# Supporting Information

## Lücker et al. 10.1073/pnas.1003860107

### SI Results

**[Fe-S] and Molybdenum Ligands in NxrA.** Both nitrite oxidoreductase (NXR) α-subunits of *Candidatus* Nitrospira defluvii (*Ca*. N. defluvii) (NxrA1 and NxrA2) contain, close to the N terminus, one cysteine-rich [Fe-S] binding motif (C-$X_3$-D-$X_3$-C-$X_{39}$-C) (Fig. S2*A*). This motif resembles the consensus [Fe-S] binding motif (H/C-$X_3$-C-$X_3$-C-$X_n$-C) of the type II group in the dimethyl sulfoxide (DMSO) reductase family of molybdopterin-binding enzymes (1, 2). In *Ca*. N. defluvii, the second cysteine residue of the consensus motif is replaced by aspartate, which can also function as a [Fe-S] ligand as shown for a ferredoxin of *Pyrococcus furiosus* (3). This aspartate residue also occurs in the phylogenetically closely related NXR-like proteins of *Ca*. K. stuttgartiensis, *Hydrogenobaculum*, and *Beggiatoa* (Fig. S2*A*). On the basis of these data, we propose a consensus [Fe-S] binding motif (H/C-$X_3$-D/C-$X_3$-C-$X_n$-C) for the type II group in the DMSO reductase family. A highly conserved aspartate residue, which functions as a molybdenum ligand in nitrate reductase A (subunit NarG) of *Escherichia coli* and most likely also in the other type II DMSO reductase-family enzymes (4), is present in both NxrA copies of *Ca*. N. defluvii (Asp278 in NxrA1) (Fig. S2*B*).

**Key Enzymes of the Reductive and Oxidative Tricarboxylic Acid Cycles.** Key enzymes for $CO_2$ fixation via the reductive tricarboxylic acid (rTCA) cycle are oxoglutarate:ferredoxin oxidoreductase (OGOR), pyruvate:ferredoxin oxidoreductase (POR), fumarate reductase (FRD), and ATP-citrate lyase (ACL). OGOR and POR usually consist of one to four distinct subunits (5, 6). The *Ca*. N. defluvii genome contains three gene clusters encoding 2-oxoacid:ferredoxin oxidoreductases that could have POR or OGOR activity. One cluster consists of the α-, β-, and fused γ/δ-subunits of a putative four-subunit POR, which is similar to homologs in *Pelobacter*, *Desulfotalea* (both *Deltaproteobacteria*), and *Ca*. K. stuttgartiensis (*Planctomycetes*). Alternatively, these three coding sequences (CDS) might represent a 2-oxoisovalerate:ferredoxin oxidoreductase, which could function in the degradation of branched-chain amino acids. Each of the other two gene clusters consists of five CDS, which are highly similar to five-subunit forms of POR and OGOR found recently in members of the phylum Aquificae, such as *Aquifex aeolicus* and *Hydrogenobacter thermophilus* (5, 7), and also in *Leptospirillum* (8). One of these clusters in *Ca*. N. defluvii is slightly more similar to OGOR, whereas the other cluster is more likely to be POR (Table S2), suggesting that both enzymes are present in *Nitrospira*.

Four CDS in the *Ca*. N. defluvii genome code for subunits of FRD or the highly similar counterpart of this enzyme in the oxidative tricarboxylic acid (oTCA) cycle, succinate dehydrogenase (SDH) (Table S2). Two of these CDS are homologous to the highly conserved fumarate- or succinate-binding flavoprotein subunit FrdA/SdhA, one is homologous to the iron-sulfur subunit FrdB/SdhB, and one is similar to subunit FrdE/SdhE. To date, five types (A–E) of FRD/SDH that differ in their subunit composition and distribution among bacteria, archaea, and eukaryotes are known (9). Subunit A is too conserved for a classification of these types, but FrdB/SdhB and FrdE/SdhE of *Ca*. N. defluvii resemble the respective components of the four-subunit type E enzymes. This type was described in archaea such as *Sulfolobus* spp. and *Acidianus ambivalens*, but occurs also in various bacteria (10). No homolog of the fourth subunit (named SdhF in type E enzymes) was found in *Ca*. N. defluvii, indicating that *Nitrospira* has a non-canonical form of FRD/SDH that is similar to the type E enzymes known from other organisms. A unique type E-like FRD/SDH exists also in *Leptospirillum*, but it seems to lack a homolog of FrdE/SdhE (11). The type E SDH of *A. ambivalens* is reversible and catalyzes both the oxidation of succinate and the reduction of fumarate (12). If the type E-like enzyme of *Ca*. N. defluvii also operates in either direction, it could function in the rTCA and the oTCA cycles. As two CDS encode FrdA/SdhA-like subunits, it is tempting to speculate that in *Ca*. N. defluvii the substrate specificity and catalytic properties of the holoenzyme depend on the respective version of subunit A.

Although both *Leptospirillum* and *Ca*. N. defluvii belong to the phylum Nitrospirae and fix $CO_2$ via the rTCA cycle, they differ in one critical step of this pathway: the cleavage of citrate to acetyl-CoA and oxaloacetate. *Ca*. N. defluvii employs ACL and encodes both subunits of ACL at one *aclBA* locus in close proximity to the five-subunit OGOR gene cluster. In contrast, *Leptospirillum* lacks ACL but uses two enzymes, citryl-CoA synthetase and citryl-CoA lyase, for cleaving citrate (11).

In the oTCA cycle, the 2-oxoglutarate dehydrogenase complex (ODH) irreversibly catalyzes the oxidative decarboxylation of 2-oxoglutarate to succinyl-CoA and $CO_2$. *Ca*. N. defluvii possesses three gene clusters coding for the E1 and E2 components and two genes encoding the E3 component of 2-oxoacid dehydrogenase complexes (Table S2). These CDS most likely represent two copies of pyruvate dehydrogenase and one 2-oxoisovalerate dehydrogenase, but probably not ODH. Despite the apparent lack of ODH, the oTCA cycle may operate if ODH is replaced by OGOR, which is present in *Ca*. N. defluvii (see above). In contrast to ODH, OGOR catalyzes a reversible reaction and thus can function in both the rTCA and oTCA cycles. For example, in *Helicobacter pylori*, which also lacks ODH, a four-subunit form of OGOR functionally replaces ODH in the oTCA cycle (13, 14). Further studies are needed to clarify whether the complete oTCA cycle is functional in *Nitrospira* and how the reductive and oxidative versions of the pathway are regulated in vivo and under different growth conditions.

**Use of Organic Substrates.** In a previous study, FISH combined with microautoradiography showed that *Nitrospira* from a sewage treatment plant used pyruvate, but not acetate, as an organic carbon source (15). Consistent with these results, no canonical acetate permease was identified in the genome of *Ca*. N. defluvii. However, the genome encodes a putative member of the GPR1/YaaH protein family (Nide1910). In yeast, one protein from this family has been identified as a candidate acetate transporter (16). If the remote homolog in *Ca*. N. defluvii indeed facilitates acetate uptake, acetate can be metabolized by activation to acetyl-CoA in the acetyl-CoA synthetase reaction and subsequent carboxylation to pyruvate, which is catalyzed by POR. Because pyruvate is a precursor for sugar biosynthesis via gluconeogenesis, carbon from exogenous acetate or pyruvate can be stored in glycogen deposits within *Ca*. N. defluvii cells.

The genome encodes a soluble formate dehydrogenase for the oxidation of formate to $CO_2$ with $NAD^+$ as electron acceptor, suggesting that *Ca*. N. defluvii can use formate as a substrate. In addition, a cluster of six CDS seems to code for a six-subunit, membrane-bound [NiFe]-hydrogenase that might be part of a formate hydrogenlyase complex (Table S2). However, the function of this putative hydrogenase remains unclear. It lacks the amino acid signatures of all known [NiFe]-hydrogenase groups (17) and the cysteine ligands of the [NiFe] center, suggesting that this enzyme does not have hydrogenase activity or belongs to a unique class of hydrogenases. Interestingly, a highly similar

enzyme found in the genome of *Ca*. K. stuttgartiensis has the cysteine ligands and the signature of group 4 membrane-associated, energy-converting, and $H_2$-evolving hydrogenases (17).

**Uptake, Secretion, and Storage.** About 5–6% of the *Ca*. N. defluvii genome consists of genes involved in transport and secretion, which comprise diverse transporter families (Table S1). Transporters for various organic nutrients (Fig. S5) support the notion that *Ca*. N. defluvii is not confined to pure autotrophy. *Ca*. N. defluvii has uptake systems for $PO_4^{3-}$, $NH_3/NH_4^+$, and $NO_2^-$ (Table S1). The predicted $NO_3^-/NO_2^-$ antiporter (Nide1382) could alternatively function as $H^+/NO_2^-$ antiporter and thus be important for resistance against elevated cytoplasmic nitrite concentrations (18). A gene coding for polyphosphate kinase was identified, which is consistent with observed polyphosphate granules in *Nitrospira moscoviensis* cells (19). The genome encodes a ferredoxin-nitrite reductase for the reduction of $NO_2^-$ to $NH_4^+$, indicating that $NO_2^-$ also serves as a nitrogen source for biosyntheses. Carbon is stored in glycogen as suggested by genes for the complete gluconeogenesis pathway, glycogen synthase, glycogen phosphorylase, and two putative glycogen-debranching enzymes (Table S2). Indeed, glycogen deposits have been observed by electron microscopy in *Nitrospira* cells (20).

The genome encodes complete type I and VI protein secretion systems and the Sec and Tat systems for protein transport to the inner membrane and periplasmic space. A number of proteins that are conserved in Gram-negative bacteria and that function in type II protein secretion or type IV pilus assembly were identified.

*Ca*. N. defluvii has a high demand for iron needed in key components of the respiratory chain, including NXR, which contains several [Fe-S] clusters and is constitutively expressed. A region consisting of 38 CDS (77.5 kbp) is dedicated to the synthesis of siderophores and the import of iron. It includes genes of five nonribosomal peptide synthetases, a polyketide synthase, a class III aminotransferase, a type II thioesterase, and a putative cyclic peptide transporter (Table S2). Recently, a gene cluster of similar composition was shown to be involved in siderophore production in the cyanobacterium *Anabaena* PCC 7120 (21). In *Ca*. N. defluvii, this region also contains genes for the ferric citrate sensor FecR, the iron uptake regulators FecI and Fur, and TonB-dependent uptake systems for $Fe^{3+}$, ferric dicitrate, and ferrichrome-type siderophores (Table S2). An additional putative $Fe^{3+}$ transporter of the ATP binding cassette (ABC) type I family is encoded at a different locus. At another location of the *Ca*. N. defluvii genome, two adjacent genes encode two highly similar (52.9%) subunits of a bacterioferritin that most likely functions in the intracellular storage of iron, but may also play roles in iron and oxygen detoxification (22). A third ferritin-like CDS was identified, which is not in proximity to these two bacterioferritin genes, and a CDS that is remotely similar to a small [2Fe-2S] ferredoxin (BFD) of *E. coli* (Table S2). BFD is thought to be involved in iron-dependent gene regulation and in the release of iron from bacterioferritin (23).

Consistent with molybdenum being another cofactor of NXR, an ABC type I transporter for molybdate was also identified.

**Stress Response and Defense.** *Ca*. N. defluvii possesses a cyanase for cyanate detoxification and a genomic locus with several genes for arsenic resistance, including an arsenite efflux transporter (ArsB), arsenate reductase (ArsC), the arsenic resistance operon regulator ArsR, and a putative arsenite *S*-adenosylmethyltransferase. Interestingly, this genomic region also encodes both subunits of arsenite oxidase (AOX), a member of the DMSO reductase family of molybdenum proteins. AOX could function in arsenite detoxification or enable *Ca*. N. defluvii to use arsenite as electron donor.

The annotation of several β-lactamase–like CDS is consistent with the previously observed resistance of enriched *Ca*. N. defluvii to moderate concentrations of ampicillin (24), whereas two putative tetracycline efflux transporters are contrary to the observed tetracycline sensitivity of *Ca*. N. defluvii (24).

The thioredoxin-dependent peroxiredoxins, which may be involved in $H_2O_2$ protection in *Ca*. N. defluvii, include glutathione peroxidase, thiol peroxidase, and a putative alkylhydroperoxidase (Table S2). Thioredoxin reductase, which is important for the regeneration of reduced thioredoxin as a prerequisite for $H_2O_2$ detoxification by peroxiredoxins, was also identified in the genome.

**Evolutionary History of *Nitrospira*.** Additional support for our hypothesis that *Nitrospira* evolved from anaerobic or microaerophilic ancestors stems from estimating genus divergence times within the Nitrospirae phylum by using 16S rRNA as a molecular clock. We are aware of the limitations of this approach (25), but noted an interesting correlation of the predicted emergence time of the genus *Nitrospira* and geochemical data. By analyzing the current sequence dataset, we found a minimal 16S rRNA similarity of 83.4% within the genus *Nitrospira*, which is considered to contain exclusively $NO_2^-$-oxidizing bacteria. On the basis of an estimated rate of 16S rRNA divergence of 1% per 50 million years (Myr) (26), the radiation of *Nitrospira* took place approximately during the past 830 Myr. The 16S rRNA similarity between the *Nitrospira* and *Leptospirillum* lineages ranges from 75.8% to 82.8% on the basis of current datasets. Almost identical values (75.8–82.6%) were determined for *Nitrospira* and *Thermodesulfovibrio*. Using these values, we estimate that the three lineages shared a common ancestor about 870–1,210 million years ago (Mya). Geochemical data indicate that a significant increase of the atmospheric and oceanic $O_2$ levels began in the late Proterozoic about 850 Mya, whereas Earth was only mildly oxygenated in the preceding $10^9$ years (27). Thus, ancient members of the phylum Nitrospirae most likely existed under conditions favoring an anaerobic or microaerophilic lifestyle. The sharp increase in $O_2$ must have resulted in new ecological niches for those chemolithotrophs that also evolved a sufficient $O_2$ tolerance. We assume that this environmental change gave rise to the lineages *Nitrospira* and *Leptospirillum* and led to their separation from the still anaerobic *Thermodesulfovibrio* lineage. It is interesting to note that the minimal 16S rRNA similarity among all known anammox lineages (83.6%) indicates that ancestral anammox bacteria and *Nitrospira* might have lived in the same era (about 830 Mya).

## SI Materials and Methods

**Genome Sequencing and Annotation.** The same DNA extraction protocol was used for all genomic libraries. Biomass from the *Ca*. N. defluvii enrichment was harvested by centrifugation, and DNA was extracted from the biomass pellet in agarose plugs as described in ref. 28. Shotgun randomly sheared DNA libraries were constructed using a fosmid vector (pCC1FOS; Epicentre Biotechnologies) and low- or high-copy plasmids [pCNS (3 kb insert) and pCDNA2.1 (6 kb insert), respectively]. Terminal clone end sequences were determined using BigDye terminator chemistry and capillary DNA sequencers (model 3730XL; Applied Biosystems) according to standard protocols established at Genoscope. A total of 99,899 Sanger reads (12,565 fosmid ends, 65,702 pCDNA2.1, and 21,632 pCNS plasmid ends) were assembled using Phrap (version 0.960731; http://www.phrap.org) and produced 39 contigs organized into one scaffold. Gap closure and manual finishing was carried out by (*i*) transposon mutagenesis of two regions and (*ii*) PCR amplification and sequencing of specific targeted regions. The complete genome sequence of *Ca*. N. defluvii contains 51,095 Sanger reads, achieving an average of 8.2-fold sequence coverage per base. Only 374 additional Sanger reads were needed during the finishing step. Genome assembly robustness was validated by fosmid coverage coherence (relative orientation and fosmid insert size of about 3,000 fosmids).

The automated analysis pipeline of the MaGe software system (29) was used for the prediction and annotation of CDS. CDS were

predicted using the software AMIGene (30) and then submitted to automatic functional annotation (29). Subsequently, the annotation of the entire genome was refined manually on the basis of the comprehensive set of data collected automatically for each CDS in the relational database "NitrospiraScope" (https://www.genoscope.cns.fr/agc/mage/wwwpkgdb/MageHome/index.php). CDS were assigned to functional categories according to the MultiFun (31) and TIGRFAM (32) functional role catalogs. Proteins with an amino acid identity ≥35% (over at least 80% of the sequence lengths) to characterized proteins in the SwissProt or TrEMBL databases were annotated as homologous to proteins with a known function. Especially in ambiguous cases, information on orthologous relationships retrieved from the clusters of orthologous groups (COG) database, protein signatures collected from the InterPro database, and enzyme profile data provided by PRIAM and HAMAP were used for a tentative functional assignment of annotated genes. CDS with an amino acid identity ≥25% (over at least 80% of the sequence lengths) to characterized proteins or signatures in the aforementioned databases were annotated as putative homologs of the respective database entries. The relatively low thresholds of 35% and 25% sequence identity were chosen to account for the large phylogenetic distance between *Ca*. N. defluvii and most other genome-sequenced microorganisms. CDS with an amino acid identity ≥25% (over at least 80% of the sequence lengths) to uncharacterized proteins were annotated as conserved proteins of unknown function. In the absence of any significant database hit, CDS were annotated as proteins of unknown function and, in the case of uncertain CDS prediction, as doubtful CDS. Finally, CDS with an amino acid identity ≥25% to any database entry over less than 50% of the length of the longer sequence were annotated as modular proteins or protein fragments, respectively. The genomic context of CDS and the functions of flanking genes, as predicted either in *Ca*. N. defluvii or in reference genomes from the PkGDB and NCBI RefSeq databases, were considered during CDS annotation on the basis of the synteny information and visualization that is provided by the MaGe software. Metabolic pathways were reconstructed with the help of the KEGG (33) and MetaCyc (34) pathway tools implemented in MaGe. The 63 COGs, which are present in all genomes in the current COG database (50 bacterial, 13 archaeal, 3 eukaryotic genomes), were identified by using the software EPPS (35) via the online interface (http://web.dmz.uni-wh.de/projects/protein_chemistry/epps/index.php).

**Phylogenetic Analyses.** Amino acid sequence databases of type II DMSO reductase-family molybdopterin cofactor-binding enzymes, of forms I–IV RubisCO and RubisCO-like proteins, and of cyt. *bd* and cyt. *bd*-like oxidases were established using the software ARB (36). Multiple protein sequence alignments were created automatically by ClustalW2 (37) and MUSCLE (38) and were manually refined by using the sequence editor included in the ARB software. Phylogenetic analyses of these proteins were performed by applying distance-matrix, maximum-parsimony, and maximum-likelihood methods: neighbor-joining (with 1,000 bootstrap iterations using the Dayhoff PAM 001 matrix as amino acid substitution model and the implementation in the ARB software package), protein parsimony (PHYLIP version 3.66 with 100 or 1,000 bootstrap iterations), and protein maximum-likelihood [PHYLIP version 3.66, PhyML (39), and TREE-PUZZLE (40) with the Dayhoff PAM 001, Whelan-Goldman, or the JTT substitution model and 1,000 bootstrap iterations using PhyML]. If applicable, N-terminal signal peptide sequences were excluded from the analyses and manually created indel filters were used. To determine the minimal 16S rRNA sequence similarities within the Nitrospirae phylum and among the anammox organisms, pairwise similarity matrices were generated, by using ARB, from all high-quality *Nitrospira* ($n$ = 206), *Leptospirillum* ($n$ = 203), *Thermodesulfovibrio* ($n$ = 28), and anammox planctomycete ($n$ = 140) sequences in the SILVA 100 16S rRNA database (released in August 2009) (41). High-quality sequences were longer than 1,399 nucleotides, had a Pintail (42) score greater than 79, and fell into the monophyletic lineages formed by each of the aforementioned groups. For sequence similarity calculations, the alignment positions 9–1,507 (*E. coli* numbering) were considered. The similarity matrices were exported to spreadsheet software (Microsoft Excel) and the minimal values were extracted for each phylogenetic lineage.

For the calculation of phylogenetic trees for each protein in the proteome, the fully automated software PhyloGenie (43) was used. The reference database for PhyloGenie was generated from the National Center for Biotechnology Information nonredundant protein database NCBI nr (44), in which taxon names were edited to remove characters that control the structure of tree files in the Newick format. The NCBI taxonomy database name file was adapted in a similar manner. The PhyloGenie software was executed for each query protein using default parameters with the following modification: -blammerparams=-taxid f. For the BLAST (45) calculations in PhyloGenie, NCBI BLAST (version 2.2.19) was used. Protein phylogenies were calculated on the basis of full or partial automatic alignments produced by the BLAMMER program included in PhyloGenie. All trees were postprocessed by an in-house script, which sorted all operational taxonomic units according to their distances in the tree to the query protein.

**Incubation of *Ca*. N. defluvii for Expression Analyses of NxrB.** *Ca*. N. defluvii enrichment biomass was starved for 11 (for mRNA analysis) or 110 (for protein analysis) d in mineral medium (24) lacking any energy source. After removing a biomass aliquot for later analysis, 300 μM $NO_2^-$ was added to the medium and the remaining biomass was further incubated for 3 (mRNA analysis) or 8 (protein analysis) d. Biomass from all samples was harvested by centrifugation and stored at −80 °C until further processing.

**Quantification of *Ca*. N. defluvii Cells in the Enrichment.** For the immunological detection of NxrB, total protein had to be extracted from similar numbers of starved or $NO_2^-$-oxidizing *Ca*. N. defluvii cells to ensure that the results were comparable between these treatments. For this purpose, the large cell clusters formed by *Ca*. N. defluvii were disintegrated by bead-beating of biomass with a Fastprep Bead-beater (BIO 101) at level 4 for 5 s. Subsequently, the biomass was harvested by centrifugation (10,000 × $g$, 20 min), and the pellet was resuspended in 1× PBS. As confirmed by FISH with a *Nitrospira*-specific probe (15), this treatment resulted in a cell suspension containing mainly planktonic *Nitrospira* cells and only a few small cell clusters. An aliquot of this suspension was stored at 4 °C for protein extraction. The remaining cell suspension was used for determining the *Nitrospira* cell density by quantitative FISH. It was diluted, paraformaldehyde-fixed according to ref. 46, and defined volumes were filtered onto polycarbonate filters (pore size 0.2 μm, diameter 47 mm, type GTTP; Millipore). The filters were washed two times in 1× PBS and double-distilled water, air dried, and stored at −20 °C. FISH of the *Ca*. N. defluvii cells on the filters was performed according to ref. 47 with the *Nitrospira*-specific 16S rRNA-targeted probes Ntspa1431 (48), Ntspa662, and Ntspa712 (15), which were 5′-labeled with Cy3 and applied simultaneously to increase the signal-to-background ratio. Following FISH, 28 images of each filtered cell suspension were recorded using a confocal laser scanning microscope (LSM 510 Meta; Zeiss), and the average *Nitrospira* cell number per image was determined by visual counting of the probe-labeled cells in each image. The *Nitrospira* cell density in the original (undiluted) cell suspension was then calculated from the average number of cells per image, the known area of one image in square micrometers as reported by the Zeiss imaging software, the known area of the polycarbonate filter, the volume of filtered cell suspension, the dilution factor, and a correction factor. The cor-

rection factor was introduced to account for the possible loss of cells from the filters during FISH. To determine this factor, filter pieces containing biomass were embedded, before or after FISH, in a mixture of the antifadent Citifluor (Citifluor) with a 1:500 diluted Sybr Green II solution (Cambrex) for fluorescent staining of the total bacterial biomass. Subsequently, 17 images of the stained total biomass were recorded (by confocal microscopy) per filter piece. This was done separately with filter pieces that had been embedded before or after FISH. The median area (in pixels) of the biomass in each set of images was measured by using the image analysis software DAIME (49). The ratio of the median biomass areas before and after FISH informed on the extent of cell loss during FISH and was the aforementioned correction factor for the calculation of cell density.

**Transcriptional Analysis of *nxrB*.** Total RNA was extracted from starved or $NO_2^-$-oxidizing *Ca*. N. defluvii enrichment biomass by using TRIzol (Invitrogen) according to the protocol recommended by the manufacturer and with the modifications described by Hatzenpichler et al. (50). After DNA digestion using DNase (Fermentas), reverse transcription of 3 μg total RNA from each treatment was carried out by using the RevertAID first-strand cDNA synthesis kit (Fermentas) according to the manufacturer's instructions. The primers Ntspa1158R (48), specific for the 16S rRNA gene of the genus *Nitrospira*, and nxrBR1237 (GTA GAT CGG CTC TTC GAC CTG), targeting both *nxrB* genes, were used for cDNA synthesis. For cDNA amplification, reaction mixtures with the primer combinations 907F/Ntspa1158R (48, 51) for the 16S rRNA gene and nxrBF916 (GAG CAG GTG GCG CTC CCG C)/nxrBR1237 for the *nxrB* genes, respectively, were prepared according the manufacturer's recommendations in a total volume of 50 μL with 2 mM $MgCl_2$ and 1.25 U of Taq polymerase (Fermentas). For both primer combinations, thermal cycling comprised initial denaturation at 95 °C for 4 min followed by 40 cycles of denaturation at 95 °C for 40 s, annealing at 58 °C for 40 s, and elongation at 72 °C for 60 s. Cycling was completed by a final elongation step at 72 °C for 10 min.

**Translational Analysis of NxrB.** Defined volumes of starved or $NO_2^-$-oxidizing *Ca*. N. defluvii cell suspensions were centrifuged (10,000 × *g* for 20 min) to harvest the biomass. On the basis of the results of quantitative FISH (see above), these aliquots contained approximately the same numbers of *Nitrospira* cells in all experiments. The cell pellet was resuspended in 5× lysis buffer [7 M urea, 2 M thiourea, 20 mg/mL amberlite, 4% (wt/vol) 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate, 40 mM Tris, 2% (vol/vol) IPGbuffer, 0.2% (wt/vol) bromophenol blue, 1% (wt/vol) dithiothreitol, 10% (wt/vol) glycerol], heated for 3 min at 90 °C, and the crude extract was loaded onto a SDS PAGE gel (12.5% polyacrylamide) with a molecular weight marker (PageRuler Prestained Protein Ladder #SM0671; Fermentas). All subsequent steps of the immunological detection of NxrB by Western blotting were performed as described for chlorite dismutase in ref. 52 with the following modifications. The polyvinylidene fluoride membrane was incubated for 30 min with the NxrB-specific monoclonal antibody Hyb 153.3 (53), which had been diluted 1:1,000 in TBS buffer (20 mM Tris, 150 mM NaCl, pH 7.5) containing 0.1% (vol/vol) Tween 20. The secondary antibody (peroxidase-conjugated goat antimouse IgG; Dianova Germany) was diluted 1:5,000 in TBS-Tween buffer. Crude cell extract from *E. coli* BL21 (DE3) expressing recombinant NxrB was used as a positive control in the Western blot experiments.

**Cloning and Heterologous Expression of NxrB.** The two identical *nxrB* genes of *Ca*. N. defluvii were PCR-amplified by using the High Fidelity PCR enzyme mix (Fermentas) according to the protocol recommended by the manufacturer. Instead of extracted genomic DNA, 2 μL of precooked *Ca*. N. defluvii enrichment biomass was added directly to the PCR mix. The applied *nxrB*-specific primers were the forward primer NXRV2 (CGA GCG CAT ATG CCA GAA GTC TAT AAC TGG), which contains an NdeI restriction site upstream of the NxrB start codon, and the reverse primer NXRR (TTA CGA GAA TTC CCC AGC CAG TTC ACG CGC TC), which contains a 5′-EcoRI restriction site. Thermal cycling comprised an initial denaturation step at 94 °C for 5 min followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 56 °C for 40 s, and elongation at 68 °C for 90 s. Cycling was completed by a final elongation step at 68 °C for 10 min. The amplicon was cloned into the vector pCR-XL-TOPO by using the TOPO XL cloning kit (Invitrogen) as recommended by the manufacturer. For heterologous expression, the amplicon was cloned into the expression vector pET21b(+), which contains a promoter for T7 RNA polymerase and a C-terminal His-tag (Novagen), by digestion of amplicon and vector with the restriction endonucleases NdeI and EcoRI followed by ligation with T4 DNA Ligase (Invitrogen) according to the manufacturer's protocol. The expression vector with the *nxrB* gene was first transformed by electroporation into *E. coli* XL1 blue cells (Stratagene). Sanger sequencing confirmed that the cloned *nxrB* gene used for heterologous expression was identical to the *nxrB* genes in the *Ca*. N. defluvii genome. For the expression of NxrB, the vector pET21b (+) containing the *nxrB* gene was transformed into *E. coli* BL21 (DE3) cells (Stratagene). The recombinant cells were grown at 37 °C under agitation (225 rpm) in liquid Luria Bertani medium. After growth up to an optical density (600 nm) of 0.8, the expression of NxrB was induced by adding isopropyl-β-D-thiogalactopyranosidase to a final concentration of 1 mM. Cells were harvested after ∼4 h by centrifugation (5,000 × *g* for 10 min), and the cell pellets were stored at −20 °C.

**Transcriptional Analysis of the Putative Cytochrome *bd*-Like Oxidase.** *Ca*. N. defluvii enrichment biomass was kept under oxic conditions with $NO_2^-$ or was starved for 14 d in mineral medium (24) lacking any energy source. $NO_2^-$ test strips (Merckoquant; Merck) were used to confirm the absence of residual $NO_2^-$ in the starved cultures. Biomass was then harvested by centrifugation, the supernatant was discarded, and the biomass was resuspended in anoxic mineral medium lacking $NO_2^-$ (starved biomass) or containing 3 mM $NO_2^-$ (nonstarved biomass). The anoxic medium had been prepared in accordance with basic principles of medium preparation for strict anaerobes as described by Widdel and Bak (54). Following one additional centrifugation and washing step, the starved and nonstarved biomass was transferred into two separate 150-mL flasks containing 40 mL of anoxic mineral medium and air-free headspace (flushed with $N_2$). Subsequently, biomass aliquots were transferred into 300-mL flasks containing 100 mL of mineral medium. For anoxic treatments, the flasks contained anoxic medium and air-free headspace. For oxic treatments, the headspace of the flasks contained air. All flasks were plugged with butyl rubber stoppers that were fixed with screw caps. To all treatments, 5 $cm^3$ of an $N_2$:$CO_2$ (80:20) gas mixture was added to provide $CO_2$ as carbon source, and all working steps were performed using strictly anoxic techniques. $NO_2^-$ was added to nonstarved cultures to a final concentration of 3 mM, whereas no $NO_2^-$ was added to the starved cultures. On the basis of these procedures, four different incubation conditions were realized: oxic with nitrite, oxic without nitrite, anoxic with nitrite, and anoxic without nitrite. For each treatment, two replicate flasks containing biomass were incubated for 5 d at 30 °C. During incubations with $NO_2^-$, the consumption of $NO_2^-$ was monitored by using test strips, and consumed $NO_2^-$ was replenished. After the incubations, the biomass from each flask was harvested by centrifugation. Total RNA was extracted according to the protocol of Lueders et al. (55) with the modification that samples were kept for 2 h on ice after the addition of polyethylene glycol. After DNA digestion using DNase (Fermentas), PCR was carried out with the

primers 341F and 518R (51, 56), which target bacterial 16S rRNA genes. All these test PCR runs were negative, confirming the absence of residual DNA in the RNA extracts. Subsequently, reverse transcription of ~250 ng total RNA from each treatment was carried out by using the RevertAID first-strand cDNA synthesis kit (Fermentas) according to the manufacturer's instructions. The following primers were used for multiplex cDNA synthesis: Nide0901R1 (CTC GGA AGC ATC GGC CTC AGG), specific for the putative cyt. *bd*-like terminal oxidase of *Ca*. N. defluvii (gene *nide0901*), and 1431R, specific for the 16S rRNA of sublineage I *Nitrospira* (48). For cDNA amplification, reaction mixtures with the primer combinations 1158Fa (modified from ref. 48; ACT GCC CAG GAT AAC GGG)/1431R for the 16S rRNA gene and Nide0901F (GGT GTC TGG GGT TAC TTC GTT)/Nide0901R2 (ACC GTA GAT GTG CCA GTG AAC) for gene *nide0901*, respectively, were prepared according the manufacturer's recommendations in a total volume of 50 μL with 2 mM $MgCl_2$ and 1.25 U of Taq polymerase (Fermentas). For both primer combinations, thermal cycling comprised initial denaturation at 95 °C for 5 min followed by 10 cycles of denaturation at 95 °C for 30 s, annealing at 70–65 °C for 30 s (−0.5 °C in each cycle), and elongation at 72 °C for 40 s. This was followed by 25 cycles of denaturation at 95 °C for 30 s, annealing at 65 °C for 30 s, and elongation at 72 °C for 40 s. Cycling was completed by a final elongation step at 72 °C for 10 min. The specific reverse transcription of the target RNAs was confirmed by Sanger sequencing of the obtained amplicons.

**Stable Carbon Isotopic Fractionation.** Highly enriched cultures of *Ca*. N. defluvii were grown in batch mode in mineral medium (24) in 5-L bottles at 28 °C in the dark. Cell suspensions were moderately stirred, and $NO_2^-$ was replenished from a 2.5 M stock solution. Incubation was performed for about 4 wk until the suspension was turbid. In the second batch culture, $\delta^{13}C_{DIC}$ was monitored over time. Duplicate samples of 25 mL were taken at different stages of growth and transferred to gas-tight tubes, fixed with one drop of 35% (vol/vol) formaldehyde, and stored at 4 °C. Biomass was harvested by centrifugation at 10,000 rpm (15,600 × *g*), washed, suspended in 0.9% NaCl, and frozen at −20 °C.

Analysis of the $\delta^{13}C$ of total dissolved inorganic carbon (DIC) in the medium was performed by headspace analysis of 0.5–1 mL of water that had reacted with $H_3PO_4$ for at least 1 h at room temperature. The headspace was subsequently analyzed ~10 times by using a Thermofinnigan Gas Bench II coupled to a Delta$^{PLUS}$ irmMS system with typical SDs of 0.3‰. Stable carbon isotope ratios were determined relative to laboratory standards calibrated on NBS-18 carbonate [International Atomic Energy Agency (IAEA)]. Differences in $\delta^{13}C_{DIC}$ of the duplicate samples were always <0.5‰. The $\delta^{13}C$ values of the biomass at the end of the batch culture incubation were determined by elemental analysis (EA)/isotope-ratio-monitoring mass spectrometry (EA/irmMS) using a Carlo Erba Flash elemental analyzer coupled to a Thermofinnigan Delta$^{PLUS}$ irmMS system with a reproducibility of ~0.1‰. Stable carbon isotope ratios were determined using laboratory standards calibrated on NBS-22 oil (IAEA).

Lipids were extracted from the harvested biomass and analyzed by gas chromatography (GC) and gas chromatography-mass spectrometry (GC/MS) and GC/isotope-ratio-monitoring/MS as described previously (57). Hydrocarbons were measured in the apolar fraction obtained from the total extract. Bacteriohopanepolyols were transformed into hopanols using periodic acid and sodium borohydride (58). The $\delta^{13}C$ values of individual lipids were corrected for added carbon from derivatization. Values are reported in the usual δ-notation against Vienna Pee Dee Belemnite and represent the average of duplicate runs.


1. Trieber CA, Rothery RA, Weiner JH (1996) Engineering a novel iron-sulfur cluster into the catalytic subunit of *Escherichia coli* dimethyl-sulfoxide reductase. *J Biol Chem* 271: 4620–4626.
2. McDevitt CA, Hugenholtz P, Hanson GR, McEwan AG (2002) Molecular analysis of dimethyl sulphide dehydrogenase from *Rhodovulum sulfidophilum*: Its place in the dimethyl sulphoxide reductase family of microbial molybdopterin-containing enzymes. *Mol Microbiol* 44:1575–1587.
3. Calzolai L, et al. (1995) $^1$H NMR investigation of the electronic and molecular structure of the four-iron cluster ferredoxin from the hyperthermophile *Pyrococcus furiosus*. Identification of Asp 14 as a cluster ligand in each of the four redox states. *Biochemistry* 34:11373–11384.
4. Jormakka M, Richardson D, Byrne B, Iwata S (2004) Architecture of NarGH reveals a structural classification of Mo-bisMGD enzymes. *Structure* 12:95–104.
5. Ikeda T, et al. (2006) Anabolic five subunit-type pyruvate:ferredoxin oxidoreductase from *Hydrogenobacter thermophilus* TK-6. *Biochem Biophys Res Commun* 340:76–82.
6. Blamey JM, Adams MW (1993) Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*.. *Biochim Biophys Acta* 1161:19–27.
7. Yun NR, Yamamoto M, Arai H, Ishii M, Igarashi Y (2002) A novel five-subunit-type 2-oxoglutalate:ferredoxin oxidoreductases from *Hydrogenobacter thermophilus* TK-6. *Biochem Biophys Res Commun* 292:280–286.
8. Goltsman DS, et al. (2009) Community genomic and proteomic analyses of chemoautotrophic iron-oxidizing "*Leptospirillum rubarum*" (group II) and "*Leptospirillum ferrodiazotrophum*" (group III) bacteria in acid mine drainage biofilms. *Appl Environ Microbiol* 75:4599–4615.
9. Lancaster CR (2002) Succinate:quinone oxidoreductases: An overview. *Biochim Biophys Acta* 1553:1–6.
10. Lemos RS, Fernandes AS, Pereira MM, Gomes CM, Teixeira M (2002) Quinol:fumarate oxidoreductases and succinate:quinone oxidoreductases: Phylogenetic relationships, metal centres and membrane attachment. *Biochim Biophys Acta* 1553:158–170.
11. Levicán G, Ugalde JA, Ehrenfeld N, Maass A, Parada P (2008) Comparative genomic analysis of carbon and nitrogen assimilation mechanisms in three indigenous bioleaching bacteria: Predictions and validations. *BMC Genomics* 9:581.
12. Gomes CM, et al. (1999) The unusual iron sulfur composition of the *Acidianus ambivalens* succinate dehydrogenase complex. *Biochim Biophys Acta* 1411:134–141.
13. Tsugawa H, et al. (2008) Alpha-ketoglutarate oxidoreductase, an essential salvage enzyme of energy metabolism, in coccoid form of *Helicobacter pylori*.. *Biochem Biophys Res Commun* 376:46–51.
14. Hughes NJ, Clayton CL, Chalk PA, Kelly DJ (1998) *Helicobacter pylori* porCDAB and oorDABC genes encode distinct pyruvate:flavodoxin and 2-oxoglutarate:acceptor oxidoreductases which mediate electron transport to NADP. *J Bacteriol* 180: 1119–1128.
15. Daims H, Nielsen JL, Nielsen PH, Schleifer KH, Wagner M (2001) In situ characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microbiol* 67:5273–5284.
16. Paiva S, Devaux F, Barbosa S, Jacq C, Casal M (2004) Ady2p is essential for the acetate permease activity in the yeast *Saccharomyces cerevisiae*. *Yeast* 21:201–210.
17. Vignais PM, Billoud B (2007) Occurrence, classification, and biological function of hydrogenases: An overview. *Chem Rev* 107:4206–4272.
18. Rowe JJ, Ubbink-Kok T, Molenaar D, Konings WN, Driessen AJ (1994) NarK is a nitrite-extrusion system involved in anaerobic nitrate respiration by *Escherichia coli*. *Mol Microbiol* 12:579–586.
19. Ehrich S, Behrens D, Lebedeva E, Ludwig W, Bock E (1995) A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Arch Microbiol* 164:16–23.
20. Watson SW, Bock E, Valois FW, Waterbury JB, Schlosser U (1986) *Nitrospira marina* gen. nov. sp. nov.: A chemolithotrophic nitrite-oxidizing bacterium. *Arch Microbiol* 144:1–7.
21. Jeanjean R, et al. (2008) A large gene cluster encoding peptide synthetases and polyketide synthases is involved in production of siderophores and oxidative stress response in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Environ Microbiol* 10: 2574–2585.
22. Carrondo MA (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J* 22:1959–1968.
23. Quail MA, et al. (1996) Spectroscopic and voltammetric characterisation of the bacterioferritin-associated ferredoxin of *Escherichia coli*. *Biochem Biophys Res Commun* 229:635–642.
24. Spieck E, et al. (2006) Selective enrichment and molecular characterization of a previously uncultured *Nitrospira*-like bacterium from activated sludge. *Environ Microbiol* 8:405–415.
25. Kuo CH, Ochman H (2009) Inferring clocks when lacking rocks: The variable rates of molecular evolution in bacteria. *Biol Direct* 4:35.
26. Ochman H, Wilson AC (1987) Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26:74–86.
27. Holland HD (2006) The oxygenation of the atmosphere and oceans. *Philos Trans R Soc Lond B Biol Sci* 361:903–915.
28. Strous M, et al. (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* 440:790–794.
29. Vallenet D, et al. (2006) MaGe: A microbial genome annotation system supported by synteny results. *Nucleic Acids Res* 34:53–65.

30. Bocs S, Cruveiller S, Vallenet D, Nuel G, Médigue C (2003) AMIGene: Annotation of MIcrobial Genes. *Nucleic Acids Res* 31:3723–3726.
31. Serres MH, Riley M (2000) MultiFun, a multifunctional classification scheme for *Escherichia coli* K-12 gene products. *Microb Comp Genomics* 5:205–222.
32. Haft DH, et al. (2001) TIGRFAMs: A protein family resource for the functional identification of proteins. *Nucleic Acids Res* 29:41–43.
33. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30.
34. Caspi R, et al. (2006) MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* 34 (Database issue):D511–D516.
35. Reichard K, Kaufmann M (2003) EPPS: Mining the COG database by an extended phylogenetic patterns search. *Bioinformatics* 19:784–785.
36. Ludwig W, et al. (2004) ARB: A software environment for sequence data. *Nucleic Acids Res* 32:1363–1371.
37. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
38. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
39. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
40. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.
41. Pruesse E, et al. (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35: 7188–7196.
42. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* 71:7724–7736.
43. Frickey T, Lupas AN (2004) PhyloGenie: Automated phylome generation and analysis. *Nucleic Acids Res* 32:5231–5238.
44. Sayers EW, et al. (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 38 (Database issue):D5–D16.
45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
46. Daims H, Stoecker K, Wagner M (2005) Fluorescence in situ hybridisation for the detection of prokaryotes. *Molecular Microbial Ecology*, eds Osborn AM, Smith CJ (Bios-Garland, Abingdon, United Kingdom), pp 213–239.
47. Glöckner FO, et al. (1996) An in situ hybridization protocol for detection and identification of planctonic bacteria. *Syst Appl Microbiol* 19:403–406.
48. Maixner F, et al. (2006) Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. *Environ Microbiol* 8:1487–1495.
49. Daims H, Lücker S, Wagner M (2006) *daime*, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* 8:200–213.
50. Hatzenpichler R, et al. (2008) A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc Natl Acad Sci USA* 105:2134–2139.
51. Lane DJ, et al. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA* 82:6955–6959.
52. Maixner F, et al. (2008) Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium '*Candidatus* Nitrospira defluvii'. *Environ Microbiol* 10:3043–3056.
53. Aamand J, Ahl T, Spieck E (1996) Monoclonal antibodies recognizing nitrite oxidoreductase of *Nitrobacter hamburgensis*, *N. winogradskyi*, and *N. vulgaris*. *Appl Environ Microbiol* 62:2352–2355.
54. Widdel F, Bak F (1992) Gram-negative mesophilic sulfate-reducing bacteria. *The Prokaryotes*, eds Balows A, Trüper HG, Dworking M, Harder W, Schleifer K-H (Springer, New York), 2nd Ed, pp 3352–3378.
55. Lueders T, Manefield M, Friedrich MW (2004) Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environ Microbiol* 6:73–78.
56. Edwards U, Rogall T, Blöcker H, Emde M, Böttger EC (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res* 17:7843–7853.
57. Schouten S, et al. (1998) Biosynthetic effects on the stable carbon isotopic compositions of algal lipids: Implications for deciphering the carbon isotopic biomarker record. *Geochim Cosmochim Acta* 62:1397–1406.
58. Rohmer M, Bouvier-Navé P, Ourisson G (1984) Distribution of hopanoid triterpenes in prokaryotes. *J Gen Microbiol* 130:1137–1150.

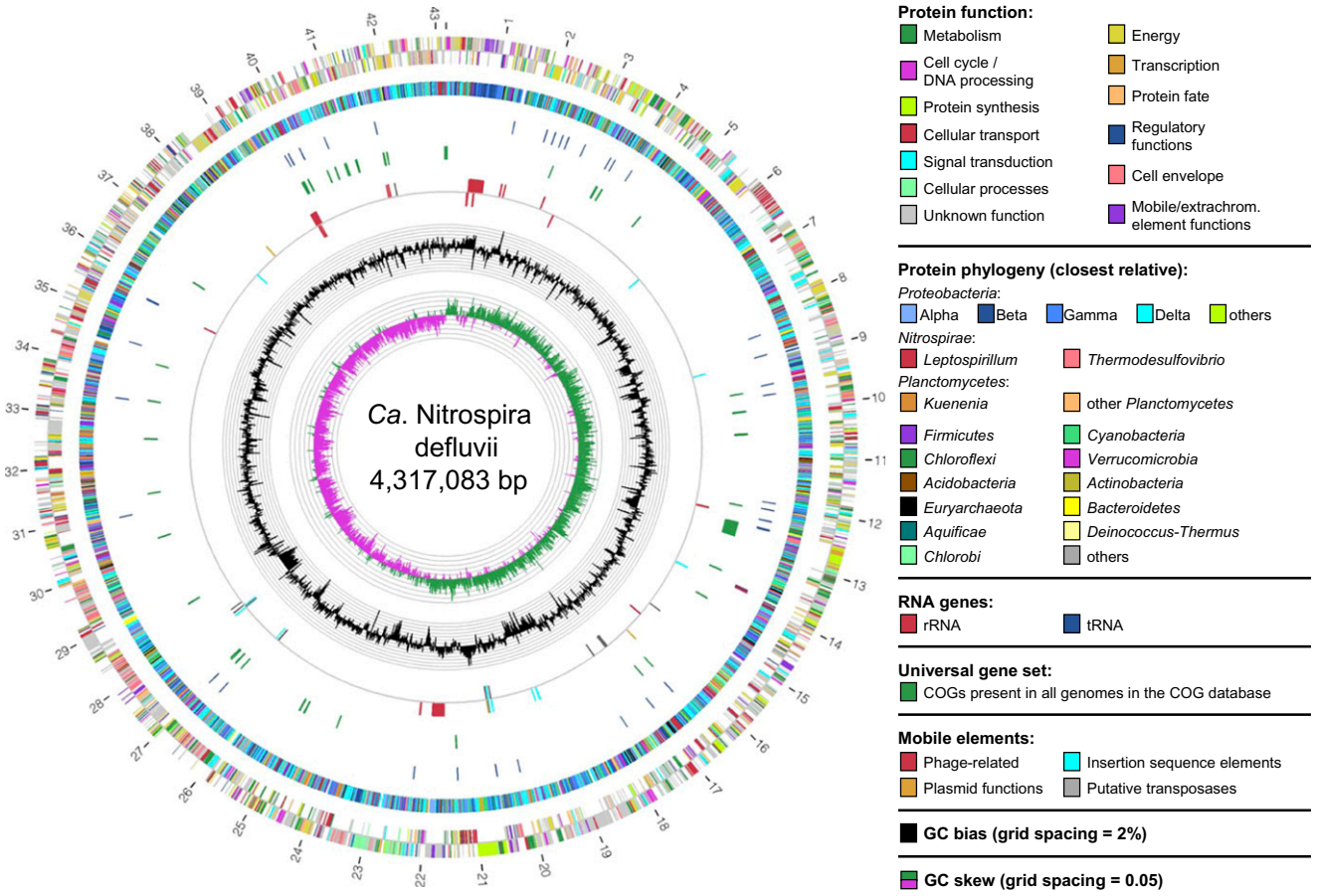


**Fig. S1.** Circular representation of the *Ca*. N. defluvii chromosome. Annotated coding sequences (rings 1 + 2), genome-wide protein phylogeny (ring 3), RNA genes (ring 4), universal clusters of orthologous groups (ring 5), mobile genetic elements (rings 6 + 7), and local nucleotide composition measures (rings 8 + 9) are shown. Very short features were enlarged to enhance visibility. Clustered genes, such as several tRNA genes, may appear as one line due to space limitations. The most closely related homolog in other sequenced genomes was determined for each protein of *Ca*. N. defluvii (ring 3) by using a genome-wide phylogenetic approach (*SI Materials and Methods*). Three larger clusters of phage-related genes (in rings 6 + 7) probably represent prophages or remnants thereof. The image was created by using the software Circos (1).

1. Krzywinski M, et al. (2009) Circos: An information aesthetic for comparative genomics. *Genome Res* 19:1639–1645.

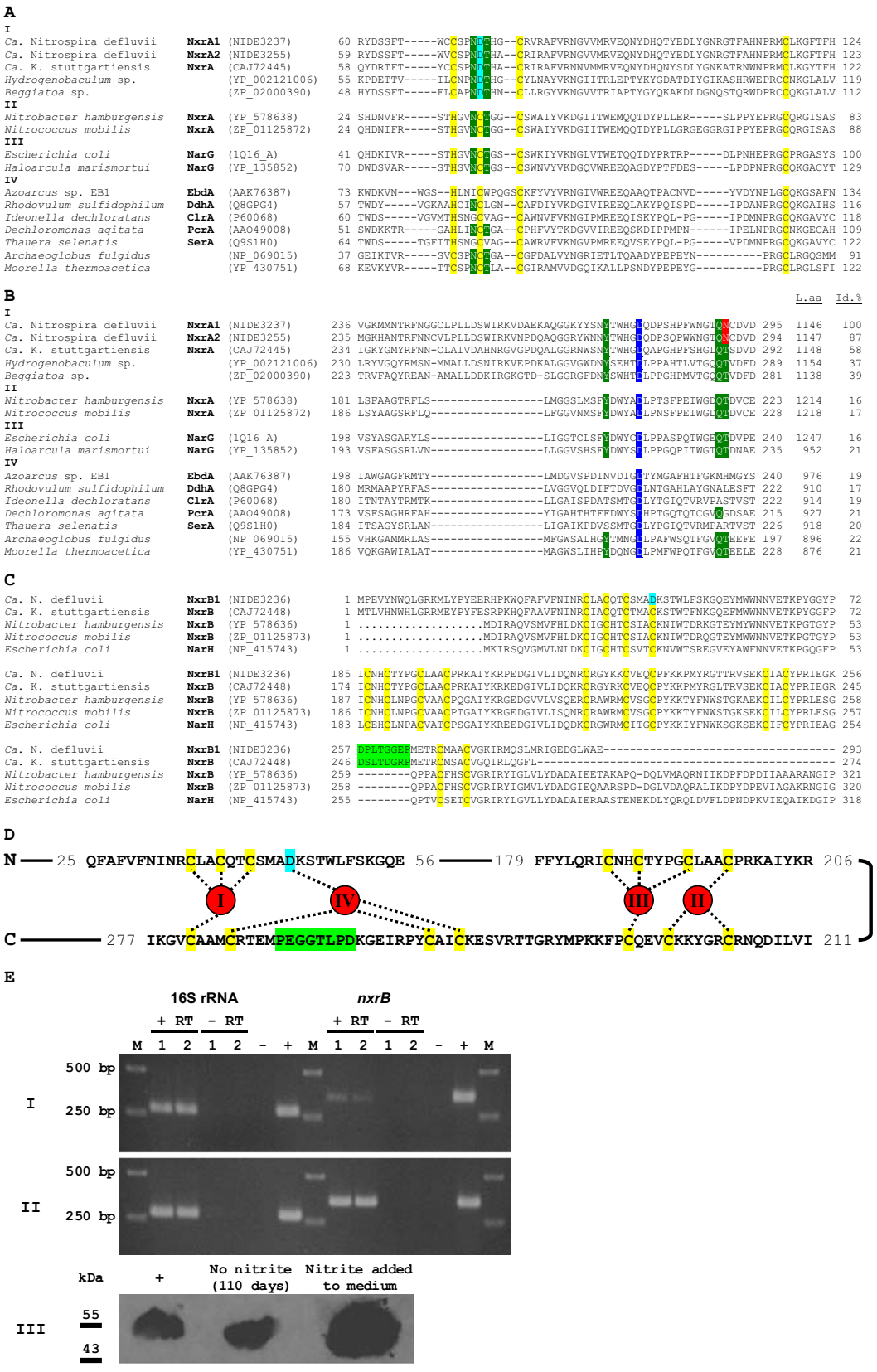


**Fig. S2.** Metal-coordinating regions in the α- and β-subunits of nitrite oxidoreductase and expression of the β-subunit. (*A* and *B*) Alignments of metal-coordinating regions in the α-subunits of selected molybdopterin-binding enzymes belonging to the type II group in the dimethyl sulfoxide (DMSO) reductase family (1, 2). Sequences are vertically grouped according to their known or putative functions: (I) and (II) nitrite oxidoreductases (Nxr) and closely related enzymes; (III) nitrate reductases (Nar); (IV) other or unknown functions (Ebd, ethylbenzene dehydrogenase; Ddh, dimethylsulfide dehydrogenase; Clr, chlorate

Legend continued on following page

reductase; Pcr, perchlorate reductase; Ser, selenate reductase). Accession numbers of the sequences are indicated in parentheses. Five signature residues, which are conserved in nitrate reductases (3) and nitrite oxidoreductases, are highlighted in green. Note that, on the basis of these residues, the yet-uncharacterized enzymes of *Archeoglobus fulgidus* and *Moorella thermoacetica* might use nitrate and/or nitrite as substrate. Asparagine (highlighted in red) replaces one threonine signature residue in both NxrA copies of *Ca*. N. defluvii. (*A*) Iron-sulfur binding center. Known [Fe-S]-binding residues in NarG of *E. coli* (2) and the homologous positions in the other sequences are highlighted in yellow. The aspartate residue, which replaces one cysteine as a putative [Fe-S] ligand in the NxrA subunits of *Ca*. N. defluvii and the related enzymes, is highlighted in cyan. Aspartate was previously shown to function as [Fe-S] ligand in a ferredoxin of *P. furiosus* (4). (*B*) Molybdenum ligand binding site. The conserved aspartate residue (highlighted in blue) acts as the molybdenum ligand in NarG of *E. coli* (2). The last two columns show the overall length of the amino acid sequences (L. aa) and amino acid sequence identities (Id.%) to NxrA1 of *Ca*. N. defluvii. (*C* and *D*) [Fe-S] cluster-coordinating regions in the β-subunits of nitrite oxidoreductases (NxrB) and nitrate reductase A of *E. coli* (NarH), which belong to the type II group in the DMSO reductase family (1, 2). (*C*) Protein sequence alignments of the relevant regions. For *Ca*. N. defluvii, only one of the two identical NxrB copies is shown. Known [Fe-S]-binding residues in NarH (5) and the homologous positions in the NxrB sequences are highlighted in yellow. An aspartate residue, which replaces one cysteine as a putative [Fe-S] ligand in both NxrB copies of *Ca*. N. defluvii, is highlighted in cyan. Aspartate was previously shown to function as [Fe-S] ligand in a ferredoxin of *P. furiosus* (4). An insertion, which is found only in the NxrB of *Ca*. N. defluvii and *Ca*. K. stuttgartiensis, is highlighted in green (see also *D*). Accession numbers of the sequences are indicated in parentheses. (*D*) Schematic of the putative [Fe-S] cluster coordination in NxrB of *Ca*. N. defluvii. Red circles represent putative [Fe-S] clusters, which are numbered I–IV as done previously for NarH of *E. coli* (5). Probable [Fe-S]-coordinating residues are highlighted in yellow and cyan and were assigned to the four [Fe-S] clusters on the basis of the known coordination pattern of the homologous residues in NarH (5). An insertion, which is found only in the region coordinating [Fe-S] cluster IV in the NxrB subunits of *Ca*. N. defluvii and *Ca*. K. stuttgartiensis, is highlighted in green (see also *C*). Numbers indicate the corresponding positions in the alignments shown in *C*. N, N terminus; C, C terminus of the protein. (*E*) Expression of nitrite oxidoreductase (β-subunit) by enriched *Ca*. N. defluvii. (I) Detection of *Ca*. N. defluvii 16S rRNA and *nxrB* mRNA after 11 d of starvation in $NO_2^-$-free mineral medium. +RT, RNA detection by reverse transcription PCR; −RT, PCR control for DNA contamination in the RNA extract; +, positive control, with use of a cloned 16S rRNA or *nxrB* gene fragment; −, negative control without nucleic acids; 1 and 2, biological replicates; M, size marker. (II) Detection of *Ca*. N. defluvii 16S rRNA and *nxrB* mRNA 3 d after addition of 300 μM $NO_2^-$ to the starved enrichment. Labels as in I. (III) Immunological detection of NxrB in enriched *Ca*. N. defluvii after 110 d of starvation in $NO_2^-$-free mineral medium and 8 d after addition of 300 μM $NO_2^-$ to the starved enrichment. Total protein extracts were prepared from similar numbers of *Ca*. N. defluvii cells. +, positive control, with NxrB of *Ca*. N. defluvii heterologously expressed in *E. coli*.


1. Kisker C, et al. (1998) A structural comparison of molybdenum cofactor-containing enzymes. *FEMS Microbiol Rev* 22:503–521.
2. Jormakka M, Richardson D, Byrne B, Iwata S (2004) Architecture of NarGH reveals a structural classification of Mo-bisMGD enzymes. *Structure* 12:95–104.
3. Martinez-Espinosa RM, et al. (2007) Look on the positive side! The orientation, identification and bioenergetics of 'Archaeal' membrane-bound nitrate reductases. *FEMS Microbiol Lett* 276:129–139.
4. Calzolai L, et al. (1995) $^1$H NMR investigation of the electronic and molecular structure of the four-iron cluster ferredoxin from the hyperthermophile *Pyrococcus furiosus*. Identification of Asp 14 as a cluster ligand in each of the four redox states. *Biochemistry* 34:11373–11384.
5. Blasco F, et al. (2001) The coordination and function of the redox centres of the membrane-bound nitrate reductases. *Cell Mol Life Sci* 58:179–193.

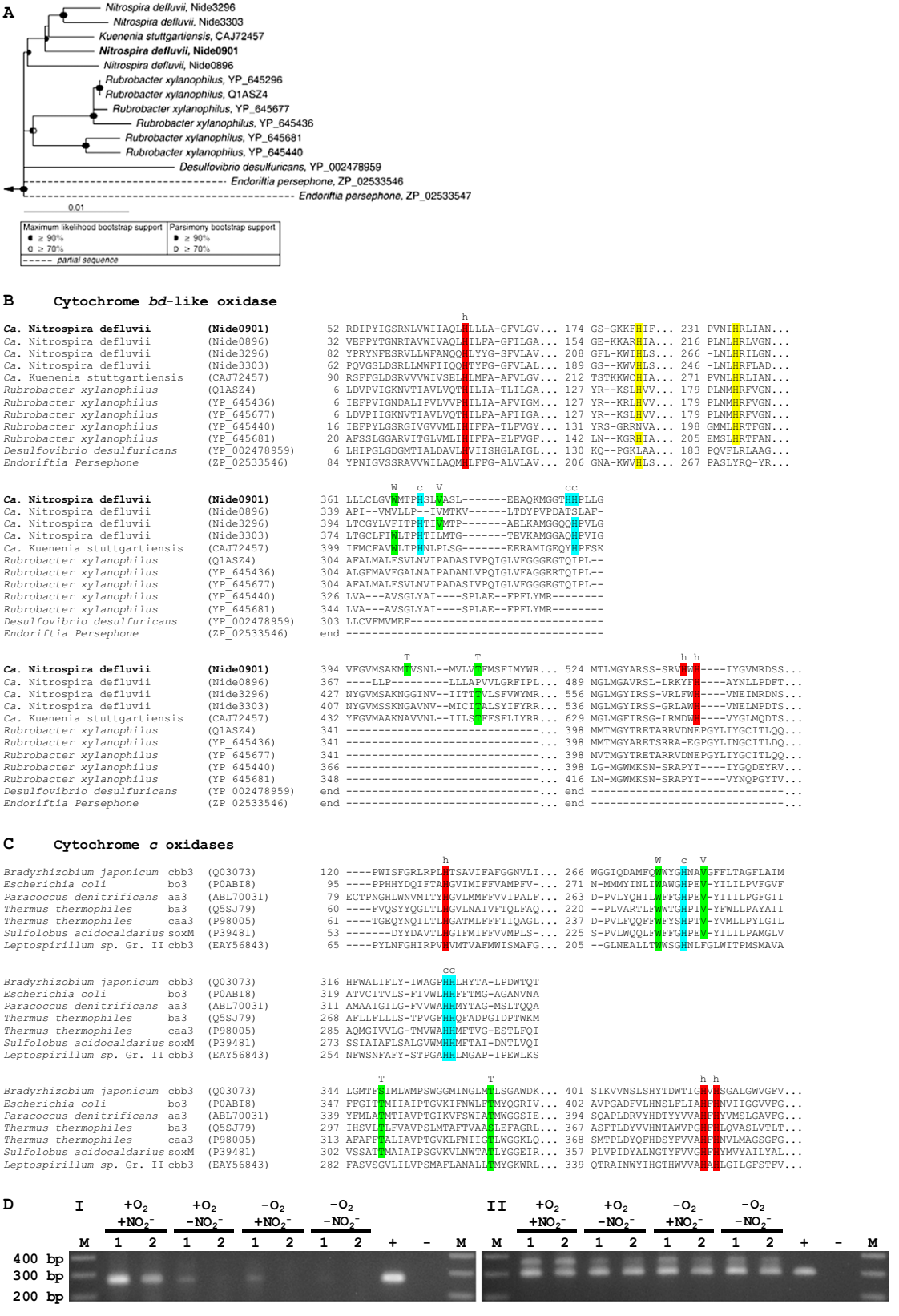


**Fig. S3.** Analysis of the putative cytochrome *bd*-like oxidase family. (*A*) Consensus tree showing the relationship of all known proteins belonging to the putative enzyme family. Tree topology is based on maximum-likelihood and maximum-parsimony methods, using a filter excluding insertions and deletions of single sequences. In total, 628 alignment positions were considered for the phylogenetic analyses. Branch lengths were derived from the maximum-likelihood tree. Bootstrap values are based on 1,000 iterations. In *Endoriftia persephone*, the N- and C-terminal parts of the protein are split in two CDS, and each part was added to the trees by using the ARB Parsimony Interactive method. Canonical cyt. *bd* oxidases were used as outgroup. The putative terminal cytochrome *c*

Legend continued on following page

oxidase of *Ca*. N. defluvii (Nide0901) is indicated in boldface. (*B*) Multiple sequence alignment of all known members of the cytochrome *bd*-like oxidase family. Alignment positions putatively involved in function as terminal cyt. *c* oxidase (1) are highlighted in color and by a letter indicating function. Red: Histidine residues involved in binding of heme groups (h). Yellow: Alternative histidines for heme interaction. Turquoise: Histidine residues involved in copper ($Cu_B$) binding (c). Amino acids conserved in all cyt. *c* oxidases are labeled green. Note that only Nide0901 contains all residues conserved also in bona fide heme-copper cyt. *c* oxidases (see also *C*). (*C*) Multiple sequence alignment of heme-copper cyt. *c* oxidases belonging to type A, B, and C. The $cbb_3$-like enzyme from *Leptospirillum* is also shown. Conserved residues are indicated as in *B*. (*D*) Expression of the putative terminal cytochrome *c* oxidase (Nide0901) by enriched *Ca.* N. defluvii. Detection of *nide0901* mRNA (I) and *Ca.* N. defluvii 16S rRNA (II) by reverse transcription PCR after 5 d of incubation under oxic ($+O_2$) or anoxic ($-O_2$) conditions in $NO_2^-$-containing ($+NO_2^-$) or $NO_2^-$-free ($-NO_2^-$) mineral media. 1 and 2, biological replicates; +, positive control with genomic DNA extracted from the enrichment; –, negative control without nucleic acids; M, size marker. In II, the 16S rRNA of *Ca*. N. defluvii is represented by the lower band as confirmed by Sanger sequencing.

1. Pereira MM, Santana M, Teixeira M (2001) A novel scenario for the evolution of haem-copper oxygen reductases. *Biochim Biophys Acta* 1505:185–208.

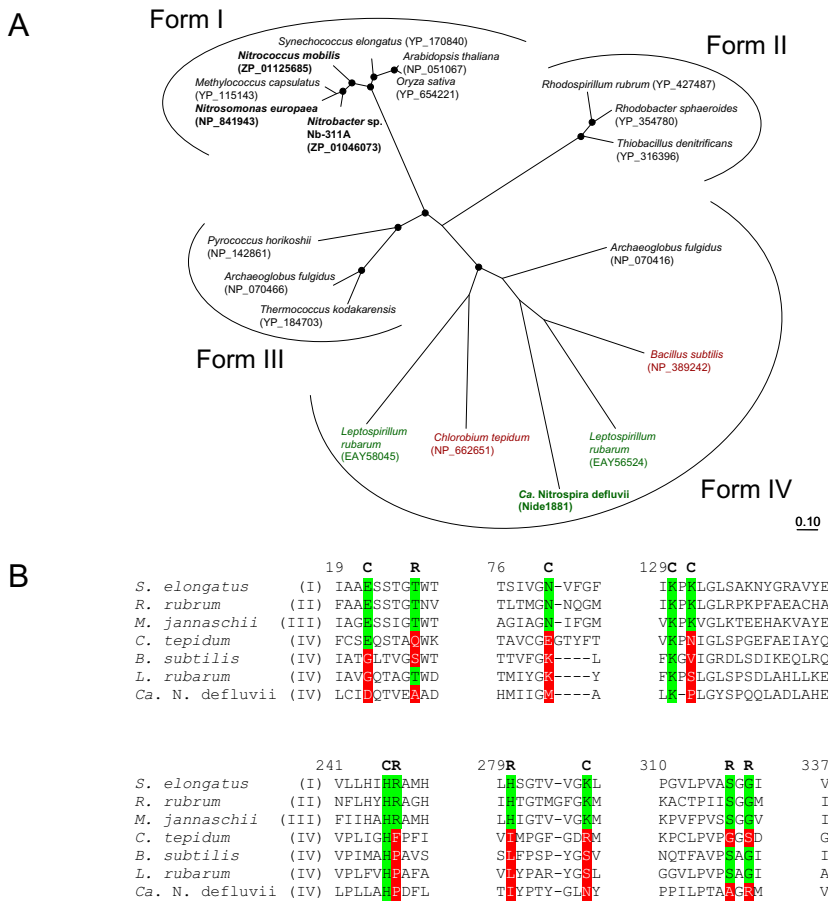


**Fig. S4.** A form IV RubisCO-like protein of *Ca*. N. defluvii. (*A*) Maximum-likelihood phylogenetic analysis of the large subunits of selected ribulose-1,5-bisphosphate carboxylases (RubisCO, forms I–III) and RubisCO-like (form IV) proteins. In total, 484 amino acid positions were considered. Names of ammonia- or nitrite-oxidizing bacteria are in boldface. Names of members of the phylum Nitrospirae are in green. Enzymes of organisms, whose names are in red, have been demonstrated to lack the carboxylating activity of bona fide RubisCO and are involved in sulfur metabolism, oxidative stress response, or methionine biosynthesis (1, 2). Black dots indicate high (> 90%) parsimony bootstrap (100 iterations) support of the respective nodes. The scale bar indicates 10% estimated sequence divergence. Sequence accession numbers are indicated in parentheses. (*B*) Partial amino acid sequence alignment of the large subunits of selected forms I–III RubisCO and form IV RubisCO-like proteins. Known active site residues of RubisCO (1) are highlighted in green, whereas substitutions at the homologous positions in form IV RubisCO-like proteins are highlighted in red. C, residues involved in the catalytic mechanism; R, residues involved in binding of ribulose-1,5-bisphosphate; GXDFXKXDE, conserved RubisCO signature motif (1). The alignment is numbered according to the *Ca*. N. defluvii (Nide1881) sequence and is based on an alignment published by Hanson and Tabita (1).

1. Hanson TE, Tabita FR (2001) A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. *Proc Natl Acad Sci USA* 98:4397–4402.
2. Ashida H, et al. (2003) A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science* 302:286–290.

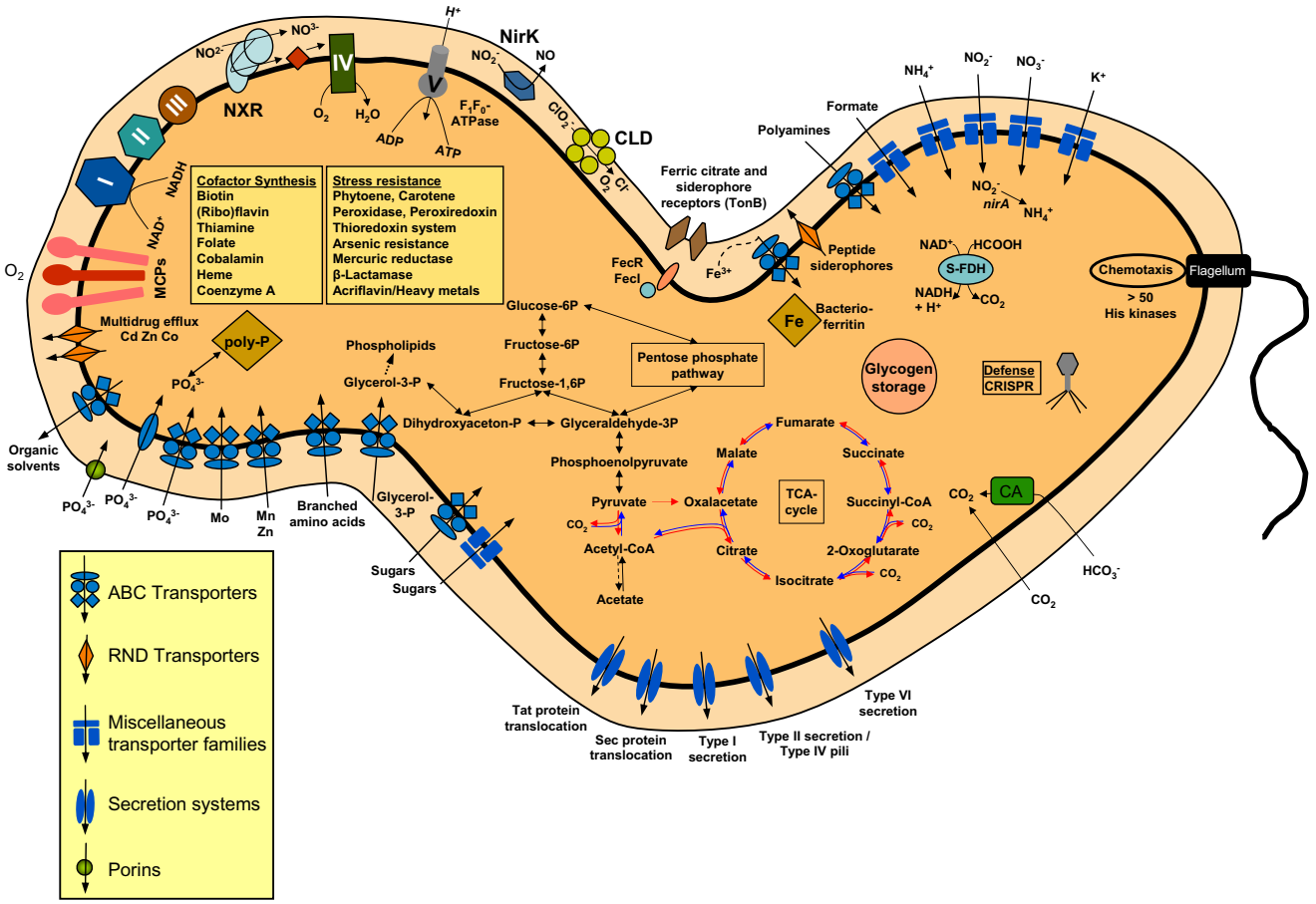


**Fig. S5.** Cell metabolic cartoon constructed from the annotation of the *Ca*. N. defluvii genome. CLD, chlorite dismutase; CA, carbonic anhydrase; MCPs, methyl-accepting chemotaxis proteins; S-FDH, soluble fumarate dehydrogenase; *nirA*, ferredoxin-nitrite reductase. Enzyme complexes of the electron transport chain are labeled by Roman numerals (see Fig. 1 in main text for details). Red arrows depict the oTCA and blue arrows the rTCA cycle, respectively.

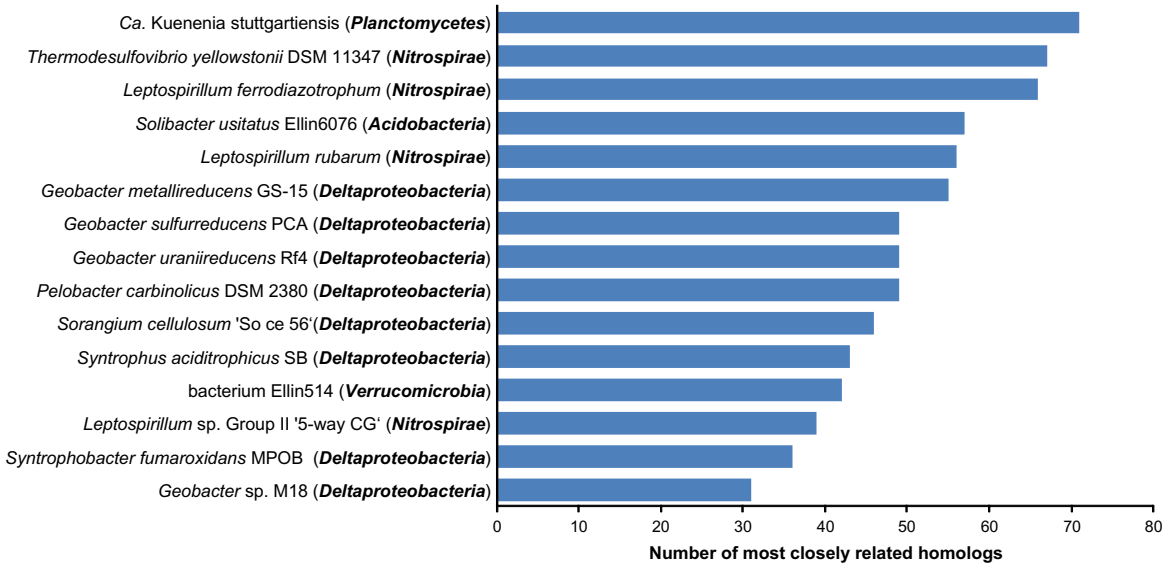


**Fig. S6.** Summarized results of a phylogenetic analysis of each protein in the *Ca*. N. defluvii proteome. The nearest phylogenetic neighbor (closest homolog) in other sequenced genomes was determined for each protein of *Ca*. N. defluvii. The graph depicts the 15 organisms that contain the highest numbers of closest homologs to *Ca*. N. defluvii per individual genome. Names of bacterial phyla (or classes of the *Proteobacteria*) are in boldface. Because only the closest homolog of each protein was considered, the analysis presented is nonredundant, and each protein of *Ca*. N. defluvii was counted only once. Closest homologs were identified in the listed 15 bacteria and in 624 other organisms (lower numbers per genome) for, in total, 3,109 proteins (72.7% of all CDS in the *Ca*. N. defluvii genome).

## Other Supporting Information Files

Table S1 (DOC)
Table S2 (DOC)
Table S3 (DOC)