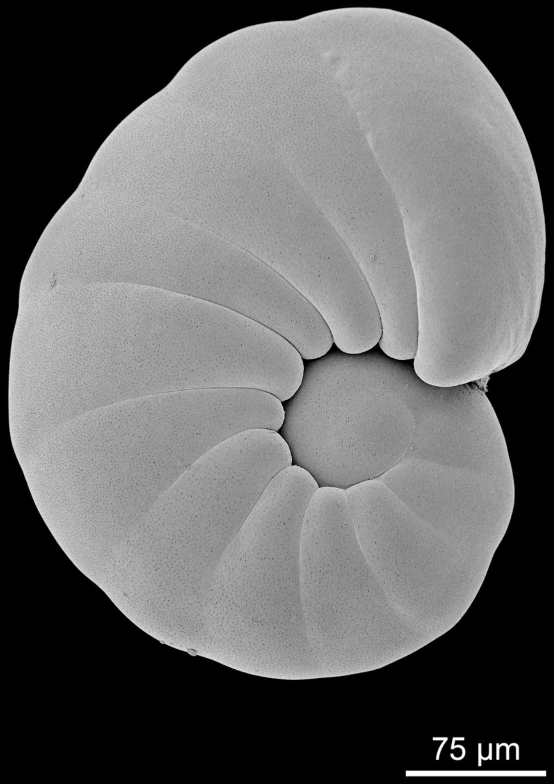


Bolivina costata

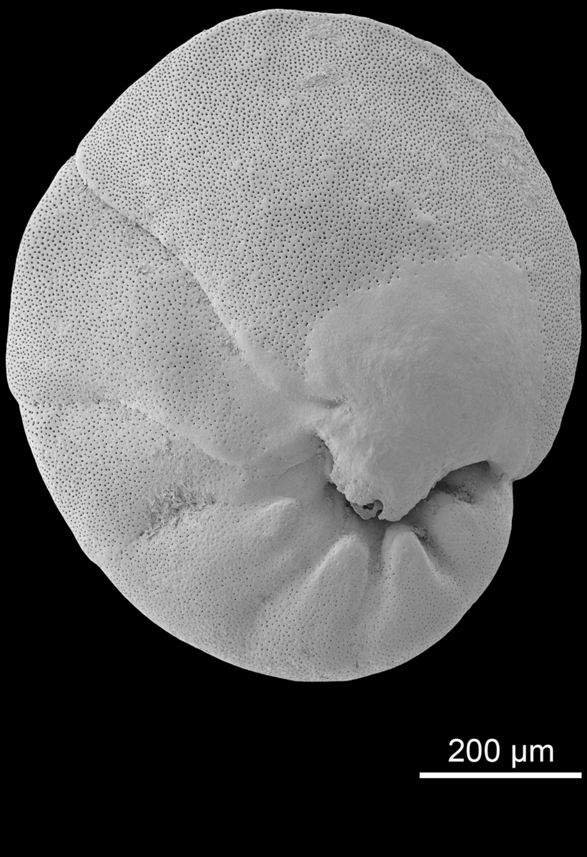


Uvigerina striata

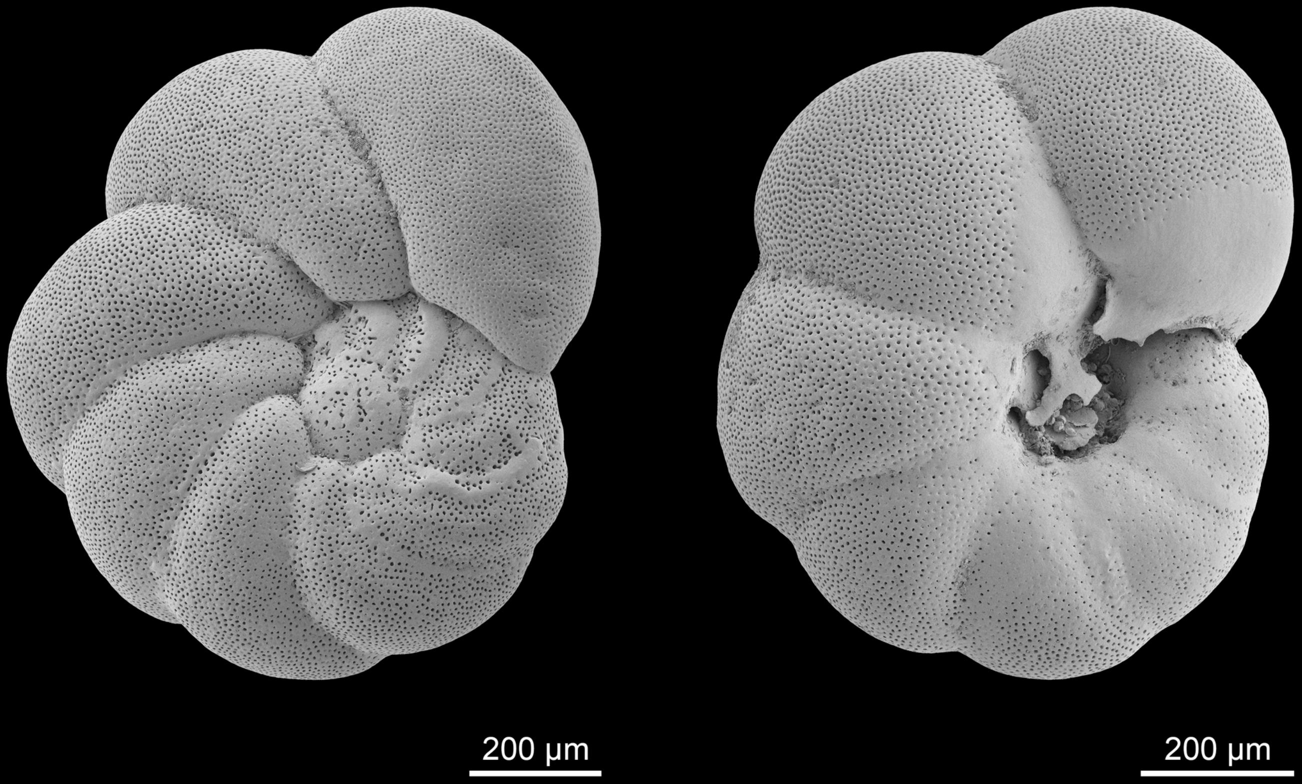
Nonionella auris



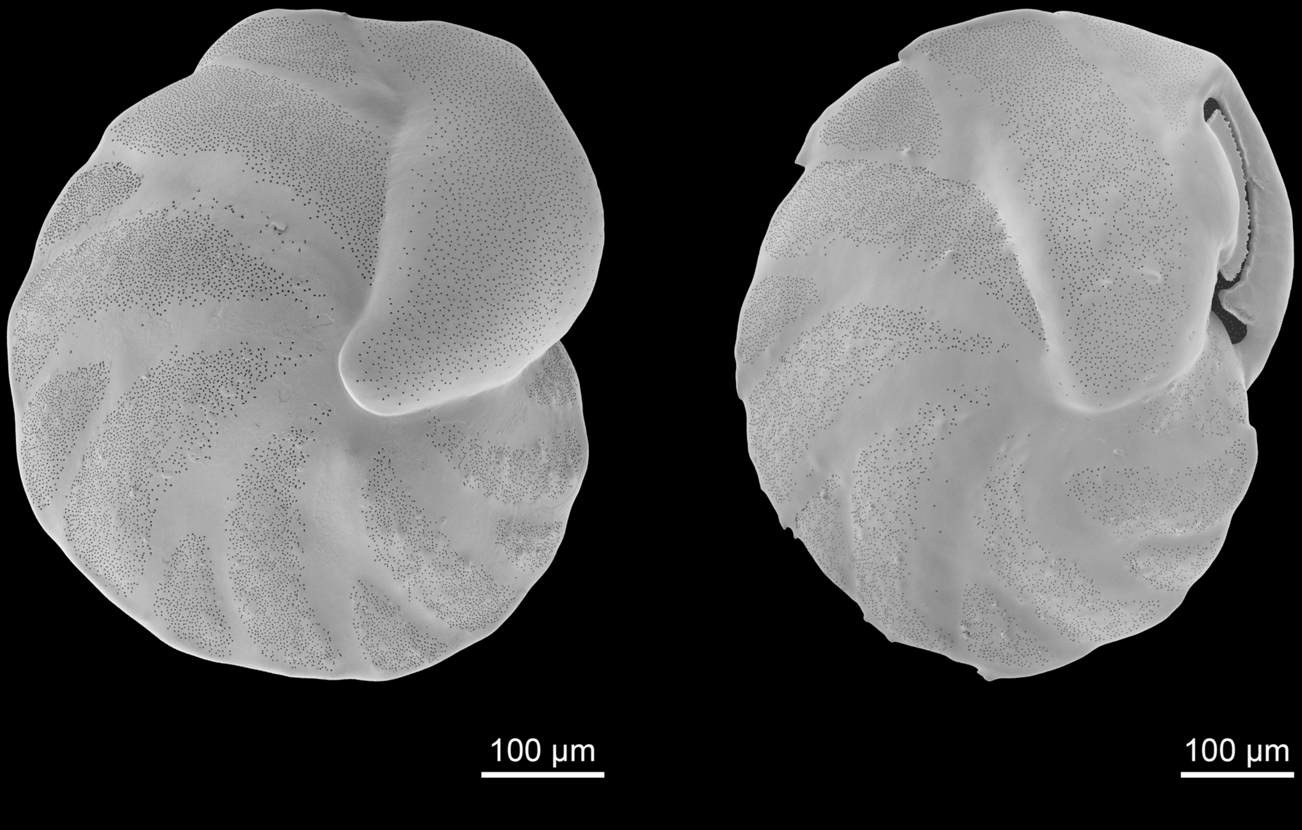
Cancris carmenensis



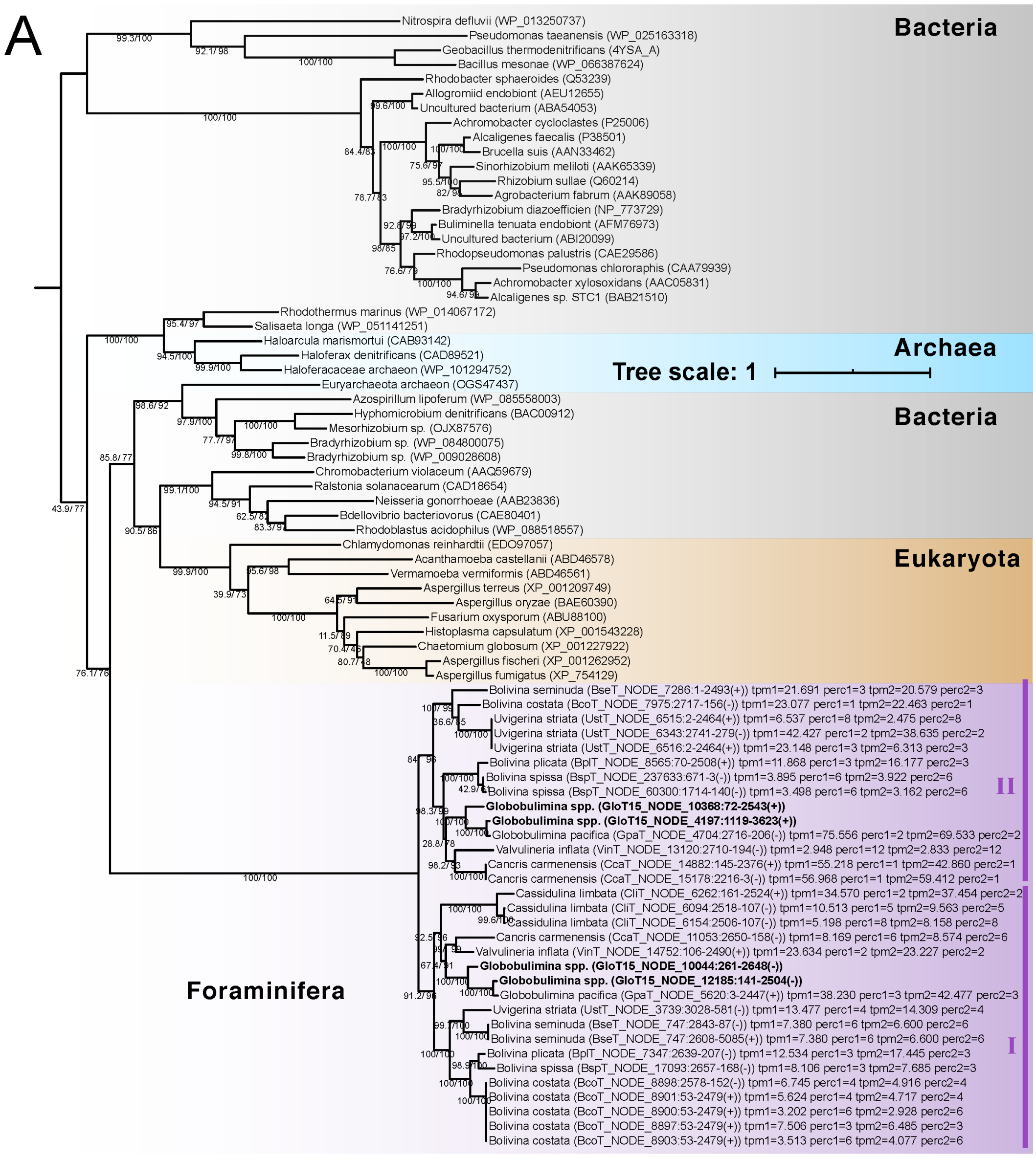
Valvulineria inflata

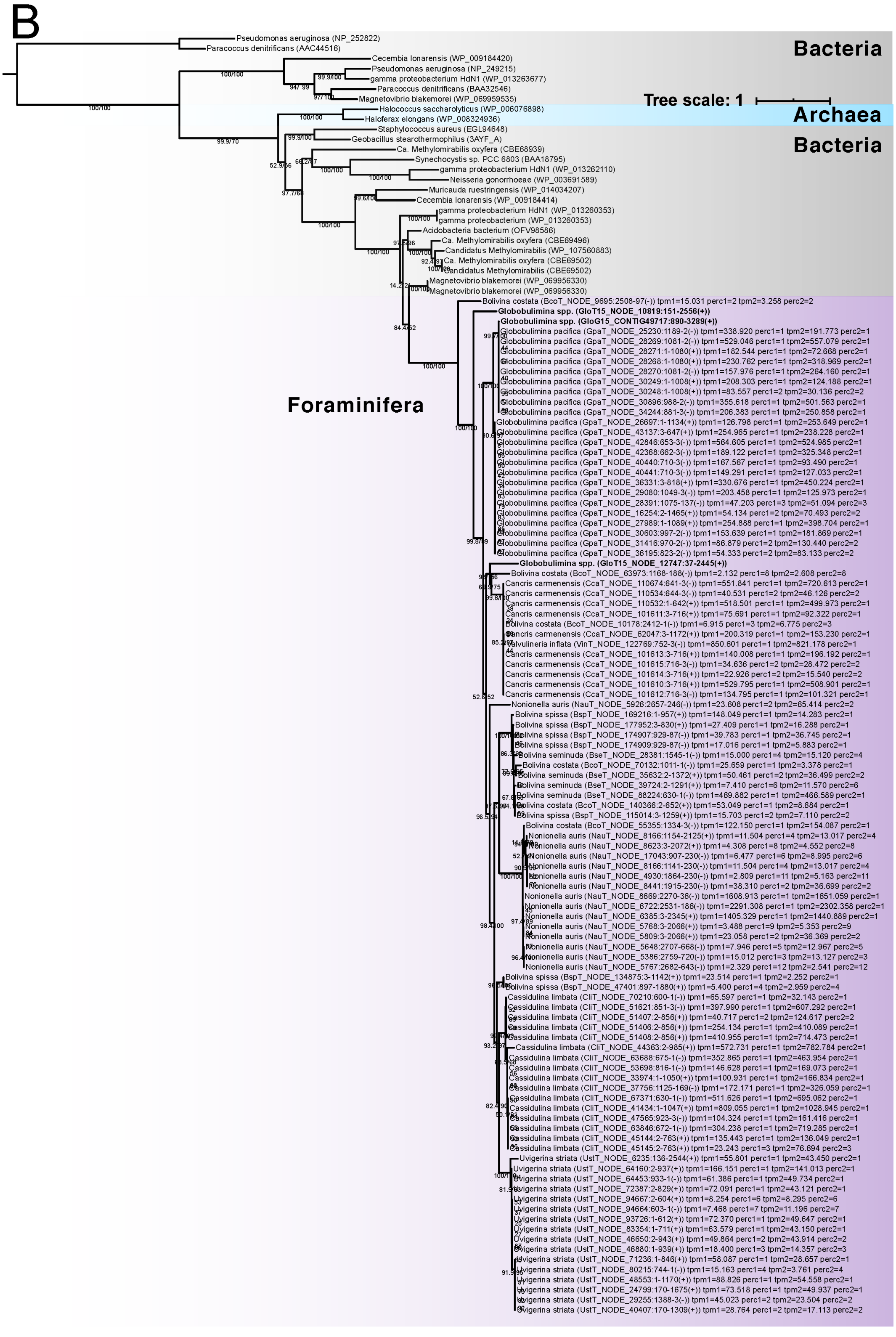


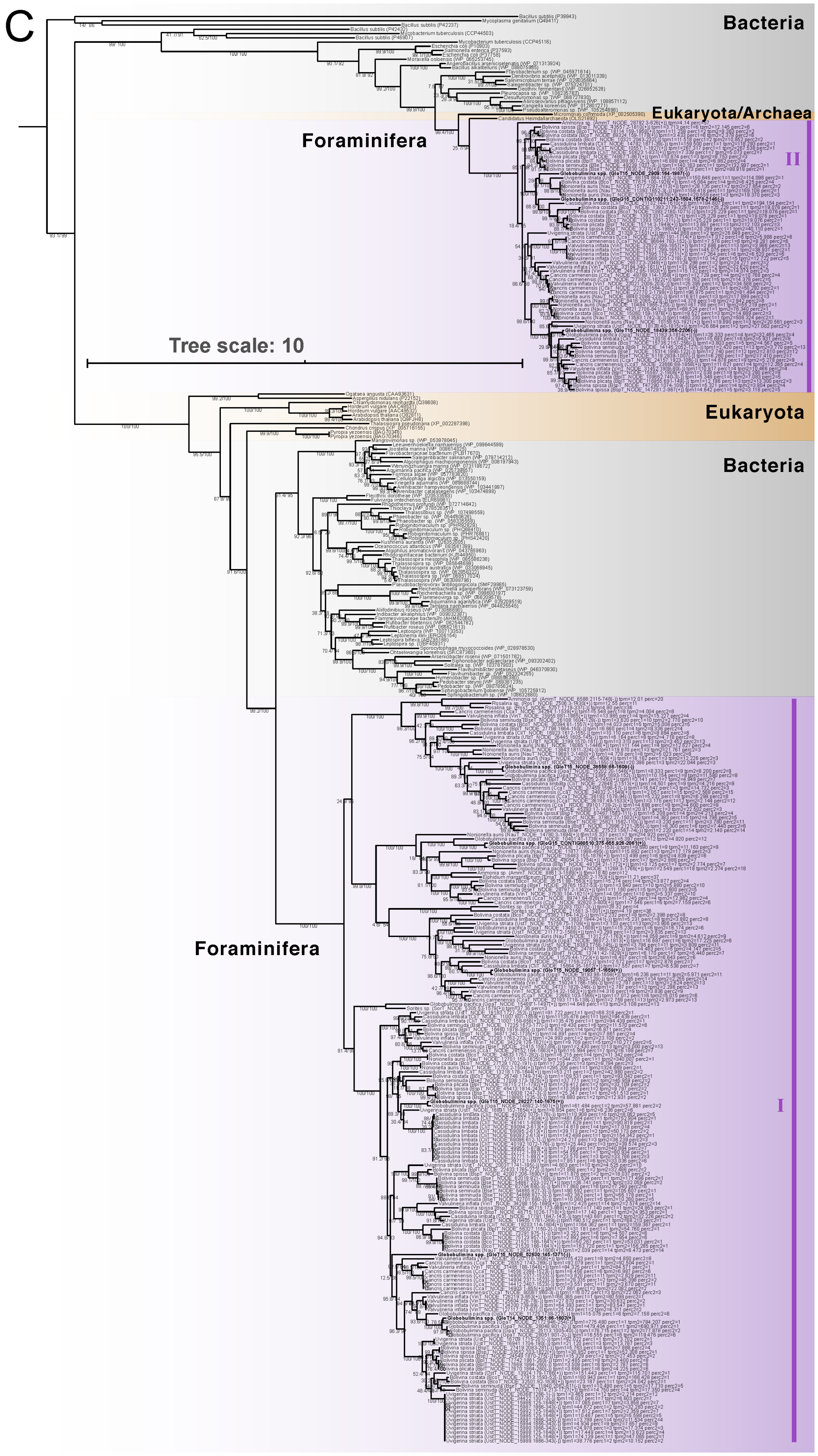
Cassidulina limbata



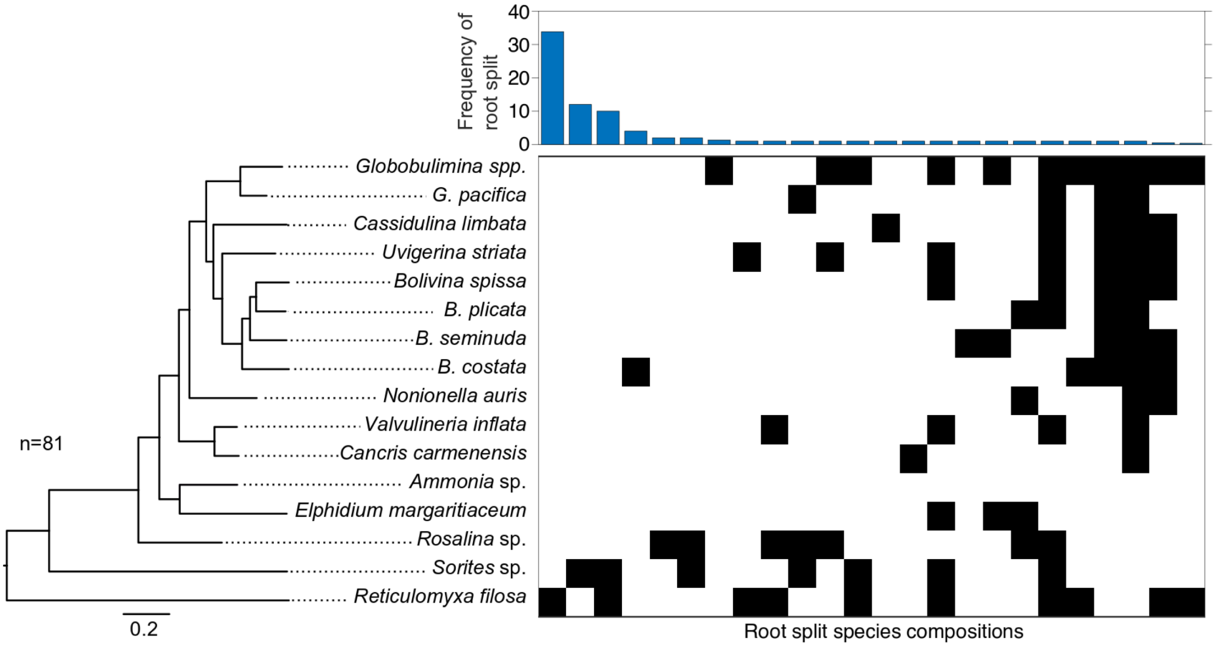
Suppl. Figure S1. Scanning electron micrographs of sampled species. Detailed views of the graphs shown in Figure 1A.



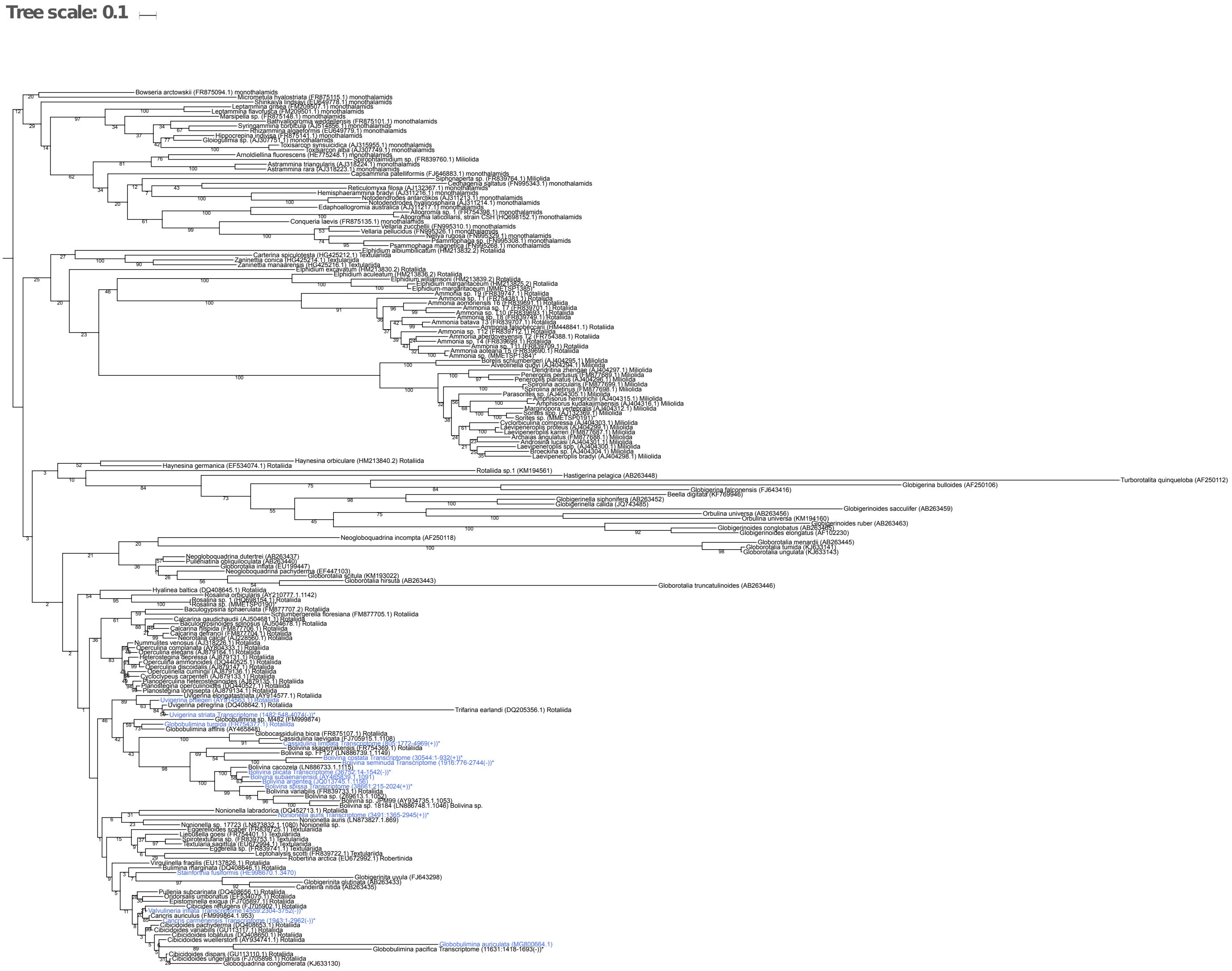

B
Bacteria
Tree scale: 1
Archaea
Bacteria
Foraminifera
Pseudomonas aeruginosa (NP_252822)
Paracoccus denitrificans (AAC44516)
Cecembia lonarensis (WP_009184420)
Pseudomonas aeruginosa (NP_249215)
gamma proteobacterium HdN1 (WP_013263677)
Paracoccus denitrificans (BAA32546)
Magnetovibrio blakemorei (WP_069959535)
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)
100/100
100/100
94/ 99
99.9/100
97/100
100/100
99.9/70
99.9/100
52.9/66
66.2/87
100/100
100/100
97.7/68
99.6/100
100/100
100/100
97.5/96
100/100
100/100
92.4/97
100/100
14.2/21
100/100
84.4/62
100/100
100/100
99.8/89
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=65.414 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=466.589 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=146.628 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=161.416 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 (CliT_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
99.8/100
85.2/97
52.6/52
100/100
86.3/92
67.8/63
96.5/94
100/100
97.4/89
96.4/100
98.4/100
93.2/97
82.4/90
50.1/81
100/100
91.9/95


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.

Dataset 1. Sequencing statistics for metagenomics & transcriptomics.

Dataset 2. List of foraminiferal homologs of denitrification proteins remaining after applying cutoffs.

Dataset 3. List of foraminiferal homologs of denitrification proteins removed by cutoffs.

Dataset 4. List of marker proteins used for phylogenetic reconstructions.

Dataset 5. 18S sequences including metadata for phylogenetic reconstructions.

Dataset 6. Proportions of family rank classification of foraminifera-associated microbial communities.

Dataset 7. High quality genomic bins of foraminifera-associated species communities.

Dataset 8. Metabolic classification of Desulfobacteracea genomes.