Supplementary Information

A novel enrichment culture highlights core features of microbial networks contributing to autotrophic Fe(II) oxidation coupled to nitrate reduction

Yu-Ming Huang[a, b], Daniel Straub[a, c], Andreas Kappler[b], Nicole Smith[a], Nia Blackwell[a] and Sara Kleindienst[a,#]

[a] Microbial Ecology, Center for Applied Geoscience, University of Tuebingen, 72076, Tuebingen, Germany

[b] Geomicrobiology, Center for Applied Geoscience, University of Tuebingen, 72076, Tuebingen, Germany

[c] Quantitative Biology Center (QBiC), University of Tuebingen, 72076, Germany

[#]**Corresponding Author:**

Jun.-Prof. Dr. Sara Kleindienst

Microbial Ecology, Center for Applied Geosciences

Eberhard-Karls-University of Tuebingen

Schnarrenbergstrasse 94-96

Tuebingen, 72076, Germany

Tel: +49 -7071-29-74701

E-mail: sara.kleindienst@uni-tuebingen.de

Supplemental Experimental Procedures

Supplemental Tables 1-4

Supplemental Figures 1-6

Supplemental Data 1-4 (individual files)

Supplemental Experimental Procedures

Sampling for cultivation

Culture BP originated from sediments of a freshwater pond (latitude 53°06'36.7"N and longitude 8°50'48.6"E) at the Max Planck Institute for Marine Microbiology, located in Bremen, Germany, sampled in September 2015. Sediments from ca. 10 cm depth were taken and stored in Schott bottles (filled up completely with half sediment and half pond water) at 4°C. The sediment was transferred into 25 mL mineral medium [Hegler et al., 2008; Blothe and Roden, 2009] with 10 mM $FeCl_2$ and 4 mM $NaNO_3$ in 58 mL serum bottles and incubated in the dark at 28°C starting in February 2016. The culture was first transferred in March 2016 with the same medium and additives under autotrophic conditions.

Cell counts

For the cell counts under autotrophic growth conditions, 200 µL of samples from the enrichment culture BP were taken and cells were fixed with 10 µL of 20% (w/v) paraformaldehyde (PFA) (Sigma-Aldrich). In addition, the sample was treated with 590 µL oxalate solution (oxic; 2.8 g ammonium oxalate, 1.5 g oxalic acid, total volume 100 mL, pH 3, filtered sterile, Sigma-Aldrich) and 200 µL ferrous ethylenediammonium sulfate (Fe-EDAS, 100 mM, anoxic, filtered into an anoxic autoclaved serum bottle, Sigma-Aldrich) to dissolve the Fe minerals prior to the cell counts (personal communication with Stefanie Becker). For the heterotrophic conditions, 1 mL of sample of the enrichment culture BP was taken and the cells were fixed with 50 µL of 20% PFA. The cell samples of heterotrophic conditions were diluted to the range of 100-1,300 cells/µL prior to the measurement. For flow cytometry, cells were stained with BacLight Green stain (Thermo Fisher Scientific, 1 mM stain/1mL sample). Cell numbers were determined with an Attune NxT Flow Cytometer (Thermo Fisher Scientific) with the setting adapted from Schmidt et al. [2020]. The results were reported as average from the measurements conducted in triplicates.

Experimental setup for meta'omics

The metagenome sample was obtained in 2019, at the 3rd day of incubation under autotrophic conditions with 10 mM of Fe(II) and 4 mM of nitrate in a 22 mM bicarbonate-buffered medium.

For the metatranscriptomics and metaproteomics analysis, conducted in 2020, both autotrophic and heterotrophic conditions were used with biological triplicates, respectively. Under autotrophic conditions, within 4 days on average 80% of Fe(II) was oxidized and 50% of nitrate was reduced. Therefore, samples under autotrophic conditions were taken at the 2nd day (remaining Fe(II) and NO_3^- was 7.1 mM and 3.7 mM, respectively; 2.91×10^6 cells/mL). Under heterotrophic conditions, in a first step a pre-culture was grown for two transfers with 10% and 1% (vol/vol) inoculum for 24 hours each, allowing more than seven generations under pre-culture conditions, considering an average doubling time of 3 hours, to avoid carryover of signals from gene and protein expression under autotrophic conditions. In a second step, the 3rd transfer under heterotrophic conditions with 1% (vol/vol) inoculum was used for the experimental setup: within 37 hours, 66% of acetate was oxidized and 96% of nitrate was reduced. Therefore, samples of the cultures grown under heterotrophic conditions were taken after approximately 18 hours. (remaining acetate, NO_3^- and NO_2^- was 3.94 mM, 2.69 mM and 0.82 mM, respectively; 2.78×10^6 cells/mL).

Biomass sampling for meta'omics

At the sampling time points, biomass of culture BP with total cell numbers ranging from 10^8 (for DNA/RNA extraction) to 10^9 cells (for protein extraction) was collected under sterile conditions on cellulose filters (EMD Millipore S-Pak mixed cellulose ester sterile filter membrane, 0.22 µm pore size, 47mm filter diameter; Millipore) using vacuum filtration. The filters were cut into pieces under sterile conditions and either stored in 15 mL falcon tubes at -80°C before proceeding with DNA and RNA extractions or stored in 50 mL falcon tubes at -80°C before proceeding with the protein extractions.

DNA/RNA co-extraction for metagenomics and metatranscriptomics

DNA/RNA co-extraction was done according to the protocol of Lueders et al. [2004] with the following modifications: two tubes of "MP Bio Lysis Matrix E" beads were added into a 15 mL falcon tube, including the filter pieces with the collected biomass. To disrupt the cells, 3.75 mL PB buffer (with 112.87 mM Na_2HPO_4 and 7.12 mM NaH_2PO_4) and 1.25 mL TNS buffer (with 500 mM Tris-HCL, 100 mM NaCl and 10% w/v SDS) were added, followed by bead beating for 4

minutes on the vortex adapter (maximum power) at room temperature (RT). The following centrifugation steps were all at maximum speed (7,197 ×g; Eppendorf Centrifuge 5430 with Rotor F35-6-30) at 4°C. All transfer steps were done on ice. The samples were centrifuged twice for 15 minutes and transferred to a new 15 mL falcon tube in between centrifugation steps, to obtain a clear supernatant. Then, the supernatant was split into new sterile 2 mL tubes (1 mL per tube) to proceed with phenol-chloroform-isoamyl alcohol and chloroform-isoamyl alcohol extractions according to Lueders et al. [2004]. Subsequently, all aqueous phases were pooled again into a new 15 mL tube for polyethylene glycol (with 30% w/v polyethylene glycol 6000 and 1.6 M NaCl) precipitation overnight. From the ethanol-washing step on, the extraction was done under a clean bench. The ethanol was removed carefully with a pipette with filter tips and the pellet was dried at RT for ca. 10 minutes. The DNA/RNA pellet was dissolved in 20-50 µl DEPC-treated water with Invitrogen™ Ambion™ RNase inhibitor (40 Unit/µl, 1 µl RNase inhibitor/ 40 µl), at RT for 30 min. All samples were stored at -80°C before sequencing.

Protein extraction for metaproteomics

Protein extraction was performed according to the "Protein extraction method B" from Spät et al. [2015] with the following modifications: 5 mL and 2 mL lysis buffer were added to the samples from autotrophic and heterotrophic conditions, respectively, to dissolve cell pellets. The samples were incubated for 10 minutes at 95°C in a water bath, vortexed briefly and chilled on ice with 2 minute-intervals in total 5 times and, subsequently, sonicated on ice for 30 seconds with an ultrasonic homogenizer (Bandelin Sonopuls) at output control 4 and 40% duty cycle. The lysate was centrifuged at 7,197 ×g for 1 minute at RT and the supernatant was then transferred into several 1.5 mL Eppendorf tubes and centrifuged at 20,817 ×g (Eppendorf Centrifuge 5430 with Rotor 30x1,5/2,0 mL) for 10 minutes at RT. Samples were pooled again into sterile 50 mL solvent-resistant tubes. After the 8:1 acetone:methanol precipitation and incubation step, the precipitate was washed with 5 mL ice-cold 80% (v/v) acetone in water and centrifuged at 7,197 ×g for 5 minutes at 4°C. The protein pellets were air-dried at RT and later dissolved in urea buffer and stored at -20°C.

Sampling, DNA/RNA co-extraction and cDNA synthesis for *in situ* analysis

In order to investigate the relative abundance of microorganisms in environmental samples, six cores were obtained in 2017. DNA and RNA were co-extracted using the method published by Griffiths et al. [2000] with the following modifications. At the first step, we used 1 g of sediment wet sample in the tubes of "MP Bio Lysis Matrix E" beads, followed by bead beating for 10 minutes on the vortex adapter (maximum power) at RT. At precipitation step, we used 2 volumes of 10% (wt/vol) polyethylene glycol 8000 with 1.2 M NaCl.
In order to synthesize cDNA from the extracted RNA samples, DNA digestion was proceeded by following a rigorous DNase treatment of the TURBO DNA-free™ kit (Invitrogen, Thermo Fisher Scientific) and reverse transcription was conducted following the user manual of SuperScript™ III Reverse Transcriptase (Invitrogen, Thermo Fisher Scientific). All centrifugation steps were performed with 16,100 ×g in an Eppendorf Centrifuge 5430 with Rotor FA-45-30.

Short-read 16S rRNA (gene) amplicon sequencing

PCR on DNA and cDNA with universal primers GTGYCAGCMGCCGCGGTAA [Parada et al., 2016] and GGACTACNVGGGTWTCTAAT [Apprill et al., 2015] fused to Illumina adapters was performed with cycling conditions as follows: 95°C for 3 min, 24 or 30 cycles of 95°C for 30 s, 55°C for 30 s, and 75°C for 30 s, and it was followed by a final elongation step at 72°C for 3 min. The quality of the purified amplicons was determined using agarose gel electrophoresis. Subsequent library preparation steps and sequencing steps of the 16S rRNA (gene) amplicons were performed on an Illumina MiSeq sequencing system (Illumina, San Diego, CA, USA) using the 2 × 250 bp MiSeq Reagent Kit v2 by Microsynth AG (Balgach, Switzerland). From 9,852 to 224,723 (average 95,455) read pairs were generated per sample in two separate sequencing runs on the same MiSeq machine, resulting in total in 6.7 million read pairs. Quality control, reconstruction of 16S rRNA (gene) sequences and taxonomic annotation was performed with nf-core/ampliseq v1.1.0 [Ewels et al., 2020; Straub et al., 2020] with Nextflow v20.10.0 [Di Tommaso et al., 2017] using containerized software with singularity v3.4.2 [Kurtzer et al., 2017]. Data from the two sequencing runs were treated initially separately by the pipeline using the option "mulipleSequencingRuns" and ASV tables were merged. Primers were trimmed, and untrimmed sequences were discarded (one sample 71%, others < 30%, on average 8.6%) with Cutadapt

v1.16 [Martin, 2011]. Adapter and primer-free sequences were imported into QIIME2 version 2018.06 [Bolyen et al., 2019], their quality was checked with demux (https://github.com/qiime2/q2-demux), and they were processed with DADA2 version 1.6.0 [Callahan et al., 2016] to eliminate PhiX contamination, trim reads (position 230 in forward reads and 170 in reverse reads), correct errors, merge read pairs, and remove PCR chimeras; ultimately, 30,936 amplicon sequencing variants (ASVs) were obtained across all samples. Alpha rarefaction curves were produced with the QIIME2 diversity alpha-rarefaction plugin, which indicated that the richness of the samples had been fully observed. A Naive Bayes classifier was fitted with 16S rRNA (gene) sequences extracted with the PCR primer sequences from the QIIME compatible, 99%-identity clustered SILVA v132 database [Pruesse et al., 2007]. ASVs were classified by taxon using the fitted classifier [Bokulich et al., 2018]. 273 ASVs that classified as chloroplasts or mitochondria were removed, totalling to < 21% (average 1.2%) relative abundance per sample, and the remaining ASVs had their abundances extracted by feature-table (https://github.com/qiime2/q2-feature-table).

Long-read 16S rRNA gene amplicon sequencing

DNA from culture BP was amplified in two rounds of amplification. At the first round, the KAPA HiFi Hot Start ReadyMix PCR Kit (KAPA BioSystems, Cape Town, South Africa) was used with primers universal for bacterial 16S rRNA gene, tailed with PacBio universal sequencing adapters (universal tags) and 5' amino modifiers (27F gcagtcgaacatgtagctgactcaggtcacAGRGTTYGATYMTGGCTCAG, 1492R tggatcacttgtgcaagcatcacatcgtagRGYTACCTTGTTACGACTT) (Biomers.net, Ulm, Germany) to amplify the long-read 16S rRNA gene from the genomic DNA extracted for culture BP. The first PCR amplification program was performed with 26 cycles: initial denaturation at 95°C for 3 minutes, denaturation at 95°C for 30 seconds, annealing at 57°C for 30 seconds and extension at 72°C for 60 seconds. Amplicons were purified with the QIAquick PCR purification kit (QIAGEN, Hilden, Germany) according to the user's manual. The second PCR was performed using the PacBio Barcoded Universal F/R Primers Plate – 96 (Pacific Biosciences, California, USA) with KAPA HiFi ReadyMix PCR Kit, followed by AMPure PB bead kit (PacBio biosciences, California, USA) purification according to the user manual. The second PCR amplification program was

performed with 20 cycles: denaturation at 95°C for 30 seconds, annealing at 57°C for 30 seconds and extension at 72°C for 60 seconds. The quality and quantity of PCR products were checked using an Agilent 2100 Bioanalyzer System (Agilent, California, USA) after both PCR rounds. SMRTbell Template Prep Kit (PacBio biosciences, California, USA) with user manual instructions was used for SMRTbell library preparation [Franzén et al., 2015]. Subsequently, 20,635 circular consensus sequencing reads were analyzed with DADA2 v1.10.0 [Callahan et al., 2016; Callahan et al., 2019] in R v 3.5.1 [Team, 2018] by sequentially orienting reads and removing primers, filtering (no ambiguous nucleotides and maximum 2 expected errors) and trimming (1000 bp to 1600 bp read length; leading to 11,978 sequences), dereplicating sequences (7,161 unique sequences), learning error rates, removing bimera de novo and finally assigning taxonomy to the detected sequences based on SILVA v132 [Callahan, 2018]. Lastly, 25 ASVs with 7,831 total counts were obtained.

Metagenome sequencing, assembly and annotation, analysis of essential single-copy genes, taxonomic binning and draft genome recovery

1 µg of DNA was used for library preparation with the TruSeq DNA PCR-Free Kit from Illumina without modifications and libraries were sequenced on the Illumina NovaSeq 6000 platform to generate 91,056,140 paired-end (2 × 150-bp) reads (27.48 giga base pairs) by CeGaT, Tuebingen, Germany. Raw read quality control, assembly, metagenome assembled genome binning and taxonomic annotation was performed with nf-core/mag v1.0.0 (https://nf-co.re/mag, DOI: 10.5281/zenodo.3589528) [Ewels et al., 2020] with Nextflow v20.04.1 [Di Tommaso et al., 2017] using containerized software with singularity v3.0.3 [Kurtzer et al., 2017]. Read quality was assessed with FastQC v0.11.8 [Andrews, 2010], quality filtering and Illumina adapter removal was performed with fastp v0.20.0 [Chen et al., 2018], and reads mapped with Bowtie2 v2.3.5 [Langmead and Salzberg, 2012] to the PhiX genome (Enterobacteria phage WA11, GCA_002596845.1, ASM259684v1) were removed. Finally, 90,303,622 processed read pairs (27.25 giga base pairs) were assembled with MEGAHIT v1.2.7 [Li et al., 2015] and the assembly was evaluated with QUAST v5.0.2 [Gurevich et al., 2013], 99.35% reads were represented in the assembly. Metagenome assembled genomes (MAGs) were binned with MetaBAT2 v2.13 [Kang et al., 2019] aided by the sequencing depth, checked for their completeness and contamination

with BUSCO v3.0.2 [Waterhouse et al., 2018] using 148 near-universal single-copy orthologues (http://busco.ezlab.org/v3/datasets/bacteria_odb9.tar.gz) selected from OrthoDB v9 [Zdobnov et al., 2017], and summary statistics were obtained with QUAST for each MAG.

Mapping of RNA sequences and differential RNA abundance analysis

Nf-core/rnaseq v1.4.2 (https://nf-co.re/rnaseq, DOI: 10.5281/zenodo.3503887) [Ewels et al., 2020] and its containerized software was used with singularity v3.0.3 [Kurtzer et al., 2017] and executed with Nextflow v20.04.1 [Di Tommaso et al., 2017]. The pipeline performed quality checks with FastQC v0.11.8 [Andrews, 2010], removed 0.1% to 3.7% base pairs per sample due to adapter contamination and trimming of low quality regions with Trim Galore! v0.6.4, removed 35% to 76% (average: 54%) rRNA sequences with SortMeRNA v2.1b [Kopylova et al., 2012], aligned with STAR v2.6.1d 88-91% and 9-92% (9%, 90%, 92%) reads for autotrophic and heterotrophic conditions, respectively, to the metagenome and finally summarized 2.7-16.8 and 0.3-27.1 million counts per sample for autotrophic and heterotrophic conditions, respectively, to genes based on the IMGAP annotation by featureCounts v1.6.4 [Liao et al., 2014]. Gene counts were used in differential abundance analysis in R v3.5.1 with DESeq2 v1.22.1 [Love et al., 2014] and a significant difference was postulated for transcripts with Benjamini and Hochberg adjusted $p \leq 0.05$.

Metaproteome analysis

Protein concentrations were determined via the Bradford assay (Bio-Rad) according to the user's manual. As for autotrophic samples, the entire protein extract was used for digestion. SDS PAGE short gel purification (Invitrogen) was run and in-gel digestion with Trypsin was conducted as described previously [Borchert et al., 2010]. Extracted peptides were desalted using C18 StageTips (Rappsilber *et al*, 2007). The entire sample for autotrophic samples and calculated 4 µg of peptides for heterotrophic samples were subjected to LC-MS/MS analysis. Peptides were separated on a 20 cm analytical column packed in house with ReprosSil-Pur C18-AQ 1.9 µm resin (Dr. Maisch GmbH (Ltd.), Ammerbuch, Germany) using Easy-nLC 1200 UHPLC (Thermo Fisher Scientific). It was coupled to an QExactive HF-X Orbitrap mass spectrometer (Thermo Fisher Scientific) with nano-electrospray source. The peptides were injected in HPLC solvent A

(0.1% formic acid) and subsequently eluted with a 127 min segmented gradient of 10-33-50-90% of HPLC solvent B (80% acetonitrile in 0.1% formic acid) at a flow rate of 200 nl/min. The 12 most intense precursor ions were selected for fragmentation with HCD, sequenced mass precursors were excluded from further fragmentation for 30 s. Target values were 3,000,000 charges for the MS scan with a resolution of 60,000 and 100,000 charges for the MS/MS fragmentation scan with a resolution of 30,000. The MS data was processed with MaxQuant software suite v.1.6.7.0. (Cox & Mann, 2008). The iBAQ and LFQ algorithms were enabled, and samples of the same treatment were matched. Database search against provided database (IMG metagenome ID: 3300036710) was performed using the Andromeda search engine (Cox *et al*, 2011). 23,626 identified peptides by MaxQuant were loaded with R package proteus v0.2.13 [Gierlinski et al., 2018] (https://github.com/bartongroup/Proteus) in R v3.6.0 [Team, 2018] (https://www.R-project.org/.) and subsequently assigned to 3,792 proteins, accumulated protein intensities were normalized by each samples median, transformed by log2, and finally differential protein abundance analysis was performed for the two conditions with three samples each. The significance level for rejecting the null hypothesis was set to 0.05.

Phylogenetic tree construction for protein or gene sequences

Phylogenetic analysis of Cyc2, MtoA/B, RbcL/S protein sequences of culture BP and related Fe(II)-oxidizing bacteria (FeOB), as well as 16S rRNA genes of culture BP, *Gallionellaceae* spp. identified in environmental samples and isolated *Gallionellaceae* spp. were conducted (Fig. S1, S2, S3, S4 and S5). Analyses were conducted in MEGA X using the Maximum Likelihood method and the Tamura-Nei model with bootstrap values of 1000 for each tree [Felsenstein, 1985; Tamura and Nei, 1993; Kumar et al., 2018]. The tree with the highest log likelihood and the percentage of trees in which the associated taxa clustered together was indicated next to the branches. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. These analyses involved 68 16S rRNA nucleotide sequences (Fig. S1), 38 RbcL, 23 RbcS, 38 Cyc2,

7 MtoA, 16 MtoB amino acid sequences, respectively (Fig. S2, S3, and S4) and 49 nucleotide sequences (Fig. S5).

Supplemental References

Andrews S. FastQC: A quality control tool for high throughput sequence data 2010.

Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015;75(2):129-37.

Blothe M, Roden EE. Composition and activity of an autotrophic Fe(II)-oxidizing, nitrate-reducing enrichment culture. Appl Environ Microbiol. 2009 Nov;75(21):6937-40.

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018 May 17;6(1):90.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019 Aug;37(8):852-57.

Borchert N, Dieterich C, Krug K, Schutz W, Jung S, Nordheim A, et al. Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. Genome Res. 2010 Jun;20(6):837-46.

Callahan B. Silva taxonomic training data formatted for DADA2 (Silva version 132). 2018.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul;13(7):581-3.

Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Res. 2019 Oct 10;47(18):e103.

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018 Sep 1;34(17):i884-i90.

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017 Apr 11;35(4):316-19.

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020 2020/03/01;38(3):276-78.

Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39(4):783-91.

Franzén O, Hu J, Bao X, Itzkowitz SH, Peter I, Bashir A. Improved OTU-picking using long-read 16S rRNA gene amplicon sequencing and generic hierarchical clustering. Microbiome. 2015 2015/10/05;3(1):43.

Gierlinski M, Gastaldello F, Cole C, Barton GJ. Proteus: an R package for downstream analysis of MaxQuant output. bioRxiv. 2018:416511.

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Appl Environ Microbiol. 2000 Dec;66(12):5488-91.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072-5.

Hegler F, Posth NR, Jiang J, Kappler A. Physiology of phototrophic iron(II)-oxidizing bacteria: implications for modern and ancient environments. FEMS Microbiol Ecol. 2008 Nov;66(2):250-60.

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.

Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012 Dec 15;28(24):3211-7.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018 Jun 1;35(6):1547-49.

Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. PLoS One. 2017;12(5):e0177459.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9.

Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015 May 15;31(10):1674-6.

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014 Apr 1;30(7):923-30.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 2014/12/05;15(12):550.

Lueders T, Manefield M, Friedrich MW. Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. Environ Microbiol. 2004 Jan;6(1):73-8.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011 2011-05-02;17(1):3.

Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2016;18(5):1403-14.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35(21):7188-96.

Schmidt C, Nikeleit V, Schaedler F, Leider A, Lueder U, Bryce C, et al. Metabolic responses of a phototrophic co-culture enriched from a freshwater sediment on changing substrate availability and its relevance for biogeochemical iron cycling. Geomicrobiol J. 2020:1-15.

Spät P, Maček B, Forchhammer K. Phosphoproteome of the cyanobacterium *Synechocystis* sp. PCC 6803 and its dynamics during nitrogen starvation. Front Microbiol. 2015;6:248.

Straub D, Blackwell N, Langarica-Fuentes A, Peltzer A, Nahnsen S, Kleindienst S. Interpretations of environmental microbial community studies are biased by the selected 16S rRNA (gene) amplicon sequencing pipeline. Front Microbiol. 2020;11:550420.

Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993 May;10(3):512-26.

Team RDC. R: A language and environment for statistical computing. Generic. 2018.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018 Mar 1;35(3):543-48.

Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 2017 Jan 4;45(D1):D744-d49.

Supplementary Tables

Table S1. Metagenome assembled genome (MAG) quality and data statistics with mapped number of significantly changed (adjusted $p \leq 0.05$) transcripts and proteins under autotrophic conditions compared to heterotrophic conditions.

	Gallionellaceae	***Noviherbaspirillum***	***Rhodoferax***	***Thiobacillus***	***Rhodoblastus***	***Ramlibacter***
Information of genome bins						
Estimated completeness (%)	93.2	90.5	98	90.6	94.6	88.5
Estimated contamination (%)	0	2	0.7	1.4	0	2
No. contigs (scaffold)	42	41	14	15	88	458
Largest contig	241555	621421	1452120	926960	324380	78674
Gene count	2374	5070	3237	3295	4459	5281
Genome size (bp)	2,408,093	5,429,401	3,368,592	3,249,831	4,511,103	5,100,867
Sequencing depth	6757	1506	121	24	17	9
GC (%)	59.05	59.52	63.91	62.7	62.96	66.61
No. of CDS	2325	5002	3177	3241	4376	5222
Taxonomy ID	96	1344552	28065	919	168658	174951
IMG Submission ID	215615	215613	221513	221511	221512	221514
IMG Genome ID	2831290873	2831285802	2840079448	2840071692	2840074988	2840082686
GOLD Analysis Project ID	Ga0394452	Ga0394450	Ga0401103	Ga0401101	Ga0401102	Ga0401104
16S rRNA gene similarity (%) of closest related isolated species*	*Ferrigenium kumadai* (96.17)	*Noviherbaspirillum autotrophicum* strain TSA66 (98.97)	*Curvibacter delicatus* strain NBRC 14919 (98.56)	*Thiobacillus thioparus strain* THI 111 (99.66)	*Rhodoblastus sphagnicola* strain RS (97.94)	*Xenophilus aerolatus* strain 5516S-2 (97.68)
Community composition (exponential phase) (%)						
Autotrophic	77.11	14.76	2.24	1.42	1.11	0.27
Heterotrophic	0.08	4.93	94.87	0	0	0
Number of significantly changed (adjusted $p \leq 0.05$) transcripts under autotrophic conditions compared to heterotrophic conditions						
Total	2396	4546	216	696	747	532
Significant-up	2396	4546	59	696	747	532
Significant-down	0	0	157	0	0	0
Number of significantly changed proteins (adjusted $p \leq 0.05$) under autotrophic conditions compared to heterotrophic conditions						
Total	12	122	103	0	1	2
Significant-up	3	96	5	0	0	1
Significant-down	9	26	98	0	1	1

*16S rRNA gene similarities are based on PacBio long-read amplicon sequencing

Table S2. Metagenome (A), metatranscriptome (B) and metaproteome (C) data statistics.

	Culture BP
(A) Metagenome	
Raw read pairs	91,056,140 (27.48 × 10^9 base pairs)
Read pairs that passed QC	90,303,622 (27.25 × 10^9 base pairs)
Genome size (10^6 base pairs, Mbp)	57,442,683
Proportion of reads assembled (%)	99.35
No. contigs (scaffold)	25083
Gene count	75943
GC (%)	64.01
No. of CDS	75105
IMG Submission ID	221510
IMG Genome ID	3300036710
GOLD Analysis Project ID	Ga0394449
rRNA	103
16S rRNA Count	34
(B) Metatranscriptome (range / average)	
Sequences [million]	6.3 – 53.3 / 25
% duplicates	94 – 99 / 98
GC%	52 - 54 / 53
% rRNA	35 – 76 / 54
% mapped to metagenome	9.5 – 92.6 / 77
mapped to metagenome [million]	0.3 – 30.2 / 10
Total no. of detected transcripts	17,471
No. of significant higher expression level transcripts under autotrophic condition	10,503
No. of significant higher expression level transcripts under heterotrophic condition	157
(C) Metaproteome (range / average)	
Analyzed MS/MS spectra [thousand]	85 – 89 / 87
Identified MS/MS spectra (%)	3.7 – 27.8 / 16
Total no. of detected proteins	3792
No. of significant higher abundant proteins under autotrophic condition	126
No. of significant higher abundant proteins under heterotrophic condition	150

Table S3. Summary of key metabolic pathways in selected NRFeOx-related metagenome assembled genomes (MAGs).

Function	*Gallionellaceae-KS*	*Gallionellaceae-BP*	*Noviherbaspirillum*	*Rhodoblastus*	*Thiobacillus*	*Rhodoferax*	*Ramlibacter*
Carbon Metabolism							
CBB reductive pentose phosphate cycle	+	+	+	+	+	(+)	+
Glycolysis (Embden-Meyerhof pathway)	+	+	+	+	+	+	+
Pyruvate oxidation to acetyl-CoA through pyruvate dehydrogenase	+	+	+	+	+	+	+
Citrate cycle (TCA cycle, Krebs cycle)	+	+	+	+	(+)	+	+
Pentose phosphate pathway	+	+	(+)	(+)	+	(+)	(+)
Glyoxylate cycle	-	(+)	+	(+)	(+)	(+)	+
Nitrogen Metabolism							
Nitrogen fixation, nitrogen → ammonia	-	-	-	(+)	-	-	-
Dissimilatory nitrate reduction	-	(+)	+	+	+	+	+
Assimilatory Nitrate reduction	-	-	(+)	(+)	(+)	(+)	(+)
Nitrification, ammonia → nitrite	-	-	-	-	-	-	-
<u>Denitrification</u>							
Nitrate → nitrite	+	-	+	+	+	+	+
Nitrite → nitric oxide	+	+	+	+	+	+	+
Nitric oxide → nitrous oxide	-	+	+	+	+	+	+
Nitrous oxide → nitrogen gas	-	-	+	+	+	+	+
Complete denitrification	No	No	Yes	Yes	Yes	Yes	Yes

Table S3 (continued). Summary of key metabolic pathways in selected NRFeOx-related metagenome assembled genomes (MAGs).

Function	*Gallionellaceae*-KS	*Gallionellaceae*-BP	*Noviherbaspirillum*	*Rhodoblastus*	*Thiobacillus*	*Rhodoferax*	*Ramlibacter*
Oxidative Phosphorylation							
Complex I, NADH: quinone oxidoreductase	+	+	+	+	+	+	(+)
Complex II, succinate dehydrogenase	+	+	+	+	+	+	+
Complex III, cytochrome bc_1 complex	+	+	+	+	+	+	+
Alternative complex III (ACIII)	-	-	-	-	-	-	-
aa_3-type cytochrome *c* oxidase	+	+	+	+	+	+	+
Cytochrome *bd* ubiquinol oxidase	-	-	(+)	(+)	(+)	-	(+)
cbb_3-type cytochrome *c* oxidase	+	+	+	+	+	+	+
F-type ATPase	+	+	+	+	+	+	+
V-type ATPase	-	-	-	-	-	-	-

Table S4. Overview of meta'omics samples from culture BP, archived in the Sequencing Read Archive (SRA, bioproject PRJNA693457) or ProteomeXchange Consortium (via the PRIDE partner repository, identifier PXD023710). All amplicon sequencing was performed with primers 515f and 805r, except SRR13504099, that was amplified with primers 27F and 1492R.

Conditions	Time-points	Replicate	Amplicon sequencing	Metagenomics	Metatranscriptomics	Metaproteomics
Autotrophic	T3	1	SRR13504099			
Autotrophic	T1	1		SRR13494970		
Autotrophic	T0	1	SRR13489090			
	T1	1	SRR13489089		SRR13494978	R23
	T1	2	SRR13489148		SRR13494977	R24
	T1	3	SRR13489137		SRR13494976	R25
	T2	1	SRR13489126			
	T2	2	SRR13489115			
	T2	3	SRR13489104			
	T3	1	SRR13489093			
	T3	2	SRR13489092			
	T3	3	SRR13489091			
Heterotrophic	T0	1	SRR13489088			
	T1	1	SRR13489087			
	T1	2	SRR13489156			
	T1	3	SRR13489155			
	T2	1	SRR13489154		SRR13494975	R26
	T2	2	SRR13489153		SRR13494974	R27
	T2	3	SRR13489152		SRR13494973	R28
	T3	1	SRR13489151			
	T3	2	SRR13489150			
	T3	3	SRR13489149			

Supplementary Figures

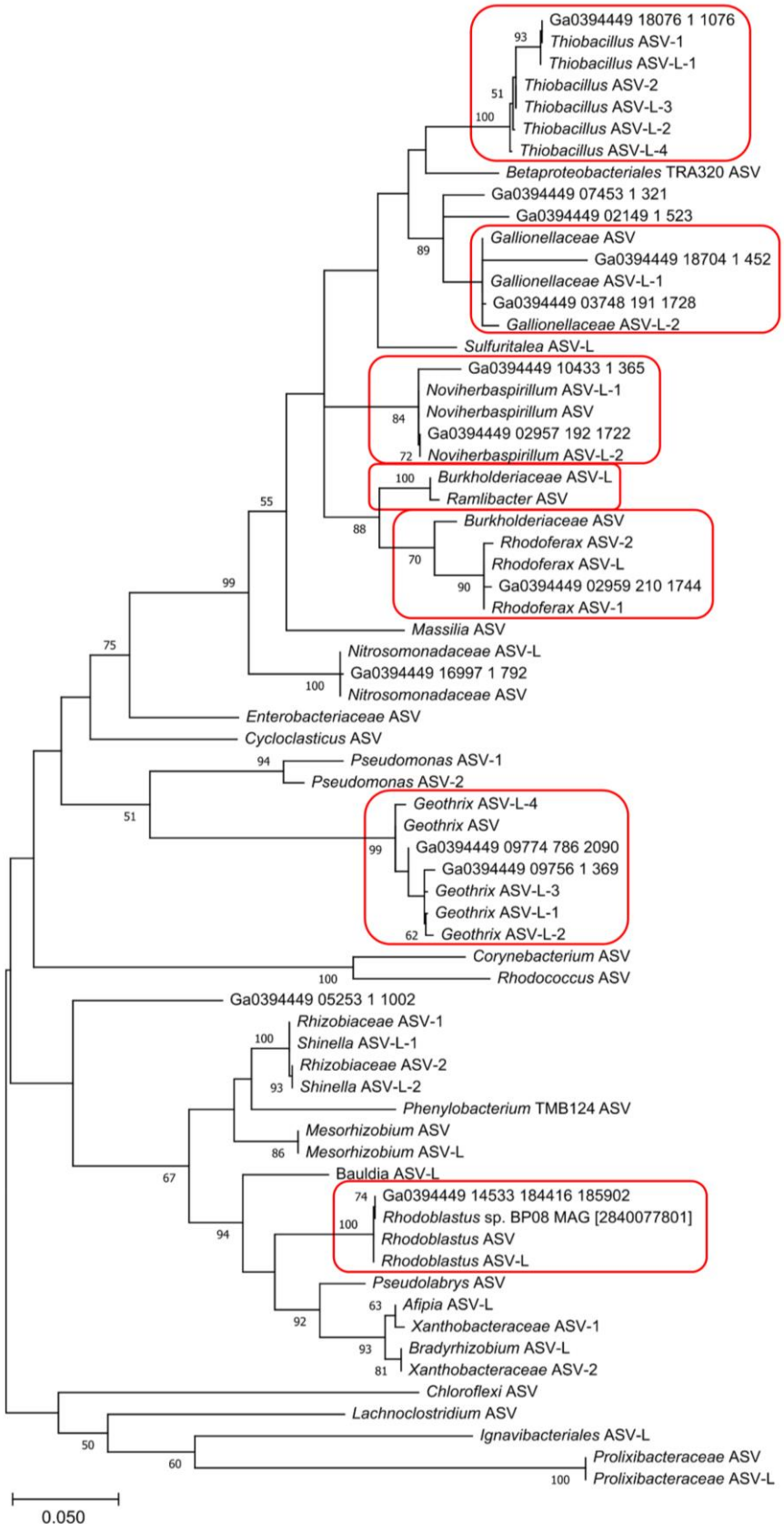


Fig. S1. Phylogenetic tree showing the relationships between 68 16S rRNA nucleotide sequences obtained by different methods. 1. Long-read ASVs ("ASV-L-x"), 2. short-read ASVs ("ASV-x"), 3. MAG sequence ("MAG"), and 4. metagenome sequences ("Ga0394449"). If there were more ASVs, the number "x" was added right after the "ASV-L-" or "ASV-". 16S rRNA sequences retrieved from the MAG and metagenome were only included if their length was ≥300 bp and if they overlapped with the V4 region of the 16S rRNA gene (i.e. with the short-read amplicons). The taxa related to the organisms mentioned in this study are marked with red boxes. Evolutionary analyses were conducted in MEGA-X and the bootstrap values were calculated on 100 replicates using the Maximum Likelihood method.

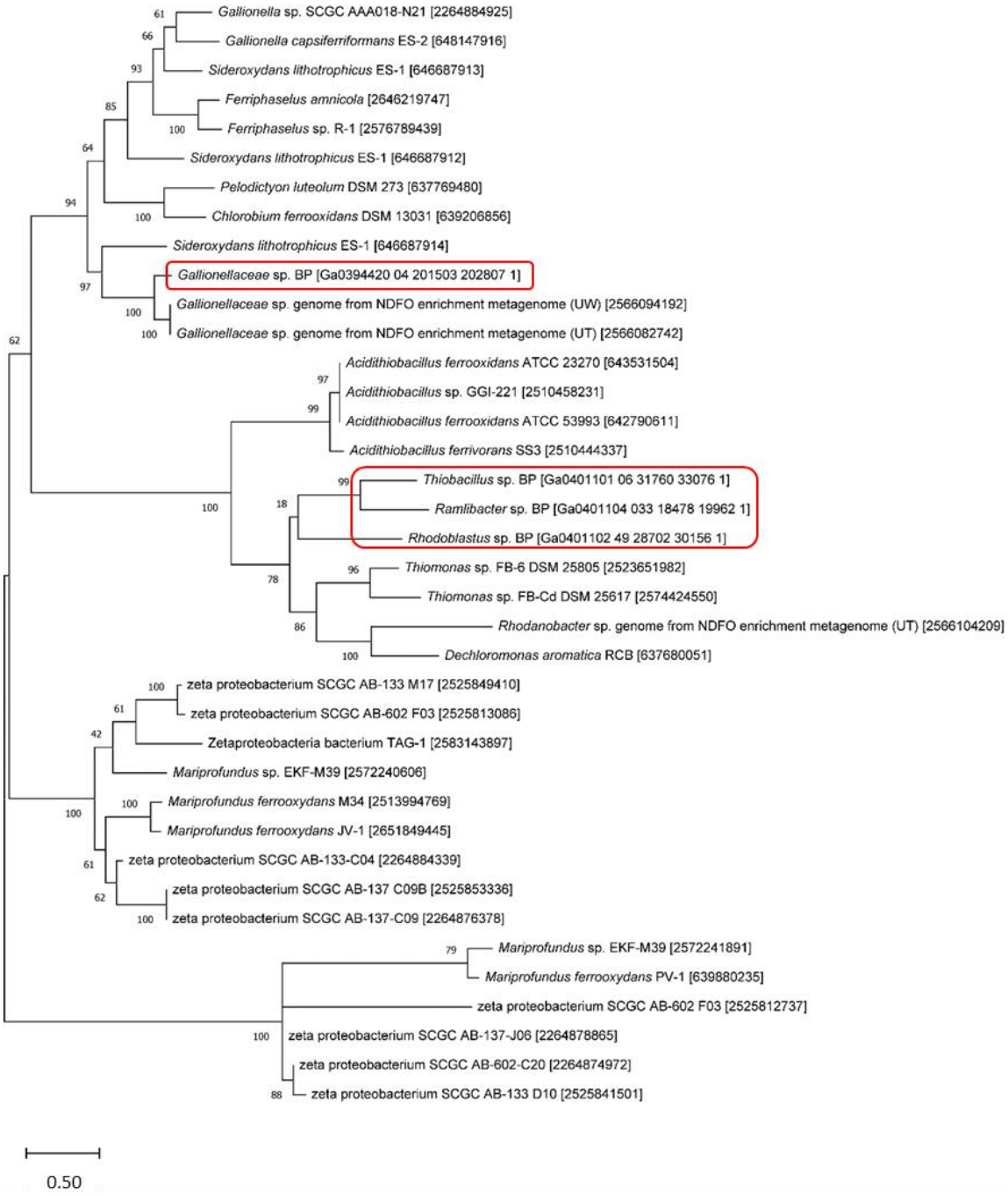


Fig. S2. Phylogenetic tree of Cyc2 from the organisms in culture BP and other Fe(II)-oxidizing bacteria based on amino acid sequences with bootstrap values, calculated on 1000 replicates using the Maximum Likelihood method. The organisms in culture BP are marked with red boxes. The scale bar indicates branch lengths measured by the number of substitutions per site. Numbers in brackets show the identifier of each gene in IMG.

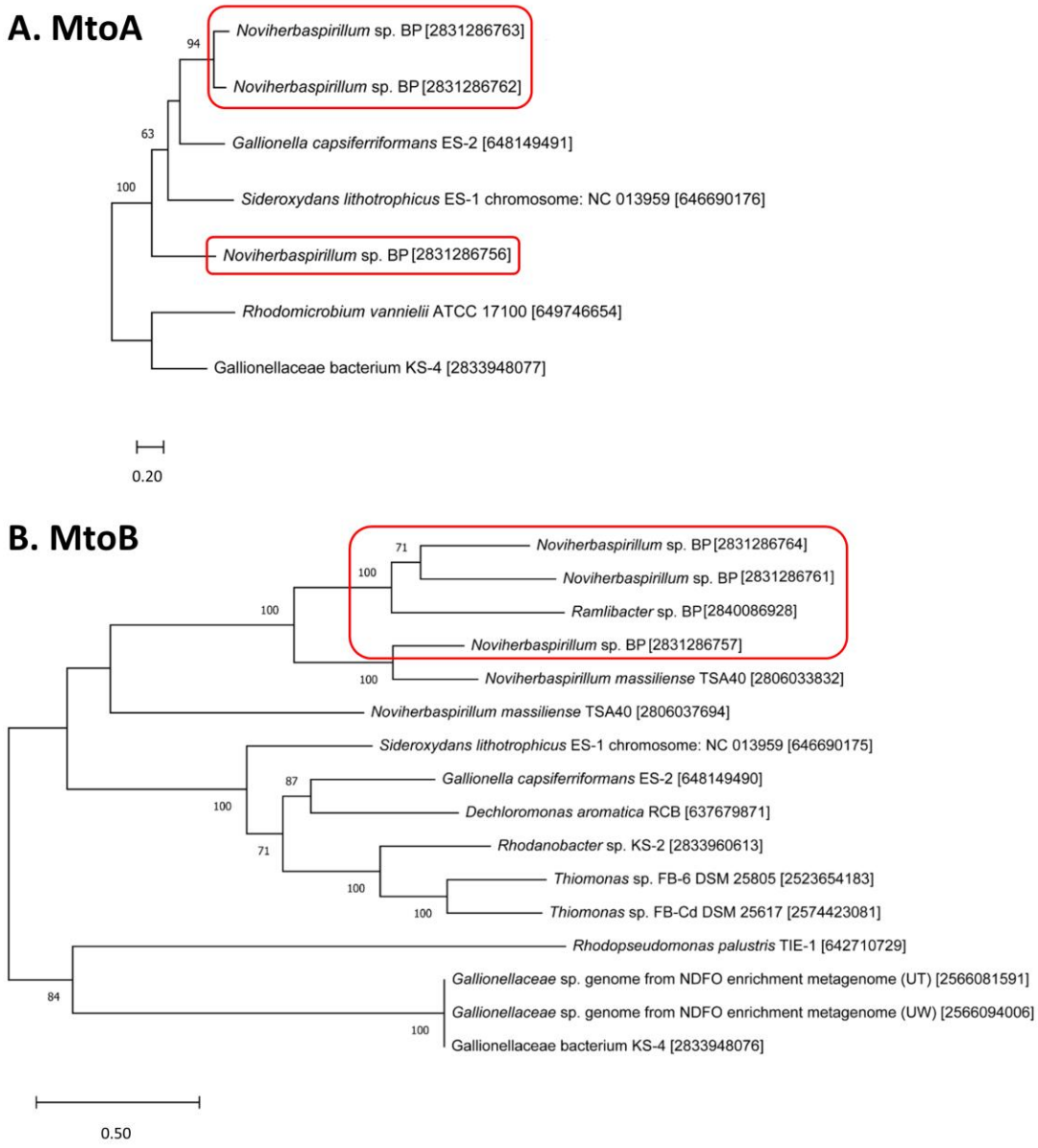
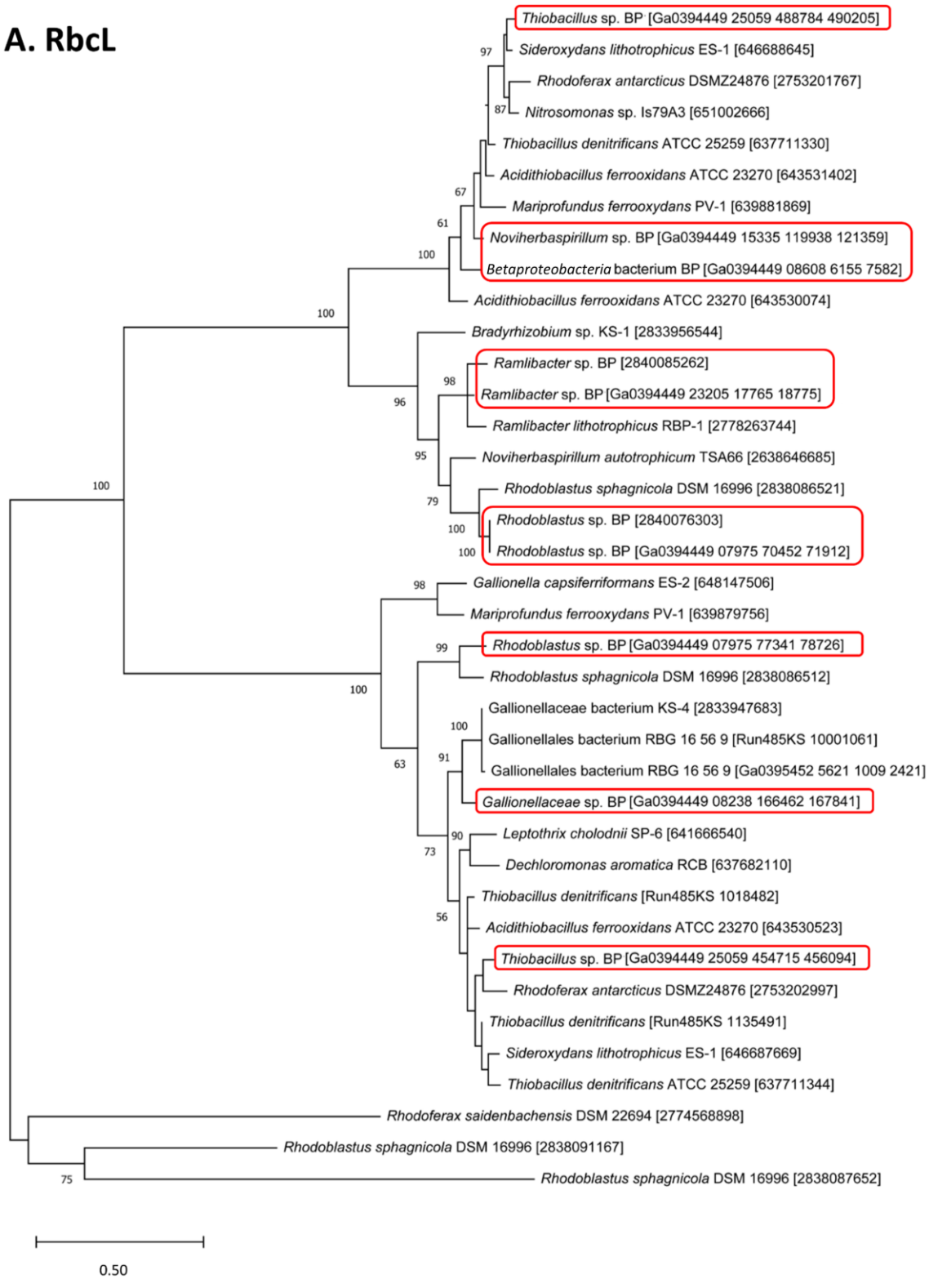


Fig. S3. Phylogenetic trees of MtoA (A) and MtoB (B) of *Noviherbaspirillum* sp. and other Fe(II)-oxidizing bacteria based on amino acid sequences with bootstrap values, calculated on 1000 replicates using the Maximum Likelihood method. The organisms in culture BP are marked with red boxes. The scale bar indicates branch lengths measured by the number of substitutions per site. Numbers in brackets show the identifier of each gene in IMG.


A. RbcL
Thiobacillus sp. BP [Ga0394449 25059 488784 490205]
Sideroxydans lithotrophicus ES-1 [646688645]
Rhodoferax antarcticus DSMZ24876 [2753201767]
Nitrosomonas sp. Is79A3 [651002666]
Thiobacillus denitrificans ATCC 25259 [637711330]
Acidithiobacillus ferrooxidans ATCC 23270 [643531402]
Mariprofundus ferrooxydans PV-1 [639881869]
Noviherbaspirillum sp. BP [Ga0394449 15335 119938 121359]
Betaproteobacteria bacterium BP [Ga0394449 08608 6155 7582]
Acidithiobacillus ferrooxidans ATCC 23270 [643530074]
Bradyrhizobium sp. KS-1 [2833956544]
Ramlibacter sp. BP [2840085262]
Ramlibacter sp. BP [Ga0394449 23205 17765 18775]
Ramlibacter lithotrophicus RBP-1 [2778263744]
Noviherbaspirillum autotrophicum TSA66 [2638646685]
Rhodoblastus sphagnicola DSM 16996 [2838086521]
Rhodoblastus sp. BP [2840076303]
Rhodoblastus sp. BP [Ga0394449 07975 70452 71912]
Gallionella capsiferriformans ES-2 [648147506]
Mariprofundus ferrooxydans PV-1 [639879756]
Rhodoblastus sp. BP [Ga0394449 07975 77341 78726]
Rhodoblastus sphagnicola DSM 16996 [2838086512]
Gallionellaceae bacterium KS-4 [2833947683]
Gallionellales bacterium RBG 16 56 9 [Run485KS 10001061]
Gallionellales bacterium RBG 16 56 9 [Ga0395452 5621 1009 2421]
Gallionellaceae sp. BP [Ga0394449 08238 166462 167841]
Leptothrix cholodnii SP-6 [641666540]
Dechloromonas aromatica RCB [637682110]
Thiobacillus denitrificans [Run485KS 1018482]
Acidithiobacillus ferrooxidans ATCC 23270 [643530523]
Thiobacillus sp. BP [Ga0394449 25059 454715 456094]
Rhodoferax antarcticus DSMZ24876 [2753202997]
Thiobacillus denitrificans [Run485KS 1135491]
Sideroxydans lithotrophicus ES-1 [646687669]
Thiobacillus denitrificans ATCC 25259 [637711344]
Rhodoferax saidenbachensis DSM 22694 [2774568898]
Rhodoblastus sphagnicola DSM 16996 [2838091167]
Rhodoblastus sphagnicola DSM 16996 [2838087652]
0.50

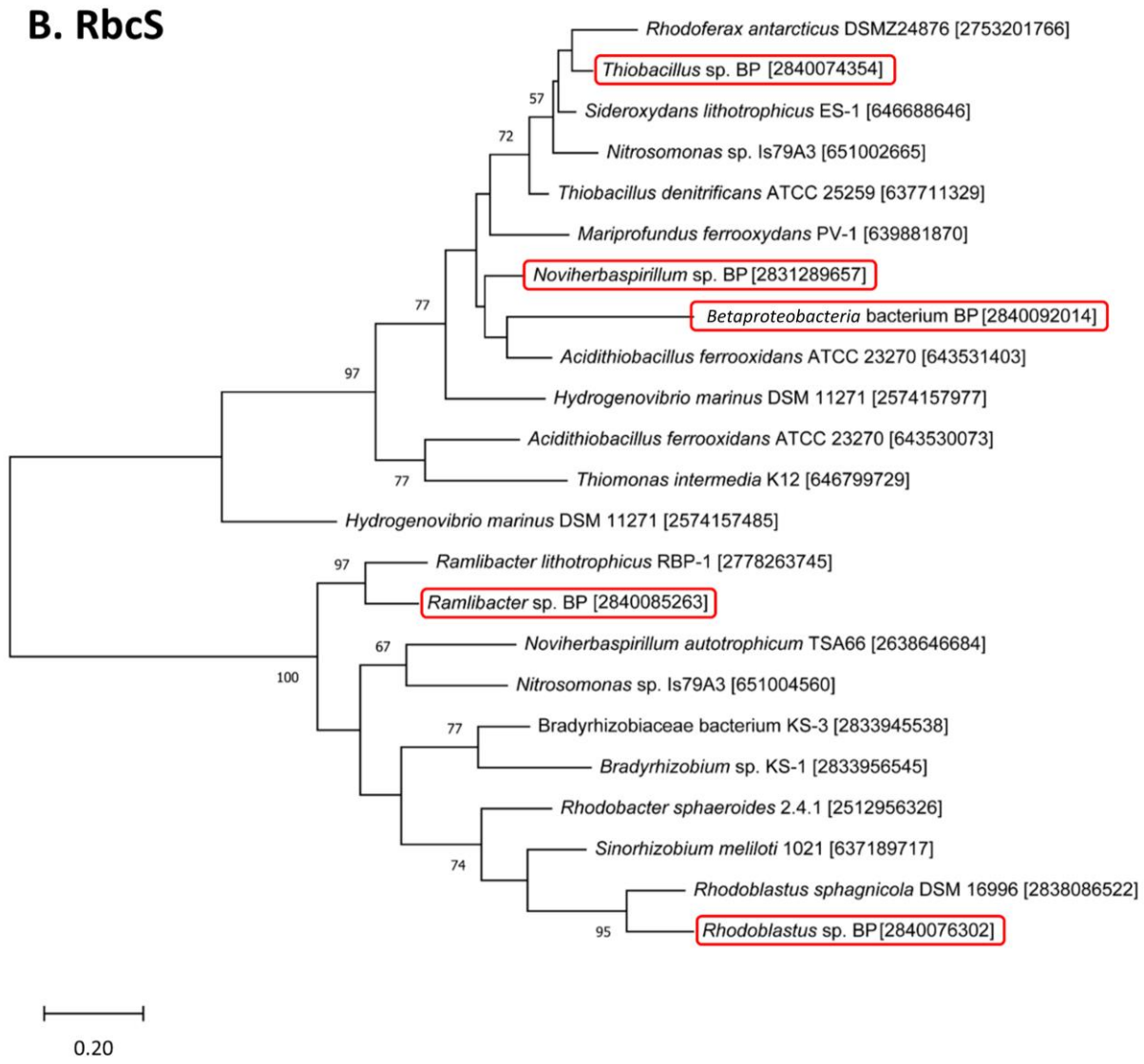


Fig. S4. Phylogenetic trees of RuBisCO from culture BP, flanking members and other Fe(II)-oxidizing bacteria: RbcL (A) and RbcS (B) based on amino acid sequences with bootstrap values, calculated on 1000 replicates using the Maximum Likelihood method. The organisms in culture BP are marked with red boxes. The scale bar indicates branch lengths measured by the number of substitutions per site. Numbers in brackets show the identifier of each gene in IMG.

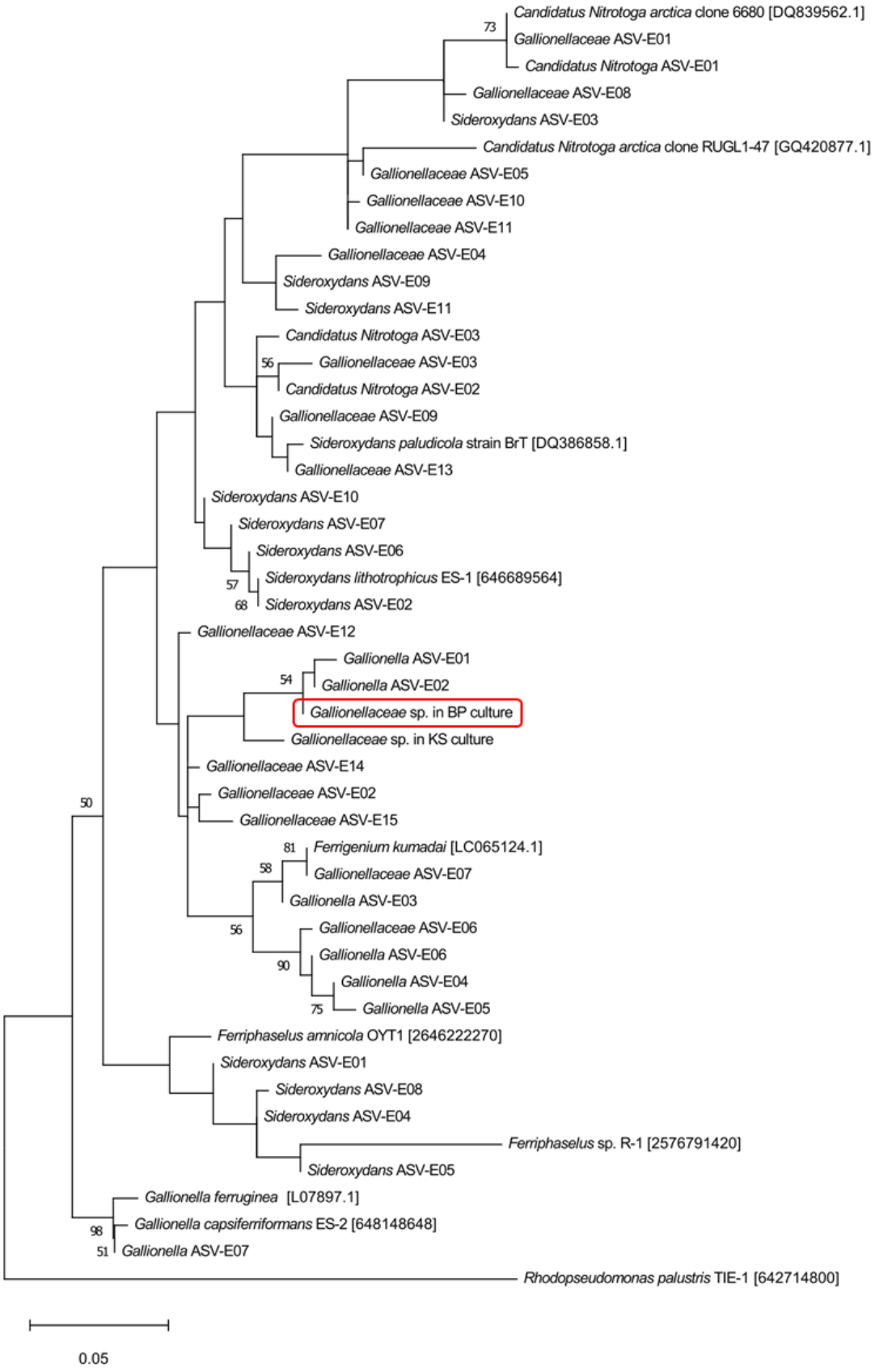
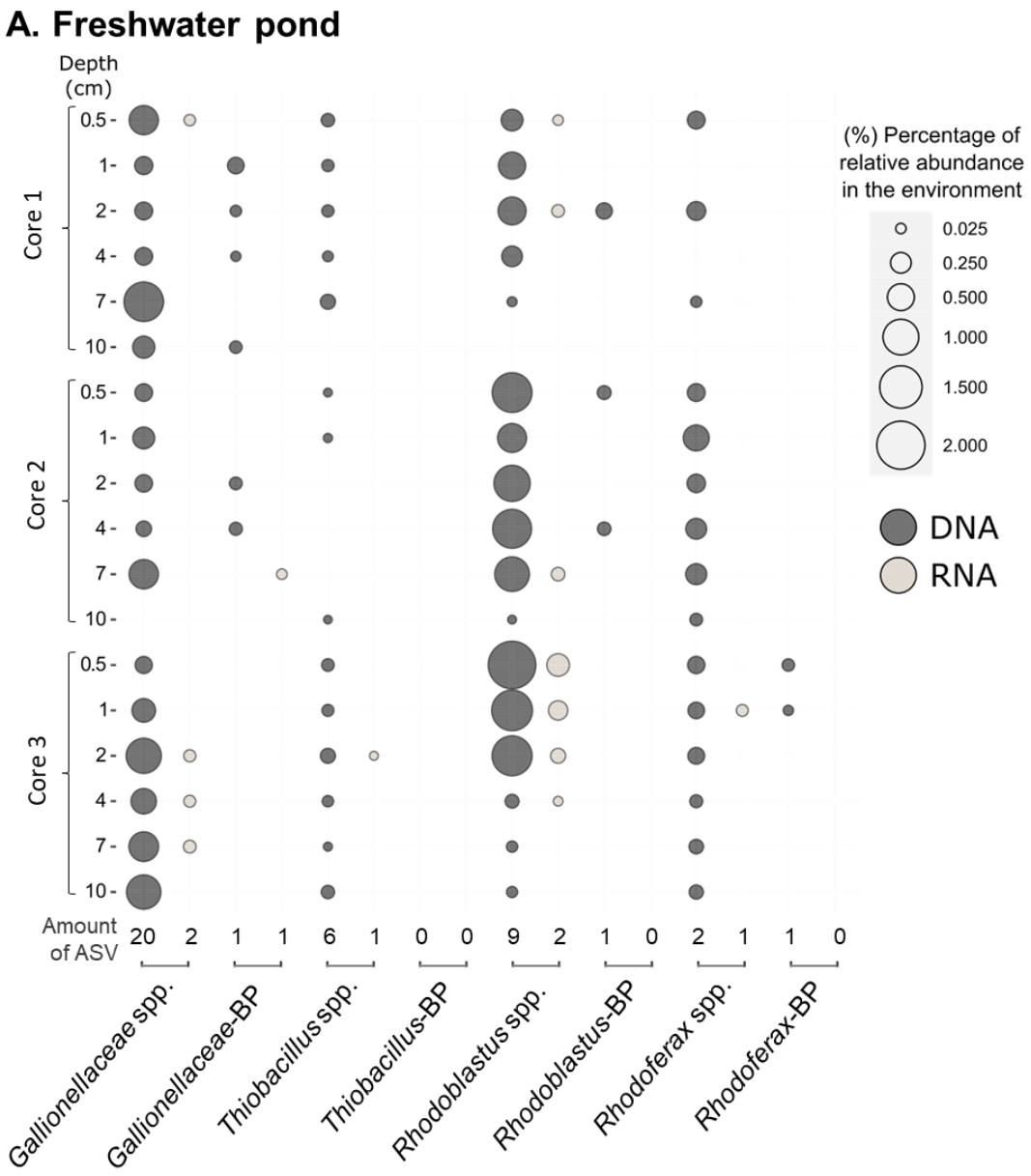

Candidatus Nitrotoga arctica clone 6680 [DQ839562.1]
73
Gallionellaceae ASV-E01
Candidatus Nitrotoga ASV-E01
Gallionellaceae ASV-E08
Sideroxydans ASV-E03
Candidatus Nitrotoga arctica clone RUGL1-47 [GQ420877.1]
Gallionellaceae ASV-E05
Gallionellaceae ASV-E10
Gallionellaceae ASV-E11
Gallionellaceae ASV-E04
Sideroxydans ASV-E09
Sideroxydans ASV-E11
Candidatus Nitrotoga ASV-E03
56
Gallionellaceae ASV-E03
Candidatus Nitrotoga ASV-E02
Gallionellaceae ASV-E09
Sideroxydans paludicola strain BrT [DQ386858.1]
Gallionellaceae ASV-E13
Sideroxydans ASV-E10
Sideroxydans ASV-E07
Sideroxydans ASV-E06
57
Sideroxydans lithotrophicus ES-1 [646689564]
68
Sideroxydans ASV-E02
Gallionellaceae ASV-E12
Gallionella ASV-E01
54
Gallionella ASV-E02
Gallionellaceae sp. in BP culture
Gallionellaceae sp. in KS culture
Gallionellaceae ASV-E14
Gallionellaceae ASV-E02
50
Gallionellaceae ASV-E15
81
Ferrigenium kumadai [LC065124.1]
58
Gallionellaceae ASV-E07
Gallionella ASV-E03
56
Gallionellaceae ASV-E06
Gallionella ASV-E06
90
Gallionella ASV-E04
75
Gallionella ASV-E05
Ferriphaselus amnicola OYT1 [2646222270]
Sideroxydans ASV-E01
Sideroxydans ASV-E08
Sideroxydans ASV-E04
Ferriphaselus sp. R-1 [2576791420]
Sideroxydans ASV-E05
Gallionella ferruginea [L07897.1]
98
Gallionella capsiferriformans ES-2 [648148648]
51
Gallionella ASV-E07
Rhodopseudomonas palustris TIE-1 [642714800]
0.05

Fig. S5. Phylogenetic tree of the *Gallionellaceae* spp. ASVs detected in the environmental sediments compared to isolated *Gallionellaceae* spp. and Fe(II)-oxidizing bacteria based on nucleotide sequences with bootstrap values, calculated on 1000 replicates using the Maximum Likelihood method. The *Gallionellaceae* sp. in culture BP is marked with red boxes. The scale bar indicates branch lengths measured by the number of substitutions per site. Numbers in brackets show accession numbers.


A. Freshwater pond
Depth (cm)
Core 1
Core 2
Core 3
0.5
1
2
4
7
10
Amount of ASV
20 2 1 1 6 1 0 0 9 2 1 0 2 1 1 0
Gallionellaceae spp.
Gallionellaceae-BP
Thiobacillus spp.
Thiobacillus-BP
Rhodoblastus spp.
Rhodoblastus-BP
Rhodoferax spp.
Rhodoferax-BP
(%) Percentage of relative abundance in the environment
0.025
0.250
0.500
1.000
1.500
2.000
DNA
RNA

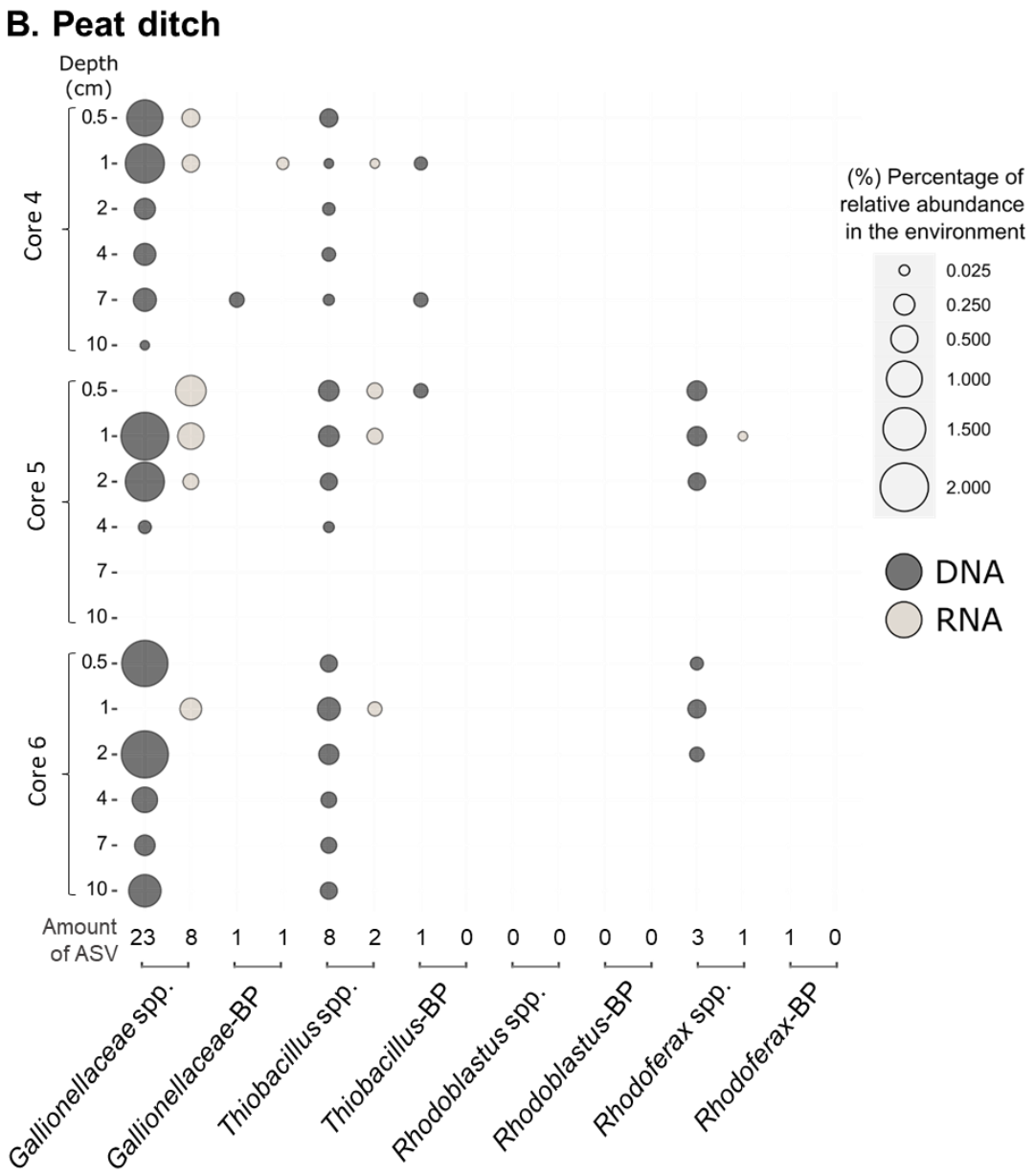


Fig. S6. Bubble plot showing the relative abundance of amplicon sequence variants (ASVs) of microbial populations from culture BP (identical ASV and genus or family level, respectively), detected in environmental sediments. Samples were taken (A) at an organic-rich freshwater pond and (B) at an organic-rich peat ditch in Bremen, Germany. Sediments were taken from three different cores per site and from 0-10 cm depth. Dark and light grey bubbles demonstrate DNA- and RNA-based short-read 16S rRNA (gene) amplicon sequencing data, respectively.