universität wien

DISSERTATION

Titel der Dissertation

Exploring the ecology and genomics
of globally important nitrite-oxidizing bacteria

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr. rer.nat.)

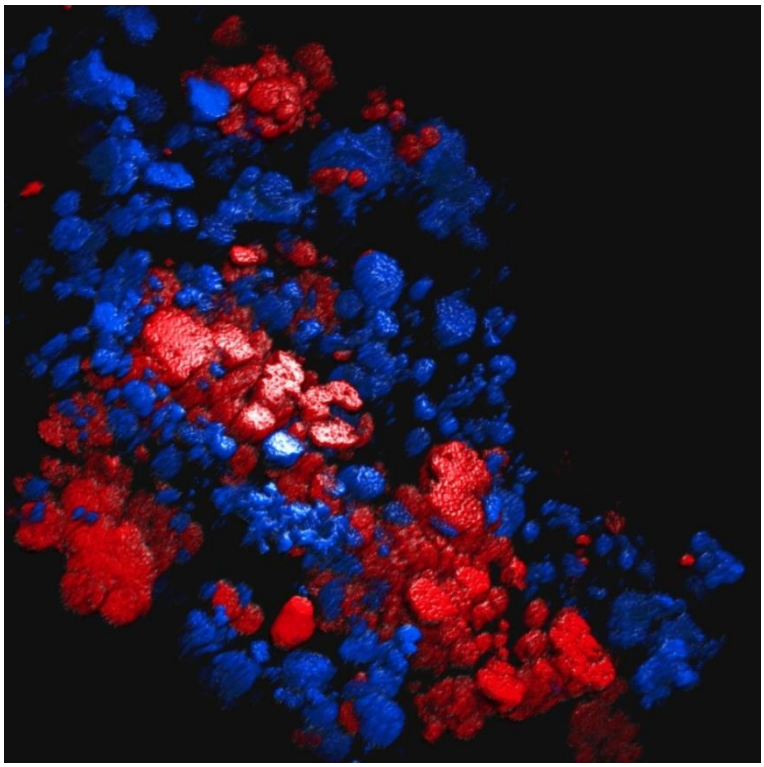
Verfasser:	Sebastian Lücker
Matrikel-Nummer:	0409616
Dissertationsgebiet (lt. Studienblatt):	444 Ökologie
Betreuerin / Betreuer:	Univ.-Prof. Dr. Michael Wagner

Wien, am 25. Oktober 2010

Contents

Chapter I

Introduction & Outline



Front: 3D visualization of a nitrifying bacterial community in a sequencing batch reactor biofilm. Nitrite-oxidizing bacteria of the genus *Nitrospira* are colored in red, betaproteobacterial ammonia-oxidizers in blue.

The nitrogen cycle

Nitrogen is one of the key elements for life. Besides constituting 78% of our atmosphere, it also is a major component of the cell's building blocks, such as nucleic acids and proteins. Nitrogen can exist in an extraordinary range of oxidation states, spanning from –III in ammonia to +V in nitrate. This versatility enables nitrogen to exist and to be transformed into a huge array of different molecules. Due to biological transformations as well as chemical instability most oxidation states of nitrogen have only a transient existence. Dinitrogen represents the chemically most inert and therefore also most frequent state. The sum of all transformation reactions forms the biogeochemical nitrogen cycle (Figure 1), with most of the steps being catalyzed exclusively by microorganisms. As the input source of the cycle, nitrogen fixing archaea (Murray and Zinder, 1984) and bacteria reduce the inert, gaseous dinitrogen to ammonia (Postgate, 1970). During assimilation, ammonia is incorporated into biomass, mainly in the form of amino acids (Zehr and Ward, 2002). From these biological nitrogen compounds, the ammonia is released through microbial ammonification, also termed mineralization (McLain and Martens, 2005). The mineralized ammonia subsequently is used for nitrification, the stepwise oxidation of ammonia to nitrate by highly specialized chemolithoautotrophic organisms (Prosser, 1989). While various heterotrophic organisms also are able to oxidize ammonia to nitrate, this heterotrophic nitrification is not coupled to energy generation (Focht and Verstraete, 1977). The nitrate formed can then either be assimilated into cell material (Bock and Wagner, 2006) or, under oxygen-limited or anoxic conditions, be used as terminal electron acceptor for anaerobic respiration. Here nitrate is, on the one hand, converted to ammonia by respiratory ammonification (Simon, 2002) and, on the other, reduced to dinitrogen by denitrification (Zumft, 1997), thus closing the cycle.

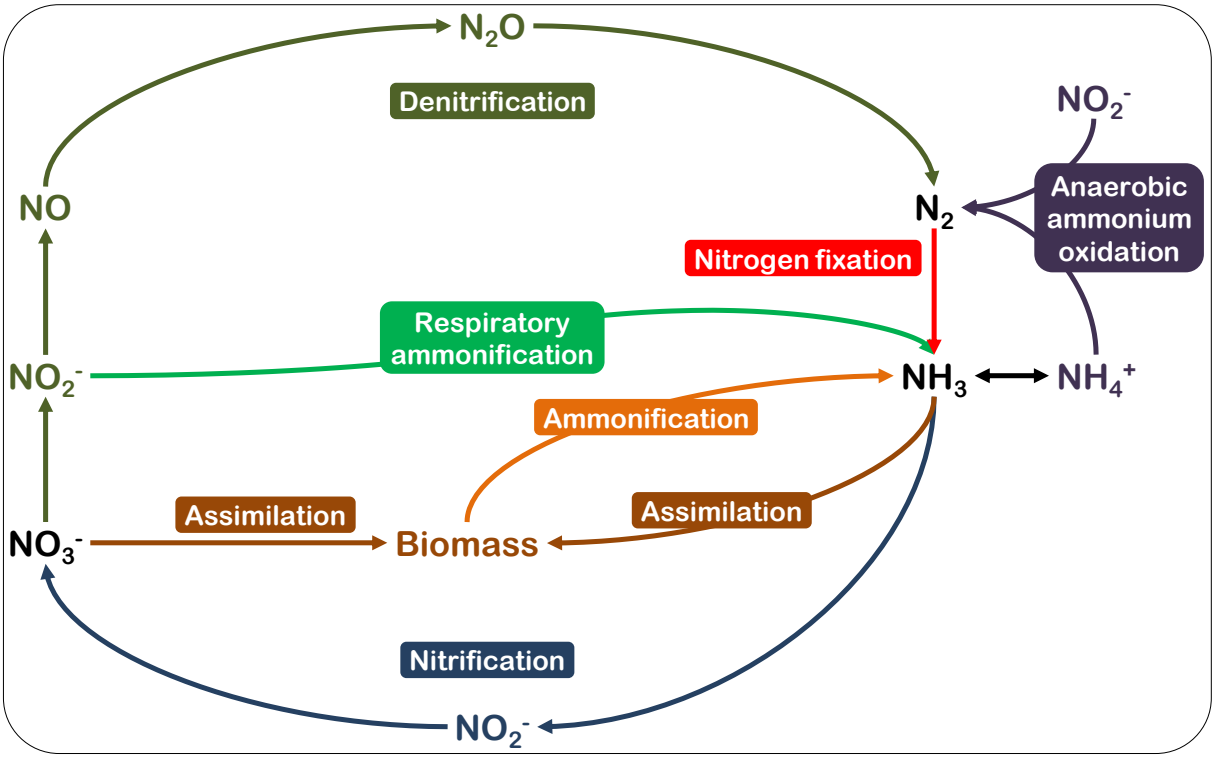


Figure 1. Schematic representation of the biogeochemical nitrogen cycle.

In some organisms only partial denitrification takes place, leading to the formation of nitric oxide or nitrous oxide (Bock *et al.*, 1995). Incomplete denitrification to nitric or nitrous oxide can also be observed in oxic habitats, where it mainly serves as nitrite detoxification system (Lipschultz *et al.*, 1981). Furthermore, some members of the phylum *Planctomycetes* realize a shortcut from ammonium to dinitrogen in anoxic environments. These bacteria are able to couple anaerobic ammonium oxidation (anammox) directly to the reduction of nitrite, an energy-yielding reaction which also results in the release of gaseous dinitrogen (Jetten *et al.*, 2009).

The role of wastewater treatment in the nitrogen cycle

Nitrogen fluxes through the global nitrogen cycle have been greatly altered by anthropogenic influences over the past decades. Nowadays, more than half of the nitrogen entering the cycle is derived from ammonia production via the Haber-Bosch process and the cultivation of nitrogen-fixing crops. Also, fossil fuel combustion leads to the formation of large amounts of nitrogen oxides (Galloway and Cowling, 2002; Klotz and Stein, 2008). Besides nitrogen-based fertilizers used in agriculture, household sewage and industrial waste are among the largest sources of anthropogenic ammonia. Release of excess ammonia into aquatic systems leads to eutrophication and hypoxia of the affected water bodies (Diaz and Rosenberg, 2008). Together with the toxicity of ammonia and nitrite on aquatic life (Camargo and Alonso, 2006), this emphasizes the importance of efficient nitrogen removal from wastewater in order to preserve environmental health.

In wastewater treatment plants, tertiary treatment achieves the goal of nitrite elimination. Here, after ammonification and removal of most carbon compounds from the wastewater, oxic and anoxic conditions are applied intermittently or subsequently to provide conditions for nitrification and denitrification, respectively. Since most denitrifying bacteria rely on the availability of organic electron donors and carbon sources, this step often is realized by admixing methanol or ethanol as external carbon and energy source, or by sludge circulation techniques, that recirculate the activated sludge from the nitrifying into the denitrifying stage which also receives carbon-rich influent wastewater. Application of tertiary treatment leads to near-complete removal of inorganic nitrogen from wastewater, either in the form of gaseous dinitrogen or by incorporation into biomass, thus preventing detrimental nutrition discharge into the attached and downstream water systems.

Nitrifying microorganisms

Autotrophic nitrification, the stepwise oxidation of ammonia to nitrate, is catalyzed by two different functional groups of microorganisms (Figure 2). First, ammonia is oxidized via hydroxylamine to nitrite by ammonia-oxidizing bacteria (AOB) (Bock *et al.*, 1991) and, as discovered more recently, archaea (AOA) (Könneke *et al.*, 2005; Francis *et al.*, 2007). Nitrite then is released and serves as substrate for nitrite-oxidizing bacteria (NOB), which further oxidize it to nitrate, the end product of aerobic nitrification.

All AOB known to date belong to the *Beta-* or *Gammaproteobacteria*. The genera *Nitrosomonas* (including *Nitrosococcus mobilis*), *Nitrosolobus*, *Nitrosospira*, and *Nitrosovibrio* (Teske *et al.*, 1994) all are affiliated with the family *Nitrosomonadaceae* within the *Betaproteobacteria* (Purkhold *et al.*, 2003). While some studies suggested combining *Nitrosolobus* and *Nitrosovibrio* with the genus *Nitrosospira* due to high 16S rRNA similarities (Aakra *et al.*, 2001), the separation into distinct genera is supported by morphological as well as molecular data (Ida *et al.*, 2005). As the only known AOB outside the *Betaproteobacteria*, the genus *Nitrosococcus* constitutes a separate branch within the *Gammaproteobacteria* (Purkhold *et al.*, 2000; Ward and O'Mullan, 2002).

Archaea known to perform ammonia oxidation all are affiliated with the recently proposed phylum *Thaumarchaeota* (Brochier-Armanet *et al.*, 2008). Existence of AOA was first proposed after the discovery of a genomic fragment from a soil metagenome which carried an archaeal 16S rRNA gene as well as genes similar to the large and small subunits of ammonia monooxygenase (*amoAB*) (Treusch *et al.*, 2005), coding for the enzyme that catalyzes the oxidation of ammonia to hydroxylamine. Shortly after, the successful isolation of an archaeon able to grow chemolithoautotrophically by oxidizing ammonia to nitrite (Könneke *et al.*, 2005) confirmed the existence of AOA. By now, additional ammonia oxidizing laboratory cultures are available (de la Torre *et al.*, 2008; Hatzenpichler *et al.*, 2008).While these organisms first were thought to be phylogenetically affiliated with the *Crenarchaeota*, the availability of genomic data (Hallam *et al.*, 2006; Spang *et al.*, 2010; Walker *et al.*, 2010) gave further support to the classification of the currently known AOA as *Thaumarchaeota* (Spang *et al.*, 2010).

In contrast to the ammonia oxidizers, so far no *Archaea* have been found to perform the oxidation of nitrite, the second step of nitrification. Bacteria gaining energy from this reaction can be divided into five phylogenetic groups (Teske *et al.*, 1994; Daims *et al.*, 2010). The genera *Nitrobacter*

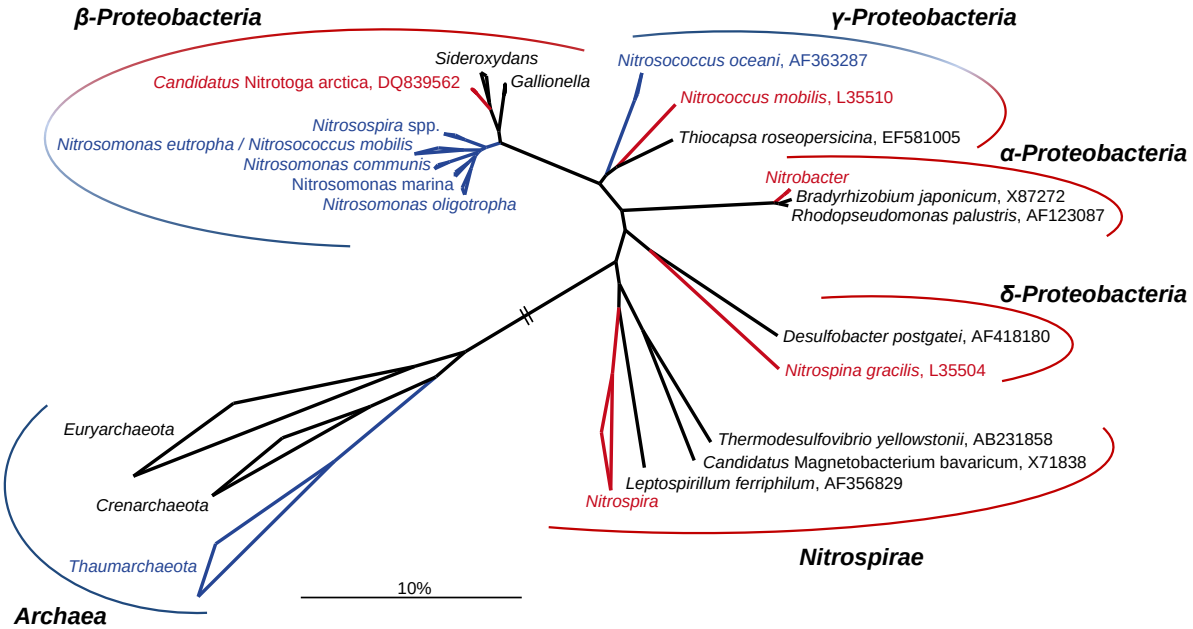


Figure 2. 16S rRNA-based phylogenetic tree reflecting the affiliations of nitrifying organisms. Nitrite-oxidizing bacteria are highlighted in red, ammonia-oxidizing archaea and bacteria in blue.

(Winogradsky, 1892; Stackebrandt *et al.*, 1988), "Nitrotoga" (Alawi *et al.*, 2007), and *Nitrococcus* (Watson and Waterbury, 1971; Teske *et al.*, 1994) belong to the *Alpha-*, *Beta-*, and *Gammaproteobacteria*, respectively. The genus *Nitrospina* (Watson and Waterbury, 1971) has been provisionally assigned to the *Deltaproteobacteria* (Teske *et al.*, 1994), but analyses of larger 16S rRNA gene sequence datasets indicate that this NOB might belong to a separate phylum (Schloss and Handelsman, 2004). Furthermore, clearly separated from the *Proteobacteria*, the genus *Nitrospira* belongs to the distinct phylum *Nitrospirae* (Ehrich *et al.*, 1995), which besides *Nitrospira* only encompasses an iron-oxidizing and a sulfate-reducing genus, *Leptospirillum* and *Thermodesulfovibrio*, respectively, as well as the magnetotactic "*Candidatus* Magnetobacterium bavaricum".

Nitrite-oxidizing bacteria

Since the discovery of the first nitrite oxidizer by Sergej Winogradsky in the late 19th century (Winogradsky, 1892) our knowledge of NOB diversity has been greatly enhanced, both by cultivation-based and molecular techniques. Considering the ecological key role of NOB, it is not surprising that members of this functional group have been detected in a wide variety of aquatic and terrestrial ecosystems, also including extreme environments like permafrost soil (Alawi *et al.*, 2007) and geothermal springs (Lebedeva *et al.*, 2005). As mentioned in the previous chapter, chemolithoautotrophic NOB known to date are affiliated with five different genera (Figure 3) (Daims *et al.*, 2010) .

(i) Either genus *Nitrococcus* and *Nitrospina* contains only one described species, *Nitrococcus mobilis* and *Nitrospina gracilis* (Watson and Waterbury, 1971), respectively. These species were isolated from oceanic samples (Watson and Waterbury, 1971) and, while according to molecular data globally distributed, appear to be restricted to marine systems (Teske *et al.*, 1994; Mincer *et al.*, 2007). *Nitrospina* is attracting new interest since, according to recent reports, communities consisting of *N. gracilis* and AOA appear to be the key organisms performing nitrification in the ocean (Lam *et al.*, 2009; Santoro *et al.*, 2010).

(ii) The candidate genus "Nitrotoga" was formed with the discovery of "*Candidatus* Nitrotoga arctica". Since this species was cultured from arctic permafrost soil (Alawi *et al.*, 2007), only limited knowledge about the functional importance and environmental distribution of this novel nitrite oxidizer has been gained. 16S rRNA sequences highly similar to "N. arctica" have been detected in aquatic and biofilm samples (Percent *et al.*, 2008; Chen *et al.*, 2009) and recently a closely related strain was enriched from activated sludge (Alawi *et al.*, 2009).

(iii) Bacteria affiliated with the genus *Nitrobacter* are the least resistant to cultivation among the fastidious and slow growing chemolithoautotrophic nitrite oxidizers and therefore represent the best studied NOB. The genus encompasses four validly described species, *N. winogradskyi* (Winslow *et al.*, 1917; Watson, 1971), *N. hamburgensis* (Bock *et al.*, 1983), *N. vulgaris* (Bock *et al.*, 1990), and *N. alkalikus* (Sorokin *et al.*, 1998). The additional species "N. agilis" (Nelson, 1931) is considered invalid

due to insufficient phenotypic difference to *N. winogradskyi* (Pan, 1971). *Nitrobacter* strains have been isolated from a wide variety of habitats, including freshwater (Bock *et al.*, 1990), marine (Starkenburg *et al.*, 2008c), and soda lake systems (Sorokin *et al.*, 1998), acidic (Hankinson and Schmidt, 1988) and neutral soil (Bock *et al.*, 1990), as well as from rock (Lebedeva *et al.*, 1978) and building stone (Mansch and Bock, 1998). Interestingly, this obvious physiological flexibility contrasts a very low phylogenetic diversity within the genus. In rRNA-based phylogenetic analyses all *Nitrobacter* species cluster closely together and display 16S rRNA gene similarities above 99% (Orso *et al.*, 1994). This low 16S rRNA gene sequence variability within the genus, taken together with the high similarity to its closest non-nitrifying relatives, *Bradyrhizobium japonicum* and *Rhodopseudomonas palustris*, indicate a recent acquisition of the nitrite oxidizing lifestyle (Orso *et al.*, 1994). Despite the relatively young age of the genus, different *Nitrobacter* isolates display a significant phenotypic as well as genetic diversity (Grundmann and Normand, 2000; Starkenburg *et al.*, 2008c) which is missed in diversity studies based on the 16S rRNA gene. One recent approach to resolve this limitation was the use of the nitrite oxidoreductase (Nxr), the key enzyme for nitrite oxidation, as functional and phylogenetic marker (Vanparys *et al.*, 2007; Wertz *et al.*, 2008). Phylogeny based on the gene encoding the α-subunit (*nxrA*) of the Nxr enzyme complex revealed the presence of multiple gene copies for the type strains analyzed The paralogous copies grouped into four distinct sequence clusters, clearly separating the four strains (Poly *et al.*, 2008). The genes from the available *Nitrobacter* genomes (Starkenburg *et al.*, 2006; Starkenburg *et al.*, 2008c) also grouped consistently in this phylogenetic analysis. Interestingly, screening of soil samples discovered additional sequence clusters of *Nitrobacter*-like *nxrA* sequences, indicating an unresolved diversity of *Nitrobacter* in the environment (Poly *et al.*, 2008).

(iv) The largest diversity, based on 16S rRNA gene data, is found within the genus *Nitrospira*. Due to the tedious approaches necessary to cultivate most of the members of this genus (Lebedeva *et al.*, 2008), the importance of these NOB has long been underestimated. After the discovery of the first *Nitrospira*, the marine *N. marina* (Watson *et al.*, 1986), almost a decade passed till the description of a second strain, *N. moscoviensis* (Ehrich *et al.*, 1995). This organism was isolated from a corroded pipeline of the Moscow heating system, thus already indicating the physiological versatility within the genus. Only with the advance of cultivation-independent molecular techniques the prevalence of *Nitrospira* in a wide variety of natural and man-made habitats became apparent (Daims *et al.*, 2001). By now, *Nitrospira* 16S rRNA genes have been detected in samples from marine (Foesel *et al.*, 2008; Santelli *et al.*, 2008) and freshwater systems (Stein *et al.*, 2001), various soils (Marilley *et al.*, 1999) and sediments (Li *et al.*, 1999; Todorov *et al.*, 2000), aquarium filters (Hovanec *et al.*, 1998), cave wall biofilms (Holmes *et al.*, 2001; Chen *et al.*, 2009; Pasic *et al.*, 2010), various thermal springs (Anitori *et al.*, 2002; Kanokratana *et al.*, 2004; Lebedeva *et al.*, 2005; Weidler *et al.*, 2007), and even are involved in symbiotic interactions with sponges (Hentschel *et al.*, 2002; Taylor *et al.*, 2007). With special significance from the applied perspective, *Nitrospira*-like organisms have been shown to be the

main nitrite oxidizers within lab-scale bioreactors (Burrell *et al.*, 1998; Schramm *et al.*, 1998) and full-scale wastewater treatment systems (Juretschko *et al.*, 1998; Daims *et al.*, 2001; Juretschko *et al.*, 2002). Facilitating studies on these biotechnologically important NOB, "*Candidatus* N. defluvii" was successfully enriched from activated sludge recently (Spieck *et al.*, 2006). On the phylogenetic level, the genus *Nitrospira* was divided in four sublineages (Daims *et al.*, 2001), grouping sequences from activated sludge and "*Candidatus* N. defluvii" in lineage I, *N. moscoviensis* with environmentally as well as some wastewater-derived sequences in lineage II, and *N. marina* and the closely related sponge symbionts in lineage IV. The marine sublineage IV can further be split into sequence cluster IVa, containing the mostly free-living marine organisms related to *N.marina*, and IVb, encompassing the sponge-derived symbiotic *Nitrospira*. One sponge-associated *Nitrospira* strain, which nevertheless is more closely related to *N. marina* than to the cluster IVb sponge symbionts, was recently enriched from the marine sponge *Aplysina aerophoba* (Off *et al.*, 2010). Sublineage III contains sequences from the aforementioned Nullarbor cave system (Holmes *et al.*, 2001) and from some deep-sea sediments (Santelli *et al.*, 2008), but lacks any cultured representative. Over the last years, the number of sublineages within the genus was extended by the successful cultivation of moderately thermophilic *Nitrospira* strains. Sublineage V was founded by "*Candidatus* N. bockiana", a strain also isolated from the Moscow heating system (Lebedeva *et al.*, 2008), accompanied by few soil-derived environmental sequences. *N. calida*, an isolate from the terrestrial geothermal spring Gorjachinsk, forms the novel sublineage VI together with an enrichment culture from the Garga hot spring (Lebedeva *et al.*, in preparation), implying that moderate thermophily is widespread within the genus. Besides these phylogenetically stable sublineages, public databases still contain additional environmentally derived 16S rRNA gene sequences that indicate the presence of an even greater diversity of *Nitrospira*-like bacteria.

Besides the aerobic, chemolithotrophic NOB, the ability to oxidize nitrite has been identified in some phototrophic bacteria (Griffin *et al.*, 2007). These organisms perform anaerobic phototrophic nitrite oxidation, where nitrite serves as electron donor for anoxygenic photosynthesis. So far, this has been described for *Rhodopseudomonas* sp. strain LQ17, an alphaproteobacterium highly similar to the purple non-sulfur bacterium *R. palustris*, as well as for three purple sulfur bacteria, *Thiocapsa* sp. strain KS1 and *T. roseopersicina* strains DSM 217 and DSM 221, within the *Gammaproteobacteria* (Figure 3) (Schott *et al.*, 2010). Noteworthy, these organisms are closely related to the aerobic NOB *Nitrobacter* and *Nitrococcus*, respectively. This relationship strengthens the theory that autotrophic nitrite oxidation within the *Alpha*- and *Gammaproteobacteria* is derived from photosynthetic ancestors (Teske *et al.*, 1994), also explaining the presence of intracytoplasmic membrane stacks in *Nitrobacter* and *Nitrococcus*. These membrane systems are absent in all other known NOB, which also lack closely related phototrophic relatives.

Anammox organisms should also be mentioned in the context of nitrite oxidation. Anammox bacteria belong to the order *Brocardiales* and are affiliated with the *Planctomycetes*. The order

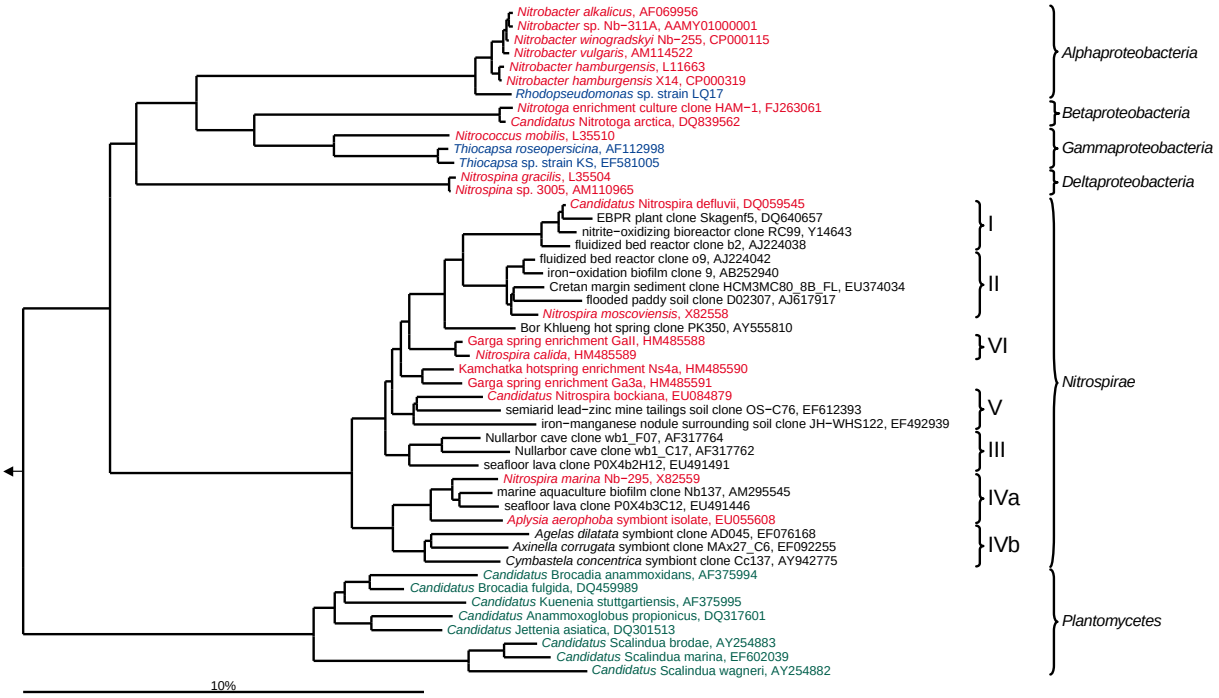


Figure 3. 16S rRNA-based phylogenetic tree reflecting the affiliation of nitrite-oxidizing and anaerobic ammonium-oxidizing bacteria. Pure cultures of nitrite-oxidizing bacteria are highlighted in red, anaerobic phototrophic nitrite-oxidizing bacteria in blue, and anaerobic ammonium-oxidizing bacteria in green.

encompasses five "*Candidatus*" anammox genera (Figure 3), "Kuenenia" (Schmid *et al.*, 2000; Strous *et al.*, 2006), "Brocadia" (Strous *et al.*, 1999; Kartal *et al.*, 2008), "Anammoxoglobus" (Kartal *et al.*, 2007), "Jettenia" (Quan *et al.*, 2008), and "Scalindua" (Kuypers *et al.*, 2003; Schmid *et al.*, 2003). These anaerobic organisms use autotrophic ammonium oxidation for energy generation, but unlike the aerobic AOA and AOB, they use nitrite as electron acceptor and thus comproportionate ammonium and nitrite to dinitrogen. At the same time, they oxidize nitrite to nitrate in order to obtain reducing equivalents needed for the reduction of CO_2 (Jetten *et al.*, 2009). For this reaction anammox organisms make use of a similar pathway as NOB. Accordingly, the genome of *Kuenenia stuttgartiensis* (Strous *et al.*, 2006) encodes for an enzyme of the molybdopterin-binding family (Rothery *et al.*, 2008) which is hypothesized to be responsible for nitrite oxidation in anammox (Jetten *et al.*, 2009). Since the family of molybdopterin-binding enzymes also contains nitrate reductases and the nitrite oxidoreductase of *Nitrobacter*, the proposed function of the gene product is strongly supported by this relationship.

The Genus Nitrospira

It can already be inferred from the immense diversity of organisms affiliated with this genus that different *Nitrospira* occupy a great variety of habitats and may play diverse roles in the respective microbial communities. In the Nullarbor cave system in Australia, for example, *Nitrospira* nutritionally maintain a complex microbial community (Holmes *et al.*, 2001). In these caves, nitrite in the water column is the only energy and CO_2 the only carbon source available. The

chemolithoautotrophic *Nitrospira* convert these substrates and serve as primary producers to support a microbial community that grows slowly, but nevertheless reaches high amounts of biomass. But *Nitrospira* can also cause problems under certain circumstances. In drinking water systems, communities of AOB and *Nitrospira* can degrade chloramines added as secondary disinfectant. This results in depletion of the disinfectant and the development of autotrophic and heterotrophic microbial communities in the drinking water system, which deteriorate drinking water quality (Regan *et al.*, 2002). *Nitrospira* also have been identified in nonchloraminated drinking water systems (Martiny *et al.*, 2005),where they can prime biofilm formation (Martiny *et al.*, 2003) and thus not only cause hygienic problems but also plugging of the distribution pipes. On the positive side, efficient removal of nitrogen from wastewater would mostly not be achieved without *Nitrospira* since they are the dominant NOB found in activated sludge (Juretschko *et al.*, 1998; Schramm *et al.*, 1998; Daims *et al.*, 2001). All this causes the question of what causes the competitive success of *Nitrospira* over the other NOB in such a wide range of habitats? Probably, it is their adaptation to low substrate concentrations. While *Nitrobacter* are regarded as *r*-strategists that can outgrow the other NOB quickly when substrate availability is not limited, *Nitrospira* have been shown to be K-strategists that display high substrate affinities but lower growth rates (Schramm *et al.*, 1999; Nogueira and Melo, 2006). The usually low nitrite concentrations in natural environments thus selectively favor growth of the K-strategists *Nitrospira*. Interestingly, similar differential adaptation of AOA and AOB to low and high ammonia concentrations has received much attention in the recent literature (Prosser and Nicol, 2008; Beman *et al.*, 2010; Di *et al.*, 2010), whereas the analogous situation of NOB has long been neglected, at least beyond the field of environmental engineering and related studies. Even within the genus *Nitrospira* adaptations to different substrate concentrations have been observed. Sublineage I *Nitrospira* from activated sludge grow at higher nitrite loads than sublineage II organisms (Maixner *et al.*, 2006), thus enabling the members of this genus to succeed at a range of substrate levels. The co-occurrence of different *Nitrospira* species in the same habitat has not only been observed for wastewater treatment plants. Even three distinct populations could be identified to coexist in the Garga hot spring in Russia (Lebedeva *et al.*, 2005), indicating an extensive and previously overlooked niche differentiation within the genus.

Nitrospira also exhibit a fascinating growth pattern. In activated sludge flocs and biofilm, they grow in dense cell aggregates (Figure 4) and form complex structures interlaced with a network of microscopic channels (Daims *et al.*, 2001). These clusters contain the cells embedded in a matrix of extracellular polymeric substances (Ehrich *et al.*, 1995) and extracellular DNA (P.H. Nielsen, personal communication), which also has been shown to be involved in biofilm formation (Allesen-Holm *et al.*, 2006; Das *et al.*, 2010; Harmsen *et al.*, 2010). Even in highly enriched cultures, *Nitrospira* form flocs and planktonic single cells are rarely observed (Spieck *et al.*, 2006). This aptitude to aggregation and floc formation suggests some means of quorum sensing, but so far no mechanism of cell to cell communication has been identified in these organisms. In nitrifying mixed populations, *Nitrospira*

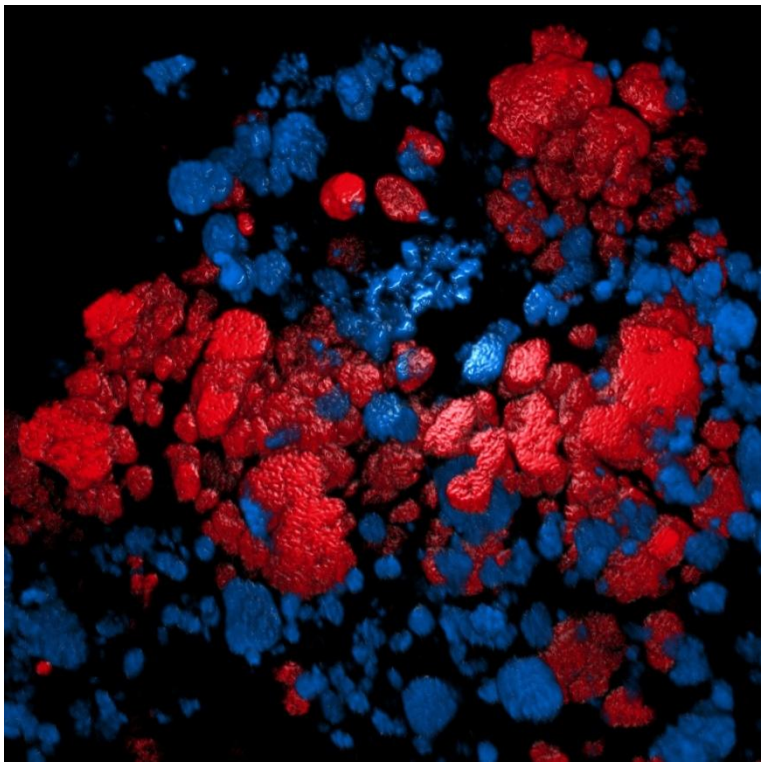


Figure 4. Confocal micrograph recorded from nitrifying sequencing batch reactor biofilm. The 3D reconstruction shows nitrite-oxidizing bacteria of the genus *Nitrospira* in red and betaproteobacterial ammonia-oxidizers in blue. The close spatial arrangement of the two functional groups within the biofilm reflects the mutualistic interactions between these microbes.

furthermore display a pronounced spatial arrangement pattern in relation to ammonia oxidizers, their partners in a mutualistic symbiosis (Juretschko *et al.*, 1998). The AOB release nitrite that serves as substrate for the nitrite oxidizers. The nitrite consumption by the NOB in turn prevents the accumulation of nitrite, which would be toxic to AOB (Stein and Arp, 1998). This interdependency is reflected by an explicit co-aggregation pattern of the two functional groups (Figure 4) (Daims *et al.*, 2006). Moreover, members of the different *Nitrospira* sublineages occur at distinct distances in relation to the AOB. The more nitrite-tolerant sublineage I *Nitrospira* clearly proliferate in closer proximity to the nitrite source than sublineage II (Maixner *et al.*, 2006), thus highlighting their potential for niche adaptation.

Physiology of NOB

As chemolithoautotrophic organisms, NOB conserve energy from the oxidation of nitrite and fix CO_2 as carbon source. Besides this common feature, little is known about the physiology of most members of this functional group. The only exceptions are some NOB affiliated with the genus *Nitrobacter*, for which also genomic data is available (Starkenburg *et al.*, 2006; Starkenburg *et al.*, 2008c). Since these organisms are fairly easy to culture in the laboratory, they often have been used as model organisms for nitrite oxidation. Physiological studies of the other NOB, irrespective of their ecological importance, have mostly been hampered by their slow growth rates and the difficulty to maintain them in pure culture.

While the physiological potential of the novel NOB "*Candidatus* Nitrotoga arctica" has not been investigated yet, some information is available for the representatives of the marine genera *Nitrococcus* and *Nitrospina*. The physiological data collected so far suggests that these organisms are obligate chemolithoautotrophs (Watson and Waterbury, 1971). While some acetate assimilation was observed for *N. mobilis*, they were not able to use organic compounds as energy or main carbon source and nitrite served as sole energy supply for CO_2 fixation. Growth of *N. gracilis* even was inhibited by the presence of organic substrates (Watson and Waterbury, 1971). *Nitrococcus* and *Nitrospina* store carbon intracellularly as glycogen and *N. mobilis* also appears to form poly-*β*-hydroxybutyrate (PHB) granules (Watson and Waterbury, 1971). The recently released genome sequence of *N. mobilis* (GenBank accession number AAOF00000000) will allow further insights into the physiology of this nitrite oxidizer. The presence of genes for the ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) (Tourova *et al.*, 2007) and a carboxysomal operon (Marin *et al.*, 2007) indicate carbon fixation via the Calvin-Benson-Bessham cycle. Further, *N. mobilis* encodes for a nitrite oxidoreductase that is highly similar to the Nxr of *Nitrobacter* (Poly *et al.*, 2008; Wertz *et al.*, 2008). The presence of these genes, together with the fact that *Nitrococcus* and *Nitrobacter* unlike the other NOB have intracytoplasmic membranes containing the Nxr system (Spieck *et al.*, 1996), suggests that *Nitrococcus* realizes energy conservation and carbon fixation via highly similar pathways as *Nitrobacter* (Bock and Wagner, 2006; Starkenburg *et al.*, 2006; Starkenburg *et al.*, 2008c).

The physiological capabilities of the members of the genus *Nitrospira* appear to be more diverse. They are aerobic chemolithoautotrophic nitrite oxidizers and form cytoplasmic glycogen or PHB as well as polyphosphate granules for carbon and phosphate storage, respectively (Watson *et al.*, 1986; Ehrich *et al.*, 1995). The use of polyphosphates also was indicated by the presence of an exopolyphosphatase gene on a genome fragment obtained from "N. defluvii" by a metagenomic approach (Maixner *et al.*, 2008). A gene coding for the β-subunit of a 2-oxoacid:ferredoxin oxidoreductase on the same contig further indicated that *Nitrospira* might be able to use pyruvate. Alternatively, the gene might be indicative for carbon fixation via the reductive citric acid cycle (Maixner *et al.*, 2008). Growth of *N. marina* is enhanced by the presence of organic substrates such as pyruvate, yeast extract, and peptone, but purely organotrophic growth has not been observed (Watson *et al.*, 1986). Uptake of pyruvate also demonstrated a mixotrophic potential (with mixotrophy defined as the ability to assimilate organic compounds as carbon sources but not as energy sources) for "*Candidatus* N. defluvii" (Daims *et al.*, 2001; Spieck *et al.*, 2006), but the full range of organic substrates utilized by this organism was hard to investigate due to the lack of a pure culture. In contrast, nitrite oxidation rates (as a measure for metabolic activity) of the moderately thermophilic strains "*Candidatus* N. bockiana" (Lebedeva *et al.*, 2008) and *N. calida* (Lebedeva *et al.*, in preparation) were not stimulated by the presence of organic substrates. An apparent lack of mixotrophy was also observed for *N. moscoviensis*, but this organism was suggested to couple

hydrogen oxidation to nitrate reduction under anoxic conditions (Ehrich *et al.*, 1995), which is so far the only indication for anaerobic respiration in the genus.

Most information about the physiology and biochemistry of NOB was obtained in studies of *Nitrobacter*, and the analyses of three *Nitrobacter* genomes (Starkenburg *et al.*, 2006; Starkenburg *et al.*, 2008c) greatly enhanced our understanding of these organisms. Some *Nitrobacter* strains display an enormous versatility, outmatching the described metabolic potential of all other known NOB. As facultative chemolithoautotrophic organisms, they gain energy from aerobic nitrite oxidation and fix carbon via the Calvin-Benson-Basham cycle (Starkenburg *et al.*, 2006). The electrons derived from nitrite are transferred via a cytochrome a_1 to cytochrome *c*, which shuttles them to the membrane-integral electron transport chain (Bock and Wagner, 2006). Here, according to metabolic models (Poughon *et al.*, 2001), most of the electrons are transferred to a cytochrome *c* oxidase of the aa_3 type and onto the terminal electron acceptor oxygen. The terminal oxidase functions as proton pump and generates the membrane potential (Poughon *et al.*, 2001; Starkenburg *et al.*, 2008a). Surprisingly, the first product for energy conservation in *Nitrobacter* was reported to be NADH, not ATP (Freitag and Bock, 1990; Bock and Wagner, 2006). This indicates that the proton motive force is used mainly for reverse electron transport to the NADH dehydrogenase complex. ATP then is formed from the NADH pool via the conventional respiratory chain with oxygen as terminal electron acceptor (Freitag and Bock, 1990; Bock and Wagner, 2006). The proton gradient generated by the cytochrome *c* oxidase is now used for ATP production by a F_0F_1-type ATP synthase (Bock and Wagner, 2006). Why ATP appears to be formed preferentially with electrons derived from NADH and not nitrite oxidation is not well understood. Interestingly, nitrate also serves as sink for electrons from the respiratory chain even under oxic conditions (Freitag and Bock, 1990). Furthermore, autotrophic growth of *Nitrobacter* also is possible with nitric oxide as electron donor. The oxidation of nitric oxide even has been reported to be more efficient than nitrite oxidation for NADH synthesis (Freitag and Bock, 1990). Still, whether this stimulation is based on an enzymatic reaction or auto-oxidation of nitrous oxide to nitrite followed by biological nitrite oxidation is under debate (Starkenburg *et al.*, 2008a). Nitric oxide might actually be an important intermediate of the nitrite oxidation pathway in *Nitrobacter* (Bock *et al.*, 1991; Poughon *et al.*, 2001). In this metabolic scheme nitrite is not only oxidized to nitrate, but also reduced to nitric oxide under aerobic conditions by a periplasmic nitrite reductase, which serves as alternative sink for the electrons derived from nitrite oxidation. The nitric oxide then diffuses through the membrane, is reoxidized on the cytoplasmic side, and serves as electron donor for the respiratory chain (Poughon *et al.*, 2001). This model, however, is based mainly on theoretical calculations and has not been experimentally verified. *Nitrobacter* can also grow on a range of organic substrates, and utilization of formate, acetate, pyruvate, α-ketoglutarate, glycerol, and *D*-lactate has been demonstrated (Delwiche and Finstein, 1965; Smith and Hoare, 1968; Bock, 1976; Starkenburg *et al.*, 2008b). While mixotrophic growth rates on these organic carbon sources are highest, *Nitrobacter* also is able to assimilate them in the absence of nitrite (Steinmüller and Bock, 1976; Bock *et al.*, 1983;

Starkenburg *et al.*, 2008b). Chemoorganotrophic growth, however, was reported to be less effective than growth under autotrophic or mixtotrophic conditions (Bock *et al.*, 1986). In agreement with the ability to grow on organic substrates, all enzymes of the citric acid cycle (Steinmüller and Bock, 1976) and genes encoding enzymes necessary for pyruvate, acetate, glycerol (Starkenburg *et al.*, 2006), and lactate metabolism (Starkenburg *et al.*, 2008b) have been identified. Further, the complete glycolysis pathway is encoded in the genomes of all *Nitrobacter* analyzed except *N. winogradskyi* (Starkenburg *et al.*, 2008c), which lacks a gene for phoshofructokinase. Utilization of hexose sugars has so far not been observed for any *Nitrobacter* strain. In cultures containing organic carbon sources and nitrite but lacking CO_2, the organic substrate cannot serve as sole carbon source (Delwiche and Finstein, 1965; Ida and Alexander, 1965), a phenomenon that has been attributed to a requirement for CO_2 reduction as sink for electrons derived from nitrite oxidation. In the absence of oxygen, *Nitrobacter* respires organic carbon by switching to partial denitrification (Kiesow, 1964). By reversing the reaction catalyzed by Nxr under anoxic conditions, nitrate is utilized as terminal acceptor for the electrons derived from substrate oxidation, and nitrite is formed (Sundermeyer-Klinger *et al.*, 1984). *Nitrobacter* can further reduce nitrite to nitric oxide by a copper-containing nitrite reductase (Freitag *et al.*, 1987). Besides nitric oxide, denitrifying cultures of *N. winogradskyi* also produce nitrous oxide (Freitag *et al.*, 1987; Ahlers *et al.*, 1990), but the mechanism of nitrous oxide formation remains unclear since genes encoding a nitric oxide reductase have not been identified in the *Nitrobacter* genomes (Starkenburg *et al.*, 2006; Starkenburg *et al.*, 2008a; Starkenburg *et al.*, 2008c).

Aims of this study

Although members of the genus *Nitrospira* are the key nitrifiers in most natural and man-made habitats, our current knowledge about the NOB still stems mainly from research performed with *Nitrobacter* cultures. Therefore, the main objective of this thesis was to provide a starting point for a thorough understanding of the ecophysiology of *Nitrospira* based on the complete genome sequence of "*Candidatus* N. defluvii". This nitrite oxidizer is central for efficient nitrogen removal from activated sludge and thus of utmost importance for wastewater treatment. Furthermore, this thesis aims at exploring the significance for wastewater treatment systems of the recently discovered candidate genus "Nitrotoga", the members of which have been overlooked in the vast majority of studies concerning nitrification so far.

Outline

Chapter I is a general introduction to the topics of this thesis. It provides basic background information on the research area, starting with an outline of the biogeochemical nitrogen cycle. The chapter then focuses on one particular functional group, the nitrite-oxidizing bacteria. For this ecologically and biotechnologically important group, a more thorough overview about phylogeny, distribution, and physiology is presented.

Chapter II describes the analysis of the complete genome sequence of "*Candidatus* Nitrospira defluvii". The study revealed fundamental differences to other known nitrite-oxidizing bacteria in the key pathways for nitrite oxidation, respiration, and autotrophic carbon fixation. The presence of the reverse tricarboxylic acid cycle for carbon fixation and the absence of classical oxygen defense mechanisms indicate a microaerophilic or anaerobic ancestor of *Nitrospira*. Furthermore, phylogenomic analysis discovered a relationship of the *Nitrospira* nitrite oxidation system to that of anaerobic ammonium-oxidizing *Planctomycetes*, reflecting an unexpected evolutionary link of two key processes of the nitrogen cycle.

In **Chapter III**, the distribution and abundance in engineered systems of members of the recently discovered candidate genus "Nitrotoga" are investigated. Screening a range of wastewater treatment plants detected the presence of *Nitrotoga*-like bacteria in approximately half of the activated sludge samples analyzed. In some plants they even were the only known nitrite-oxidizing bacteria. A statistically significant co-aggregation with ammonia-oxidizing bacteria lends further support to the proposed function of *Nitrotoga*-like organisms as novel, previously overlooked nitrite oxidizers in wastewater treatment systems.

Chapter IV contains a short summery of the presented studies in English and German.

References

Aakra, A., J. Utaker, A. Pommerening-Roser, H. Koops and I. Nes (2001). Detailed phylogeny of ammonia-oxidizing bacteria determined by rDNA sequences and DNA homology values. *Int J Syst Evol Microbiol* **51**(6): 2021-2030.

Ahlers, B., W. König and E. Bock (1990). Nitrite reductase activity in *Nitrobacter vulgaris*. *FEMS Microbiol Lett* **67**(1-2): 121-126.

Alawi, M., A. Lipski, T. Sanders, E. M. Pfeiffer and E. Spieck (2007). Cultivation of a novel cold-adapted nitrite oxidizing betaproteobacterium from the Siberian Arctic. *ISME J* **1**(3): 256-264.

Alawi, M., S. Off, M. Kaya and E. Spieck (2009). Temperature influences the population structure of nitrite-oxidizing bacteria in activated sludge. *Environ Microbiol Reports* **1**(3): 184-190.

Allesen-Holm, M., K. B. Barken, L. Yang, M. Klausen, J. S. Webb, S. Kjelleberg, S. Molin, M. Givskov and T. Tolker-Nielsen (2006). A characterization of DNA release in *Pseudomonas aeruginosa* cultures and biofilms. *Mol Microbiol* **59**(4): 1114-1128.

Anitori, R. P., C. Trott, D. J. Saul, P. L. Bergquist and M. R. Walter (2002). A culture-independent survey of the bacterial community in a radon hot spring. *Astrobiology* **2**(3): 255-270.

Beman, J. M., R. Sachdeva and J. A. Fuhrman (2010). Population ecology of nitrifying archaea and bacteria in the Southern California Bight. *Environ Microbiol* **12**(5): 1282-1292.

Bock, E. (1976). Growth of *Nitrobacter* in the presence of organic matter. II. Chemoorganotrophic growth of *Nitrobacter agilis*. *Arch Microbiol* **108**(3): 305-312.

Bock, E., H. Sundermeyer-Klinger and E. Stackebrandt (1983). New facultative lithoautotrophic nitrite-oxidizing bacteria. *Arch Microbiol* **136**(4): 281-284.

Bock, E., H. P. Koops and H. Harm (1986). Cell biology of nitrifiers. *In* J. I. Prosser (ed.), Nitrification. Washington, D.C., IRL. **Vol. 20:** 17-38.

Bock, E., H.-P. Koops, U. C. Möller and M. Rudert (1990). A new facultatively nitrite oxidizing bacterium, *Nitrobacter vulgaris* sp. nov. *Arch Microbiol* **153**(2): 105-110.

Bock, E., H.-P. Koops, B. Ahlers and H. Harms (1991). Oxidation of inorganic nitrogen compounds as energy source. *In* A. Balows, H. G. Truper, M. Dworkin, W. Harder and K.-H. Schleifer (ed.), The Prokaryotes, 2nd ed. New York, Springer-Verlag.

Bock, E., I. Schmidt, R. Stüven and D. Zart (1995). Nitrogen loss caused by denitrifying *Nitrosomonas* cells using ammonium or hydrogen as electron donors and nitrite as electron acceptor. *Arch Microbiol* **163**(1): 16-20.

Bock, E. and M. Wagner (2006). Oxidation of Inorganic Nitrogen Compounds as an Energy Source. *In* M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer and E. Stackebrandt (ed.), The Prokaryotes. New York, Springer New York. **Vol. 2: Ecophysiology and Biochemistry** 457-495.

Brochier-Armanet, C., B. Boussau, S. Gribaldo and P. Forterre (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6**(3): 245-252.

Burrell, P. C., J. Keller and L. L. Blackall (1998). Microbiology of a Nitrite-Oxidizing Bioreactor. *Appl Environ Microbiol* **64**(5): 1878-1883.

Camargo, J. A. and Á. Alonso (2006). Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: A global assessment. *Environ Internat* **32**(6): 831-849.

Chen, Y., L. Wu, R. Boden, A. Hillebrand, D. Kumaresan, H. Moussard, M. Baciu, Y. Lu and J. Colin Murrell (2009). Life without light: Microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in Movile Cave. *ISME J* **3**(9): 1093-1104.

Daims, H., J. L. Nielsen, P. H. Nielsen, K. H. Schleifer and M. Wagner (2001). *In situ* characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microbiol* **67**(11): 5273-5284.

Daims, H., S. Lücker and M. Wagner (2006). *daime*, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* **8**(2): 200-213.

Daims, H., S. Lücker, D. Le Paslier and M. Wagner (2010). Diversity, Environmental Genomics, and Ecophysiology of Nitrite-Oxidizing Bacteria. *In* B. B. Ward, D. J. Arp and M. G. Klotz (ed.), Nitrification. Washington, DC, ASM Press**:** in press.

Das, T., P. K. Sharma, H. J. Busscher, H. C. van der Mei and B. P. Krom (2010). Role of Extracellular DNA in Initial Bacterial Adhesion and Surface Aggregation. *Appl Environ Microbiol* **76**(10): 3405-3408.

de la Torre, J. R., C. B. Walker, A. E. Ingalls, M. Konneke and D. A. Stahl (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol* **10**(3): 810-818.

Delwiche, C. C. and M. S. Finstein (1965). Carbon and Energy Sources for the Nitrifying Autotroph *Nitrobacter*. *J Bacteriol* **90**(1): 102-107.

Di, H. J., K. C. Cameron, J. P. Shen, C. S. Winefield, M. O'Callaghan, S. Bowatte and J. Z. He (2010). Ammonia-oxidizing bacteria and archaea grow under contrasting soil nitrogen conditions. *FEMS Microbiol Ecol* **72**(3): 386-394.

Diaz, R. J. and R. Rosenberg (2008). Spreading Dead Zones and Consequences for Marine Ecosystems. *Science* **321**(5891): 926-929.

Ehrich, S., D. Behrens, E. Lebedeva, W. Ludwig and E. Bock (1995). A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Arch Microbiol* **164**(1): 16-23.

Focht, D. D. and W. Verstraete (1977). Biochemical ecology of nitrification and denitrification. *Adv Microb Ecol* **1**: 135–214.

Foesel, B. U., A. Gieseke, C. Schwermer, P. Stief, L. Koch, E. Cytryn, J. R. de la Torre, J. van Rijn, D. Minz, H. L. Drake and A. Schramm (2008). *Nitrosomonas* Nm143-like ammonia oxidizers and *Nitrospira marina*-like nitrite oxidizers dominate the nitrifier community in a marine aquaculture biofilm. *FEMS Microbiol Ecol* **63**(2): 192-204.

Francis, C. A., J. M. Beman and M. M. Kuypers (2007). New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *ISME J* **1**(1): 19-27.

Freitag, A., M. Rudert and E. Bock (1987). Growth of *Nitrobacter* by dissimilatoric nitrate reduction. *FEMS Microbiol Lett* **48**(1-2): 105-109.

Freitag, A. and E. Bock (1990). Energy conservation in *Nitrobacter*. *FEMS Microbiol Lett* **66**(1-3): 157-162.

Galloway, J. N. and E. B. Cowling (2002). Reactive Nitrogen and The World: 200 Years of Change. *AMBIO: A Journal of the Human Environment* **31**(2): 64-71.

Griffin, B. M., J. Schott and B. Schink (2007). Nitrite, an Electron Donor for Anoxygenic Photosynthesis. *Science* **316**(5833): 1870-.

Grundmann, G. L. and P. Normand (2000). Microscale diversity of the genus *Nitrobacter* in soil on the basis of analysis of genes encoding rRNA. *Appl Environ Microbiol* **66**(10): 4543-4546.

Hallam, S. J., K. T. Konstantinidis, N. Putnam, C. Schleper, Y. Watanabe, J. Sugahara, C. Preston, J. de la Torre, P. M. Richardson and E. F. DeLong (2006). Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proc Natl Acad Sci USA* **103**(48): 18296-18301.

Hankinson, T. R. and E. L. Schmidt (1988). An acidophilic and a neutrophilic *Nitrobacter* strain isolated from the numerically predominant nitrite-oxidizing population of an acid forest soil. *Appl Environ Microbiol* **54**(6): 1536-1540.

Harmsen, M., M. Lappann, S. Knochel and S. Molin (2010). Role of Extracellular DNA during Biofilm Formation by *Listeria monocytogenes*. *Appl Environ Microbiol* **76**(7): 2271-2279.

Hatzenpichler, R., E. V. Lebedeva, E. Spieck, K. Stoecker, A. Richter, H. Daims and M. Wagner (2008). A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc Natl Acad Sci USA* **105**(6): 2134-2139.

Hentschel, U., J. Hopke, M. Horn, A. B. Friedrich, M. Wagner, J. Hacker and B. S. Moore (2002). Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl Environ Microbiol* **68**(9): 4431-4440.

Holmes, A. J., N. A. Tujula, M. Holley, A. Contos, J. M. James, P. Rogers and M. R. Gillings (2001). Phylogenetic structure of unusual aquatic microbial formations in Nullarbor caves, Australia. *Environ Microbiol* **3**(4): 256-264.

Hovanec, T. A., L. T. Taylor, A. Blakis and E. F. Delong (1998). *Nitrospira*-like bacteria associated with nitrite oxidation in freshwater aquaria. *Appl Environ Microbiol* **64**(1): 258-264.

Ida, S. and M. Alexander (1965). Permeability of *Nitrobacter agilis* to Organic Compounds. *J Bacteriol* **90**(1): 151-156.

Ida, T., M. Kugimiya, M. Kogure, R. Takahashi and T. Tokuyama (2005). Phylogenetic relationships among ammonia-oxidizing bacteria as revealed by gene sequences of glyceraldehyde 3-phosphate dehydrogenase and phosphoglycerate kinase. *J Biosci Bioeng* **99**(6): 569-576.

Jetten, M. S., L. Niftrik, M. Strous, B. Kartal, J. T. Keltjens and H. J. Op den Camp (2009). Biochemistry and molecular biology of anammox bacteria. *Crit Rev Biochem Mol Biol* **44**(2-3): 65-84.

Juretschko, S., G. Timmermann, M. Schmid, K. H. Schleifer, A. Pommerening-Roser, H. P. Koops and M. Wagner (1998). Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. *Appl Environ Microbiol* **64**(8): 3042-3051.

Juretschko, S., A. Loy, A. Lehner and M. Wagner (2002). The microbial community composition of a nitrifying-denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. *Syst Appl Microbiol* **25**(1): 84-99.

Kanokratana, P., S. Chanapan, K. Pootanakit and L. Eurwilaichitr (2004). Diversity and abundance of *Bacteria* and *Archaea* in the Bor Khlueng Hot Spring in Thailand. *J Basic Microbiol* **44**(6): 430-444.

Kartal, B., J. Rattray, L. A. van Niftrik, J. van de Vossenberg, M. C. Schmid, R. I. Webb, S. Schouten, J. A. Fuerst, J. S. Damste, M. S. Jetten and M. Strous (2007). "*Candidatus* Anammoxoglobus propionicus" a new propionate oxidizing species of anaerobic ammonium oxidizing bacteria. *Syst Appl Microbiol* **30**(1): 39-49.

Kartal, B., L. van Niftrik, J. Rattray, J. L. van de Vossenberg, M. C. Schmid, J. Sinninghe Damste, M. S. Jetten and M. Strous (2008). "*Candidatus* Brocadia fulgida": an autofluorescent anaerobic ammonium oxidizing bacterium. *FEMS Microbiol Ecol* **63**(1): 46-55.

Kiesow, L. (1964). On the Assimilation of Energy from Inorganic Sources in Autotrophic Forms of Life. *Proc Natl Acad Sci USA* **52**: 980-988.

Klotz, M. G. and L. Y. Stein (2008). Nitrifier genomics and evolution of the nitrogen cycle. *FEMS Microbiol Lett* **278**(2): 146-156.

Könneke, M., A. E. Bernhard, J. R. de la Torre, C. B. Walker, J. B. Waterbury and D. A. Stahl (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**(7058): 543-546.

Kuypers, M. M., A. O. Sliekers, G. Lavik, M. Schmid, B. B. Jorgensen, J. G. Kuenen, J. S. Sinninghe Damste, M. Strous and M. S. Jetten (2003). Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422**(6932): 608-611.

Lam, P., G. Lavik, M. M. Jensen, J. van de Vossenberg, M. Schmid, D. Woebken, D. Gutiérrez, R. Amann, M. S. M. Jetten and M. M. M. Kuypers (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA* **106**(12): 4752-4757.

Lebedeva, E. V., N. N. Lialikova and I. Bugel'skii (1978). Participation of nitrifying bacteria in the disintegration of serpentinous ultrabasic rock. *Mikrobiologiia* **47**(6): 1101-1107.

Lebedeva, E. V., M. Alawi, C. Fiencke, B. Namsaraev, E. Bock and E. Spieck (2005). Moderately thermophilic nitrifying bacteria from a hot spring of the Baikal rift zone. *FEMS Microbiol Ecol* **54**(2): 297-306.

Lebedeva, E. V., M. Alawi, F. Maixner, P. G. Jozsa, H. Daims and E. Spieck (2008). Physiological and phylogenetic characterization of a novel lithoautotrophic nitrite-oxidizing bacterium, "*Candidatus* Nitrospira bockiana". *Int J Syst Evol Microbiol* **58**(Pt 1): 242-250.

Lebedeva, E. V., S. Off, S. Zumbrägel, M. Kruse, A. Shagzhina, S. Lücker, F. Maixner, A. Lipski, H. Daims and E. Spieck (in preparation). Isolation and characterization of a moderately thermophilic nitrite-oxidizing bacterium from a geothermal spring.

Li, L., C. Kato and K. Horikoshi (1999). Bacterial diversity in deep-sea sediments from different depths. *Biodiv Conserv* **8**(5): 659-677.

Lipschultz, F., O. C. Zafiriou, S. C. Wofsy, M. B. McElroy, F. W. Valois and S. W. Watson (1981). Production of NO and N2O by soil nitrifying bacteria. *Nature* **294**(5842): 641-643.

Maixner, F., D. R. Noguera, B. Anneser, K. Stoecker, G. Wegl, M. Wagner and H. Daims (2006). Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. *Environ Microbiol* **8**(8): 1487-1495.

Maixner, F., M. Wagner, S. Lücker, E. Pelletier, S. Schmitz-Esser, K. Hace, E. Spieck, R. Konrat, D. Le Paslier and H. Daims (2008). Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium "*Candidatus* Nitrospira defluvii". *Environ Microbiol* **10**(11): 3043-3056.

Mansch, R. and E. Bock (1998). Biodeterioration of natural stone with special reference to nitrifying bacteria. *Biodegradation* **9**(1): 47-64.

Marilley, L., U. A. Hartwig and M. Aragno (1999). Influence of an Elevated Atmospheric CO2 Content on Soil and Rhizosphere Bacterial Communities Beneath *Lolium perenne* and *Trifolium repens* under Field Conditions. *Microb Ecol* **38**(1): 39-49.

Marin, B., E. Nowack, G. Glockner and M. Melkonian (2007). The ancestor of the *Paulinella chromatophore* obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like gamma-proteobacterium. *BMC Evol Biol* **7**(1): 85.

Martiny, A. C., T. M. Jorgensen, H.-J. Albrechtsen, E. Arvin and S. Molin (2003). Long-Term Succession of Structure and Diversity of a Biofilm Formed in a Model Drinking Water Distribution System. *Appl Environ Microbiol* **69**(11): 6899-6907.

Martiny, A. C., H.-J. Albrechtsen, E. Arvin and S. Molin (2005). Identification of Bacteria in Biofilm and Bulk Water Samples from a Nonchlorinated Model Drinking Water Distribution System: Detection of a Large Nitrite-Oxidizing Population Associated with Nitrospira spp. *Appl Environ Microbiol* **71**(12): 8611-8617.

McLain, J. E. T. and D. A. Martens (2005). Nitrous oxide flux from soil amino acid mineralization. *Soil Biol Biochem* **37**(2): 289-299.

Mincer, T. J., M. J. Church, L. T. Taylor, C. Preston, D. M. Karl and E. F. DeLong (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol* **9**(5): 1162-1175.

Murray, P. A. and S. H. Zinder (1984). Nitrogen fixation by a methanogenic archaebacterium. *Nature* **312**(5991): 284-286.

Nelson, D. H. (1931). Isolation and characterisation of *Nitrosomonas* and *Nitrobacter*. *Zentr. Bakt. Parasitenk. (II. Abt.)* **83**: 280-311.

Nogueira, R. and L. F. Melo (2006). Competition between *Nitrospira* spp. and *Nitrobacter* spp. in nitrite-oxidizing bioreactors. *Biotechnol Bioeng* **95**(1): 169-175.

Off, S., M. Alawi and E. Spieck (2010). Enrichment and Physiological Characterization of a Novel *Nitrospira*-Like Bacterium Obtained from a Marine Sponge. *Appl Environ Microbiol* **76**(14): 4640-4646.

Orso, S., M. Gouy, E. Navarro and P. Normand (1994). Molecular phylogenetic analysis of *Nitrobacter* spp. *Int J Syst Bacteriol* **44**(1): 83-86.

Pan, P. H. (1971). Lack of distinction between *Nitrobacter agilis* and *Nitrobacter winogradskyi*. *J Bacteriol* **108**(3): 1416-1418.

Pasic, L., B. Kovce, B. Sket and B. Herzog-Velikonja (2010). Diversity of microbial communities colonizing the walls of a Karstic cave in Slovenia. *FEMS Microbiol Ecol* **71**(1): 50-60.

Percent, S. F., M. E. Frischer, P. A. Vescio, E. B. Duffy, V. Milano, M. McLellan, B. M. Stevens, C. W. Boylen and S. A. Nierzwicki-Bauer (2008). Bacterial community structure of acid-impacted lakes: What controls diversity? *Appl Environ Microbiol* **74**(6): 1856-1868.

Poly, F., S. Wertz, E. Brothier and V. Degrange (2008). First exploration of *Nitrobacter* diversity in soils by a PCR cloning-sequencing approach targeting functional gene *nxrA*. *FEMS Microbiol Ecol* **63**(1): 132-140.

Postgate, J. R. (1970). Biological nitrogen fixation. *Nature* **226**(5240): 25-27.

Poughon, L., C. G. Dussap and J. B. Gros (2001). Energy model and metabolic flux analysis for autotrophic nitrifiers. *Biotechnol Bioeng* **72**(4): 416-433.

Prosser, J. I. (1989). Autotrophic nitrification in bacteria. *Adv Microb Physiol* **30**: 125-181.

Prosser, J. I. and G. W. Nicol (2008). Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* **10**(11): 2931-2941.

Purkhold, U., A. Pommerening-Roser, S. Juretschko, M. C. Schmid, H. P. Koops and M. Wagner (2000). Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amoA* sequence analysis: implications for molecular diversity surveys. *Appl Environ Microbiol* **66**(12): 5368-5382.

Purkhold, U., M. Wagner, G. Timmermann, A. Pommerening-Roser and H. P. Koops (2003). 16S rRNA and *amoA*-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: extension of the dataset and proposal of a new lineage within the nitrosomonads. *Int J Syst Evol Microbiol* **53**(Pt 5): 1485-1494.

Quan, Z. X., S. K. Rhee, J. E. Zuo, Y. Yang, J. W. Bae, J. R. Park, S. T. Lee and Y. H. Park (2008). Diversity of ammonium-oxidizing bacteria in a granular sludge anaerobic ammonium-oxidizing (anammox) reactor. *Environ Microbiol* **10**(11): 3130-3139.

Regan, J. M., G. W. Harrington and D. R. Noguera (2002). Ammonia- and Nitrite-Oxidizing Bacterial Communities in a Pilot-Scale Chloraminated Drinking Water Distribution System. *Appl Environ Microbiol* **68**(1): 73-81.

Rothery, R. A., G. J. Workun and J. H. Weiner (2008). The prokaryotic complex iron-sulfur molybdoenzyme family. *Biochim Biophys Acta* **1778**(9): 1897-1929.

Santelli, C. M., B. N. Orcutt, E. Banning, W. Bach, C. L. Moyer, M. L. Sogin, H. Staudigel and K. J. Edwards (2008). Abundance and diversity of microbial life in ocean crust. *Nature* **453**(7195): 653-656.

Santoro, A. E., K. L. Casciotti and C. A. Francis (2010). Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ Microbiol* **12**(7): 1989-2006.

Schloss, P. D. and J. Handelsman (2004). Status of the microbial census. *Microbiol Mol Biol Rev* **68**(4): 686-691.

Schmid, M., U. Twachtmann, M. Klein, M. Strous, S. Juretschko, M. Jetten, J. W. Metzger, K. H. Schleifer and M. Wagner (2000). Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Syst Appl Microbiol* **23**(1): 93-106.

Schmid, M., K. Walsh, R. Webb, W. I. Rijpstra, K. van de Pas-Schoonen, M. J. Verbruggen, T. Hill, B. Moffett, J. Fuerst, S. Schouten, J. S. Damste, J. Harris, P. Shaw, M. Jetten and M. Strous (2003). "*Candidatus* Scalindua brodae", sp. nov., "*Candidatus* Scalindua wagneri", sp. nov., two new species of anaerobic ammonium oxidizing bacteria. *Syst Appl Microbiol* **26**(4): 529-538.

Schott, J., B. M. Griffin and B. Schink (2010). Anaerobic phototrophic nitrite oxidation by *Thiocapsa* sp. strain KS1 and *Rhodopseudomonas* sp. strain LQ17. *Microbiology* **156**(Pt 8): 2428-2437.

Schramm, A., D. De Beer, M. Wagner and R. Amann (1998). Identification and activities *in situ* of *Nitrosospira* and *Nitrospira* spp. as dominant populations in a nitrifying fluidized bed reactor. *Appl Environ Microbiol* **64**(9): 3480-3485.

Schramm, A., D. de Beer, J. C. van den Heuvel, S. Ottengraf and R. Amann (1999). Microscale distribution of populations and activities of *Nitrosospira* and *Nitrospira* spp. along a macroscale gradient in a nitrifying bioreactor: Quantification by *in situ* hybridization and the use of microsensors. *Appl Environ Microbiol* **65**(8): 3690-3696.

Simon, J. (2002). Enzymology and bioenergetics of respiratory nitrite ammonification. *FEMS Microbiol Rev* **26**(3): 285-309.

Smith, A. J. and D. S. Hoare (1968). Acetate Assimilation by *Nitrobacter agilis* in Relation to Its "Obligate Autotrophy". *J. Bacteriol.* **95**(3): 844-855.

Sorokin, D. Y., G. Muyzer, T. Brinkhoff, J. G. Kuenen and M. S. Jetten (1998). Isolation and characterization of a novel facultatively alkaliphilic *Nitrobacter* species, *N. alkalicus* sp. nov. *Arch Microbiol* **170**(5): 345-352.

Spang, A., R. Hatzenpichler, C. Brochier-Armanet, T. Rattei, P. Tischler, E. Spieck, W. Streit, D. A. Stahl, M. Wagner and C. Schleper (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum *Thaumarchaeota*. *Trends Microbiol* **18**(8): 331-340.

Spieck, E., J. Aamand, S. Bartosch and E. Bock (1996). Immunocytochemical detection and location of the membrane-bound nitrite oxidoreductase in cells of *Nitrobacter* and *Nitrospira*. *FEMS Microbiol Lett* **139**(1): 71-76.

Spieck, E., C. Hartwig, I. McCormack, F. Maixner, M. Wagner, A. Lipski and H. Daims (2006). Selective enrichment and molecular characterization of a previously uncultured *Nitrospira*-like bacterium from activated sludge. *Environ Microbiol* **8**(3): 405-415.

Stackebrandt, E., R. G. E. Murray and H. G. Truper (1988). Proteobacteria classis nov., a Name for the Phylogenetic Taxon That Includes the "Purple Bacteria and Their Relatives". *Int J Syst Bacteriol* **38**(3): 321-325.

Starkenburg, S. R., P. S. Chain, L. A. Sayavedra-Soto, L. Hauser, M. L. Land, F. W. Larimer, S. A. Malfatti, M. G. Klotz, P. J. Bottomley, D. J. Arp and W. J. Hickey (2006). Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *Appl Environ Microbiol* **72**(3): 2050-2063.

Starkenburg, S. R., D. J. Arp and P. J. Bottomley (2008a). Expression of a putative nitrite reductase and the reversible inhibition of nitrite-dependent respiration by nitric oxide in *Nitrobacter winogradskyi* Nb-255. *Environ Microbiol* **10**(11): 3036-3042.

Starkenburg, S. R., D. J. Arp and P. J. Bottomley (2008b). D-Lactate metabolism and the obligate requirement for CO_2 during growth on nitrite by the facultative lithoautotroph *Nitrobacter hamburgensis*. *Microbiology* **154**(Pt 8): 2473-2481.

Starkenburg, S. R., F. W. Larimer, L. Y. Stein, M. G. Klotz, P. S. Chain, L. A. Sayavedra-Soto, A. T. Poret-Peterson, M. E. Gentry, D. J. Arp, B. Ward and P. J. Bottomley (2008c). Complete genome sequence of *Nitrobacter hamburgensis* X14 and comparative genomic analysis of species within the genus Nitrobacter. *Appl Environ Microbiol* **74**(9): 2852-2863.

Stein, L. Y. and D. J. Arp (1998). Loss of ammonia monooxygenase activity in *Nitrosomonas europaea* upon exposure to nitrite. *Appl Environ Microbiol* **64**(10): 4098-4102.

Stein, L. Y., M. T. La Duc, T. J. Grundl and K. H. Nealson (2001). Bacterial and archaeal populations associated with freshwater ferromanganous micronodules and sediments. *Environ Microbiol* **3**(1): 10-18.

Steinmüller, W. and E. Bock (1976). Growth of *Nitrobacter* in the presence of organic matter. I. Mixotrophic growth. *Arch Microbiol* **108**(3): 299-304.

Strous, M., J. A. Fuerst, E. H. Kramer, S. Logemann, G. Muyzer, K. T. van de Pas-Schoonen, R. Webb, J. G. Kuenen and M. S. Jetten (1999). Missing lithotroph identified as new planctomycete. *Nature* **400**(6743): 446-449.

Strous, M., E. Pelletier, S. Mangenot, T. Rattei, A. Lehner, M. W. Taylor, M. Horn, H. Daims, D. Bartol-Mavel, P. Wincker, V. Barbe, N. Fonknechten, D. Vallenet, B. Segurens, C. Schenowitz-Truong, C. Medigue, A. Collingro, B. Snel, B. E. Dutilh, H. J. Op den Camp, C. van der Drift, I. Cirpus, K. T. van de Pas-Schoonen, H. R. Harhangi, L. van Niftrik, M. Schmid, J. Keltjens, J. van de Vossenberg, B. Kartal, H. Meier, D. Frishman, M. A. Huynen, H. W. Mewes, J. Weissenbach, M. S. Jetten, M. Wagner and D. Le Paslier (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**(7085): 790-794.

Sundermeyer-Klinger, H., W. Meyer, B. Warninghoff and E. Bock (1984). Membrane-bound nitrite oxidoreductase of *Nitrobacter*: evidence for a nitrate reductase system. *Arch Microbiol* **140**(2): 153-158.

Taylor, M. W., R. Radax, D. Steger and M. Wagner (2007). Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* **71**(2): 295-347.

Teske, A., E. Alm, J. M. Regan, S. Toze, B. E. Rittmann and D. A. Stahl (1994). Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria. *J Bacteriol* **176**(21): 6623-6630.

Todorov, J. R., A. Y. Chistoserdov and J. Y. Aller (2000). Molecular analysis of microbial communities in mobile deltaic muds of Southeastern Papua New Guinea. *FEMS Microbiol Ecol* **33**(2): 147-155.

Tourova, T. P., E. M. Spiridonova, I. A. Berg, N. V. Slobodova, E. S. Boulygina and D. Y. Sorokin (2007). Phylogeny and evolution of the family *Ectothiorhodospiraceae* based on comparison of 16S rRNA, *cbbL* and *nifH* gene sequences. *Int J Syst Evol Microbiol* **57**(10): 2387-2398.

Treusch, A. H., S. Leininger, A. Kletzin, S. C. Schuster, H. P. Klenk and C. Schleper (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol* **7**(12): 1985-1995.

Vanparys, B., E. Spieck, K. Heylen, L. Wittebolle, J. Geets, N. Boon and P. De Vos (2007). The phylogeny of the genus *Nitrobacter* based on comparative rep-PCR, 16S rRNA and nitrite oxidoreductase gene sequence analysis. *Systematic and Spplied Microbiology* **30**(4): 297-308.

Walker, C. B., J. R. de la Torre, M. G. Klotz, H. Urakawa, N. Pinel, D. J. Arp, C. Brochier-Armanet, P. S. Chain, P. P. Chan, A. Gollabgir, J. Hemp, M. Hugler, E. A. Karr, M. Konneke, M. Shin, T. J. Lawton, T. Lowe, W. Martens-Habbena, L. A. Sayavedra-Soto, D. Lang, S. M. Sievert, A. C. Rosenzweig, G. Manning and D. A. Stahl (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci USA* **107**(19): 8818-8823.

Ward, B. B. and G. D. O'Mullan (2002). Worldwide Distribution of *Nitrosococcus oceani*, a Marine Ammonia-Oxidizing Gammaproteobacterium, Detected by PCR and Sequencing of 16S rRNA and *amoA* Genes. *Appl Environ Microbiol* **68**(8): 4153-4157.

Watson, S. W. (1971). Taxonomic Considerations of the Family *Nitrobacteraceae* Buchanan: Requests for Opinions. *Int J Syst Bacteriol* **21**(3): 254-270.

Watson, S. W. and J. B. Waterbury (1971). Characteristics of two marine nitrite oxidizing bacteria, *Nitrospina gracilis* nov. gen. nov. sp. and *Nitrococcus mobilis* nov. gen. nov. sp. *Arch Microbiol* **77**(3): 203-230.

Watson, S. W., E. Bock, F. W. Valois, J. B. Waterbury and U. Schlosser (1986). *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. *Arch Microbiol* **144**(1): 1-7.

Weidler, G. W., M. Dornmayr-Pfaffenhuemer, F. W. Gerbl, W. Heinen and H. Stan-Lotter (2007). Communities of archaea and bacteria in a subsurface radioactive thermal spring in the Austrian Central Alps, and evidence of ammonia-oxidizing Crenarchaeota. *Appl Environ Microbiol* **73**(1): 259-270.

Wertz, S., F. Poly, X. Le Roux and V. Degrange (2008). Development and application of a PCR-denaturing gradient gel electrophoresis tool to study the diversity of *Nitrobacter*-like *nxrA* sequences in soil. *FEMS Microbiol Ecol* **63**(2): 261-271.

Winogradsky, S. (1892). Contributions a la morphologie des organismes de la nitrification. *Arch Sci Biol* **1**: 88-137.

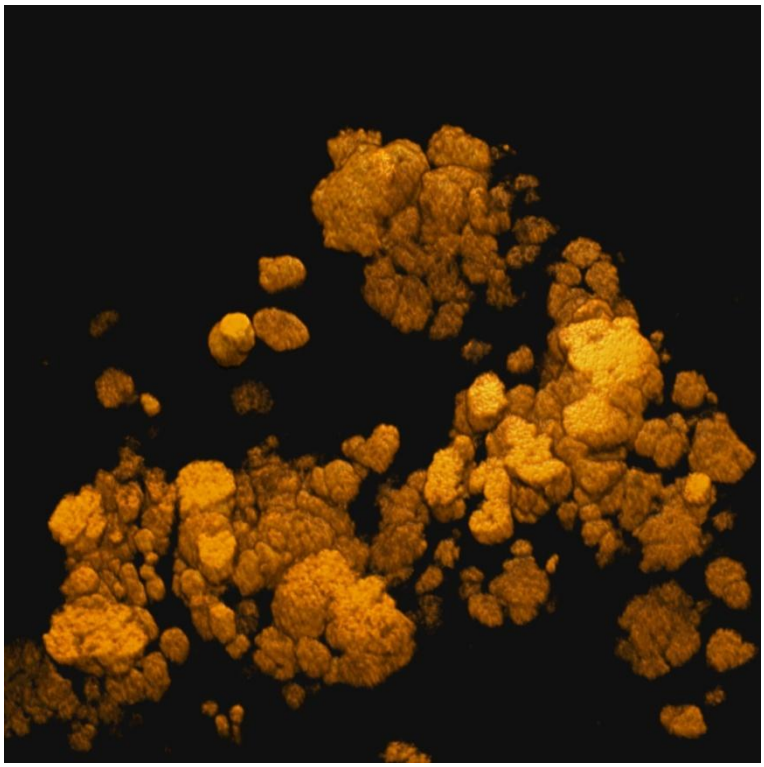
Winslow, C. E., J. Broadhurst, R. E. Buchanan, C. Krumwiede, L. A. Rogers and G. H. Smith (1917). The Families and Genera of the Bacteria: Preliminary Report of the Committee of the Society of American Bacteriologists on Characterization and Classification of Bacterial Types. *J Bacteriol* **2**(5): 505-566.

Zehr, J. P. and B. B. Ward (2002). Nitrogen Cycling in the Ocean: New Perspectives on Processes and Paradigms. *Appl Environ Microbiol* **68**(3): 1015-1024.

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

Chapter II

A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria



Front: 3D visualization of a *Nitrospira* community in an activated sludge floc. The cell clusters were stained by fluorescence *in situ* hybridization with a genus *Nitrospira*-specific probe and colored artificially by digital image processing.

A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria

Sebastian Lücker[a], Michael Wagner[a], Frank Maixner[a,1], Eric Pelletier[b,c,d], Hanna Koch[a], Benoit Vacherie[b], Thomas Rattei[e], Jaap S. Sinninghe Damsté[f], Eva Spieck[g], Denis Le Paslier[b,c,d], and Holger Daims[a,2]

[a]Department of Microbial Ecology, Ecology Centre, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria; [b]Commissariat à l'Energie Atomique, Genoscope, 91000 Evry, France; [c]Centre National de la recherche Scientifique, UMR8030, 91000 Evry, France; [d]Université d'Evry-Val-d'Essonne, 91000 Evry, France; [e]Department for Computational Systems Biology, Ecology Centre, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria; [f]Department of Marine Organic Biogeochemistry, Royal Netherlands Institute for Sea Research, P.O. Box 59, 1790 AB Den Burg, The Netherlands; [g]University of Hamburg, Biozentrum Klein Flottbek, Mikrobiologie und Biotechnologie, Ohnhorststr. 18, 22609 Hamburg, Germany

[1]Present address: Institute for Mummies and the Iceman, EURAC research, Viale Druso 1, 39100 Bolzano, Italy
[2]To whom correspondence should be addressed.

Author contributions: S.L., M.W., D.L.P., and H.D. designed research; S.L., F.M., H.K., B.V., and J.S.S.D. performed research; T.R. and E.S. contributed new reagents/analytic tools; S.L., F.M., E.P., T.R., and H.D. analyzed data; and S.L., M.W., and H.D. wrote the paper.

Data deposition: The genome sequence reported in this paper has been deposited in the GenBank database (accession no. FP929003).

Running title: Genome analysis of "*Candidatus* Nitrospira defluvii"

Published in:

Proceedings of the National Academy of Sciences of the United States of America (2010), Vol. 107 (30), p.13479 – 13484

www.pnas.org/cgi/doi/10.1073/pnas.1003860107

Abstract

Nitrospira are barely studied and mostly uncultured nitrite-oxidizing bacteria, which are according to molecular data among the most diverse and widespread nitrifiers in natural ecosystems and biological wastewater treatment. Here, environmental genomics was used to reconstruct the complete genome of "*Candidatus* Nitrospira defluvii" from an activated sludge enrichment culture. Based on this first deciphered *Nitrospira* genome and on experimental data, we show that *Ca*. N. defluvii differs dramatically from other known nitrite oxidizers in the key enzyme nitrite oxidoreductase (NXR), the composition of the respiratory chain, and the pathway used for autotrophic carbon fixation, suggesting multiple independent evolution of chemolithoautotrophic nitrite oxidation. Adaptations of *Ca*. N. defluvii to substrate-limited conditions include an unusual periplasmic NXR, which is constitutively expressed, and pathways for the transport, oxidation and assimilation of simple organic compounds that allow a mixotrophic lifestyle. The reverse tricarboxylic acid cycle as pathway for CO_2 fixation and the lack of most classical defence mechanisms against oxidative stress suggest that *Nitrospira* evolved from microaerophilic or even anaerobic ancestors. Unexpectedly, comparative genomic analyses indicate functionally significant lateral gene transfer events between the genus *Nitrospira* and anaerobic ammonium-oxidizing planctomycetes, which share highly similar forms of NXR and other proteins reflecting that two key processes of the nitrogen cycle are evolutionary connected.

Introduction

Nitrification, the microbially catalyzed sequential oxidation of ammonia *via* nitrite to nitrate, is a key process of the biogeochemical nitrogen cycle and of biological wastewater treatment. The second step of nitrification is carried out by chemolithoautotrophic nitrite-oxidizing bacteria (NOB), which are phylogenetically heterogeneous (Teske *et al.*, 1994) and occur in a wide range of aquatic and terrestrial ecosystems. Most studies on the physiology of NOB used pure cultures of *Nitrobacter*, which belong to the *Alphaproteobacteria (Teske et al., 1994)*, and complete genome sequences from NOB are available for three *Nitrobacter* strains (Starkenburg *et al.*, 2006; Starkenburg *et al.*, 2008b) and the marine gammaproteobacterium *Nitrococcus mobilis* (GenBank acc. no. AAOF00000000). However, cultivation-independent molecular methods revealed that *Nitrospira*, forming a deeply branching lineage in the bacterial phylum *Nitrospirae* (Ehrich *et al.*, 1995), are by far the most diverse and abundant NOB (Daims *et al.*, 2001). Besides their wide distribution in natural habitats such as soils (Freitag *et al.*, 2005), sediments (Altmann *et al.*, 2003), the oceans (Watson *et al.*, 1986), and hot springs (Lebedeva *et al.*, 2005), members of the genus *Nitrospira* are the predominant NOB in wastewater treatment plants (Daims *et al.*, 2001) and thus belong to the microorganisms most relevant for biotechnology.

The immense ecological and technical significance of *Nitrospira* contrasts our scarce knowledge about these bacteria. As the majority of *Nitrospira* are uncultured, and the available cultures are difficult to maintain, only few studies have addressed their ecology and physiology (e.g., Schramm *et al.*, 1999; Daims *et al.*, 2001; Maixner *et al.*, 2006). Furthermore, except of one 137 kbp contig (Maixner *et al.*, 2008), genomic sequences from *Nitrospira* have not been obtained yet. This situation has been highly unsatisfactory, because deeper insight into the biology of these elusive NOB is crucial for a better understanding of nitrogen cycling in natural and engineered systems.

Recently, a *Nitrospira* strain was enriched from activated sludge and partly characterized (Spieck *et al.*, 2006). This organism, tentatively named "*Candidatus* Nitrospira defluvii", belongs to *Nitrospira* sublineage I that is most important for sewage treatment (Daims *et al.*, 2001) but has no representative in pure culture. Here, the complete genome of *Ca*. N. defluvii was reconstructed from a metagenomic library of the enrichment. More than two decades after *Nitrospira* were discovered (Watson *et al.*, 1986), we provide an analysis of a *Nitrospira* genome with previously unmatched insight into the biology of *Nitrospira*, show striking differences in key metabolic pathways between *Nitrospira* and other NOB, and change the current perception on the evolution of NO_2^- oxidation.

Results and Discussion

Genome Reconstruction. Quantitative FISH had shown that the NO_2^--oxidizing enrichment consisted of 86% of *Ca*. N. defluvii and did not contain other known NOB (Spieck *et al.*, 2006). The complete genome of *Ca*. N. defluvii was reconstructed from this enrichment by an environmental genomics approach similar to that used for inferring the genome sequence of the anaerobic

ammonium-oxidizing bacterium ("anammox" organism) "*Candidatus* Kuenenia stuttgartiensis" (Strous *et al.*, 2006). The completeness and correct assembly of the *Nitrospira* genome was indicated by the retrieval of all 63 clusters of orthologous groups of proteins (COGs), which are present in all genomes in the current COG database (Fig. S1), by lack of suspicious redundancy in gene content, and by the presence of all essential genes in key biosynthetic pathways. The low frequency of single nucleotide polymorphisms (about one per 500 kbp) strongly suggests that the enrichment culture contained only one *Nitrospira* strain. Key features of the genome are summarized in Table S1 and Fig. S1. About 30% of the predicted coding sequences (CDS) have no homologs in other organisms, reflecting the distant relationship of *Nitrospira* to other bacteria and the lack of genome sequences from the genus *Nitrospira* in public databases. Furthermore, only two lineages within the phylum *Nitrospirae* have been explored on a genomic level. The closest genome-sequenced relatives of *Ca*. N. defluvii belong to the genus *Leptospirillum* and are aerobic acidophilic iron oxidizers (Tyson *et al.*, 2004; Levican *et al.*, 2008; Goltsman *et al.*, 2009). In addition, the genome sequence of the anaerobic sulphate reducer *Thermodesulfovibrio yellowstonii* (GenBank acc. no. NC_011296) also belonging to the *Nitrospirae*, is publicly available.

Nitrite Oxidation and Energy Metabolism. The key enzyme for NO_2^- oxidation by NOB is nitrite oxidoreductase (NXR), which shuttles two electrons per oxidized NO_2^- into the electron transport chain. In *Nitrobacter*, NXR is an iron-sulphur molybdoprotein (Meincke *et al.*, 1992) located at the inner cell membrane and at intracytoplasmic membranes (ICM). The reaction catalyzed by this NXR is reversible, so that the enzyme also reduces NO_3^- with electrons derived from organic compounds. Depending on the applied purification method, this NXR was found to consist of two (Meincke *et al.*, 1992) or three subunits with a supposed $\alpha_2\beta_2\gamma_1$ stoichiometry (Sundermeyer-Klinger *et al.*, 1984). The α-subunit (NxrA) is thought to contain the substrate-binding site with the molybdopterin cofactor (Mo-co) (Sundermeyer-Klinger *et al.*, 1984; Meincke *et al.*, 1992), whereas the β-subunit (NxrB) with [Fe-S] clusters probably channels electrons from the α- to the γ-subunit or directly to the membrane-integral electron transport chain (Kirstein and Bock, 1993).

Nitrospira are gram-negative bacteria lacking ICM (Watson *et al.*, 1986). Although no NO_3^--reducing activity has been demonstrated yet for their nitrite-oxidizing system, the term NXR is used here to be consistent with established terminology (Starkenburg *et al.*, 2006). First insight into the nature of the *Nitrospira* NXR was obtained by studying a pure culture of *Nitrospira moscoviensis* (Spieck *et al.*, 1998). Four major proteins were detected in membrane fractions showing a high NO_2^--oxidizing activity *in vitro*. Antibodies originally raised against NxrB of *Nitrobacter* bound to one of these proteins, which was designated the NxrB of *N. moscoviensis* (Spieck *et al.*, 1998). Another protein with an apparent molecular mass of 130 kDa resembled the NxrA of *Nitrobacter* (115 – 130 kDa). The other two proteins were not further characterized. The NXR of *N. moscoviensis* was

also shown to contain molybdenum and to be located at the inner cell membrane, where it faces the periplasmic space (Spieck *et al.*, 1998).

The genome of *Ca*. N. defluvii was screened for CDS with a predicted molecular mass resembling the NxrA and NxrB of *N. moscoviensis* and similarity to known NO_2^-/NO_3^--binding molybdoenzymes, such as the NXR of *Nitrobacter* or bacterial nitrate reductases (NARs). Two candidates were identified for each NxrA and NxrB (Table S3). The genes are co-localized in two clusters (*nxrA1B1* and *nxrA2B2*), which are separated by 17 other CDS from each other. The amino acid identities are 86.6% for the two NxrA and 100% for the two NxrB copies (the *nxrB* genes are identical except for a synonymous single base substitution). NxrA1 and NxrA2 contain binding motifs for one [Fe-S] cluster and for molybdenum, which are indicative of the type II group in the dimethyl sulfoxide (DMSO) reductase family of Mo-co binding enzymes (SI Results and Fig. S2*A*, *B*). Five residues, which are conserved in the α-subunits of NARs and in the NxrA of *Nitrobacter* and *Nitrococcus*, have been proposed to interact with NO_2^-/NO_3^- or to affect the conformation of the substrate entry channel (Martinez-Espinosa *et al.*, 2007). Except for one threonine, which is replaced by asparagine (Fig. S2*B*), these residues are conserved in both NxrA copies of *Ca*. N. defluvii, suggesting that the α-subunit contains the substrate-binding site. Consistent with the periplasmic orientation of NXR in *N. moscoviensis* (Spieck *et al.*, 1998), NxrA1 and NxrA2 of *Ca*. N. defluvii contain an N-terminal twin-arginine motif for export *via* the twin-arginine protein translocation (Tat) pathway.

Both NxrB copies of *Ca*. N. defluvii lack a predicted signal peptide, but may be co-translocated with NxrA into the periplasm by a "hitchhiker" mechanism as proposed for the β-subunits of other periplasmic Mo-co binding enzymes (e. g., McDevitt *et al.*, 2002). Four cysteine-rich binding motifs for [Fe-S] clusters, which occur also in NxrB of *Nitrobacter* and *Nitrococcus*, were identified (Fig. S2*C*, *D*). Homologous [Fe-S] clusters mediate intramolecular electron transfer in nitrate reductase A of *E. coli* (Blasco *et al.*, 2001).

All NxrA and NxrB copies of *Ca*. N. defluvii lack transmembrane helices, although NXR is membrane-associated in *Nitrospira* (Spieck *et al.*, 1998). Theoretically, the α/β complex might cluster with a membrane-bound terminal oxidase that receives electrons from NXR. However, other enzymes in the DMSO reductase family contain an additional membrane-integral γ-subunit, which is the membrane anchor of the holoenzyme and channels electrons between the β-subunit and the electron transport chain *via* one or two hemes (Rothery *et al.*, 2008). Four proteins encoded by *Ca*. N. defluvii could be heme-containing subunits of NXR (Table S3). Each has one transmembrane domain and an N-terminal signal peptide for translocation *via* the Sec-pathway. The largest candidate (66.7 kDa) is a *c*-type cytochrome with two predicted heme binding sites. The other three proteins are smaller (29.7 – 34.3 kDa) and remotely similar to the γ-subunit of chlorate reductase, which contains one *b*-type heme (Thorell *et al.*, 2003). These genes are not in direct proximity of the *nxrAB* clusters, but the predicted molecular masses of their products resemble the two uncharacterized major proteins from

N. moscoviensis membrane extracts (62 and 29 kDa) (Spieck *et al.*, 1998). Their biological functions and the exact composition of NXR await experimental clarification.

The sequenced *Nitrobacter* genomes encode a peptidyl-prolyl *cis-trans* isomerase (NxrX) proposed to assist in the folding of NXR (Starkenburg *et al.*, 2006; Starkenburg *et al.*, 2008b). *Ca*. N. defluvii lacks a homolog of NxrX, but one CDS is similar to chaperones involved in the assembly of other DMSO reductase-family enzymes *(Thorell et al., 2003)*. It is located directly upstream of one putative membrane-integral NXR subunit (Table S3) and could play a role in NXR maturation.

Based on biochemical (Spieck *et al.*, 1998) and genomic data, for *Nitrospira* a membrane-bound periplasmic NXR is proposed that consists of at least two subunits (Fig. 1). High-potential electrons from NO_2^- are probably transferred to cytochrome (cyt.) *c* like in *Nitrobacter* (Sundermeyer-Klinger *et al.*, 1984), and then to a terminal cyt. *c* oxidase (Fig. 1). In *Nitrobacter*, the terminal oxidase is of the *aa3*-type (Starkenburg *et al.*, 2008b). The lack of detectable cyt. *a* in *Nitrospira* cultures (Watson *et al.*, 1986; Ehrich *et al.*, 1995) and of genes coding for *a*-type cytochromes in *Ca*. N. defluvii implies that *Nitrospira* possess a different type of terminal oxidase. Intriguingly, the genome does not encode any known heme-copper oxidase, which could transfer electrons from cyt. *c* to O_2. However, *Ca*. N. defluvii has a heterodimeric cyt. *bd* quinol oxidase (genes *cydA* and *cydB*; Table S3) that could

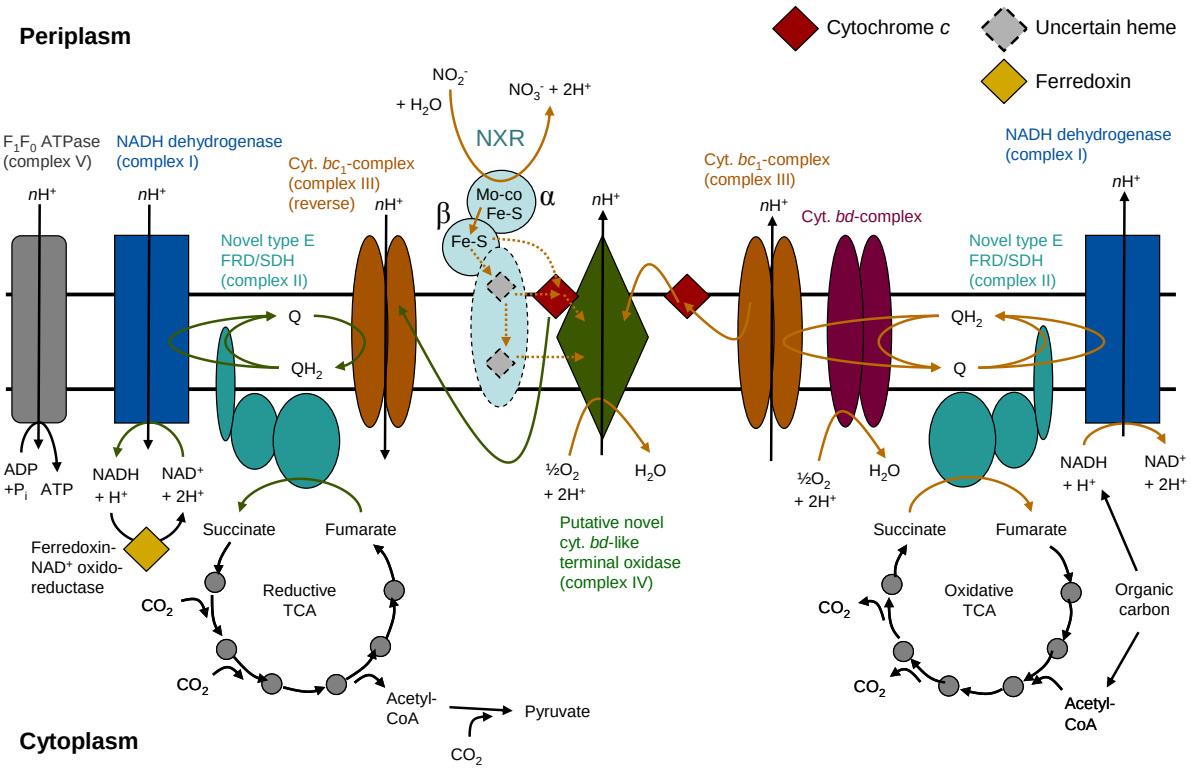


Figure 1. Schematic illustration of the genome-based model of energy metabolism in *Ca*. N. defluvii. Orange arrows indicate electron flow in the oxidative branches of the electron transport chain, whereas green arrows indicate reverse electron transport from NO_2^- to NAD^+. Stippled black lines point out that the membrane-integral subunit of NXR is uncertain. Stippled orange arrows show hypothetical possibilities for electron flow from NXR to the putative cyt. *c*-oxidase. nH^+ indicates that the number of translocated protons is unknown, because the H^+/e^- ratio of the respective complexes has not been determined for Nitrospira. FRD, fumarate reductase; SDH, succinate dehydrogenase. See main text for the definitions of other abbreviations, and refer to Table S3 for a list of the involved proteins.

receive electrons derived from low-potential donors, such as organic carbon, *via* the quinol pool (Fig. 1). The genome contains four additional CDS that resemble the CydA subunit of cyt. *bd* oxidases, but can be distinguished from the canonical proteins by phylogenetic analysis (Fig. S3*A*). We refer to these uncharacterized proteins as putative "cyt. *bd*-like oxidases". They contain 14 predicted transmembrane helices and several histidines that may serve as heme ligands (Fig. S3*B*). Interestingly, one of these CDS (Nide0901) also contains a putative copper (Cu_B) binding site (Fig. S3*B*). This motif is characteristic for the binuclear center of heme-copper cyt. *c* oxidases and it is thus tempting to speculate that Nide0901 could replace the lacking canonical heme-copper oxidases in *Nitrospira* (Fig. 1). The proposed function of Nide0901 as terminal oxidase gains further support from transcriptional analysis. High levels of *nide0901* mRNA were detected in presence of the electron donor NO_2^- and the terminal electron acceptor O_2, whereas the transcription of this gene decreased markedly in absence of these substrates (Fig. S3*D*). An alternative to a membrane-bound terminal oxidase would be a soluble cytoplasmic O_2 reductase, but this is not supported by the genomic data.

The genome-based model of energy metabolism in *Ca*. N. defluvii comprises a branched respiratory chain for NO_2^- oxidation, for the use of low-potential electron donors such as organic substrates, and for reverse electron transport (Fig. 1). In addition, two copper-containing nitrite reductases (NirK; Table S3) were identified. NirK forms NO from NO_2^- in denitrifying organisms, including other nitrifiers (e.g., Starkenburg *et al.*, 2008a). Although denitrification by *Ca*. N. defluvii has not been experimentally demonstrated, the *nirK* genes indicate that this organism may denitrify NO_2^-, for example by using organic substrates as electron donor. If NXR works reversibly in *Nitrospira*, denitrification could also start from NO_3^-. Other denitrification genes were not found. In *Nitrobacter*, NO may function in reverse electron transport (Bock *et al.*, 1991) and electron flux regulation (Starkenburg *et al.*, 2008a). It remains unclear whether NO plays similar physiological roles in *Nitrospira*.

Expression of NXR. To test whether NO_2^- induces the expression of NXR, RNA was extracted from enrichment biomass during starvation in NO_2^--free medium and after addition of NO_2^-, and *nxrB* mRNA was analyzed by reverse transcription (RT)-PCR. Interestingly, a low level of *nxrB* mRNA was detected after starvation for 11 days in NO_2^--free medium (Fig. S2*E*). Addition of NO_2^- led to an increased transcription of *nxrB*, whereas the level of 16S rRNA from *Ca*. N. defluvii did not change markedly (Fig. S2*E*). NxrB protein was detected even after starvation in NO_2^--free medium for 110 days, and its level increased markedly upon addition of NO_2^- (Fig. S2*E*). These results support the annotation of NXR. The constitutive expression of NXR should enable *Ca*. N. defluvii to use NO_2^-, whose concentration usually is low and fluctuates in natural habitats, immediately after this energy source becomes available.

Autotrophy. NOB of the genus *Nitrobacter* (Starkenburg *et al.*, 2006) and, based on genomic data, also *Nitrococcus* use the Calvin-Benson-Bassham (CBB) cycle for CO_2 fixation. The key enzymes of this pathway are ribulose-1,5-bisphosphate carboxylase (RubisCO) and ribulose-5-phosphate kinase. *Nitrospira* also grow chemolithoautotrophically on NO_2^- and CO_2 (Ehrich *et al.*, 1995; Spieck *et al.*, 2006), but their pathway for CO_2 fixation was not identified previously. *Ca*. N. defluvii encodes a form IV RubisCO-like protein (Fig. S4*A*) lacking functional key residues of canonical RubisCO (Fig. S4*B*). In *Bacillus subtilis*, a form IV RubisCO-like protein has no *bona fide* carboxylating activity (Ashida *et al.*, 2003). The absence of other genes similar to RubisCO and of ribulose-5-phosphate kinase suggests that the CBB cycle does not operate in *Ca*. N. defluvii. Instead, all genes of the reductive tricarboxylic acid (rTCA) cycle are present, including the key enzymes ATP-citrate lyase and 2-oxoglutarate:ferredoxin oxidoreductase (OGOR), and also pyruvate:ferredoxin oxidoreductase (POR) (Table S3 and SI Results).

Operation of the rTCA cycle in *Ca*. N. defluvii was confirmed by the small carbon isotopic fractionation factor (ε) between biomass and CO_2 of 2-6‰ (Table S2), typical for the rTCA cycle (Quandt *et al.*, 1977). Furthermore, the abundant (ca. 80% of all fatty acids) and characteristic straight-chain fatty acid for *Ca*. N. defluvii, $C_{16:1}$ ω5 (Spieck *et al.*, 2006), was 3-6‰ enriched relative to the biomass, whereas isoprenoid lipids were ca. 4‰ depleted (Table S2). This trend of more enriched straight-chain lipids is unusual for almost all carbon fixation pathways except for the rTCA cycle (van der Meer *et al.*, 1998).

As POR and OGOR generally are O_2-sensitive enzymes (Campbell *et al.*, 2006), the rTCA cycle is mainly found in anaerobic organisms, and its presence in an aerobic nitrifier seems surprising. However, this pathway is functional in some microaerophilic autotrophs such as *Hydrogenobacter thermophilus* (Shiba *et al.*, 1985), and it was identified in *Leptospirillum* genomes (Levican *et al.*, 2008; Goltsman *et al.*, 2009). *H. thermophilus* has two isozymes of OGOR, a two-subunit enzyme needed under anoxic conditions and a more O_2-tolerant novel five-subunit form, which mainly supports aerobic growth (Yamamoto *et al.*, 2006), and it also has an unusual five-subunit POR (Ikeda *et al.*, 2006). Highly similar five-subunit OGOR and POR in *Ca*. N. defluvii (SI Results) and *Leptospirillum* (Levican *et al.*, 2008) may allow the rTCA cycle to function in these aerobic members of the *Nitrospirae* phylum. Thus, on the basis of genomic and isotopic data, *Nitrospira* fix CO_2 *via* the rTCA cycle and represent the only nitrifier for which this pathway has been detected.

Use of Organic Substrates. *Ca*. N. defluvii and *Nitrospira marina* benefit from simple organic compounds in nitrite media (Watson *et al.*, 1986; Spieck *et al.*, 2006), and uncultured *Nitrospira* in sewage plants take up pyruvate (Daims *et al.*, 2001). However, it is unknown whether *Nitrospira* use organic substrates only as carbon sources or also for energy generation. Interestingly, the *Ca*. N. defluvii genome encodes pathways for the catabolic degradation and for the assimilation of acetate, pyruvate, and formate (Fig. S5, SI Results) and candidate genes were found for the degradation of

branched amino acids. As the Embden-Meyerhof-Parnas pathway is complete, *Ca*. N. defluvii should be able to metabolize hexose sugars. This is consistent with carbon being stored as glycogen (SI Results). Two of the three sequenced *Nitrobacter* genomes also contain the complete glycolysis pathway (Starkenburg *et al.*, 2008b), but growth of *Nitrobacter* on sugars has not been reported. Whether *Ca*. N. defluvii can take up and use sugars should mainly depend on functional sugar transport systems. The genome indeed contains putative sugar transporters (Table S1), but their function remains to be determined.

The oxidative tricarboxylic acid (oTCA) cycle shares most enzymes with the rTCA cycle except of citrate synthase and the 2-oxoglutarate dehydrogenase complex (ODH). *Ca*. N. defluvii encodes citrate synthase but apparently lacks ODH, which may however be replaced by OGOR (Table S3 and SI Results). A complete oTCA cycle was reported for *Nitrobacter* (Bock *et al.*, 1991), indicating that this pathway is not unusual in NOB.

Purely heterotrophic growth of *Nitrospira* has not been observed yet. However, if all potentially involved genes are functional, *Ca*. N. defluvii benefits from a mixotrophic lifestyle using organic compounds from sewage in addition to NO_2^- and CO_2.

Stress Response and Defense. *Ca*. N. defluvii is exposed to a plethora of potentially toxic substances in sewage. Accordingly, the genome encodes multidrug efflux systems and transporters for heavy metals, organic solvents, and antimicrobials (Table S1), and it contains genes for cyanate and arsenic resistance (Table S3, SI Results). As shown previously (Maixner *et al.*, 2008), *Ca*. N. defluvii has a functional chlorite dismutase that could degrade ClO_2^- in polluted environments, chlorinated activated sludge, or in the proximity of chlorate-reducing microbes. Most intriguingly, *Ca*. N. defluvii lacks key genes for protection from reactive oxygen species (ROS) present in most aerobic organisms. No catalase, superoxide dismutase, and superoxide reductase was found. Two cyt. *c* peroxidases and several thioredoxin-dependent peroxiredoxins could function as H_2O_2 scavengers (SI Results, Table S3). Protection from O_2^- and also H_2O_2 might be conferred by manganese [Mn(II)] (Horsburgh *et al.*, 2002). Indeed, the required permease for Mn import was identified in the genome. Bacterioferritin and carotenoids (Table S3) could also contribute to protection from radicals and ROS. Moreover, the intracellular O_2 level could be kept low by the canonical cyt. *bd* oxidase. Homologs in other organisms have a high affinity to O_2 and contribute to oxidative stress protection (Das *et al.*, 2005). Growth of *Nitrospira* in biofilms and flocs (e.g., Spieck *et al.*, 2006) could offer additional protection from ambient O_2.

Ca. N. defluvii carries one region of clustered, regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (*cas*) genes for phage defense (Barrangou *et al.*, 2007). The CRISPR repeats of *Ca*. N. defluvii show no sequence similarity to those of *Leptospirillum* groups II and III, which also differ in their Cas proteins (Goltsman *et al.*, 2009), suggesting that this defense mechanism was independently acquired by different members of the *Nitrospirae* phylum.

Ecophysiology and Evolutionary History of *Nitrospira*. The NXRs of *Nitrobacter*, *Nitrococcus* and *Nitrospira* differ in their subcellular localization and phylogenetic position within the DMSO reductase family. The NXRs of *Nitrobacter* and *Nitrococcus* are closely related to NARs. They are associated with the cytoplasmic membrane and ICM with the active site facing the cytoplasm (Spieck and Bock, 2005). The unique NXR of *Nitrospira* does not cluster with the NARs (Fig. 2). It is also attached to the cytoplasmic membrane, but is oriented towards the periplasmic space (Spieck *et al.*, 1998; and this study). The periplasmic orientation should be energetically advantageous, because proton release by NO_2^- oxidation in the periplasm and concomitant proton consumption by O_2 reduction in the cytoplasm contribute to the membrane potential (Fig. 1). Furthermore, only a cytoplasmic NXR requires the transport of NO_2^- and NO_3^- in opposite directions across the inner membrane. Accordingly, putative NO_2^-/NO_3^- transporters are found in all sequenced *Nitrobacter* genomes (Starkenburg *et al.*, 2008b) and in *Nitrococcus*. Their substrate affinities and turnover rates could be limiting factors for NO_2^- oxidation by these NOB. This and the catalytic properties of NXR could explain the relatively high apparent $K_m(NO_2^-)$ value of *Nitrobacter* (Schramm *et al.*, 1999). In contrast, the predicted NO_2^- and NO_3^- transporters of *Ca*. N. defluvii (Table S1) most likely play no role in nitrite oxidation but are required only for nitrogen assimilation and resistance against excess nitrite (SI Results).

Consistent with the predicted advantages of their periplasmic NXR, *Nitrospira* are better adapted to low NO_2^- concentrations (Schramm *et al.*, 1999; Maixner *et al.*, 2006), which also were key to the selection against co-existing *Nitrobacter* during enrichment (Spieck *et al.*, 2006). As NO_2^- rarely

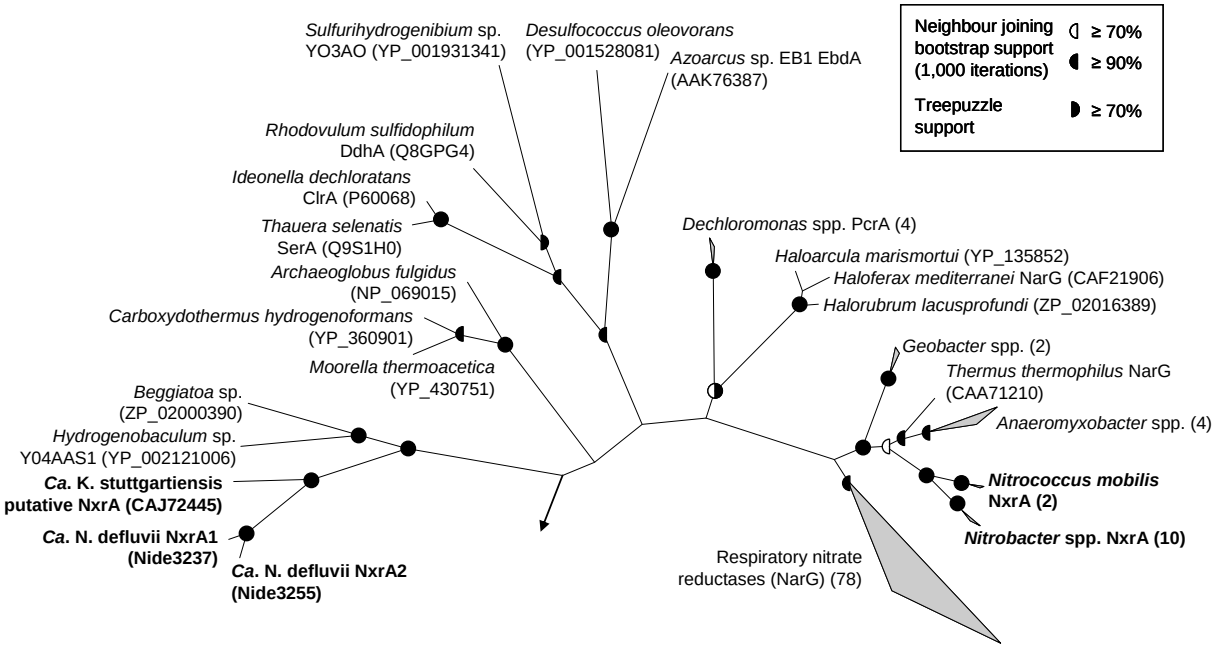


Figure 2. Maximum likelihood tree showing the phylogenetic positioning of selected type II enzymes of the DMSO reductase family. For phylogenetic analysis of the catalytic (α) subunits, 1,308 amino acid positions were considered. Names of validated enzymes are indicated (Nxr, nitrite oxidoreductase, printed bold; Nar, membrane-bound respiratory nitrate reductase; Pcr, perchlorate reductase; Ebd, ethylbenzene dehydrogenase; Ddh, dimethylsulfide dehydrogenase; Clr, chlorate reductase; Ser, selenate reductase). Numbers in parentheses give the number of sequences within a group or the accession number, respectively.

accumulates in natural environments, the highly efficient use of this substrate most likely is a main reason for the competitive success and wide natural distribution of *Nitrospira*.

The use of different key enzymes and pathways (e. g., CO_2 fixation) by *Nitrospira* in contrast to the proteobacterial NOB *Nitrobacter* and *Nitrococcus* suggests that chemolithoautotrophic NO_2^- oxidation evolved independently in these lineages. Based on the close phylogenetic affiliation of *Nitrobacter* and *Nitrococcus* to phototrophic *Proteobacteria*, which also possess ICM, Teske *et al*. (Teske *et al.*, 1994) hypothesized that these NOB were derived from phototrophic ancestors. Indeed, a recently isolated anaerobic phototroph, which uses NO_2^- as electron donor, is closely related to *Nitrococcus* (Griffin *et al.*, 2007). A cytoplasmically oriented NXR would probably be no disadvantage for phototrophic NOB where the membrane potential is sustained mainly by light-driven cyclic electron flow. The orientation of NXR may not easily be reversed, because it intimately affects the interaction with downstream components of the electron transport chain. Hence, the conservation of a cytoplasmic NXR during the transition from phototrophy to chemolithotrophy could explain the orientation of NXR in *Nitrobacter* and *Nitrococcus*. In contrast and consistent with the absence of ICM in *Nitrospira*, no phototrophic relative of *Nitrospira* is known and we hypothesize that the capability to gain energy from NO_2^- oxidation has evolved in this lineage from an anaerobic non-phototrophic ancestor. An anaerobic or microaerophilic origin of *Nitrospira* would be consistent with the rTCA cycle, the presence of the anaerobic cobalamin biosynthesis pathway (Table S3), and the lack of classical defense mechanisms against ROS. Additional support for this hypothesis stems from estimating genus divergence times within the *Nitrospirae* phylum by using 16S rRNA as molecular clock (SI Results). Extant *Nitrospira* are active at low dissolved O_2 levels in bioreactors and might still prefer hypoxic conditions (Park and Noguera, 2008).

Intriguingly, comparative genomics revealed an unexpected evolutionary link between *Nitrospira* and anammox organisms. For example, the closest homolog of the NXR of *Ca*. N. defluvii was found in *Ca*. K. stuttgartiensis (Fig. 2). NO_2^- oxidation is an integral step of the anammox metabolism where it replenishes the electron transport system (Strous *et al.*, 2006), and this NXR-like protein is the only candidate for a NO_2^--oxidizing enzyme in the *Kuenenia* genome. Its α-subunit contains the signature residues of NO_2^-/NO_3^--binding molybdoenzymes (Martinez-Espinosa *et al.*, 2007) (Fig. S2*A*, *B*). The NXRs of *Nitrospira* and *Kuenenia* are highly similar (amino acid identities are 57.4 – 57.7% for the α- and 62.5% for the β-subunits) and form a monophyletic lineage in the tree of type II enzymes of the DMSO reductase family (Fig. 2). In addition, both *Ca*. N. defluvii and *Ca*. K. stuttgartiensis have the putative chaperone for NXR assembly in analogy to NxrX of *Nitrobacter*. *Ca*. K. stuttgartiensis also has a putative cyt. *bd*-like oxidase, which is the closest relative of the four cyt. *bd*-like oxidases of *Ca*. N. defluvii (Fig. S3*A*). Interestingly, its gene is located in close proximity to *nxrA*, *nxrB*, two putative membrane subunits of NXR, and the chaperone in the *Kuenenia* genome (Fig. 3). The same region contains a monoheme cyt. *c*-like protein and three proteins of unknown function, which also have highly similar homologs in *Ca*. N. defluvii (Fig. 3). Thus, both organisms share a set of highly similar

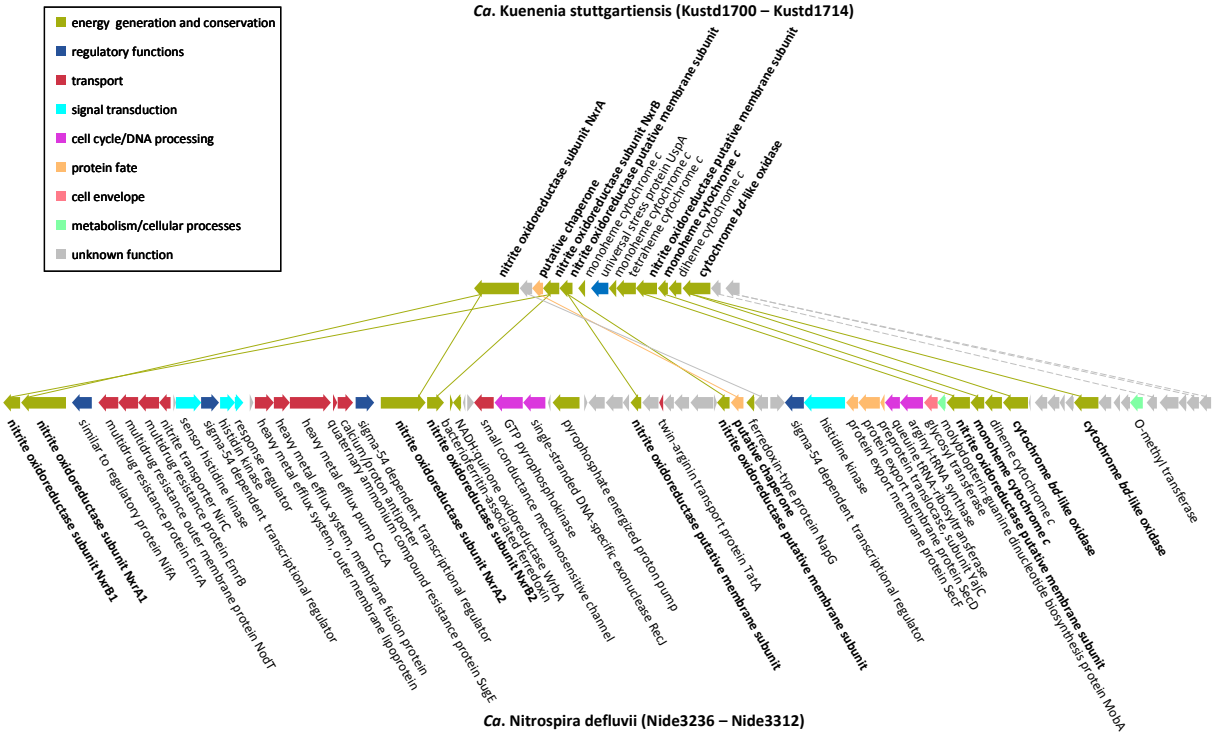


Figure 3. Schematic representation of the genomic regions in *Ca*. K. stuttgartiensis and *Ca*. N. defluvii, which contain shared genes coding for NXR, putative cyt. *bd*-like oxidases and electron carriers, and proteins of unknown function. Genes, which are connected by solid lines, have their closest homolog in either organism based on protein phylogeny. The predicted functions of their products are printed bold. Genes connected by stippled lines are similar in the two organisms, but are not the closest relatives in the respective phylogenetic protein trees, which were calculated in the course of phylome analysis. Predicted CDS and connecting lines are colored according to functional classes. CDS and intergenic regions are drawn to scale.

proteins that function in NO_2^- oxidation and probably in electron transport and respiration, and these genes are clustered as small metabolic island in the anammox genome. As anammox organisms are planctomycetes and consequently not closely related to the *Nitrospirae* (Strous *et al.*, 2006) these observations are strongly indicative of horizontal gene transfer (HGT) that established NXR and the other proteins in both lineages. Consistent with a fundamental importance of the transferred genes for the basic metabolism of *Nitrospira* and anammox, this HGT apparently occurred early during the evolution of these lineages as no remarkable deviation in GC content or codon usage of the respective genes was observed in either organism.

To explore further the influence of vertical and HGT on the evolutionary history of *Ca*. N. defluvii, we calculated phylogenies for each protein of *Ca*. N. defluvii and identified the organism encoding the respective most closely related homolog (Fig. S6). Most remarkably, in this analysis *Ca*. K. stuttgartiensis was the single organism that shared the highest number of closest homologs (71 hits) with *Ca*. N. defluvii and thus exchanged, compared to all other organisms for which genome sequences are available, most genes with *Nitrospira via* HGT. Surprisingly, the 71 hits even exceed the numbers of best hits with members of the *Nitrospirae* phylum, namely *Thermodesulfovibrio* (67 hits) and different *Leptospirillum* strains (39 – 66 hits). These findings illustrate a surprisingly small

set of closest related homologs in the *Nitrospirae*, most likely reflecting the dramatically different ecological niches inhabited by the genera affiliated with this phylum.

Taken together, the metagenome sequence of *Ca*. N. defluvii revealed that this globally important nitrite oxidizer differs fundamentally in its enzymatic repertoire (unusual NXR and putative novel terminal oxidase) and metabolic pathways (rTCA for autotrophy) from all other known nitrifiers, but strikingly exploits almost the same gene repertoire for NO_2^- oxidation as the anammox organism *Ca*. K. stuttgartiensis. The unique genomic features of *Nitrospira* already provided some well supported hypotheses for its competitive success in most nitrifying ecosystems and suggested that *Nitrospira* are well adapted to hypoxic environmental niches, where nitrite oxidation has rarely been studied until now. From an applied perspective, the lack of common protection mechanisms against oxidative stress in *Nitrospira* implies that a good aeration control is crucial for maintaining stable and active populations of these organisms in engineered systems.

Materials and Methods

Genomic Sequencing and Annotation. Metagenome sequencing and the reconstruction of the whole *Ca*. N. defluvii genome were carried out by Genoscope (SI Methods). The MaGe software system (Vallenet *et al.*, 2006) was used for the prediction, automatic annotation, and manual annotation refinement of all CDS as described in SI Methods.

Phylogenetic Analyses. Amino acid sequences of type II DMSO reductase-family enzymes, of RubisCO and RubisCO-like proteins, and of cyt. *bd* and cyt. *bd*-like oxidases were aligned and phylogenetic trees were computed by using ARB (Ludwig, 2004). For the calculation of phylogenetic trees for each protein in the proteome, PhyloGenie (Frickey and Lupas, 2004) was utilized. For details, see SI Methods.

Expression Analysis of NxrB and the Putative Terminal Cyt. *c* Oxidase (Nide0901). *Ca*. N. defluvii enrichment biomass was incubated in mineral media with or without NO_2^- and, for Nide0901, also under oxic or anoxic conditions as described in SI Methods. Following total RNA extraction, 16S rRNA of *Nitrospira* and *nxrB* or *nide0901* transcripts were detected by RT- PCR (SI Methods). Translation of NxrB was shown by Western blotting with a monoclonal antibody that binds to the NxrB of *Nitrospira* (Spieck *et al.*, 1998) (SI Methods).

Stable Carbon Isotopic Fractionation. The isotopic fraction of *Ca*. N. defluvii was measured following methods published earlier (Schouten *et al.*, 2004) (SI Methods).

Acknowledgments

We thank Lisa Stein for analyses of *nirK* genes, and Peter Bottomley, Jim Hemp, Jim Prosser and Andreas Schramm for helpful discussions. Christiane Dorninger, Alexander Galushko, Christian Baranyi, Jan Dolinšek, Patrick Tischler, Irene Rijpstra, and Michiel Kienhuis are acknowledged for technical support. This work was supported by the Vienna Science and Technology Fund (Wiener Wissenschafts-, Forschungs-, und Technologiefonds, Grant LS 216 to H.D., S.L. and F.M., and Grant LS09-40 to H.D., H.K. and S.L.), the Austrian Research Fund (Fonds zur Förderung der Wissenschaftlichen Forschung, Grant S10002-B17 to H.D., M.W., S.L. and F.M.), and the German Research Foundation (Deutsche Forschungsgemeinschaft, Grant SP 667/3-1 to E.S.).

References

Altmann, D., P. Stief, R. Amann, D. De Beer and A. Schramm (2003). *In situ* distribution and activity of nitrifying bacteria in freshwater sediment. *Environ Microbiol* **5**(9): 798-803.

Ashida, H., Y. Saito, C. Kojima, K. Kobayashi, N. Ogasawara and A. Yokota (2003). A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science* **302**(5643): 286-290.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**(5819): 1709-1712.

Blasco, F., B. Guigliarelli, A. Magalon, M. Asso, G. Giordano and R. A. Rothery (2001). The coordination and function of the redox centres of the membrane-bound nitrate reductases. *Cell Mol Life Sci* **58**(2): 179-193.

Bock, E., H. P. Koops, H. Harms and B. Ahlers (1991). The biochemistry of nitrifying organisms. *In* J. M. Shively and L. L. Barton (ed.), Variations in Autotrophic Life. London, Academic Press**:** 171-200.

Campbell, B. J., A. S. Engel, M. L. Porter and K. Takai (2006). The versatile *Epsilonproteobacteria*: Key players in sulphidic habitats. *Nat Rev Microbiol* **4**(6): 458-468.

Daims, H., J. L. Nielsen, P. H. Nielsen, K. H. Schleifer and M. Wagner (2001). *In situ* characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microbiol* **67**(11): 5273-5284.

Das, A., R. Silaghi-Dumitrescu, L. G. Ljungdahl and D. M. Kurtz, Jr. (2005). Cytochrome *bd* oxidase, oxidative stress, and dioxygen tolerance of the strictly anaerobic bacterium *Moorella thermoacetica*. *J Bacteriol* **187**(6): 2020-2029.

Ehrich, S., D. Behrens, E. Lebedeva, W. Ludwig and E. Bock (1995). A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Arch Microbiol* **164**: 16-23.

Freitag, T. E., L. Chang, C. D. Clegg and J. I. Prosser (2005). Influence of inorganic nitrogen management regime on the diversity of nitrite-oxidizing bacteria in agricultural grassland soils. *Appl Environ Microbiol* **71**(12): 8323-8334.

Frickey, T. and A. N. Lupas (2004). PhyloGenie: Automated phylome generation and analysis. *Nucl Acids Res* **32**(17): 5231-5238.

Goltsman, D. S., V. J. Denef, S. W. Singer, N. C. VerBerkmoes, M. Lefsrud, R. S. Mueller, G. J. Dick, C. L. Sun, K. E. Wheeler, A. Zemla, B. J. Baker, L. Hauser, M. Land, M. B. Shah, M. P. Thelen, R. L. Hettich and J. F. Banfield (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**(13): 4599-4615.

Griffin, B. M., J. Schott and B. Schink (2007). Nitrite, an electron donor for anoxygenic photosynthesis. *Science* **316**(5833): 1870.

Horsburgh, M. J., S. J. Wharton, M. Karavolos and S. J. Foster (2002). Manganese: Elemental defence for a life with oxygen. *Trends Microbiol* **10**(11): 496-501.

Ikeda, T., T. Ochiai, S. Morita, A. Nishiyama, E. Yamada, H. Arai, M. Ishii and Y. Igarashi (2006). Anabolic five subunit-type pyruvate:ferredoxin oxidoreductase from *Hydrogenobacter thermophilus* TK-6. *Biochem Biophys Res Comm* **340**(1): 76-82.

Kirstein, K. and E. Bock (1993). Close genetic relationship between *Nitrobacter hamburgensis* nitrite oxidoreductase and *Escherichia coli* nitrate reductases. *Arch Microbiol* **160**(6): 447-453.

Lebedeva, E. V., M. Alawi, C. Fiencke, B. Namsaraev, E. Bock and E. Spieck (2005). Moderately thermophilic nitrifying bacteria from a hot spring of the Baikal rift zone. *FEMS Microbiol Ecol* **54**(2): 297-306.

Levican, G., J. A. Ugalde, N. Ehrenfeld, A. Maass and P. Parada (2008). Comparative genomic analysis of carbon and nitrogen assimilation mechanisms in three indigenous bioleaching bacteria: Predictions and validations. *BMC Genomics* **9**: 581.

Ludwig, W. (2004). ARB: A software environment for sequence data. *Nucl Acids Res* **32**: 1363-1371.

Maixner, F., D. R. Noguera, B. Anneser, K. Stoecker, G. Wegl, M. Wagner and H. Daims (2006). Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. *Environ Microbiol* **8**(8): 1487-1495.

Maixner, F., M. Wagner, S. Lücker, E. Pelletier, S. Schmitz-Esser, K. Hace, E. Spieck, R. Konrat, D. Le Paslier and H. Daims (2008). Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium "*Candidatus* Nitrospira defluvii". *Environ Microbiol* **10**(11): 3043-3056.

Martinez-Espinosa, R. M., E. J. Dridge, M. J. Bonete, J. N. Butt, C. S. Butler, F. Sargent and D. J. Richardson (2007). Look on the positive side! The orientation, identification and bioenergetics of "Archaeal" membrane-bound nitrate reductases. *FEMS Microbiol Lett* **276**(2): 129-139.

McDevitt, C. A., P. Hugenholtz, G. R. Hanson and A. G. McEwan (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**(6): 1575-1587.

Meincke, M., E. Bock, D. Kastrau and P. M. H. Kroneck (1992). Nitrite oxidoreductase from *Nitrobacter hamburgensis*: Redox centers and their catalytic role. *Arch Microbiol* **158**: 127-131.

Park, H. D. and D. R. Noguera (2008). *Nitrospira* community composition in nitrifying reactors operated with two different dissolved oxygen levels. *J Microbiol Biotechnol* **18**(8): 1470-1474.

Quandt, L., G. Gottschalk, H. Ziegler and W. Stichler (1977). Isotope discrimination by photosynthetic bacteria. *FEMS Microbiol Lett* **1**(3): 125-128.

Rothery, R. A., G. J. Workun and J. H. Weiner (2008). The prokaryotic complex iron-sulfur molybdoenzyme family. *Biochim Biophys Acta Biomembr* **1778**(9): 1897-1929.

Schouten, S., M. Strous, M. M. Kuypers, W. I. Rijpstra, M. Baas, C. J. Schubert, M. S. Jetten and J. S. Sinninghe Damste (2004). Stable carbon isotopic fractionations associated with inorganic carbon fixation by anaerobic ammonium-oxidizing bacteria. *Appl Environ Microbiol* **70**(6): 3785-3788.

Schramm, A., D. de Beer, J. C. van den Heuvel, S. Ottengraf and R. Amann (1999). Microscale distribution of populations and activities of *Nitrosospira* and *Nitrospira* spp. along a macroscale gradient in a nitrifying bioreactor: Quantification by in situ hybridization and the use of microsensors. *Appl Environ Microbiol* **65**(8): 3690-3696.

Shiba, H., T. Kawasumi, Y. Igarashi, T. Kodama and Y. Minoda (1985). The CO_2 assimilation via the reductive tricarboxylic-acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. *Arch Microbiol* **141**(3): 198-203.

Spieck, E., S. Ehrich, J. Aamand and E. Bock (1998). Isolation and immunocytochemical location of the nitrite-oxidizing system in *Nitrospira moscoviensis*. *Arch Microbiol* **169**(3): 225-230.

Spieck, E. and E. Bock (2005). The lithoautotrophic nitrite-oxidizing bacteria. *In* J. T. Staley, D. R. Boone, D. J. Brenneret al (ed.), Bergey's Manual of Systematic Bacteriology. New York, Springer Science+Business Media. **2:** 149-153.

Spieck, E., C. Hartwig, I. McCormack, F. Maixner, M. Wagner, A. Lipski and H. Daims (2006). Selective enrichment and molecular characterization of a previously uncultured *Nitrospira*-like bacterium from activated sludge. *Environ Microbiol* **8**(3): 405-415.

Starkenburg, S. R., P. S. Chain, L. A. Sayavedra-Soto, L. Hauser, M. L. Land, F. W. Larimer, S. A. Malfatti, M. G. Klotz, P. J. Bottomley, D. J. Arp and W. J. Hickey (2006). Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *Appl Environ Microbiol* **72**(3): 2050-2063.

Starkenburg, S. R., D. J. Arp and P. J. Bottomley (2008a). Expression of a putative nitrite reductase and the reversible inhibition of nitrite-dependent respiration by nitric oxide in *Nitrobacter winogradskyi* Nb-255. *Environ Microbiol* **10**(11): 3036-3042.

Starkenburg, S. R., F. W. Larimer, L. Y. Stein, M. G. Klotz, P. S. Chain, L. A. Sayavedra-Soto, A. T. Poret-Peterson, M. E. Gentry, D. J. Arp, B. Ward and P. J. Bottomley (2008b). Complete genome sequence of *Nitrobacter hamburgensis* X14 and comparative genomic analysis of species within the genus *Nitrobacter*. *Appl Environ Microbiol* **74**(9): 2852-2863.

Strous, M., E. Pelletier, S. Mangenot, T. Rattei, A. Lehner, M. W. Taylor, M. Horn, H. Daims, D. Bartol-Mavel, P. Wincker, V. Barbe, N. Fonknechten, D. Vallenet, B. Segurens, C. Schenowitz-Truong, C. Medigue, A. Collingro, B. Snel, B. E. Dutilh, H. J. Op den Camp, C. van der Drift, I. Cirpus, K. T. van de Pas-Schoonen, H. R. Harhangi, L. van Niftrik, M. Schmid, J. Keltjens, J. van de Vossenberg, B. Kartal, H. Meier, D. Frishman, M. A. Huynen, H. W. Mewes, J. Weissenbach, M. S. Jetten, M. Wagner and D. Le Paslier (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**(7085): 790-794.

Sundermeyer-Klinger, H., W. Meyer, B. Warninghoff and E. Bock (1984). Membrane-bound nitrite oxidoreductase of *Nitrobacter*: Evidence for a nitrate reductase system. *Arch Microbiol* **140**: 153-158.

Teske, A., E. Alm, J. M. Regan, T. S., B. E. Rittmann and D. A. Stahl (1994). Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria. *J Bacteriol* **176**(21): 6623-6630.

Thorell, H. D., K. Stenklo, J. Karlsson and T. Nilsson (2003). A gene cluster for chlorate metabolism in *Ideonella dechloratans*. *Appl Environ Microbiol* **69**(9): 5585-5592.

Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar and J. F. Banfield (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43.

Vallenet, D., L. Labarre, Z. Rouy, V. Barbe, S. Bocs, S. Cruveiller, A. Lajus, G. Pascal, C. Scarpelli and C. Medigue (2006). MaGe: A microbial genome annotation system supported by synteny results. *Nucl Acids Res* **34**(1): 53-65.

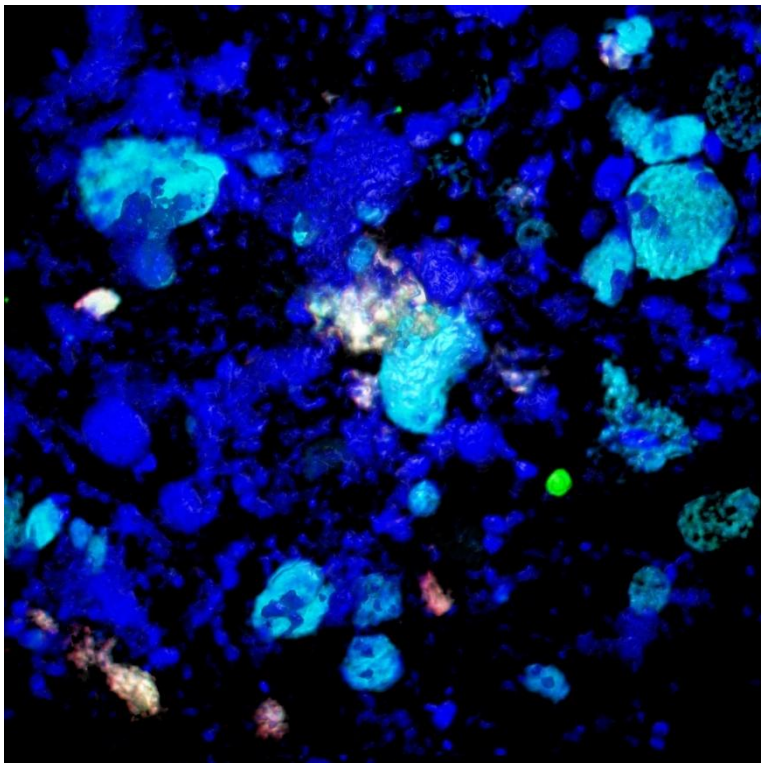
van der Meer, M. T. J., S. Schouten and J. S. S. Damste (1998). The effect of the reversed tricarboxylic acid cycle on the ^{13}C contents of bacterial lipids. *Org Geochem* **28**(9-10): 527-533.

Watson, S. W., E. Bock, F. W. Valois, J. B. Waterbury and U. Schlosser (1986). *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. *Arch Microbiol* **144**: 1-7.

Yamamoto, M., H. Arai, M. Ishii and Y. Igarashi (2006). Role of two 2-oxoglutarate:ferredoxin oxidoreductases in *Hydrogenobacter thermophilus* under aerobic and anaerobic conditions. *FEMS Microbiol Lett* **263**(2): 189-193.

Chapter III

Nitrotoga-related bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater treatment plants



Front: 3D visualization of a bacterial community in nitrifying activated sludge. The cell clusters were stained by fluorescence *in situ* hybridization and colored artificially by digital image processing. *Nitrotoga*-like organisms are shown in white, AOB in cyan, all other bacteria in blue.

Nitrotoga-related bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater treatment plants

Sebastian Lücker[1], Christiane Dorninger[1], Eva Spieck[2], Michael Wagner[1], and Holger Daims[1*]

[1]Department of Microbial Ecology, Ecology Centre, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria; [2]University of Hamburg, Biocenter Klein Flottbek, Microbiology and Biotechnology, Ohnhorststrasse 18, 22609 Hamburg, Germany

[*]Corresponding author

Running title: *Nitrotoga*-like bacteria are novel key NOB in WWTPs

Manuscript in preparation

Abstract

Over the last years numerous studies have shown members of the genus *Nitrospira* to be the most important nitrite-oxidizing bacteria (NOB) in wastewater treatment plants (WWTPs). Recently, the novel nitrite oxidizer "*Candidatus* Nitrotoga arctica" was identified in permafrost soil, and close relatives have been enriched from wastewater. Still, little is known about their *in situ* distribution, abundance, and spatial localization in natural and engineered systems. Therefore, we developed new *Nitrotoga*-specific PCR primers and FISH probes and applied these to a range of activated sludge samples from full-scale WWTPs. By using PCR, *Nitrotoga* could be detected in more than half of the samples analyzed, whereas FISH revealed a high abundance of *Nitrotoga* only in some of the plants. Surprisingly, while in most systems *Nitrotoga* coexisted with *Nitrospira*, they constituted the sole detectable NOB in one WWTP. Quantification revealed that they accounted for nearly 2% of the total bacterial community in this reactor, a number resembling the abundance of *Nitrospira* in other wastewater treatment systems. Furthermore, statistical analysis of spatial distribution patterns in activated sludge flocs revealed that *Nitrotoga* co-aggregated with ammonia oxidizers of the *Nitrosomonas oligotropha* lineage, strongly suggesting that they indeed were responsible for nitrite oxidation. Both temperature and nitrogen load appear to be environmental factors determining the abundance and ecological success of *Nitrotoga*. In conclusion, we have demonstrated that *Nitrotoga* are important nitrite oxidizers in full-scale WWTPs, often co-existing with *Nitrospira* but occasionally representing the only known NOB population.

Introduction

The anthropogenic release of large amounts of nitrogen has detrimental effects on the environment such as the eutrophication of inland and coastal water bodies, leading to algal blooms and hypoxia (Diaz and Rosenberg, 2008). As high concentrations of ammonia and nitrite are also toxic for many organisms, their discharge into the environment must be regulated to protect ecosystems and drinking water supplies (Camargo and Alonso, 2006; Conley *et al.*, 2009). Besides agriculture, human sewage is one of the largest sources of inorganic nitrogen, in particular ammonia. Prior to water discharge into natural systems, tertiary treatment in wastewater treatment plants (WWTPs) achieves the elimination of excess nitrogen by providing conditions for aerobic nitrification, the oxidation of ammonia to nitrate, and anaerobic denitrification, the subsequent reduction of nitrate to gaseous dinitrogen. Nitrification is a two-step process where ammonia-oxidizing bacteria (AOB) or archaea first oxidize ammonia to nitrite, which then is further transformed to nitrate by nitrite-oxidizing bacteria (NOB).

Before cultivation-independent molecular methods became available, members of the alphaproteobacterial genus *Nitrobacter* were thought to be mainly responsible for nitrite oxidation in sewage treatment plants, because the isolation of these NOB from activated sludge was straightforward (e.g., (Henze *et al.*, 1997). This view changed radically when molecular tools revealed that yet uncultured *Nitrospira*, but not *Nitrobacter*, are the key NOB in most engineered systems (Juretschko *et al.*, 1998; Schramm *et al.*, 1998; Daims *et al.*, 2001b). Since this discovery, research on NOB in WWTPs has focused mainly on *Nitrospira*, which belong to a distinct bacterial phylum, display a considerable phylogenetic diversity, and possess genetic and physiological features that clearly distinguish them from other known NOB (Daims *et al.*, 2001a; Gieseke *et al.*, 2005; Maixner *et al.*, 2006; Foesel *et al.*, 2008; Maixner *et al.*, 2008; Lücker *et al.*, 2010). However, novel nitrite oxidizers are still being discovered (e.g., (Schott *et al.*, 2010). Only recently, Alawi *et al.* (2007) enriched a novel nitrite-oxidizing betaproteobacterium, "*Candidatus* Nitrotoga arctica" (*Ca*. N. arctica), from permafrost soil. To date, little is known about the natural distribution and physiology of *Nitrotoga*. The organism grows at relatively low temperatures between 4°C and 17°C and prefers nitrite concentrations as low as 0.3 mM (Alawi *et al.*, 2007), which is approximately one order of magnitude below the nitrite concentrations usually applied to cultivate *Nitrospira* isolates (Lebedeva *et al.*, 2008) and even two orders below the concentrations used to grow *Nitrobacter* (Prosser, 1989).

Interestingly, the recently reported enrichment of *Nitrotoga* from a full-scale WWTP (Alawi *et al.*, 2009) raises the question of whether these novel NOB might be relevant for nitrite oxidation in engineered systems. In this context the enrichments provide hints but no clear answer, because very few *Nitrotoga* cells could already be sufficient as inoculum for a successful enrichment culture, whereas their *in situ* numbers may be low and thus irrelevant for the nitrification process in the system. In this study, we investigated whether *Nitrotoga* might be hitherto overlooked key nitrifiers in full-scale sewage treatment systems. For this purpose, we applied the full-cycle rRNA approach (Amann *et al.*, 1995) and developed new *Nitrotoga*-specific PCR primers and oligonucleotide probes

for fluorescence *in situ* hybridization (FISH) based on in-depth phylogenetic analyses of the new candidate genus *Nitrotoga*. These new cultivation-independent molecular tools were then applied to detect, visualize, and quantify *Nitrotoga* in nitrifying full-scale WWTPs and to investigate their spatial distribution patterns relative to AOB within activated sludge flocs.

Material and Methods

Activated Sludge Sampling and Fixation. Activated sludge samples were obtained from full-scale sequencing batch reactors (SBRs) (Irvine *et al.*, 1989) operated with or without differential internal cycling (DIC) (Holm, 2003), from conventional activated sludge basins, fixed bed reactors, and a membrane filtration plant. The selected WWTPs were located in Germany and Switzerland and treated municipal wastewater, which in some cases was mixed with industrial sewage or animal rendering waste (Table 1).

For FISH analysis, activated sludge samples were fixed with paraformaldehyde according to Daims *et al*. (2005). Fixed biomass was stored at -20°C. Unfixed samples for DNA extraction were harvested by centrifugation (10 000 rpm, 10 min, 4°C) and stored at -20°C.

DNA Extraction, PCR, and Cloning of 16S rRNA Genes. Genomic DNA was extracted using the PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Inc, Carlsbad, CA, USA) according to the manufacturer's instructions. For the amplification of 16S rRNA genes, reaction mixtures with the genus *Nitrotoga*-specific primer combination S-G-Ntoga-0122-a-S-19 (Ntoga122F, 5'-ATA TCG GAA CGT ACC CGG A-3') and S-G-Ntoga-1422-a-A-18 (Ntoga1422R, 5'-GCT GCT TCT GGT AGA ACC-3') were prepared according the manufacturer's recommendations in a total volume of 50 µl with 2 mM $MgCl_2$, 0.5 µM of each primer, and 1.25 U of Taq polymerase (Fermentas, St. Leon-Rot, Germany). PCR cycling consisted of an initial denaturation step at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 65°C for 30 s, and elongation at 72°C for 1 min 30 s, and was completed by a final elongation step at 72°C for 10 min. The presence and purity of amplicons were confirmed by agarose gel electrophoresis. Cloning and sequencing of amplified 16S rRNA genes were performed as described elsewhere (Juretschko *et al.*, 1998). Cloned genes to be sequenced were selected based on different restriction fragment length polymorphism (RFLP) patterns. For RFLP, 5 µl of M13 PCR product were digested with 1 µl of the restriction enzyme MspI (Fermentas, St. Leon-Rot, Germany) and 1 µl of buffer Tango at 37°C for 3 h. Fragment patterns were separated and visualized by gel electrophoresis using a 2.5% (w/v) agarose gel.

Phylogenetic Analysis. Recently deposited near full-length *Nitrotoga*-like 16S rRNA gene sequences (≥1450 nucleotides and ≥96% nucleic acid similarity to *Ca*. N. arctica, accession no. DQ839562) not yet included in the SILVA 16S rRNA database release SSURef_96 (Pruesse *et al.*, 2007) were retrieved from the NCBI nucleotide collection by BLAST (Altschul *et al.*, 1990). Chimeric

sequences were identified using Pintail (Ashelford *et al.*, 2005) and removed from the dataset. All sequences were imported into the ARB software (Ludwig, 2004) and automatically aligned. Alignments were manually refined. Phylogenetic analyses were performed in ARB using neighbor-joining, maximum-parsimony and maximum-likelihood methods with a 50% conservation filter for the family *Gallionellaceae*, which resulted in 1473 informative positions. Only near full-length sequences (>1320 nucleotides) were included in tree calculations. A consensus tree was constructed with branch lengths adopted from a maximum likelihood (AxML) tree. Bootstrap values were estimated using the neighbor-joining and maximum-likelihood algorithms with 1000 iterations. The candidate genus *Nitrotoga* was defined to include all sequences with similarities >98% to *Ca*. N. arctica, which also grouped stably in all treeing methods applied. The exact phylogenetic position of sequences with similarities between 95% and 98% to *Ca*. N. arctica could not be assigned unambiguously, and thus such sequences were excluded from our definition of the candidate genus *Nitrotoga*.

Probe and Primer Design. FISH probes and PCR primers were designed and evaluated using the ARB probe design function and the updated SILVA SSURef_96 database. Optimal hybridization conditions for newly developed probes were determined as described previously (Daims *et al.*, 1999). To determine optimal PCR conditions for the new *Nitrotoga*-targeting primer pair, temperature gradient PCR was performed. Probes used for FISH were 5'-labeled with the dyes 5(6)-Carboxyfluorescein-N-hydroxysuccinimide ester (FLUOS), Cy3, or Cy5. Labeled probes, unlabeled competitors, and PCR primers were ordered from Thermo Scientific (Thermo Fisher Scientific Inc., Waltham, MA, USA). All FISH probes used in this study, along with probe details, are listed in Table 2.

FISH, Microscopy, and Digital Image Analysis. Aliquots of PFA-fixed biomass were spotted onto microscope slides and FISH was performed as specified elsewhere (Daims *et al.*, 2005). Hybridized samples were embedded in Citifluor AF1 (Citifluor Ltd., Leicester, UK) prior to microscopic observation. Probe-conferred fluorescence was recorded on a LSM 510 confocal laser scanning microscope (CLSM, Carl Zeiss AG, Oberkochen, Germany) equipped with one argon ion (450 to 514 nm) and two helium neon lasers (543 and 633 nm) for the detection of FLUOS, Cy3, and Cy5, respectively. For determining probe dissociation profiles (Daims *et al.*, 1999), highly enriched *Ca*. N. arctica biomass was hybridized to the respective probe (Table 2) and 10 images per formamide concentration were recorded for subsequent image analysis. For quantifying relative biovolume fractions (Schmid *et al.*, 2000; Daims and Wagner, 2007), the activated sludge samples were hybridized to probe FGall221b and the EUB338 probe mix, and 40 images of each probe signal were taken at random spatial positions. For analyzing spatial distribution patterns of *Nitrotoga*-like bacteria within activated sludge flocs (Daims *et al.*, 2006), cells were stained by probes FGall221b,

Cluster6a192, and Bet42a, and 30 images of these probe signals were recorded at random positions. All digital image analysis tasks were carried out by using the software *daime* (Daims et al., 2006).

Table 1. Characteristics of the analyzed wastewater treatment plants

WWTP	Reactor type	Type of treated sewage	Detection of *Nitrotoga*[a,b]		*Nitrospira* sublineage[c]	Temp. [°C][b]	Influent [mg/l][b]	Effluent [mg/l][b]			Sampling date (2007)
			PCR	FISH			NH_4^+	NH_4^+	NO_2^-	NO_3^-	
Altmannstein	SBR	municipal	-	+	I + II	7	54.7	9.18	0.48	0.72	March 24
Ampfing	SBR	municipal, slaughter, and dairy waste	-	-	II	13	nd	0.1	0.04	3.24	March 26
Bad Zwischenahn	DIC-SBR	municipal and industrial	+	+	-	16	60	0.25	0.15	6.5	May 23
Bruchmühlen	DIC-SBR	municipal	+	+	I	15	36	0.53	0.09	4.53	May 22
Deuz	DIC-SBR	municipal	+	+	I + II	13	nd	0.33	0.09	3.46	May 21
Hettstedt	DIC-SBR	municipal and external activated sludge	+	-	I + II	15	56	12.35	0.24	3	May 24
Huntlosen	DIC-SBR	municipal	-	-	I + II	17	68	0.13	0.03	2.2	May 23
Ingolstadt	SBR	activated sludge drainage	+	-	I + II	27	856	0.3	<0.1	20.4	May 09
Kraftisried	single-stage activated sludge basin	animal rendering	+	-	I + II	7	397.5	35.3	6.2	17.4	January 29
Langenzenn	SBR	municipal	+	+	-	9	21.25	7.96	0.42	3.1	March 14
Lyss (ARA)	fixed bed reactor	municipal	-	-	I	12	20	1	0.1	18	January 29
Lyss (GZM)	membrane filtration plant	animal rendering	+	-	I + II	30	700	<1	<0.5	14	January 29
Oberding	fixed bed reactor	animal rendering	+	-	I + II	26	450	<1	<0.5	4	January 29
Plattling	two-stage activated sludge basin	animal rendering	nd	-	I + II	30	750	1	<0.5	3	January 29
Radeburg	DIC-SBR	municipal	nd	+	I	14	nd	0	0.05	3.3	May 24
Rosenheim	SBR	municipal	-	-	I + II	36	970	nd	nd	nd	May 30
Seefeld	SBR	municipal	+	-	I	nd	8.32	1.59	nd	1.73	March 28
Spenge	DIC-SBR	municipal	+	+	I	14	24	<0.2	0.05	1.38	May 22
Waldsassen	SBR	municipal and industrial	+	+	I	9	18.5	<0.1	nd	3.45	March 27
Weisstal	DIC-SBR	municipal	-	-	I + II	nd	nd	0	0.02	4.4	May 21

[a] +, *Nitrotoga* detectable; -, no *Nitrotoga* detected.
[b] nd, not determined.
[c] I, *Nitrospira* sublineage I detected by FISH; II, *Nitrospira* sublineage II detected by FISH; -, no *Nitrospira* detected.

Table 2. FISH probes used in this study

Probe full name[a]	Short name	Sequence 5'-3'	Binding position[b]	FA%[c]	Target group	Coverage[d]		Reference
						Probe	Comb.	
S-G-Ntoga-0122-a-A-19	Ntoga122	TCC GGG TAC GTT CCG ATA T	122 - 140	40	genus *Nitrotoga*	96.9%	96.9%	this study
cS-G-Ntoga-0122-a-A-19[e]	c1Ntoga122	TCW GGG TAC GTT CCG ATA T	122 - 140		-	-		this study
cS-G-Ntoga-0122-b-A-19[e]	c2Ntoga122	TCY GGG TAC GTT CCG ATG T	122 - 140		-	-		this study
S-F-Gall-0178-a-A-18	FGall178	TCC CCC TYA GGG CAT ATG	178 - 195	30	family *Gallionellaceae*	88.7%	88.7%	this study
cS-F-Gall-0178-a-A-18[e]	cFGall178	TCC CCC TYA GGG CKT ATG	178 - 195		-	-		this study
S-F-Gall-0221-a-A-18	FGall221a	TAT CGG CCA CTC CGA AAG	221 - 238	30	family *Gallionellaceae*	62.0%	92.3%	this study
cS-F-Gall-0221-a-A-18[e]	c1FGall221a	TAT CGG CCA CTC CTA AAG	221 - 238		-	-		this study
S-F-Gall-0221-b-A-18[f]	FGall221b	TAT CGG CCG CTC CGA AAA	221 - 238	30	genus *Nitrotoga*	97.4%		this study
					family *Gallionellaceae*	30.3%		
S-*-Gall-0438-a-A-18	FGall438	GTT TTC TTT CCG GCT GAA	438 - 455	25	genus *Nitrotoga*	92.3%	92.3%	this study
cS-*-Gall-0438-a-A-18[e]	c1FGall438	GAT TTC TTT CCG GCT GAA	438 - 455		*Zoogloea* spp.	-		this study
cS-*-Gall-0438-b-A-18[e]	c2FGall438	GTT TTC TTC CCG GCT GAA	438 - 455		*Thaurea*/*Dechloromonas* spp.	-		this study
cS-*-Gall-0438-c-A-18[e]	c3FGall438	GTT TTC TTT CCG TCT GAA	438 - 455		*Azoarcus* spp.	-		this study
S-G-Ntoga-1424-a-A-18	Ntoga1424	CTA GCT GCT TCT GGT AGA A	1424 - 1442	20	genus Nitrotoga	90.9%	90.9%	this study
cS-G-Ntoga-1424-a-A-18[e]	c1Ntoga1424	CTA ACT GCT TCT GGT AGA A	1424 - 1442		*Sterolibacterium*/*Dechlorosoma* spp.	-		this study
cS-G-Ntoga-1424-b-A-18[e]	c2Ntoga1424	CTA GCT GCT TCT GGT ACA A	1424 - 1442		*Acidothiobacillus* spp.	-		this study
S-*-Nsm6a-0192-a-A-20	Cluster6a192	CTT TCG ATC CCC TAC TTT CC	192 - 211	35	*Nitrosomonas oligotropha* lineage	nd	nd	(Adamczyk *et al.*, 2003)
cS-*-Nsm6a-0192-a-A-20[e]	cCluster6a192	CTT TCG ATC CCC TGC TTT CC	192 - 211		*Nitrosomonas eutropha* lineage	-		(Adamczyk *et al.*, 2003)
S-D-Bact-0338-a-A-18	EUB338	GCT GCC TCC CGT AGG AGT	338 - 355	0 - 50	most *Bacteria*	nd	nd	(Amann *et al.*, 1990)
S-*-BactP-0338-a-A-18	EUB338 II	GCA GCC ACC CGT AGG TGT	338 - 355	0 - 50	order *Planctomycetales*	nd		(Daims *et al.*, 1999)
S-*-BactV-0338-a-A-18	EUB338 III	GCT GCC ACC CGT AGG TGT	338 - 355	0 - 50	order *Verrucomicrobiales*	nd		(Daims *et al.*, 1999)
S-G-Ntspa-0662-a-A-18	Ntspa662	GGA ATT CCG CGC TCC TCT	662 - 679	35	genus *Nitrospira*	nd	nd	(Daims *et al.*, 2001)
cS-G-Ntspa-0662-a-A-18[e]	cNtspa662	GGA ATT CCG CTC TCC TCT	662 - 679		-	-		(Daims *et al.*, 2001)
S-G-Nbac-1035-a-A-18	Nit3	CCT GTG CTC CAT GCT CCG	1035 - 1052	40	genus *Nitrobacter*	nd		(Wagner *et al.*, 1996)
cS-G-Nbac-1035-a-A-18[e]	cNit3	CCT GTG CTC CAG GCT CCG	1035 - 1052		-	-		(Wagner *et al.*, 1996)
S-*-Ntspa-1151-a-A-20	Ntspa1151	TTC TCC TGG GCA GTC TCT CC	1151 - 1170	35	*Nitrospira* sublineage II	nd		(Maixner *et al.*, 2006)
S-*-Ntspa-1431-a-A-18	Ntspa1431	TTG GCT TGG GCG ACT TCA	1431 - 1448	35	*Nitrospira* sublineage I	nd		(Maixner *et al.*, 2006)
L-C-bProt-1027-a-A-17	Bet42a	GCC TTC CCA CTT CGT TT	1027 - 1043[g]	35	class *Betaproteobacteria*	nd		(Manz *et al.*, 1992)
L-C-gProt-1027-a-A-17[h]	Gam42a	GCC TTC CCA CAT CGT TT	1027 - 1043[g]	35	class *Gammaproteobacteria*	nd		(Manz *et al.*, 1992)

[a] Probe nomenclature as described by (Alm *et al.*, 1996).
[b] Probe binding position according to the *E. coli* 16S rRNA gene numbering.
[c] Percent formamide added to the hybridization buffer for optimal hybridization conditions.
[d] Group coverage is calculated from the number of sequences within respective probe target group as defined in Figure 1. Column "Probe" lists the coverage of single probes, column "Comb." that of probe combinations.
[e] Competitor probes were added to the hybridization unlabeled and in equimolar amounts to increase hybridization specificity.
[f] Probe can be used alone for detection of the genus *Nitrotoga* or in combination with probe FGall221a for detection of the family *Gallionellaceae*.
[g] Probe binds to the 23S rRNA.
[h] Probe Gam42a was used as unlabeled competitor for probe Bet42a.

Results and discussion

Phylogeny of the Candidate Genus *Nitrotoga.* *Ca*. N. arctica (Alawi *et al.*, 2007) represents the new candidate genus *Nitrotoga*, which comprises the only known nitrite oxidizers affiliated with the family *Gallionellaceae* (Skerman *et al.*, 1980) of the order *Gallionellales* (Weiss *et al.*, 2007), and the only known NOB in the *Betaproteobacteria*. Besides *Nitrotoga*, the *Gallionellaceae* comprise two genera of iron-oxidizing organisms, *Gallionella* (Henrici and Johnson, 1935; Garrity *et al.*, 2005) and *Sideroxydans* (Weiss *et al.*, 2007), some of which have the potential to couple Fe(II)-oxidation to nitrate reduction (Blöthe and Roden, 2009).

According to 16S rRNA phylogeny, the candidate genus *Nitrotoga* contains several organisms derived from a diverse range of habitats (Figure 1). *Nitrotoga*-like sequences were detected in soil (Alawi *et al.*, 2007; Sattin *et al.*, 2009), different wastewater treatment systems (Kong *et al.*, 2007; Alawi *et al.*, 2009; Maestre *et al.*, 2009), river biofilm, sediment, and water samples (Brümmer *et al.*,

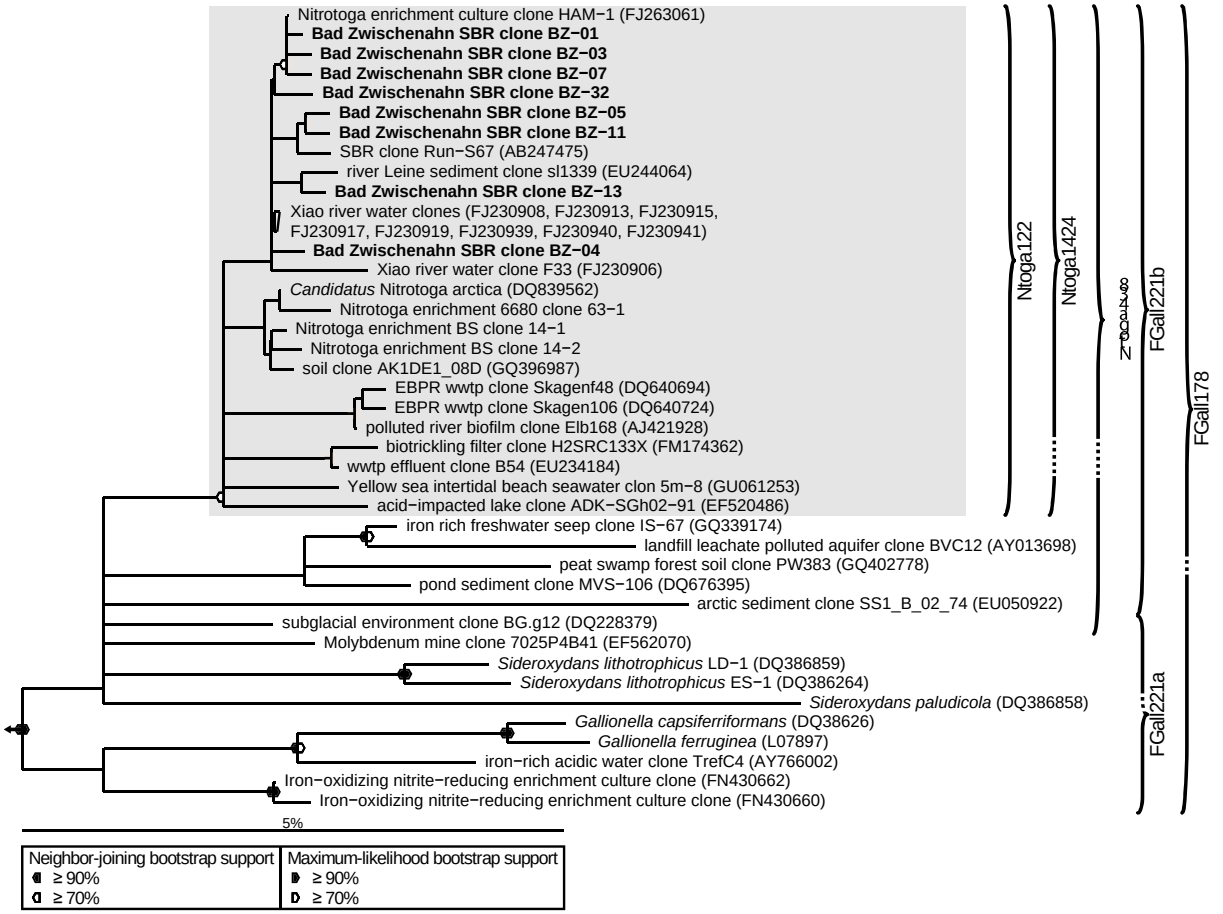


Figure 1. Phylogenetic analysis showing the affiliation of the candidate genus *Nitrotoga* to selected members of the family *Gallionellaceae*. Shown is a consensus tree based on neighbor-joining, maximum-likelihood, and maximum-parsimony calculations including all nearly full-length 16S rRNA gene sequences related to *Ca*. N. arctica. Branch lengths were estimated by the AxML algorithm. Neighbor-joining and maximum-likelihood bootstrap values ≥70% and ≥90% are indicated by empty and half-filled diamonds, respectively. All bootstrap values are based on 1000 iterations. Sequences obtained in this study are printed in bold. The candidate genus *Nitrotoga* is indicated by a gray box. Target groups and group coverage of newly designed FISH probes (Table 2) are designated by braces. Sequences not targeted by the respective probe due to mismatches in the probe-binding region are indicated by dashed lines interrupting the respective brace. The scale bar corresponds to 5% estimated sequence divergence.

2003), acid-impacted lake water (Percent *et al.*, 2008), cave-derived water and mat samples (Chen *et al.*, 2009), and in Yellow Sea intertidal beach seawater (Chun *et al.*, unpublished) (Figure 1). The exact phylogenetic affiliation of additional environmental 16S rRNA gene sequences, which have similarities of 95.8 to 97.9% to *Ca*. N. arctica, could not be resolved and these organisms might represent a yet unexplored diversity within the *Gallionellaceae*. The apparently wide distribution of *Nitrotoga*-like organisms indicates that these putative novel NOB contribute to nitrification in a great variety of habitat types. However, their *in situ* function in these systems as well as their competitive success compared to other NOB such as *Nitrospira* and *Nitrobacter* remains to be elucidated.

Probe Design and Evaluation. FISH probes for the specific *in situ* detection of *Nitrotoga*-like bacteria (Table 2) were designed according to the "multiple probe approach" (Ludwig *et al.*, 1998), ensuring the unambiguous identification of *Nitrotoga* cells by phylogenetically nested probes. The new oligonucleotide probe set hierarchically targets the candidate genus *Nitrotoga* and all members of the family *Gallionellaceae* at different phylogenetic levels (Figure 1). Probes Ntoga122 and Ntoga1424 were designed to target the candidate genus only, whereas probe Ntoga438 also includes some sequences of uncertain affiliation, which have a high sequence similarity (95.8 to 97.9%) to *Ca*. N. arctica. Probe FGall178 and the FGall221a+b probe mixture target all known members of the family *Gallionellaceae*.

When tested on highly enriched *Ca.* N. arctica biomass, all probes except Ntoga1424 yielded fluorescence signals of high intensities, irrespective of the dye used for probe labeling. Hybridization to probe Ntoga1424 resulted in weak signal intensities only. *In silico* evaluation indicated a high specificity of all probes, with only one (FGall178), two (FGall221a+b and Ntoga438), three (Ntoga122), and five (Ntoga1424) non-target organisms having the respective perfect-match probe binding site. Some non-target *Betaproteobacteria* contained only one base mismatch in the probe target regions and the probes should therefore be used together with the respective unlabeled competitor oligonucleotides as listed in Table 2. To ensure the unambiguous identification of *Nitrotoga* in the activated sludge samples, hierarchically nested probes were always applied simultaneously and were combined with probe Bet42a or the EUB probe mix detecting the *Betaproteobacteria* or the domain *Bacteria*, respectively.

Occurrence of *Nitrotoga*-like NOB in WWTPs. Utilizing the new genus *Nitrotoga*-specific primers, 18 WWTPs were screened by PCR for the presence of *Nitrotoga*-like bacteria. These plants (Table 1) included SBRs, DIC-SBRs, and different conventional activated sludge systems. PCR amplicons of the expected length were obtained for 12 samples. This indicated that the novel *Nitrotoga*-like NOB were present in most of the screened WWTPs, which represented different types of reactors and operational strategies (Table 1). The nine positively tested SBRs received municipal sewage mixed with different amounts of industrial wastewaters, whereas the other three systems

treated animal rendering waste. In one additional SBR sample *Nitrotoga* was detected by FISH but not by PCR (see below). To rule out failure of PCR due to, for example, the presence of inhibitory substances in the template DNA extract, the PCR was repeated using general bacterial 16S rRNA gene-targeted primers (data not shown). Since this resulted in amplicons of the expected size, failure to detect *Nitrotoga* in this sample may indicate incomplete coverage of the genus by the new PCR primers. Indeed, the reverse primer Ntoga1422R has one base mismatch to sequences derived from a biotrickling filter (accession no. FM174362) and from WWTP effluent (accession no. EU234184), which are phylogenetically affiliated with the candidate genus *Nitrotoga* (Figure 1).

From one of the PCR-positive SBR samples, the WWTP Bad Zwischenahn, a *Nitrotoga*-specific 16S rRNA clone library was established. Of the 32 randomly chosen clones, 27 contained an insert of the expected size, whereas the other five vectors were insert negative. These were subjected to RFLP analysis, which revealed four different restriction patterns. Sequencing of one to three clones per RFLP pattern resulted in eight nearly full-length (1322 bp) 16S rRNA gene sequences. Phylogenetic analyses revealed that all of these sequences grouped within the candidate genus *Nitrotoga*, confirming primer specificity. Although all clones shared high sequence similarities ≥99%, they formed several sub-clusters within the candidate genus (Figure 1). These sub-clusters might reflect an unexpected microdiversity of closely related and co-existing *Nitrotoga* or the presence of multiple *rrn* operons in *Nitrotoga* genomes. Alternatively, the observed sequence differences could be artifacts introduced by PCR or sequencing errors.

FISH revealed the presence of *Nitrotoga*-like organisms in six of the 18 WWTPs screened by PCR, as well as in two additional SBRs, one of which was screened but did not yield a PCR amplicon (Table 1). In these samples, the cells grew in dense clusters of heterogeneous shape located within the sludge flocs (Figure 2). The cells in these aggregates were irregularly shaped rods or cocci, resembling the morphologies described for *Ca.* N. arctica (Alawi *et al.*, 2007). All cells stained by the family *Gallionellaceae*-specific probes were also detected by the genus *Nitrotoga*-targeted oligonucleotides, indicating that (*i*) the genus-specific probes designed in this study detected all *Nitrotoga*-like organisms in the sludge samples, and (*ii*) no other known members of the *Gallionellaceae* were present. All cells stained by the new probes were also detected by probe Bet42a, confirming the affiliation of these organisms to the *Betaproteobacteria* and further supporting the specificity of the new probe set.

Although PCR detected *Nitrotoga*-like bacteria in some plants operated at elevated temperatures, all WWTPs that harbored *Nitrotoga* in sufficiently high quantities for detection by FISH (10^3-10^4 cells per ml) were operated at lower temperatures between 7 and 16°C (Table 1). This observation is fully consistent with the optimal temperature range for the growth of enriched *Nitrotoga* cultures (Alawi *et al.*, 2009). Hence, temperature seems to be one major factor affecting the growth of *Nitrotoga*-like NOB in full-scale WWTPs.

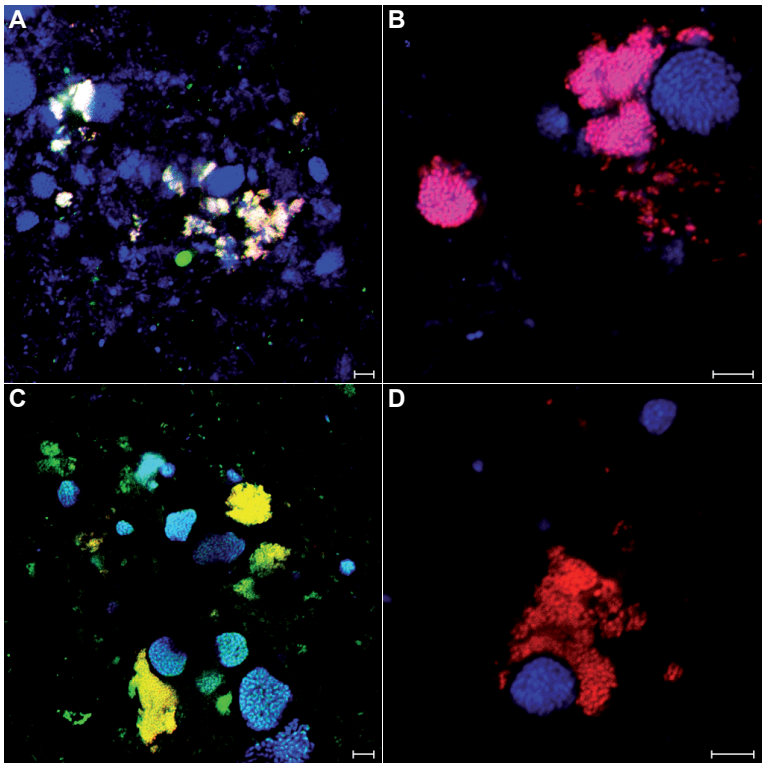


Figure 2. In situ detection of *Nitrotoga*-like bacteria in activated sludge by FISH. (A) Confocal micrograph of *Nitrotoga* cell aggregates stained by probes FGall221b (red), FGall178 (green), and EUB338mix (blue). *Nitrotoga* appear white due to overlay of all probe signals. (B) *Nitrotoga* detected by probes FGall221b (red) and EUB338mix (blue) at high magnification. (C and D) Simultaneous detection of *Nitrotoga* and AOB cell clusters. (C) Confocal micrograph of cell aggregates hybridized to probes FGall221 (red), Cluster6a192 (blue) and Bet42a (green). *Nitrotoga* appear in yellow, AOB in cyan. (D) Aggregates detected by probes FGall221 (red) and Cluster6a192 (blue) at high magnification. Note the close vicinity of *Nitrotoga* and the AOB, reflecting their mutualistic symbiotic relationship. The scale bar in all micrographs equals 5 µm.

Positive PCR results could not be confirmed by FISH for three of the SBRs and for three animal rendering plants (Table 1). In these samples, the cellular ribosome content of *Nitrotoga* could have been below the detection limit of the applied FISH protocol (1400 ribosomes per cell) (Hoshino *et al.*, 2008) due to a low metabolic activity. PCR might also have amplified DNA derived from dead cells or naked DNA in these samples. Finally, considering that PCR is approximately tenfold more sensitive than FISH (Amann *et al.*, 1995), the abundance of *Nitrotoga* might have been too low for FISH detection in these activated sludges. Such low cell densities of *Nitrotoga* would imply that these NOB were not functionally important and probably were allochthonous organisms unable to establish large stable populations in these WWTPs. Interestingly, four of these six plants treated animal rendering waste or activated sludge drainage water containing high loads of ammonia (Table 1). Thus, the NOB living in these plants might be confronted with locally and/or temporarily high concentrations of nitrite resulting from ammonia oxidation. For example, in another sequencing batch biofilm reactor also receiving high ammonia loads, peak nitrite concentrations were above 3.5 mM during the operational cycles (Daims *et al.*, 2001b). Previous research demonstrated that *Nitrotoga*-like NOB enriched from activated sludge thrived at much lower nitrite concentrations (0.3 mM) (Alawi *et al.*, 2009), and the tolerance limit of *Ca*. N. arctica was found to be at 1.2 mM of nitrite (Alawi *et al.*, 2007). Three of the

four plants, which received high ammonia loads and lacked high amounts of *Nitrotoga*, were operated at high temperatures, but the temperature in one plant (Kraftisried) was as low as 7°C (Table 1). Hence, both temperature and substrate concentration appear to be selective factors, with high nitrogen loads suppressing *Nitrotoga* despite a permissive temperature. Future research should confirm this hypothesis and elucidate whether additional parameters, such as sewage composition or inhibitory compounds, can select for or against *Nitrotoga*-like NOB in WWTPs.

Quantification of *Nitrotoga*-like Bacteria in Activated Sludge. In most of the analyzed WWTPs, *Nitrotoga*-like bacteria coexisted with NOB of the genus *Nitrospira* (Table 1). In the samples from the WWTPs Langenzenn and Bad Zwischenahn, however, no known NOB except *Nitrotoga* was detected. Whereas the sludge from Langenzenn contained only few *Nitrotoga* cell clusters, the activated sludge from Bad Zwischenahn harbored large amounts of *Nitrotoga*. Quantification by image analysis revealed that in this plant the *Nitrotoga* population constituted approximately 2% of the total bacterial biovolume. As this number resembles the abundance of *Nitrospira* in other full-scale WWTPs (Juretschko *et al.*, 1998; Egli *et al.*, 2003; Hall *et al.*, 2003), the abundance of *Nitrotoga* should be sufficient for maintaining nitrite oxidation in the WWTP Bad Zwischenahn.

Spatial Co-Localization of *Nitrotoga*-like NOB with AOB. NOB and AOB are partners in a mutualistic symbiosis (Stein and Arp, 1998) where AOB oxidize ammonia to nitrite, which then serves as substrate for NOB. Nitrite consumption by NOB also prevents the accumulation of nitrite that would otherwise be toxic to AOB. Therefore, the two functional groups are strongly interdependent and, accordingly, for *Nitrospira* a close spatial co-aggregation with AOB has been observed in nitrifying sludge and biofilm samples (Juretschko *et al.*, 1998; Okabe *et al.*, 1999; Schramm *et al.*, 1999; Maixner *et al.*, 2006). The spatial arrangement patterns of microbial populations in complex environmental samples can be analyzed by a combination of FISH, image analysis, and spatial statistics (Daims *et al.*, 2006). This quantitative approach confirmed the co-localization of *Nitrospira* and AOB in WWTPs (Daims *et al.*, 2006) but also revealed surprisingly complex co-aggregation patterns for different *Nitrospira* and AOB (Maixner *et al.*, 2006).

To test whether *Nitrotoga*-like bacteria also co-localize with AOB, their spatial distribution patterns in activated sludge were analyzed by the aforementioned method. In the sludge sample from Bad Zwischenahn, all known AOB were detected by probe Cluster6a192, indicating their affiliation with the *Nitrosomonas oligotropha* lineage. Visual observation already indicated that most *Nitrotoga*-like cell clusters occurred in close vicinity of AOB within the sludge flocs, sometimes even growing around the AOB cell aggregates (Figure 2). Quantitative analysis confirmed a pronounced co-aggregation of AOB and *Nitrotoga*-like bacteria at distances below 50 µm between the cell clusters (Figure 3). The degree of clustering was highest at distances between 2 and 40 µm with two local maxima, one at 12 and the more pronounced one at 26 µm. These two distinct peaks might reflect the

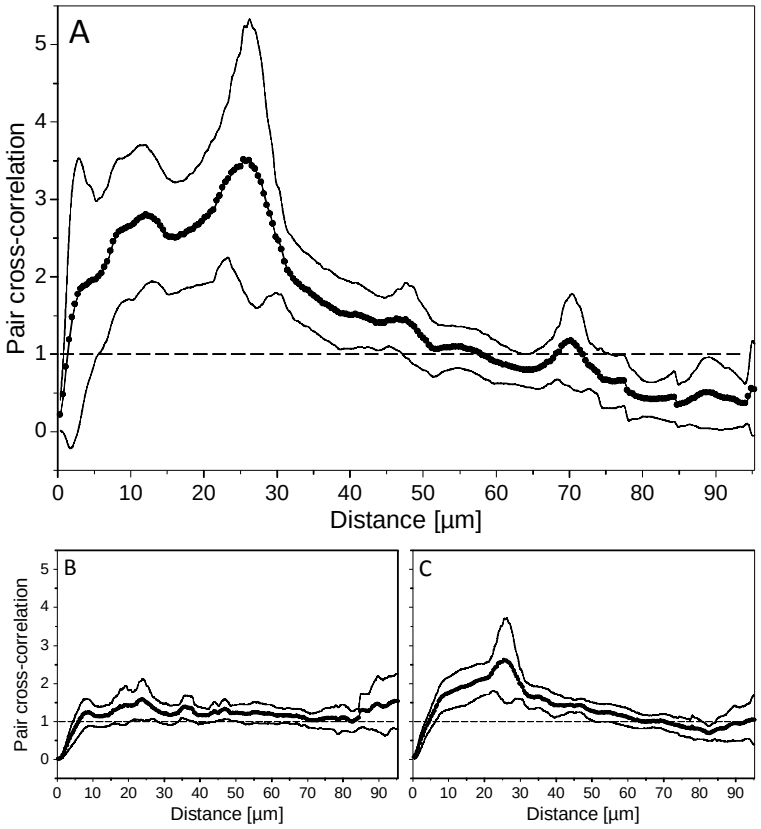


Figure 3. Statistical analyses of the spatial arrangement of *Nitrotoga*-like bacteria in the activated sludge sample from the Bad Zwischenahn SBR relative to (A) AOB, (B) all other *Betaproteobacteria* excluding AOB, and (C) all other *Betaproteobacteria* including AOB. Black circles depict the mean pair cross-correlation function, and the upper and lower lines delimit 95% confidence intervals. Values >1 indicate co-aggregation, values <1 repulsion, and values =1 (dashed horizontal line) random distribution at the corresponding distance (Daims *et al.*, 2006).

micro-diversity observed in the clone library and indicate the presence of at least two *Nitrotoga* subpopulations, which grow at different distances from the AOB. A similar spatial distribution pattern relative to AOB has been observed for NOB of the genus *Nitrospira* sublineage I and II, which have different nitrite concentration optima and thus occurred at different distances to the source of nitrite, the AOB (Maixner *et al.*, 2006).

Activated sludge is flocculated, and the observed co-aggregation of *Nitrotoga* and AOB could theoretically be due to floc structure and size instead of a true biological interaction of these organisms. To verify this, the spatial distribution pattern of *Nitrotoga* relative to all non-nitrifying *Betaproteobacteria* was also recorded. Prior to this analysis, the *Nitrotoga* and AOB probe signals (both groups are *Betaproteobacteria*) were digitally subtracted from the Bet42a probe signal. This resulted in images containing only betaproteobacterial cells which did not belong to any known nitrifying population and were not expected to have any specific functional link to *Nitrotoga*. Subsequently, the spatial arrangement of *Nitrotoga* relative to these non-nitrifying *Betaproteobacteria* was quantified. This test resulted in pair cross-correlation values close to one over the whole range of analyzed distances (Figure 3), indicating the absence of co-aggregation but random distribution of *Nitrotoga* and non-nitrifying *Betaproteobacteria* in the flocs. When only the *Nitrotoga* probe signal

was subtracted from the Bet42a signal, the presence of AOB within the remaining Bet42a-defined population led again to the detection of co-aggregation at distances between 6 and 50 µm (Figure 3).

At distances beyond 50 µm, the values of the pair cross-correlation function were around and below one. This indicates that *Nitrotoga* did not frequently occur without AOB in their close vicinity, but low pair cross-correlation values at large distances could also be due to the limited floc size of the sludge (Daims *et al.*, 2006).

Altogether, these data confirm that *Nitrotoga*-like NOB specifically co-aggregated with AOB within short distances between the cell clusters of these populations, strongly supporting our assumption that *Nitrotoga* functioned as nitrite oxidizers in the full-scale WWTP Bad Zwischenahn.

Conclusions

The development and application of specific FISH probes and PCR primers revealed a high frequency of bacteria related to *Ca.* N. arctica in a surprisingly large number of activated sludge samples. In one exemplary WWTP, the high abundance of these novel NOB and their co-aggregation with AOB, along with the absence of all other known nitrite oxidizers, strongly suggest that *Nitrotoga* were the key organisms responsible for nitrite oxidation. Environmental factors determining the abundance and ecological success of *Nitrotoga* in WWTPs appear to be both temperature and nitrogen load, but this and the possible influence of other parameters remains to be investigated in detail.

This study demonstrates that the recently discovered *Nitrotoga* are important nitrite oxidizers in full-scale engineered systems, where they often co-exist with *Nitrospira* but occasionally represent the only known NOB populations. Thus, *Nitrotoga* should be included in studies of nitrification in WWTPs in addition to *Nitrospira* and *Nitrobacter*, and the new *Nitrotoga*-specific PCR primers and FISH probes are the tools needed to detect, visualize and quantify these NOB *in situ*. Clearly, encompassing insight into the microbiology of nitrification will require further research on *Nitrotoga*, reaching from environmental distribution surveys to functional, genomic, and post-genomic analyses. Especially in the context of wastewater treatment, future work should address the yet unknown ecophysiological traits of *Nitrotoga*, their competition and co-existence with other NOB, their sensitivity or resilience to disturbances during reactor operation, and their implications for the still frequently encountered problems and failures of the nitrification process.

Acknowledgments

We thank Anneliese Müller and Christian Baranyi for excellent technical assistance. Niels Holm, Dieter Schreff, Uwe Temper, and the technical staff at the WWTPs are acknowledged for providing the activated sludge samples. This work was supported by the Vienna Science and Technology Fund (WWTF) grant LS09-040 to C.D., H.D., and S.L., the Austrian Research Fund (FWF) grants I44-B06 and S10002-B17 to H.D., M.W., and S.L., and the German Research Foundation (DFG) grant SP 667/3-1 to E.S.

References

Adamczyk, J., M. Hesselsoe, N. Iversen, M. Horn, A. Lehner, P. H. Nielsen, M. Schloter, P. Roslev and M. Wagner (2003). The isotope array, a new tool that employs substrate-mediated labeling of rRNA for determination of microbial community structure and function. *Appl Environ Microbiol* **69**(11): 6875-6887.

Alawi, M., A. Lipski, T. Sanders, E. M. Pfeiffer and E. Spieck (2007). Cultivation of a novel cold-adapted nitrite oxidizing betaproteobacterium from the Siberian Arctic. *ISME J* **1**(3): 256-264.

Alawi, M., S. Off, M. Kaya and E. Spieck (2009). Temperature influences the population structure of nitrite-oxidizing bacteria in activated sludge. *Environ Microbiol Reports* **1**(3): 184-190.

Alm, E. W., D. B. Oerther, N. Larsen, D. A. Stahl and L. Raskin (1996). The oligonucleotide probe database. *Appl Environ Microbiol* **62**(10): 3557-3559.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic local alignment search tool. *J Mol Biol* **215**(3): 403-410.

Amann, R. I., B. J. Binder, R. J. Olson, S. W. Chisholm, R. Devereux and D. A. Stahl (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**(6): 1919-1925.

Amann, R. I., W. Ludwig and K. H. Schleifer (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**(1): 143-169.

Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones and A. J. Weightman (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**(12): 7724-7736.

Blöthe, M. and E. E. Roden (2009). Composition and Activity of an Autotrophic Fe(II)-Oxidizing, Nitrate-Reducing Enrichment Culture. *Appl Environ Microbiol* **75**(21): 6937-6940.

Brümmer, I. H., A. Felske and I. Wagner-Dobler (2003). Diversity and seasonal variability of *Betaproteobacteria* in biofilms of polluted rivers: analysis by temperature gradient gel electrophoresis and cloning. *Appl Environ Microbiol* **69**(8): 4463-4473.

Camargo, J. A. and Á. Alonso (2006). Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: A global assessment. *Environ Internat* **32**(6): 831-849.

Chen, Y., L. Wu, R. Boden, A. Hillebrand, D. Kumaresan, H. Moussard, M. Baciu, Y. Lu and J. Colin Murrell (2009). Life without light: Microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in Movile Cave. *ISME J* **3**(9): 1093-1104.

Conley, D. J., H. W. Paerl, R. W. Howarth, D. F. Boesch, S. P. Seitzinger, K. E. Havens, C. Lancelot and G. E. Likens (2009). ECOLOGY: Controlling Eutrophication: Nitrogen and Phosphorus. *Science* **323**(5917): 1014-1015.

Daims, H., A. Bruhl, R. Amann, K. H. Schleifer and M. Wagner (1999). The domain-specific probe EUB338 is insufficient for the detection of all *Bacteria*: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**(3): 434-444.

Daims, H., J. L. Nielsen, P. H. Nielsen, K. H. Schleifer and M. Wagner (2001a). *In situ* characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microbiol* **67**(11): 5273-5284.

Daims, H., U. Purkhold, L. Bjerrum, E. Arnold, P. A. Wilderer and M. Wagner (2001b). Nitrification in sequencing biofilm batch reactors: Lessons from molecular approaches. *Water Sci Technol* **43**(3): 9-18.

Daims, H., K. Stoecker and M. Wagner (2005). Fluorescence *in situ* hybridization for the detection of prokaryotes. *In* A. M. Osborn and C. J. Smith (ed.), Molecular Microbial Ecology. New York, Taylor & Francis Group**:** 213-239.

Daims, H., S. Lücker and M. Wagner (2006). *daime*, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* **8**(2): 200-213.

Daims, H. and M. Wagner (2007). Quantification of uncultured microorganisms by fluorescence microscopy and digital image analysis. *Appl Microbiol Biotechnol* **75**(2): 237-248.

Diaz, R. J. and R. Rosenberg (2008). Spreading Dead Zones and Consequences for Marine Ecosystems. *Science* **321**(5891): 926-929.

Egli, K., C. Langer, H. R. Siegrist, A. J. Zehnder, M. Wagner and J. R. van der Meer (2003). Community analysis of ammonia and nitrite oxidizers during start-up of nitritation reactors. *Appl Environ Microbiol* **69**(6): 3213-3222.

Foesel, B. U., A. Gieseke, C. Schwermer, P. Stief, L. Koch, E. Cytryn, J. R. de la Torre, J. van Rijn, D. Minz, H. L. Drake and A. Schramm (2008). *Nitrosomonas* Nm143-like ammonia oxidizers and *Nitrospira marina*-like nitrite oxidizers dominate the nitrifier community in a marine aquaculture biofilm. *FEMS Microbiol Ecol* **63**(2): 192-204.

Garrity, G. M., J. A. Bell and T. Lilburn (2005). Family III. *Gallionellaceae* Henrici and Johnson 1935b. *In* D. Y. Brenner, N. R. Krieg and J. T. Staley (ed.), Bergey's Manual of Systematic Bacteriology, Second Edition. New York, Springer. **Volume Two: The *Proteobacteria*:** 880-886.

Gieseke, A., J. L. Nielsen, R. Amann, P. H. Nielsen and D. de Beer (2005). *In situ* substrate conversion and assimilation by nitrifying bacteria in a model biofilm. *Environ Microbiol* **7**(9): 1392-1404.

Hall, S. J., J. Keller and L. L. Blackall (2003). Microbial quantification in activated sludge: the hits and misses. *Water Sci Technol* **48**(3): 121-126.

Henrici, A. T. and D. E. Johnson (1935). Studies of Freshwater Bacteria: II. Stalked Bacteria, a New Order of *Schizomycetes*. *J Bacteriol* **30**(1): 61-93.

Henze, M., P. Harremoes, J. l. C. Jansen and E. Arvin (1997). Wastwater Treatment - Biological and Chemical Processes. Berlin, Springer.

Holm, N. (2003). Process for the discontinuous purification and installation for carrying out this process. E. P. Office. Germany. **EP 0 834 474 B1:** 13.

Hoshino, T., L. S. Yilmaz, D. R. Noguera, H. Daims and M. Wagner (2008). Quantification of target molecules needed to detect microorganisms by fluorescence *in situ* hybridization (FISH) and catalyzed reporter deposition-FISH. *Appl Environ Microbiol* **74**(16): 5068-5077.

Irvine, R. L., L. H. Ketchum and T. Asano (1989). Sequencing batch reactors for biological wastewater treatment. *Crit Rev Environ Contr* **18**(4): 255 - 294.

Juretschko, S., G. Timmermann, M. Schmid, K. H. Schleifer, A. Pommerening-Roser, H. P. Koops and M. Wagner (1998). Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. *Appl Environ Microbiol* **64**(8): 3042-3051.

Kong, Y., Y. Xia, J. L. Nielsen and P. H. Nielsen (2007). Structure and function of the microbial community in a full-scale enhanced biological phosphorus removal plant. *Microbiology* **153**(12): 4061-4073.

Lebedeva, E. V., M. Alawi, F. Maixner, P. G. Jozsa, H. Daims and E. Spieck (2008). Physiological and phylogenetic characterization of a novel lithoautotrophic nitrite-oxidizing bacterium, "*Candidatus* Nitrospira bockiana". *Int J Syst Evol Microbiol* **58**(Pt 1): 242-250.

Lücker, S., M. Wagner, F. Maixner, E. Pelletier, H. Koch, B. Vacherie, T. Rattei, J. S. S. Damsté, E. Spieck, D. Le Paslier and H. Daims (2010). A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci USA* **107**(30): 13479-13484.

Ludwig, W., R. Amann, E. Martinez-Romero, W. Schönhuber, S. Bauer, A. Neef and K.-H. Schleifer (1998). rRNA based identification and detection systems for *Rhizobia* and other bacteria. *Plant and Soil* **204**(1): 1-19.

Ludwig, W. (2004). ARB: A software environment for sequence data. *Nucl Acids Res* **32**: 1363-1371.

Maestre, J. P., R. Rovira, X. Gamisans, K. A. Kinney, M. J. Kirisits, J. Lafuente and D. Gabriel (2009). Characterization of the bacterial community in a biotrickling filter treating high loads of H_2S by molecular biology tools. *Water Sci Technol* **59**(7): 1331-1337.

Maixner, F., D. R. Noguera, B. Anneser, K. Stoecker, G. Wegl, M. Wagner and H. Daims (2006). Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. *Environ Microbiol* **8**(8): 1487-1495.

Maixner, F., M. Wagner, S. Lücker, E. Pelletier, S. Schmitz-Esser, K. Hace, E. Spieck, R. Konrat, D. Le Paslier and H. Daims (2008). Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium "*Candidatus* Nitrospira defluvii". *Environ Microbiol* **10**(11): 3043-3056.

Manz, W., R. Amann, W. Ludwig, M. Wagner and K.-H. Schleifer (1992). Phylogenetic Oligodeoxynucleotide Probes for the Major Subclasses of *Proteobacteria*: Problems and Solutions. *Syst Appl Microbiol* **15**: 593-600.

Okabe, S., H. Satoh and Y. Watanabe (1999). *In situ* analysis of nitrifying biofilms as determined by *in situ* hybridization and the use of microelectrodes. *Appl Environ Microbiol* **65**(7): 3182-3191.

Percent, S. F., M. E. Frischer, P. A. Vescio, E. B. Duffy, V. Milano, M. McLellan, B. M. Stevens, C. W. Boylen and S. A. Nierzwicki-Bauer (2008). Bacterial community structure of acid-impacted lakes: What controls diversity? *Appl Environ Microbiol* **74**(6): 1856-1868.

Prosser, J. I. (1989). Autotrophic nitrification in bacteria. *Adv Microb Physiol* **30**: 125-181.

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies and F. O. Glockner (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl Acids Res* **35**(21): 7188-7196.

Sattin, S. R., C. C. Cleveland, E. Hood, S. C. Reed, A. J. King, S. K. Schmidt, M. S. Robeson, N. Ascarrunz and D. R. Nemergut (2009). Functional shifts in unvegetated, perhumid, recently-deglaciated soils do not correlate with shifts in soil bacterial community composition. *J Microbiol* **47**(6): 673-681.

Schmid, M., U. Twachtmann, M. Klein, M. Strous, S. Juretschko, M. Jetten, J. W. Metzger, K. H. Schleifer and M. Wagner (2000). Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Syst Appl Microbiol* **23**(1): 93-106.

Schott, J., B. M. Griffin and B. Schink (2010). Anaerobic phototrophic nitrite oxidation by *Thiocapsa* sp. strain KS1 and *Rhodopseudomonas* sp. strain LQ17. *Microbiology* **156**(Pt 8): 2428-2437.

Schramm, A., D. De Beer, M. Wagner and R. Amann (1998). Identification and activities *in situ* of *Nitrosospira* and *Nitrospira* spp. as dominant populations in a nitrifying fluidized bed reactor. *Appl Environ Microbiol* **64**(9): 3480-3485.

Schramm, A., D. de Beer, J. C. van den Heuvel, S. Ottengraf and R. Amann (1999). Microscale distribution of populations and activities of *Nitrosospira* and *Nitrospira* spp. along a macroscale gradient in a nitrifying bioreactor: Quantification by *in situ* hybridization and the use of microsensors. *Appl Environ Microbiol* **65**(8): 3690-3696.

Skerman, V. B. D., V. McGowan and P. H. A. Sneath (1980). Approved lists of bacterial names. *Int J Syst Bacteriol* **30**: 225-420.

Stein, L. Y. and D. J. Arp (1998). Loss of ammonia monooxygenase activity in *Nitrosomonas europaea* upon exposure to nitrite. *Appl Environ Microbiol* **64**(10): 4098-4102.

Wagner, M., G. Rath, H.-P. Koops, J. Flood and R. Amann (1996). *In situ* analysis of nitrifying bacteria in sewage treatment plants. *Water Sci Technol* **34**(1-2): 237-244.

Weiss, J., J. Rentz, T. Plaia, S. Neubauer, M. Merrill-Floyd, T. Lilburn, C. Bradburne, J. P. Megonigal and D. Emerson (2007). Characterization of Neutrophilic Fe(II)-Oxidizing Bacteria Isolated from the Rhizosphere of Wetland Plants and Description of *Ferritrophicum radicicola* gen. nov. sp. nov., and *Sideroxydans paludicola* sp. nov. *Geomicrobiol J* **24**: 559-570.

Chapter IV

Summary

Zusammenfassung

Summary

The anthropogenic release of large amounts of inorganic nitrogen has detrimental effects on the environment, leading to the eutrophication of inland and coastal water bodies. Moreover, particularly ammonia and nitrite are toxic for most organisms at high concentrations. Within the biogeochemical nitrogen cycle, these water-soluble nitrogen compounds are transformed via a range of intermediates and finally released as harmless dinitrogen gas. This process of nitrogen elimination can be divided in two main parts. During aerobic nitrification ammonium is oxidized to nitrate, which consecutively is reduced to dinitrogen by anaerobic denitrification. The two-step process of nitrification is performed exclusively by microorganisms. First, ammonia-oxidizing archaea and bacteria convert ammonia to nitrite. The end-product nitrite is released and serves as substrate for nitrite-oxidizing bacteria (NOB), the second functional group involved in nitrification. NOB are highly specialized organisms that grow chemolithoautotrophically with nitrite as energy and CO_2 as sole carbon source. Among the five phylogenetic groups known to contain NOB, the members of the alphaproteobacterial genus *Nitrobacter* still are the best studied representatives since they are least resilient to cultivation. Consequentially, a relatively large amount of information about the physiology and genomics of *Nitrobacter* is available. Adversely, the advance of molecular techniques revealed that members of the genus *Nitrospira* constitute the dominant NOB in most habitats. Despite their obvious ecological and biotechnological importance, however, knowledge about *Nitrospira* ecophysiology and genomics still is scarce, primarily due to the resistance of these fastidious organisms to grow in pure culture.

To overcome this limitation, the complete genome sequence of "*Candidatus* Nitrospira defluvii" was reconstructed using a metagenomic approach. The genome analysis presented in the first part of this thesis revealed fundamental differences of *Ca*. N. defluvii to the other known NOB in the key enzymes for nitrification, the composition of the respiratory chain, and the pathway for autotrophic carbon fixation. These surprising findings suggest that chemolithoautotrophic nitrite oxidation evolved independently and multiple times in these organisms. Adaptation of *Ca*. N. defluvii to limited substrate concentrations is mediated by a periplasmic, constitutively expressed nitrite oxidoreductase system. Further, pathways for transport, oxidation, and assimilation of simple organic carbon sources allow for a mixotrophic lifestyle. The presence of the reverse tricarboxylic acid cycle for CO_2 fixation and the absence of most classical oxygen defense mechanisms indicate a microaerophilic or anaerobic ancestor of the genus *Nitrospira*. Unexpectedly, whole-genome phylogenomic analysis discovered an evolutionary link of the *Nitrospira* nitrite-oxidizing system to anaerobic ammonium-oxidizing *Planctomycetes*, reflecting an unexpected evolutionary link of two key processes of the nitrogen cycle.

Numerous studies over the last decades have demonstrated *Nitrospira* to be the dominant nitrite oxidizer in wastewater treatment plants (WWTPs). Recently, "*Candidatus* Nitrotoga arctica" was described and close relatives have been enriched from activated sludge. However, little is known about the *in situ* distribution, abundance, and spatial localization of these novel NOB. Therefore, in the second part of this thesis new *Nitrotoga*-specific PCR primers and FISH probes were developed and a range of WWTPs was screened in order to explore the importance of these NOB for engineered systems. Surprisingly, *Ca*. N. arctica-like bacteria were detected in approximately half of the samples analyzed and even constituted the sole known NOB in some plants. For one of these, quantification revealed a high abundance of *Nitrotoga* resembling counts of *Nitrospira* in other wastewater treatment systems. Furthermore, a statistically significant spatial co-aggregation with ammonia-oxidizing bacteria of the *Nitrosomonas oligotropha* lineage lends additional support to the proposed function as novel, previously overlooked NOB in full-scale WWTPs. Both temperature and nitrogen load appear to be environmental factors determining the abundance and ecological success of *Nitrotoga*.

Zusammenfassung

Die anthropogene Freisetzung großer Mengen anorganischen Stickstoffes hat fatale Auswirkungen für die Umwelt, da sie Eutrophierung inländischer und küstennaher Gewässer verursacht. Des Weiteren sind besonders Ammoniak und Nitrit in hohen Konzentrationen giftig für fast alle Lebensformen. Im Rahmen des biogeochemischen Stickstoffkreislaufes werden diese wasserlöslichen Stickstoffverbindungen über eine Reihe von Intermediaten umgesetzt und zu guter Letzt als harmloser Distickstoff ausgestoßen. Dieser Prozess der Stickstoffeliminierung kann in zwei Reaktionsabläufe unterteilt werden. Zunächst wird in der aeroben Nitrifikation der Ammoniak zu Nitrit oxidiert, welches im Folgenden durch die anaerobe Denitrifikation zu Distickstoff reduziert wird. Die zwei an der Nitrifikation beteiligten Reaktionen werden dabei ausschließlich von Mikroorganismen katalysiert. Als erstes wandeln ammoniakoxidierende Archaeen und Bakterien den Ammoniak zu Nitrit um. Das Endprodukt Nitrit wird freigesetzt und dient der zweiten in die Nitrifikation involvierten funktionellen Gruppe, den nitritoxidierenden Bakterien (NOB) als Substrat. NOB stellen Spezialisten dar, die chemolithoautotroph mit Nitrit als Energie- und CO_2 als einziger Kohlenstoffquelle wachsen. Von den fünf phylogenetischen Gruppen, welche NOB enthalten, sind die Mitglieder des alphaproteobakteriellen Genus *Nitrobacter* mit Abstand die am besten erforschten Vertreter dieser funktionellen Gilde, da sie relativ gut in Reinkultur zu bringen sind. Daher ist für diese Organismen verhältnismäßig viel Information über ihre Physiologie und ihr genomisches Potential verfügbar. Dennoch wurde durch die Entwicklung und den Einsatz moderner molekularer Methoden deutlich, dass es Mitglieder des Genus *Nitrospira* sind, die in den meisten Habitaten die dominanten NOB darstellen. Doch trotz dieser eindeutigen ökologischen und biotechnologischen Schlüsselrolle ist unser Wissen über die Ökophysiologie sowie das genomische Potential von *Nitrospira* immer noch gering, vor allem da sich diese extrem anspruchsvollen Bakterien sehr resistent dagegen zeigen, in Reinkultur zu wachsen.

Um diese unzufriedenstellende Situation zu verbessern wurde von „*Candidatus* Nitrospira defluvii“ die vollständige Genomsequenz mittels eines metagenomischen Ansatzes rekonstruiert. Die im ersten Teil dieser Arbeit präsentierte Analyse dieses Genoms zeigte fundamentale Unterschiede zwischen *Ca*. N. defluvii und allen anderen bekannten NOB in den Genen der Nitritoxidation, Atmungskette und autotrophen Kohlenstofffixierung auf. Diese unerwartete Erkenntnis deutet auf eine voneinander unabhängige, mehrfache Entstehung der chemolithoautotrophen Nitritoxidation in diesen Mikroorganismen hin. Außerdem erlauben Transport-, Abbau- und Assimilierungssysteme für einfache organische Kohlenstoffquellen *Nitrospira* eine mixotrophe Lebensweise. Die Anwesenheit des reduktiven Tricarbonsäurezyklus für die CO_2-Fixierung sowie das Fehlen typischer Sauerstoffstress-Resistenzmechanismen deuten auf eine Abstammung von mikroaerophilen oder anaeroben Vorfahren hin. Phylogenomische Analysen deckten ferner eine Verwandtschaft des Nitritoxidationssystems von *Nitrospira* mit

denen anaerober ammoniumoxidierender *Planctomyceten* auf, welche auf eine unerwartete evolutive Verbindung dieser zwei Schlüsselprozesse des Stickstoffkreislaufs hinweist.

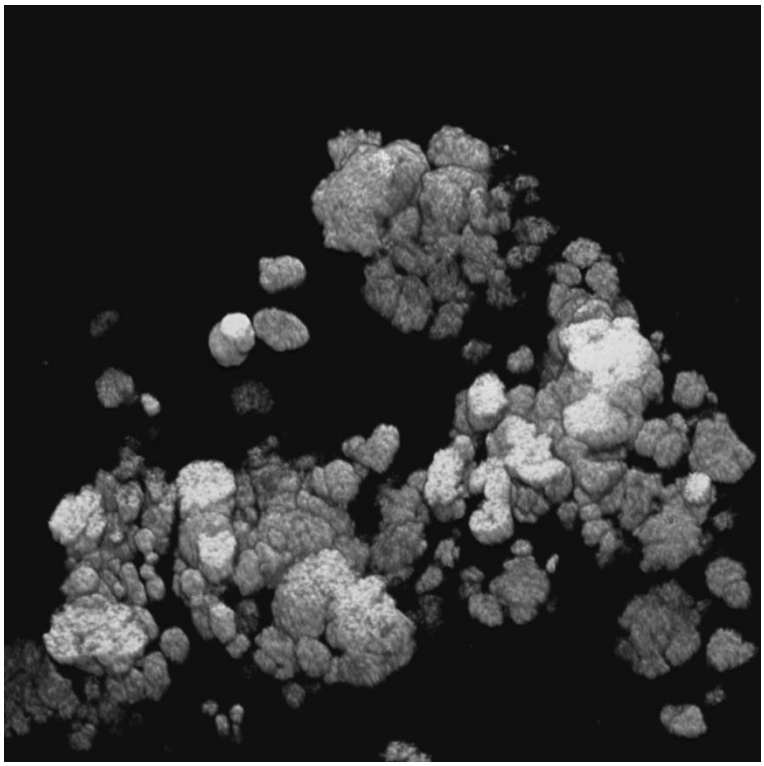
Zahlreiche Studien in den letzten Jahrzehnten konnten demonstrieren, dass *Nitrospira* der dominante Nitritoxidierer in Abwasserkläranlagen ist. Vor kurzen wurde nun „*Candidatus* Nitrotoga arctica“ beschrieben und es gelang, einen nah verwandten Organismus aus Abwasser anzureichern. Bislang ist jedoch wenig über diese neuen NOB hinsichtlich Verbreitung, Abundanz und Wachstumsweise sowie über ihre Bedeutung für technische Anwendungen bekannt. Aus diesem Grund wurden im zweiten Teil dieser Arbeit neue *Nitrotoga*-spezifische PCR Primer und FISH Sonden entwickelt und zur Untersuchung von Proben unterschiedlicher Kläranlagen eingesetzt. Das Ergebnis zeigte erstaunlicherweise, dass *Nitrotoga*-ähnliche Bakterien in ungefähr der Hälfte der analysierten Anlagen vorhanden sind und in einigen sogar die einzig bekannten NOB darstellen. Eine Quantifizierung von *Nitrotoga* in einer Probe deckte Mengenverhältnisse auf, welche denen von *Nitrospira* in anderen Abwasserreinigungssystemen entspricht. Ihre Rolle als neue, bislang in Kläranlagen übersehene NOB wird zusätzlich durch eine statistisch signifikante räumliche Koaggregation mit ammoniakoxidierenden Bakterien aus der *Nitrosomonas oligotropha*-Linie bekräftigt. Sowohl Temperatur als auch Stickstoffbelastung des Abwassers scheinen Faktoren zu sein, welche die Abundanz und den ökologischen Erfolg von *Nitrotoga* bestimmen.

Appendix

Supplementary Information

Acknowledgments

Curriculum Vitae



Front: 3D visualization of a *Nitrospira* community in an activated sludge floc. The cell clusters were stained by fluorescence *in situ* hybridization with a genus *Nitrospira*-specific probe and colored artificially by digital image processing.

A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria

Sebastian Lücker[a], Michael Wagner[a], Frank Maixner[a,1], Eric Pelletier[b,c,d], Hanna Koch[a], Benoit Vacherie[b], Thomas Rattei[e], Jaap S. Sinninghe Damsté[f], Eva Spieck[g], Denis Le Paslier[b,c,d], and Holger Daims[a,2]

[a]Department of Microbial Ecology, Ecology Centre, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria; [b]Commissariat à l'Energie Atomique, Genoscope, 91000 Evry, France; [c]Centre National de la recherche Scientifique, UMR8030, 91000 Evry, France; [d]Université d'Evry-Val-d'Essonne, 91000 Evry, France; [e]Department for Computational Systems Biology, Ecology Centre, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria; [f]Department of Marine Organic Biogeochemistry, Royal Netherlands Institute for Sea Research, P.O. Box 59, 1790 AB Den Burg, The Netherlands; [g]University of Hamburg, Biozentrum Klein Flottbek, Mikrobiologie und Biotechnologie, Ohnhorststr. 18, 22609 Hamburg, Germany

[1]Present address: Institute for Mummies and the Iceman, EURAC research, Viale Druso 1, 39100 Bolzano, Italy
[2]To whom correspondence should be addressed.

Supplementary Information

Published in:

Proceedings of the National Academy of Sciences of the United States of America (2010), Vol. 107 (30), p.13479 – 13484

www.pnas.org/cgi/doi/10.1073/pnas.1003860107

SI Results

[Fe-S] and Molybdenum Ligands in NxrA. Both nitrite oxidoreductase (NXR) α subunits of *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 (Trieber *et al.*, 1996; McDevitt *et al.*, 2002). In *Ca*. N. defluvii the second cystein residue of the consensus motif is replaced by aspartate, which can also function as [Fe-S] ligand as shown for a ferredoxin of *Pyrococcus furiosus* (Calzolai *et al.*, 1995). This aspartate residue also occurs in the phylogenetically closely related NXR-like proteins of *Ca*. K. stuttgartiensis, *Hydrogenobaculum*, and *Beggiatoa* (Fig. S2*A*). Based on these data, we propose a new consensus [Fe-S] binding motif (H/C-X_3-D/C-X_3-C-X_n-C) for the type II group in the dimethyl sulfoxide (DMSO) reductase family. A highly conserved aspartate residue, which functions as molybdenum ligand in nitrate reductase A (subunit NarG) of *E. coli* and most likely also in the other type II DMSO reductase-family enzymes (Jormakka *et al.*, 2004), is present in both NxrA copies of *Ca*. N. defluvii (Asp278 in NxrA1) (Fig. S2*B*).

Key Enzymes of the Reductive (rTCA) and Oxidative (oTCA) Tricarboxylic Acid Cycles. Key enzymes for CO_2 fixation *via* the 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 1 – 4 distinct subunits (Blamey and Adams, 1993; Ikeda *et al.*, 2006). 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 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* (Yun *et al.*, 2002; Ikeda *et al.*, 2006), and also in *Leptospirillum* (Goltsman *et al.*, 2009). 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 S3), 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 oTCA cycle, succinate dehydrogenase (SDH) (Table S3). 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 are known that differ in their subunit composition and distribution among bacteria, archaea and eukaryotes (Lancaster, 2002). 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 (Lemos *et al.*, 2002). 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 (Levican *et al.*, 2008). The type E SDH of *A. ambivalens* is reversible and catalyzes both the oxidation of succinate and the reduction of fumarate (Gomes *et al.*, 1999). If the type E-like enzyme of *Ca*. N. defluvii also operates in either direction, it could function in the rTCA and the oTCA cycle. 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, which is 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 (Levican *et al.*, 2008).

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 S3). 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 catalyses 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 (Hughes *et al.*, 1998; Tsugawa *et al.*, 2008). 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 organic carbon source (Daims *et al.*, 2001). 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 (Paiva *et al.*, 2004). 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. As 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 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 S3). However, the function of this putative hydrogenase remains unclear. It lacks the amino acid signatures of all known [NiFe]-hydrogenase groups (Vignais and Billoud, 2007) 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 (Vignais and Billoud, 2007).

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 (Rowe *et al.*, 1994). A gene coding for polyphosphate kinase was identified, which is consistent with observed polyphosphate granules in *N. moscoviensis* cells (Ehrich *et al.*, 1995). The genome encodes a ferredoxin-nitrite reductase for the reduction of NO_2^- to NH_4^+, indicating that NO_2^- also serves as 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 S3). Indeed, glycogen deposits have been observed by electron microscopy in *Nitrospira* cells (Watson *et al.*, 1986).

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 were identified that are conserved in gram-negative bacteria and function in type II protein secretion or type IV pilus assembly.

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 non-ribosomal peptide synthetases, a polyketide synthase, a class III aminotransferase, a type II thioesterase, and a putative cyclic peptide transporter (Table S3). Recently,

a gene cluster of similar composition was shown to be involved in siderophore production in the cyanobacterium *Anabaena* PCC 7120 (Jeanjean *et al.*, 2008). 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 S3). 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 (Carrondo, 2003). 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 S3). BFD is thought to be involved in iron-dependent gene regulation and in the release of iron from bacterioferritin (Quail *et al.*, 1996).

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 beta-lactamase-like CDS is consistent with the previously observed resistance of enriched *Ca*. N. defluvii to moderate concentrations of ampicillin (Spieck *et al.*, 2006), whereas two putative tetracycline efflux transporters are contrary to the observed tetracycline sensitivity of *Ca*. N. defluvii (Spieck *et al.*, 2006).

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 S3). Thioredoxin reductase, which is important for the regeneration of reduced thioredoxin as 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 molecular clock. We are aware of the limitations of this approach (Kuo and Ochman, 2009), but noticed 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. Based on an estimated rate of 16S rRNA

divergence of 1% per 50 million years (Myr) (Ochman and Wilson, 1987), 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% based on current datasets. Almost identical values (75.8 – 82.6%) were determined for *Nitrospira* and *Thermodesulfovibrio*. Taking these values we estimate that the three lineages shared a common ancestor about 870 – 1,210 Myr ago. Geochemical data indicate that a significant increase of the atmospheric and oceanic O_2 levels began in the late Proterozoic about 850 Myr ago, whereas Earth was only mildly oxygenated in the preceding 10^9 years (Holland, 2006). 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 Myr ago).

SI 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 (Strous *et al.*, 2006). 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 (a) transposon mutagenesis of two regions and (b) 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 (Vallenet *et al.*, 2006) was used for the prediction and annotation of CDS. CDS were predicted using the software AMIGene (Bocs *et al.*, 2003) and then submitted to automatic functional annotation (Vallenet *et al.*, 2006). Subsequently, the annotation of the entire genome was refined manually based on 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 (Serres and Riley, 2000) and TIGRFAM (Haft *et al.*, 2001) functional role catalogues. 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 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 based on the synteny information and visualization that is provided by the MaGe software. Metabolic pathways were reconstructed with help of the KEGG (Kanehisa and Goto, 2000) and MetaCyc (Caspi *et al.*, 2006) 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 (Reichard and Kaufmann, 2003) *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 Mo-co 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 (Ludwig, 2004). Multiple protein sequence alignments were created automatically by ClustalW2 (Larkin *et al.*, 2007) and MUSCLE (Edgar, 2004) 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: neighbour-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 (Guindon and Gascuel, 2003), and TREE-PUZZLE (Schmidt *et al.*, 2002) 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) (Pruesse *et al.*, 2007). High-quality sequences were longer than 1,399 nucleotides, had a Pintail (Ashelford *et al.*, 2005) 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 (Frickey and Lupas, 2004) was utilized. The reference database for PhyloGenie was generated from the non-redundant protein database NCBI nr (Sayers *et al.*, 2010), 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 (Altschul *et al.*, 1990) calculations in PhyloGenie, NCBI BLAST (version 2.2.19) was used. Protein phylogenies were calculated based on full or partial automatic alignments produced by the BLAMMER program included in PhyloGenie. All trees were post-processed 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) days in mineral medium (Spieck *et al.*, 2006) 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) days. 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 sec. Subsequently, the biomass was harvested by centrifugation (10,000 g, 20 min) and the pellet was re-suspended in 1x phosphate-buffered saline (PBS). As confirmed by fluorescence *in situ* hybridization (FISH) with a *Nitrospira*-specific probe (Daims *et al.*, 2001), this treatment resulted in a cell suspension containing mainly planktonic *Nitrospira* cells and only 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 (Daims *et al.*, 2005), and defined volumes were filtered onto polycarbonate filters (pore size 0.2 µm; diameter 47 mm; type GTTP; Millipore, USA). The filters were washed two times in 1x 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 (Glöckner *et al.*, 1996) with the *Nitrospira*-specific 16S rRNA-targeted probes Ntspa1431 (Maixner *et al.*, 2006), Ntspa662, and Ntspa712 (Daims *et al.*, 2001), which were 5'-labelled 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, Germany), 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 μm^2 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 correction 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 Citiflour (Citifluor, United Kingdom) with a 1:500 diluted Sybr Green II solution (Cambrex, USA) 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 (Daims *et al.*, 2006). 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, USA) according to the protocol recommended by the manufacturer and with the modifications described by Hatzenpichler *et al.* (Hatzenpichler *et al.*, 2008). After DNA digestion using DNase (Fermentas, Germany), reverse transcription of 3 µg total RNA from each treatment was carried out by using the RevertAID first strand cDNA synthesis kit (Fermentas, Germany) according to the manufacturer's instructions. The primers Ntspa1158R (Maixner *et al.*, 2006), 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 (Lane *et al.*, 1985; Maixner *et al.*, 2006) 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, Germany). 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 sec, annealing at 58°C for 40 sec, and elongation at 72°C for 60 sec. 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 (10,000g; 20 min) to harvest the biomass. Based on 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 5x lysis buffer [7 M urea, 2 M thiourea, 20 mg/ml amberlite, 4% (w/v) 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate, 40 mM TRIS, 2% (v/v) IPGbuffer, 0.2% (w/v) bromphenol blue, 1% (w/v) dithiothreitol, 10% (w/v) 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, Germany). All subsequent steps of the immunological detection of NxrB by Western blotting were performed as described for chlorite dismutase in (Maixner *et al.*, 2008) with the following modifications. The polyvinylidene fluoride (PVDF) membrane was incubated for 30 min with the NxrB-specific monoclonal antibody Hyb 153.3 (Aamand *et al.*, 1996), which had been diluted 1:1,000 in TBS buffer (20 mM TRIS, 150 mM NaCl, pH 7.5) containing 0.1% (v/v) 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 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, Germany) 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 reaction 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 sec, annealing at 56°C for 40 sec, and elongation at 68°C for 90 sec. 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, USA) 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, Germany), by digestion of amplicon and vector with the restriction endonucleases NdeI and EcoRI followed by ligation with T4 DNA Ligase (Invitrogen, USA) 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, Germany). 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, Germany). 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 (IPTG) to a final concentration of 1 mM. Cells were harvested after approximately 4 h by centrifugation (5,000 g; 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 days in mineral medium (Spieck *et al.*, 2006) lacking any energy source. NO_2^- test strips (Merckoquant, Merck,

Germany) 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^- (non-starved biomass). The anoxic medium had been prepared in accordance to basic principles of medium preparation for strict anaerobes as described by Widdel and Bak (Widdel and Bak, 1992). Following one additional centrifugation and washing step, the starved and non-starved 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 non-starved cultures to a final concentration of 3 mM, whereas no NO_2^- was added to the starved cultures. Based on 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 days 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.* (Lueders *et al.*, 2004) with the modification that samples were kept for 2 h on ice after the addition of polyethylene glycol. After DNA digestion using DNase (Fermentas, Germany), PCR was carried out with the primers 341F and 518R (Lane *et al.*, 1985; Edwards *et al.*, 1989), 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 ca. 250 ng total RNA from each treatment was carried out by using the RevertAID first strand cDNA synthesis kit (Fermentas, Germany) 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* (Maixner *et al.*, 2006). For cDNA amplification, reaction mixtures with the primer combinations 1158Fa [modified from (Maixner *et al.*, 2006); 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, Germany). 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 sec, annealing at 70-65°C for 30 sec (-0.5°C in each cycle), and elongation at 72°C for 40 sec. This was followed by 25 cycles of denaturation at 95°C for 30 sec, annealing at 65°C for 30 sec, and

elongation at 72°C for 40 sec. 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 (Spieck *et al.*, 2006) 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 weeks 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% (v/v) formaldehyde, and stored at 4°C. Biomass was harvested by centrifugation at 10,000 rpm (15,600 g), washed and 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 using a Thermofinnigan Gas Bench II coupled to a Delta^{PLUS} irmMS system with typical standard deviations of 0.3‰. Stable carbon isotope ratios were determined relative to laboratory standards calibrated on NBS-18 carbonate (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 ca. 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 (Schouten *et al.*, 1998). Hydrocarbons were measured in the apolar fraction obtained from the total extract. Bacteriohopanepolyols were transformed into hopanols using periodic acid and sodium borohydride (Rohmer *et al.*, 1984). The $\delta^{13}C$ values of individual lipids were corrected for added carbon from derivatization. Values are reported in the usual delta notation against Vienna Pee Dee Belemnite and represent the average of duplicate runs.

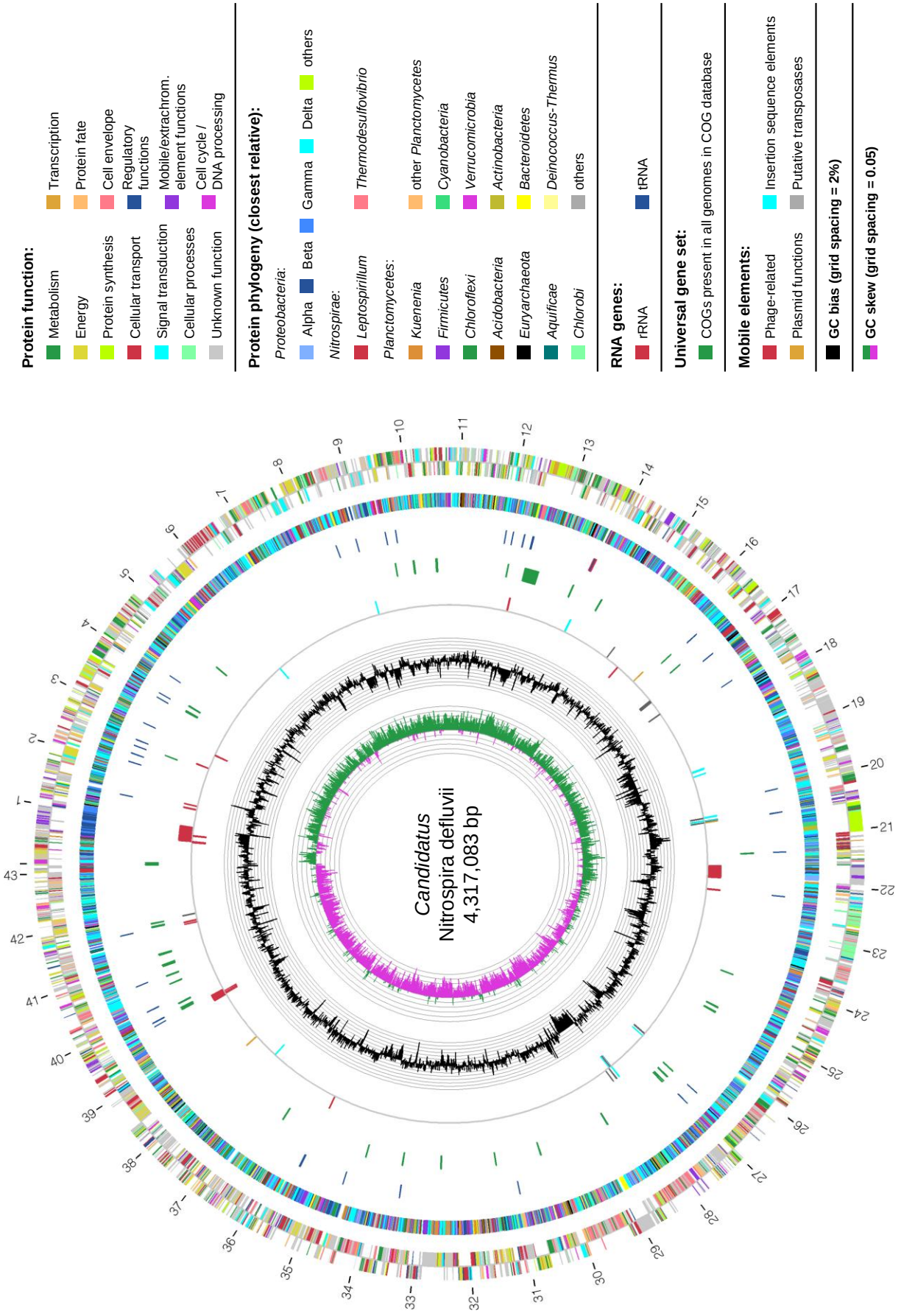

Protein function:
Metabolism
Transcription
Energy
Protein fate
Protein synthesis
Cell envelope
Cellular transport
Regulatory functions
Signal transduction
Mobile/extrachrom. element functions
Cellular processes
Cell cycle / DNA processing
Unknown function
Protein phylogeny (closest relative):
Proteobacteria:
Alpha
Beta
Gamma
Delta
others
Nitrospirae:
Leptospirillum
Thermodesulfovibrio
Planctomycetes:
Kuenenia
other Planctomycetes
Firmicutes
Cyanobacteria
Chloroflexi
Verrucomicrobia
Acidobacteria
Actinobacteria
Euryarchaeota
Bacteroidetes
Aquificae
Deinococcus-Thermus
Chlorobi
others
RNA genes:
rRNA
tRNA
Universal gene set:
COGs present in all genomes in COG database
Mobile elements:
Phage-related
Insertion sequence elements
Plasmid functions
Putative transposases
GC bias (grid spacing = 2%)
GC skew (grid spacing = 0.05)
Candidatus
Nitrospira defluvii
4,317,083 bp

Figure S1 (left page): 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 (COGs) (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 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 (Krzywinski *et al*., 2009)

A

```
I
Ca. Nitrospira defluvii      NxrA1 (NIDE3237)        60 RYDSSFT-----WCCSPNDTHG--CRVRAFVRNGVVMRVEQNYDHQTYEDLYGNRGTFAHNPRMCLKGFTFH 124
Ca. Nitrospira defluvii      NxrA2 (NIDE3255)        59 RYDSSFT-----WVCSPNDTHA--CRIRAFVRNGVVMRVEQNYDHQTYEDLYGNRGTFAHNPRMCLKGFTFH 123
Ca. K. stuttgartiensis       NxrA  (CAJ72445)        58 QYDRTFT-----YCCSPNDTHA--CRIRAFVRNNVMMRVEQNYDHQNYSDLYGNKATRNWNPRMCLKGYTFH 122
Hydrogenobaculum sp.               (YP_002121006)  55 KPDETTV-----ILCNPNDTHG--CYLNAYVKNGIITRLEPTYKYGDATDIYGIKASHRWEPRCCNKGLALV 119
Beggiatoa sp.                      (ZP_02000390)   48 HYDSSFT-----FLCAPNDTHN--CLLRGYVKNGVVTRIAPTYGYQKAKDLDGNQSTQRWDPRCCQKGLALV 112
II
Nitrobacter hamburgensis     NxrA  (YP_578638)        24 SHDNVFR-----STHGVNCTGG--CSWAIYVKDGIITWEMQQTDYPLLER-----SLPPYEPRGCQRGISAS  83
Nitrococcus mobilis          NxrA  (ZP_01125872)      24 QHDNIFR-----STHGVNCTGG--CSWAIYVKDGIITWEMQQTDYPLLGRGEGGRGIPPYEPRGCQRGISAS  88
III
Escherichia coli             NarG  (1Q16_A)          41 QHDKIVR-----STHGVNCTGS--CSWKIYVKNGLVTWETQQTDYPRTRP-----DLPNHEPRGCPRGASYS 100
Haloarcula marismortui       NarG  (YP_135852)       70 DWDSVAR-----STHSVNCTGS--CSWNVYVKDGQVWREEQAGDYPTFDES-----LPDPNPRGCQKGACYT 129
IV
Azoarcus sp. EB1             EbdA  (AAK76387)        73 KWDKVN---WGS--HLNICWPQGSCKFYVYVRNGIVWREEQAAQTPACNVD-----YVDYNPLGCQKGSAFN 134
Rhodovulum sulfidophilum     DdhA  (Q8GPG4)          57 TWDY-----VGKAAHCINCLGN--CAFDIYVKDGIVIREEQLAKYPQISPD-----IPDANPRGCQKGAIHS 116
Ideonella dechloratans       ClrA  (P60068)          60 TWDS-----VGVMTHSNGCVAG--CAWNVFVKNGIPMREEQISKYPQL-PG-----IPDMNPRGCQKGAVYC 118
Dechloromonas agitata        PcrA  (AA049008)        51 SWDKKTR-----GAHLINCTGA--CPHFVYTKDGVVIREEQSKDIPPMPN------IPELNPRGCNKGECAH 109
Thauera selenatis            SerA  (Q9S1H0)          64 TWDS-----TGFITHSNGCVAG--CAWRVFVKNGVPMREEQVSEYPQL-PG-----VPDMNPRGCQKGAVYC 122
Archaeoglobus fulgidus             (NP_069015)     37 GEIKTVR-----SVCSPNCTGA--CGFDALVYNGRIETLTQAADYPEPEYN----------PRGCLRGQSMM  91
Moorella thermoacetica             (YP_430751)     68 KEVKYVR-----TTCSPNCTLA--CGIRAMVVDGQIKALLPSNDYPEPEYG----------PRGCLRGLSFI 122
```

B

```
                                                                                                                    L.aa  Id.%
I
Ca. Nitrospira defluvii      NxrA1 (NIDE3237)       236 VGKMMNTRFNGGCLPLLDSWIRKVDAEKAQGGKYYSNYTWHGDQDPSHPFWNGTCNCDVD 295  1146   100
Ca. Nitrospira defluvii      NxrA2 (NIDE3255)       235 MGKHANTRFNNCVLPLLDSWIRKVNPDQAQGGRYWNNYTWHGDQDPSQPWWNGTCNCDVD 294  1147    87
Ca. K. stuttgartiensis       NxrA  (CAJ72445)       234 IGKYGMYRFNN-CLAIVDAHNRGVGPDQALGGRNWSNYTWHGDQAPGHPFSHGLQTSDVD 292  1148    58
Hydrogenobaculum sp.               (YP_002121006)  230 LRYVGQYRMSN-MMALLDSNIRKVEPDKALGGVGWDNYSEHTDLPPAHTLVTGQQTVDFD 289  1154    37
Beggiatoa sp.                      (ZP_02000390)   223 TRVFAQYREAN-AMALLDDKIRGKGTD-SLGGRGFDNYSWHTDLPPGHPMVTGQQTVDFD 281  1138    39
II
Nitrobacter hamburgensis     NxrA  (YP_578638)       181 LSFAAGTRFLS-----------------LMGGSLMSFYDWYADLPTSFPEIWGDQTDVCE 223  1214    16
Nitrococcus mobilis          NxrA  (ZP_01125872)     186 LSYAAGSRFLQ-----------------LFGGVNMSFYDWYADLPNSFPEIWGDQTDVCE 228  1218    17
III
Escherichia coli             NarG  (1Q16_A)         198 VSYASGARYLS-----------------LIGGTCLSFYDWYCDLPPASPQTWGEQTDVPE 240  1247    16
Haloarcula marismortui       NarG  (YP_135852)      193 VSFASGSRLVN-----------------LLGGVSHSFYDWYSDLPPGQPITWGTQTDNAE 235   952    21
IV
Azoarcus sp. EB1             EbdA  (AAK76387)       198 IAWGAGFRMTY-----------------LMDGVSPDINVDICRTYMGAFHTFGKMHMGYS 240   976    19
Rhodovulum sulfidophilum     DdhA  (Q8GPG4)         180 MRMAAPYRFAS-----------------LVGGVQLDIFTDVGDLNTGAHLAYGNALESFT 222   910    17
Ideonella dechloratans       ClrA  (P60068)         180 ITNTAYTRMTK-----------------LLGAISPDATSMTGDLYTGIQTVRVPASTVST 222   914    19
Dechloromonas agitata        PcrA  (AA049008)       173 VSFSAGHRFAH-----------------YIGAHTHTFFDWYSDHPTGQTQTCGVCGDSAE 215   927    21
Thauera selenatis            SerA  (Q9S1H0)         184 ITSAGYSRLAN-----------------LIGAIKPDVSSMTGDLYPGIQTVRMPARTVST 226   918    20
Archaeoglobus fulgidus             (NP_069015)     155 VHKGAMMRLAS-----------------MFGWSALHGYTMNGDLPAFWSQTFGVQTEEFE 197   896    22
Moorella thermoacetica             (YP_430751)     186 VQKGAWIALAT-----------------MAGWSLIHPYDQNGDLPMFWPQTFGVQTEELE 228   876    21
```

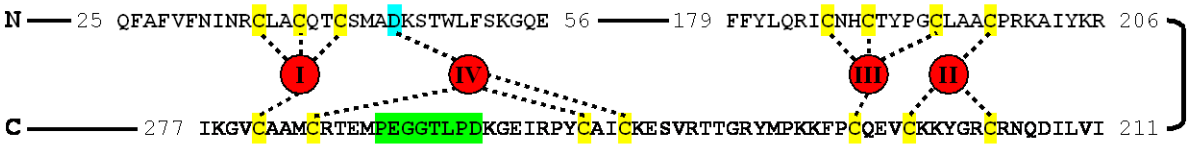
C

```
Ca. N. defluvii              NxrB1 (NIDE3236)         1 MPEVYNWQLGRKMLYPYEERHPKWQFAFVFNINRCLACQTCSMADKSTWLFSKGQEYMWWNNVETKPYGGYP  72
Ca. K. stuttgartiensis       NxrB  (CAJ72448)         1 MTLVHNWHLGRRMEYPYFESRPKHQFAAVFNINRCIACQTCTMACKSTWTFNKGQEFMWWNNVETKPYGGFP  72
Nitrobacter hamburgensis     NxrB  (YP_578636)        1 ...................MDIRAQVSMVFHLDKCIGCHTCSIACKNIWTDRKGTEYMYWNNVETKPGTGYP  53
Nitrococcus mobilis          NxrB  (ZP_01125873)      1 ...................MDIRAQVSMVFHLDKCIGCHTCSIACKNIWTDRQGTEYMWWNNVETKPGTGYP  53
Escherichia coli             NarH  (NP_415743)        1 ...................MKIRSQVGMVLNLDKCIGCHTCSVTCKNVWTSREGVEYAWFNNVETKPGQGFP  53

Ca. N. defluvii              NxrB1 (NIDE3236)       185 ICNHCTYPGCLAACPRKAIYKRPEDGIVLIDQNRCRGYKKCVEQCPFKKPMYRGTTRVSEKCIACYPRIEGK 256
Ca. K. stuttgartiensis       NxrB  (CAJ72448)       174 ICNHCTYPGCLAACPRKAIYKRKEDGIVLIDQKRCRGYRKCVEQCPYKKPMYRGLTRVSEKCIACYPRIEGR 245
Nitrobacter hamburgensis     NxrB  (YP_578636)      187 ICNHCLNPGCVAACPQGAIYKRGEDGVVLVSQERCRAWRMCVSGCPYKKTYFNWSTGKAEKCILCYPRLESG 258
Nitrococcus mobilis          NxrB  (ZP_01125873)    186 ICNHCLNPGCVAACPTGAIYKRGEDGIVLISQNRCRAWRMCVSGCPYKKTYFNWSTGKSEKCILCYPRLESG 257
Escherichia coli             NarH  (NP_415743)      183 LCEHCLNPACVATCPSGAIYKREEDGIVLIDQDKCRGWRMCITGCPYKKIYFNWKSGKSEKCIFCYPRIEAG 254

Ca. N. defluvii              NxrB1 (NIDE3236)       257 DPLTGGEPMETRCMAACVGKIRMQSLMRIGEDGLWAE---------------------------------- 293
Ca. K. stuttgartiensis       NxrB  (CAJ72448)       246 DSLTDGRPMETRCMSACVGQIRLQGFL------------------------------------------- 274
Nitrobacter hamburgensis     NxrB  (YP_578636)      259 --------QPPACFHSCVGRIRYIGLVLYDADAIEETAKAPQ-DQLVMAQRNIIKDPPDPDIIAAARANGIP 321
Nitrococcus mobilis          NxrB  (ZP_01125873)    258 --------QPPACFHSCVGRIRYIGMVLYDADGIEQAARSPD-DGLVDAQRALIKDPYDPEVIAGAKRNGIG 320
Escherichia coli             NarH  (NP_415743)      255 --------QPTVCSETCVGRIRYLGVLLYDADAIERAASTENEKDLYQRQLDVFLDPNDPKVIEQAIKDGIP 318
```

D



E

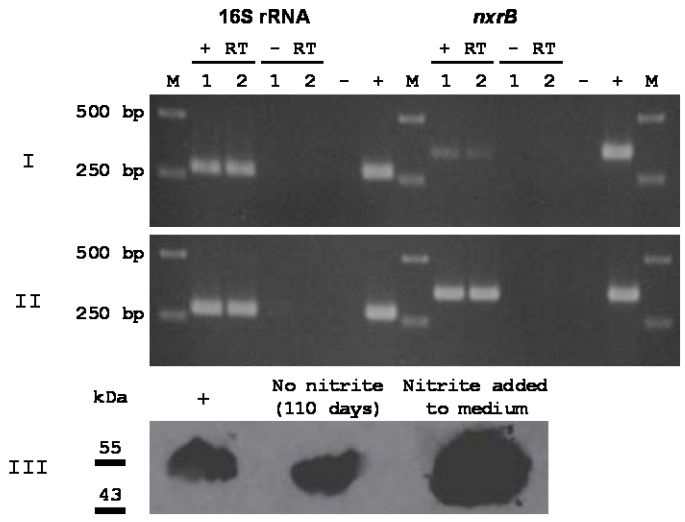


Figure S2 (left page): Metal-coordinating regions in the α- and β-subunits of nitrite oxidoreductase and expression of the β-subunit.

(*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 (Kisker *et al.*, 1998; Jormakka *et al.*, 2004). 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 reductase; Pcr, perchlorate reductase; Ser, selenate reductase). Accession numbers of the sequences are indicated in brackets. Five signature residues, which are conserved in nitrate reductases (Martinez-Espinosa *et al.*, 2007) and nitrite oxidoreductases, are highlighted green. Please note that, based on these residues, the yet uncharacterized enzymes of *A. fulgidus* and *M. thermoacetica* might use nitrate and/or nitrite as substrate. Asparagine (highlighted 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* (Jormakka *et al.*, 2004) and the homologous positions in the other sequences are highlighted 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 cyan. Aspartate was previously shown to function as [Fe-S] ligand in a ferredoxin of *Pyrococcus furiosus* (Calzolai *et al.*, 1995).

(*B*) Molybdenum ligand binding site. The conserved aspartate residue (highlighted blue) acts as molybdenum ligand in NarG of *E. coli* (Jormakka *et al.*, 2004). 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 dimethyl sulfoxide (DMSO) reductase family (Kisker *et al.*, 1998; Jormakka *et al.*, 2004).

(*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 (Blasco *et al.*, 2001) and the homologous positions in the NxrB sequences are highlighted yellow. An aspartate residue, which replaces one cysteine as a putative [Fe-S] ligand in both NxrB copies of *Ca*. N. defluvii, is highlighted cyan. Aspartate was previously shown to function as [Fe-S] ligand in a ferredoxin of *Pyrococcus furiosus* (Calzolai *et al.*, 1995). An insertion, which is found only in the NxrB of *Ca*. N. defluvii and *Ca*. K. stuttgartiensis, is highlighted green (see also panel *D*). Accession numbers of the sequences are indicated in brackets.

(*D*) Schematic representation 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* (Blasco *et al.*, 2001). Probable [Fe-S]-coordinating residues are highlighted yellow and cyan, and were assigned to the four [Fe-S] clusters based on the known coordination pattern of the homologous residues in NarH (Blasco *et al.*, 2001). 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 green (see also panel *C*). Numbers indicate the corresponding positions in the alignments shown in panel *C*. N=N-terminus, C=C-terminus of the protein.

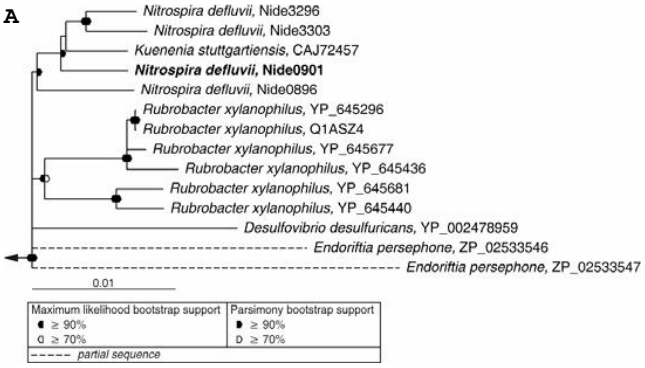
(*E*) Expression of nitrite oxidoreductase (β-subunit) by enriched *Ca*. N. defluvii.

(I) Detection of *Ca*. N. defluvii 16S rRNA and *nxrB* mRNA after 11 days 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 three days 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 days of starvation in NO_2^--free mineral medium and 8 days 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*.

A



B **Cytochrome *bd*-like oxidase**

```
                                                  h
Ca. Nitrospira defluvii          (Nide0901)      52 RDIPYIGSRNLVWIIAQLHLLLA-GFVLGV... 174 GS-GKKFHIF... 231 PVNIHRLIAN...
Ca. Nitrospira defluvii          (Nide0896)      32 VEFPYTGNRTAVWIVAQLHILFA-GFILGA... 154 GE-KKARHIA... 216 PLNLHRLVGN...
Ca. Nitrospira defluvii          (Nide3296)      82 YPRYNFESRVLLWFANQQHLYYG-SFVLAV... 208 GFL-KWIHLS... 266 -LNLHRILGN...
Ca. Nitrospira defluvii          (Nide3303)      62 PQVGSLDSRLLMWFIIQQHTYFG-GFVLAL... 189 GS--KWVHLS... 246 -LNLHRFLAD...
Ca. Kuenenia stuttgartiensis     (CAJ72457)      90 RSFFGLDSRVVVWIVSELHLMFA-AFVLGV... 212 TSTKKWCHIA... 271 PVNLHRLIAN...
Rubrobacter xylanophilus         (Q1ASZ4)         6 LDVPVIGKNVTIAVLVQTHILIA-TLILGA... 127 YR--KSLHVV... 179 PLNMHRFVGN...
Rubrobacter xylanophilus         (YP_645436)      6 IEFPVIGNDALIPVLVVPHILIA-AFVIGM... 127 YR--KRLHVV... 179 PLNMHRFVGN...
Rubrobacter xylanophilus         (YP_645677)      6 LDVPIIGKNVTIAVLVQTHILFA-AFIIGA... 127 YR--KSLHVV... 179 PLNMHRFVGN...
Rubrobacter xylanophilus         (YP_645440)     16 IEFPYLGSRGIVGVVMLIHIFFA-TLFVGY... 131 YRS-GRRNVA... 198 GMMLHRTFGN...
Rubrobacter xylanophilus         (YP_645681)     20 AFSSLGGARVITGLVMLIHIFFA-ELFVGF... 142 LN--KGRHIA... 205 EMSLHRTFAN...
Desulfovibrio desulfuricans      (YP_002478959)   6 LHIPGLGDGMTIALDAVLEVIISHGLAIGL... 130 KQ--PGKLAA... 183 PQVFLRLAAG...
Endoriftia Persephone            (ZP_02533546)   84 YPNIGVSSRAVVWILAQMHLFFG-ALVLAV... 206 GNA-KWVHLS... 267 PASLYRQ-YR...

                                                W   c  V                    cc
Ca. Nitrospira defluvii          (Nide0901)     361 LLLCLGVWMTPHSLVASL-------EEAQKMGGTHHPLLG
Ca. Nitrospira defluvii          (Nide0896)     339 API--VMVLLP--IVMTKV------LTDYPVPDATSLAF-
Ca. Nitrospira defluvii          (Nide3296)     394 LTCGYLVFITPHTIVMTP-------AELKAMGGQQHPVLG
Ca. Nitrospira defluvii          (Nide3303)     374 LTGCLFIWLTPHTILMTG-------TEVKAMGGAQHPVIG
Ca. Kuenenia stuttgartiensis     (CAJ72457)     399 IFMCFAVWLTPHNLPLSG-------EERAMIGEQYHPFSK
Rubrobacter xylanophilus         (Q1ASZ4)       304 AFALMALFSVLNVIPADASIVPQIGLVFGGGEGTQIPL--
Rubrobacter xylanophilus         (YP_645436)    304 ALGFMAVFGALNAIPADANLVPQIGLVFAGGERTQIPL--
Rubrobacter xylanophilus         (YP_645677)    304 AFALMALFSVLNVIPADASIVPQIGLVFGGGEGTQIPL--
Rubrobacter xylanophilus         (YP_645440)    326 LVA---AVSGLYAI----SPLAE--FPFLYMR-------
Rubrobacter xylanophilus         (YP_645681)    344 LVA---AVSGLYAI----SPLAE--FPFLYMR-------
Desulfovibrio desulfuricans      (YP_002478959) 303 LLCVFMVMEF-----------------------------
Endoriftia Persephone            (ZP_02533546)  end ---------------------------------------

                                                T              T                    h h
Ca. Nitrospira defluvii          (Nide0901)     394 VFGVMSAKMTVSNL--MVLVTFMSFIMYWR... 524 MTLMGYARSS-SRVHWH----IYGVMRDSS...
Ca. Nitrospira defluvii          (Nide0896)     367 ----LLP---------LLLAPVVLGRFIPL... 489 MGLMGAVRSL-LRKYFH----AYNLLPDFT...
Ca. Nitrospira defluvii          (Nide3296)     427 NYGVMSAKNGGINV--IITTTVLSFVWYMR... 556 MGLMGYIRSS-VRLFWH----VNEIMRDNS...
Ca. Nitrospira defluvii          (Nide3303)     407 NYGVMSSKNGAVNV--MICITALSYIFYRR... 536 MGLMGYIRSS-GRLAWH----VNELMPDTS...
Ca. Kuenenia stuttgartiensis     (CAJ72457)     432 YFGVMAAKNAVVNL--IILSIFFSFLIYRR... 629 MGLMGFIRSG-LRMDWH----VYGLMQDTS...
Rubrobacter xylanophilus         (Q1ASZ4)       341 -----------------------------... 398 MMTMGYTRETARRVDNEPGYLIYGCITLQQ...
Rubrobacter xylanophilus         (YP_645436)    341 -----------------------------... 398 MMTMGYARETSRRA-EGPGYLINGCITLDQ...
Rubrobacter xylanophilus         (YP_645677)    341 -----------------------------... 398 MVTMGYTRETARRVDNEPGYLIYGCITLQQ...
Rubrobacter xylanophilus         (YP_645440)    366 -----------------------------... 398 LG-MGWMKSN-SRAPYT----IYGQDEYRV...
Rubrobacter xylanophilus         (YP_645681)    348 -----------------------------... 416 LN-MGWMKSN-SRAPYT----VYNQPGYTV...
Desulfovibrio desulfuricans      (YP_002478959) end ---------------------------------... end ---------------------------------...
Endoriftia Persephone            (ZP_02533546)  end ---------------------------------... end ---------------------------------...
```

C **Cytochrome *c* oxidases**

```
                                                  h                                W   c  V
Bradyrhizobium japonicum  cbb3   (Q03073)       120 ----PWISFGRLRPLMTSAVIFAFGGNVLI... 266 WGGIQDAMFQWWYGHNAVGFFLTAGFLAIM
Escherichia coli          bo3    (P0ABI8)        95 ----PPHHYDQIFTAMGVIMIFFVAMPFV-... 271 N-MMMYINLIMAWGHPEV-YILILPVFGVF
Paracoccus denitrificans  aa3    (ABL70031)      79 ECTPNGHLWNVMITYMGVLMMFFVVIPALF... 263 D-PVLYQHILWFFGHPEV-YIIILPGFGII
Thermus thermophiles      ba3    (Q5SJ79)        60 ----FVQSYYQGLTLMGVLNAIVFTQLFAQ... 220 --PLVARTLFWWTGHPIV-YFWLLPAYAII
Thermus thermophiles      caa3   (P98005)        61 ----TGEQYNQILTLMGATMLFFFIIQAGL... 237 D-PVLFQQFFWFYSHPTV-YVMLLPYLGIL
Sulfolobus acidocaldarius soxM   (P39481)        53 -------DYYDAVTLMGIFMIFFVVMPLS-... 225 S-PVLWQQLFWFFGHPEV-YILILPAMGLV
Leptospirillum sp. Gr. II cbb3   (EAY56843)      65 ----PYLNFGHIRPVHVMTVAFMWISMAFG... 205 --GLNEALLTWWSGHNLFGLWITPMSMAVA

                                                  cc
Bradyrhizobium japonicum  cbb3   (Q03073)       316 HFWALIFLY-IWAGPHHLHYTA-LPDWTQT
Escherichia coli          bo3    (P0ABI8)       319 ATVCITVLS-FIVWLHHFFTMG-AGANVNA
Paracoccus denitrificans  aa3    (ABL70031)     311 AMAAIGILG-FVVWAHHMYTAG-MSLTQQA
Thermus thermophiles      ba3    (Q5SJ79)       268 AFLLFLLLS-TPVGFHHQFADPGIDPTWKM
Thermus thermophiles      caa3   (P98005)       285 AQMGIVVLG-TMVWAHHMFTVG-ESTLFQI
Sulfolobus acidocaldarius soxM   (P39481)       273 SSIAIAFLSALGVWMHHMFTAI-DNTLVQI
Leptospirillum sp. Gr. II cbb3   (EAY56843)     254 NFWSNFAFY-STPGAHHLMGAP-IPEWLKS

                                                T              T                    h h
Bradyrhizobium japonicum  cbb3   (Q03073)       344 LGMTFSIMLWMPSWGGMINGLMTLSGAWDK... 401 SIKVVNSLSHYTDWTIGWVHSGALGWVGFV...
Escherichia coli          bo3    (P0ABI8)       347 FFGITTMIIAIPTGVKIFNWLFTMYQGRIV... 402 AVPGADFVLHNSLFLIAWFHNVIIGGVVFG...
Paracoccus denitrificans  aa3    (ABL70031)     339 YFMLATMTIAVPTGIKVFSWIATMWGGSIE... 394 SQAPLDRVYHDTYYVVAHFHYVMSLGAVFG...
Thermus thermophiles      ba3    (Q5SJ79)       297 IHSVLTLFVAVPSLMTAFTVAASLEFAGRL... 367 ASFTLDYVVHNTAWVPGHFHLQVASLVTLT...
Thermus thermophiles      caa3   (P98005)       313 AFAFFTALIAVPTGVKLFNIIGTLWGGKLQ... 368 SMTPLDYQFHDSYFVVAHFHNVLMAGSGFG...
Sulfolobus acidocaldarius soxM   (P39481)       302 VSSATTMAIAIPSGVKVLNWTATLYGGEIR... 357 PLVPIDYALNGTYFVVGHFHYMVYAILYAL...
Leptospirillum sp. Gr. II cbb3   (EAY56843)     282 FASVSGVLILVPSMAFLANALLTMYGKWRL... 339 QTRAINWYIHGTHWVVAHAHLGILGFSTFV...
```

D

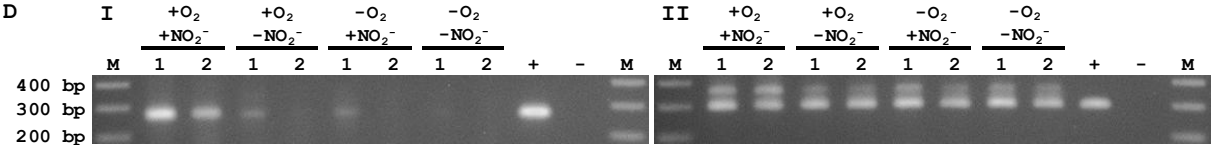
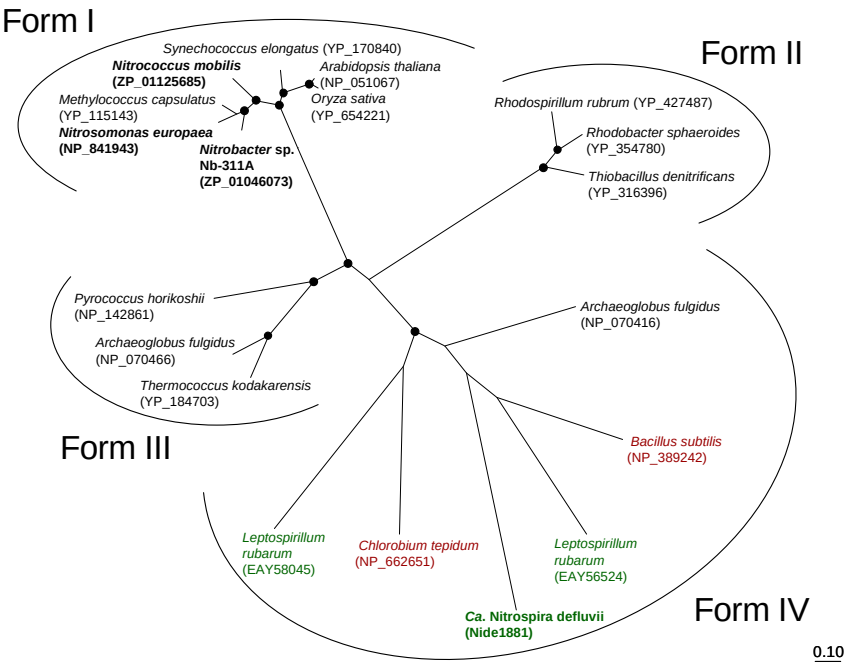


Figure S3 (left page): Analysis of the putative cytochrome *bd*-like oxidase family.
(*A*) Consensus tree showing the relationship of all known proteins belonging to the novel putative enzyme family. Tree topology is based on maximum likelihood and maximum parsimony methods, using a filter excluding 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 *E. 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 novel terminal cytochrome *c* oxidase of *Ca*. N. defluvii (Nide0901) is indicated in bold.
(*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 (Pereira *et al*., 2001) 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. Please note that only Nide0901 contains all residues conserved also in *bona fide* heme-copper cyt. *c* oxidases (see also panel *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 panel *B*.
(*D*) Expression of the putative novel 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 days 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.

A



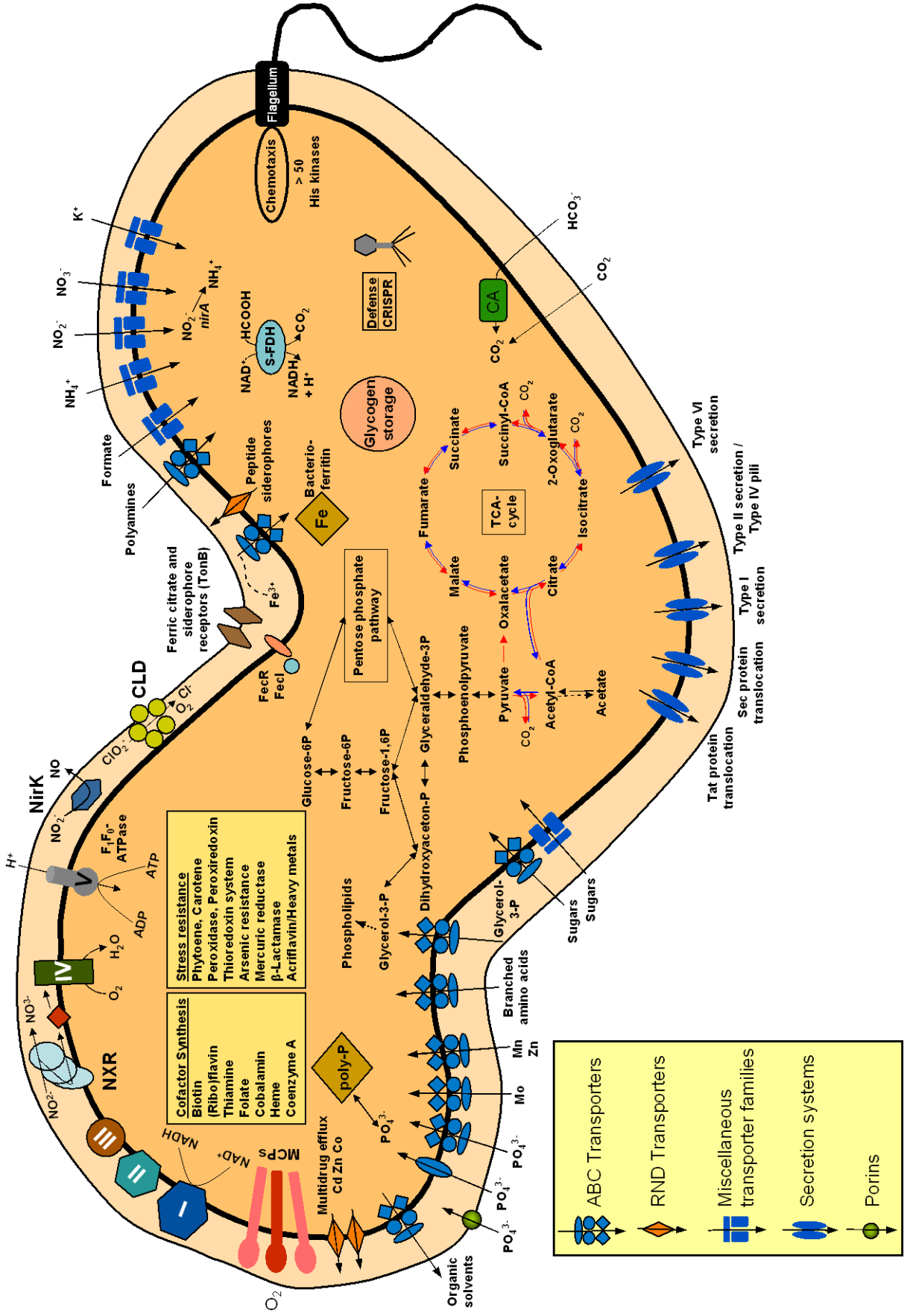
B

```
                          19  C     R        76    C          129C C                    GXDFXKXDE
S. elongatus      (I)  IAAESSTGTWT      TSIVGN-VFGF      IKPKLGLSAKNYGRAVYECLRGGLDFTKDDEN
R. rubrum        (II)  FAAESSTGTNV      TLTMGN-NQGM      IKPKLGLRPKPFAEACHAFWLGG-DFIKNDEP
M. jannaschii   (III)  IAGESSIGTWT      AGIAGN-IFGM      VKPKVGLKTEEHAKVAYEAWVGGVDLVKDDEN
C. tepidum       (IV)  FCSEQSTAQWK      TAVCGEGTYFT      VKPNIGLSPGEFAEIAYQSWLGGLDIAKDDEM
B. subtilis      (IV)  IATGLTVGSWT      TTVFGK----L      FKGVIGRDLSDIKEQLRQQALGGVDLIKDDEI
L. rubarum       (IV)  IAVGQTAGTWD      TMIYGK----Y      FKPSLGLSPSDLAHLLKETAEAGLDTIKDDEI
Ca. N. defluvii  (IV)  LCIDQTVEAAD      HMIIGM----A      LK-PLGYSPQQLADLAHEFALGEVDLIKDDQS

                          241    CR       279R       C     310       R R     337   RR
S. elongatus      (I)  VLLHIHRAMH      LHSGTV-VGKL      PGVLPVASGGI      VLQFGGGTLG
R. rubrum        (II)  NFLHYHRAGH      IHTGTMGFGKM      KACTPIISGGM      ILTAGGGAFG
M. jannaschii   (III)  FIIHAHRAMH      LHIGTV-VGKM      KPVFPVSSGGV      IIQAGGGVHG
C. tepidum       (IV)  VPLIGHFPFI      VIMPGF-GDRM      KPCLPVPGGSD      GFVPGRGVFG
B. subtilis      (IV)  VPIMAHFAVS      SLFPSP-YGSV      NQTFAVPSAGI      IINAGGGVHG
L. rubarum       (IV)  VPLFVHFAFA      VLYPAR-YGSL      GGVLPVPSAGI      ALNAGTGIMD
Ca. N. defluvii  (IV)  LPLLAHFDFL      THYPTY-GLNY      PPILPTAAGRM      VFILGSDLRR
```

Figure 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 printed bold. Names of members of the phylum *Nitrospirae* are labeled green. Enzymes of organisms, whose names are labeled 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 (Hanson and Tabita, 2001; Ashida *et al.*, 2003). Black dots indicate high (>90%) parsimony bootstrap (100 iterations) support of the respective nodes. The scale indicates 10% estimated sequence divergence. Sequence accession numbers are indicated in brackets.
(B) Partial amino acid sequence alignment of the large subunits of selected form I-III RubisCO and form IV RubisCO-like proteins. Known active site residues of RubisCO (1) are highlighted green, whereas substitutions at the homologous positions in form IV RubisCO-like proteins are highlighted red. C, residues involved in the catalytic mechanism; R, residues involved in binding of ribulose-1,5-bisphosphate; GXDFXKXDE, conserved RubisCO signature motif (Hanson and Tabita, 2001). The alignment is numbered according to the *Ca*. N. defluvii (Nide1881) sequence and is based upon an alignment published by Hanson and Tabita (2001).

Figure S5 (right page): Cell metabolic cartoon constructed from the annotation of the *Ca*. N. defluvii genome. Abbreviations not used in the text are 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.


Flagellum
Chemotaxis
> 50 His kinases
Defense CRISPR
Glycogen storage
S-FDH
HCOOH
NAD+
NADH + H+
CO2
CA
HCO3-
NO3-
NO2-
nirA
NH4+
K+
Formate
Polyamines
Peptide siderophores
Bacterio-ferritin
Fe
Fe3+
Ferric citrate and siderophore receptors (TonB)
FecR
FecI
CLD
Cl-
O2
ClO2-
NirK
NO
NO2-
H+
F1F0-ATPase
ATP
ADP
V
IV
H2O
O2
NO3-
NO2-
NXR
III
II
I
NADH
NAD+
MCPs
O2
Stress resistance
Phytoene, Carotene
Peroxidase, Peroxiredoxin
Thioredoxin system
Arsenic resistance
Mercuric reductase
β-Lactamase
Acriflavin/Heavy metals
Cofactor Synthesis
Biotin
(Ribo)flavin
Thiamine
Folate
Cobalamin
Heme
Coenzyme A
poly-P
PO43-
Multidrug efflux
Cd Zn Co
Organic solvents
Pentose phosphate pathway
Glucose-6P
Fructose-6P
Fructose-1,6P
Glyceraldehyde-3P
Dihydroxyaceton-P
Glycerol-3-P
Phospholipids
Glycerol-3-P
Phosphoenolpyruvate
Pyruvate
Acetyl-CoA
Acetate
CO2
TCA-cycle
Fumarate
Succinate
Succinyl-CoA
2-Oxoglutarate
Isocitrate
Malate
Oxalacetate
Citrate
Sugars
Sugars
Branched amino acids
Mn
Zn
Mo
Type VI secretion
Type II secretion / Type IV pili
Type I secretion
Sec protein translocation
Tat protein translocation
ABC Transporters
RND Transporters
Miscellaneous transporter families
Secretion systems
Porins

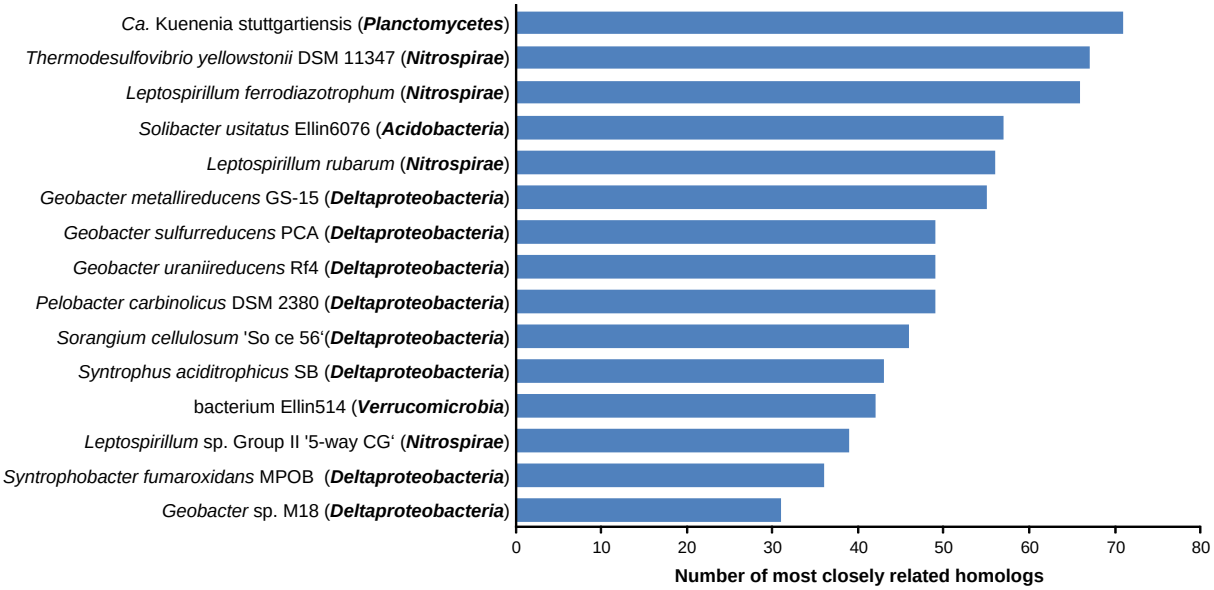


Figure S6: Summarized results of a phylogenetic analysis of each protein in the *Ca*. N. defluvii proteome. The nearest phylogenetic neighbour (closest homolog) in other sequenced genomes was determined for each protein of *Ca*. N. defluvii. The graph depicts the 15 organisms that contain the highest numbers of closest homologs to *Ca*. N. defluvii per individual genome. Names of bacterial phyla (or classes of the *Proteobacteria*) are in boldface. As only the closest homolog of each protein was considered, the presented analysis is non-redundant 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).

Table S1. Overview of key features of the *Ca*. Nitrospira defluvii genome.

Genome size	4,317,083 bp
Average G+C content	59.03%
Number of genomic objects [CDS, fragment CDS, (r,t)RNA]	4,321
Number of coding sequences (CDS)	4,272
CDS with predicted functions	2,148 (50.3%)
rRNA operons	1
tRNA genes	46
Coding density	89.45%
Repeated regions	2.29%
Transposon-related genes (including fragments)	46

Clusters of orthologous groups (COG) automated classification

Functional category		CDS	CDS (%)
D	Cell cycle control, cell division, chromosome partitioning	54	1.26
M	Cell wall/membrane/envelope biogenesis	321	7.51
N	Cell motility	131	3.07
O	Posttranslational modification, protein turnover, chaperones	194	4.54
T	Signal transduction mechanisms	300	7.02
U	Intracellular trafficking, secretion, and vesicular transport	129	3.02
V	Defense mechanisms	106	2.48
W	Extracellular structures	1	0.02
B	Chromatin structure and dynamics	1	0.02
J	Translation, ribosomal structure and biogenesis	187	4.38
K	Transcription	186	4.35
L	Replication, recombination and repair	223	5.22
C	Energy production and conversion	266	6.23
E	Amino acid transport and metabolism	301	7.05
F	Nucleotide transport and metabolism	72	1.69
G	Carbohydrate transport and metabolism	168	3.93
H	Coenzyme transport and metabolism	149	3.49
I	Lipid transport and metabolism	119	2.79
P	Inorganic ion transport and metabolism	237	5.55
Q	Secondary metabolites biosynthesis, transport and catabolism	121	2.83

Transporter families and their functions

Transport system[a]	Functions[b]	No. of genes[c]
ABC and ABC-II	Branched amino acids, Fe^{3+}, Mn, Mo, PO_4^{3-}, SO_4^{2-}/thiosulphate, polyamines, toluene, multidrug resistance, macrolides, cyclic peptides, lipoproteins, lipopolysaccharides, cobalamin, sugars, unknown	87
RND	Multidrug resistance, heavy metal efflux (Co/Zn/Cd), siderophore export, unknown	36
MFS	Multidrug resistance, tetracycline efflux, muropeptides, NO_3^-, sugars, unknown	16
TonB-dependent	Ferrichrome-like siderophores, ferric dicitrate, Fe^{3+}, unknown	24
OMP	PO_4^{3-}, anions, unknown	8
MscS / MscL	Ion flux	5
CPA-1 / CPA-2	Flux of $Na^+/H^+/K^+$	4
P-ATPase	Flux of divalent cations ($Cu^{2+}/Mg^{2+}/Ca^{2+}$)	4
Miscellaneous	$Na^+/Ca^{2+}/H^+/K^+$, NH_4^+ (Amt/MEP/Rh family), NO_2^-/formate, Ni/Co efflux, As efflux, Zn^{2+}, Mg^{2+}, PO_4^{3-}, amino acids, polysaccharides, lipopolysaccharides, C_4-dicarboxylate/Na^+/H^+ symporter, oligopeptides, multidrug resistance, quaternary ammonium compounds, Na^+/solute symporter, unknown	42

[a] ABC, ATP-binding cassette; RND, resistance-nodulation-cell division; MFS, major facilitator superfamily; OMP, outer membrane protein; MscS/MscL, small/large conductance mechanosensitive ion channel; CPA, cation:proton antiporter.

[b] Includes functions proposed for putative transporters based on their genomic localization and/or on limited sequence similarities to transporters characterized in other organisms.

[c] Number of genes encoding transporters, including putative transporters.

Table S2. Stable carbon isotope data and fractionation factors obtained from enrichment cultures of *Ca*. Nitrospira defluvii.

	$\delta^{13}C$ (‰ vs. VPDB)								ε^c (‰ vs. VPDB)		
	DIC	CO_2[a]	biomass	**I**[b]	**II**	**III**	**IV**	**V**	CO_2-biomass	CO_2-**I**	CO_2-isopr.
Batch culture 1	-4.5[d]	-13.1	-18.7	-12.9	-16.8	n.d.	n.d.	n.d.	5.7	-0.2	n.d.
Batch culture 2	-5.4[e]	-14.0	-16.4	-13.8	-15.9	-20.5	-22.3	-19.0	2.4	-0.2	6.7

[a] Calculated according to the equation given by (Mook *et al.*, 1974). The temperature in the cultures was 28°C.

[b] Key: **I** = $C_{16:1}$ ω5 fatty acid, **II** = $C_{16:0}$ fatty acid, **III** = squalene, **IV** = hop-22(29)-ene, **V** = C_{32} hopanol formed after Rohmer degradation.

[c] ε CO_2-biomass =(((1000+δ^{13}C CO_2)/(1000+δ^{13}C biomass))-1)*1000; ε biomass-lipid =(((1000+δ^{13}C CO_2)/(1000+δ^{13}C lipid))-1)*1000. For the ε between CO_2 and the isoprenoid lipids the average values of the three isoprenoid lipids (**III** – **V**) were taken.

[d] Measured at the time of harvesting the culture.

[e] Average of six determinations at different phases of the culture. $\delta^{13}C_{DIC}$ varied from -3.9 to -6.8 ‰.

VPDB = Vienna Pee Dee Belemnite.

DIC = dissolved inorganic carbon.

n.d. = not determined.

Table S3. *Ca*. Nitrospira defluvii proteins with predicted functions in key metabolic pathways

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
Nitrogen metabolism						
	Nitrite metabolism					
	Nitrite oxidoreductase, putative membrane subunit		Nide0904	Ideonella dechloratans (P60000: 239/275, 3-238, 28%)	Candidatus Kuenenia stuttgartiensis (Q1PZE8: 535/275, 281-525, 33%)	
nirA	Ferredoxin-nitrite reductase	1.7.7.1	Nide1367	Synechococcus elongatus (P39661: 512/531, 11-501, 45%)	Trichodesmium sp. WH 9601 (Q9RA39: 510/531, 11-501, 45%)	
nirK	copper-containing Nitrite reductase (NO forming)	1.7.2.1	Nide2534	Neisseria meningitidis serogroup A (Q9JTB8: 386/323, 77-339, 38%)	Herminiimonas arsenicoxydans (A4GA15: 309/323, 4-296, 57%)	
nxrB	Nitrite oxidoreductase, beta subunit	1.7.99.4	Nide3236	Rhodovulum sulfidophilum (Q8GPG3: 325/429, 3-292, 35%)	Candidatus Kuenenia stuttgartiensis (Q1PZD5: 410/429, 1-408, 63%)	
nxrA	Nitrite oxidoreductase, alpha subunit	1.7.99.4	Nide3237	Thauera selenatis (Q9S1H0: 918/1146, 239-649, 24%)	Candidatus Kuenenia stuttgartiensis (Q1PZD8: 1148/1146, 1-1146, 57%)	
nxrA	Nitrite oxidoreductase, alpha subunit	1.7.99.4	Nide3255	Ideonella dechloratans (P60068: 914/1147, 10-751, 23%)	Candidatus Kuenenia stuttgartiensis (Q1PZD8: 1148/1147, 1-1146, 58%)	
nxrB	Nitrite oxidoreductase, beta subunit	1.7.99.4	Nide3256	Rhodovulum sulfidophilum (Q8GPG3: 325/429, 3-292, 35%)	Candidatus Kuenenia stuttgartiensis (Q1PZD5: 410/429, 1-408, 63%)	
	Nitrite oxidoreductase, putative membrane subunit		Nide3271	Thauera selenatis (Q9S1G7: 239/277, 39-237, 28%)	Candidatus Kuenenia stuttgartiensis (Q1PZE8: 535/277, 274-500, 29%)	
	Nitrite oxidoreductase, putative membrane subunit		Nide3278	Ideonella dechloratans (P60000: 239/317, 5-153, 31%)	Candidatus Kuenenia stuttgartiensis (Q1PZD4: 322/317, 54-316, 33%)	
	putative Chaperone protein		Nide3279	Photobacterium profundum (Q6LS21: 216/325, 94-212, 22%)	Candidatus Kuenenia stuttgartiensis (Q1PZD6: 269/325, 21-243, 35%)	
napG	Ferredoxin-type protein NapG		Nide3280	Escherichia coli (P0AAL3: 231/201, 4-217, 35%)	Candidatus Kuenenia stuttgartiensis (Q1PWN6: 202/201, 11-200, 40%)	might be involved in electron transfer to or from NXR
	Nitrite oxidoreductase, putative membrane subunit		Nide3293	Thauera selenatis (Q9S1G7: 239/594, 120-237, 26%)	Candidatus Kuenenia stuttgartiensis (Q1PZE8: 535/594, 57-522, 31%)	contains 2 cytochrome c binding regions
nirK	copper-containing Nitrite reductase (NO forming)	1.7.2.1	Nide4252	Neisseria gonorrhoeae (Q02219: 392/321, 71-346, 38%)	Herminiimonas arsenicoxydans (A4GA15: 309/321, 3-296, 53%)	
	Ammonia metabolism					
	Nitroreductase	1.-.-.-	Nide0047	Mycobacterium smegmatis (P41401: 147/201, 3-1146, 38%)	Cyanothece sp. PCC 7425 (B8HKS7: 214/201, 13-205, 59%)	
gcvP	Glycine dehydrogenase, glycine cleavage system P protein	1.4.4.2	Nide0312	Thermosynechococcus elongatus (Q8DII3: 954/961, 13-948, 67%)	Gloeobacter violaceus (Q7NP12: 998/961, 40-991, 67%)	
gcvT	Aminomethyltransferase, glycine cleavage system T protein	2.1.2.10	Nide0320	Halothermothrix orenii (B8D1D7: 357/369, 1-356, 48%)	Anaeromyxobacter dehalogenans (Q2IQD4: 360/369, 4-359, 49%)	
gdhA	Glutamate dehydrogenase	1.4.1.3	Nide0440	Thermotoga maritima (P96110: 416/419, 7-416, 57%)	Planctomyces maris (A6CBA5: 552/419, 137-552, 58%)	
argF	Ornithine carbamoyltransferase	2.1.3.3	Nide0484	Bacillus stearothermophilus (Q9ZB62: 331/326, 9-311, 58%)	Geobacillus kaustophilus (Q5L1U9: 312/326, 9-312, 61%)	can act as Carbamate kinase (EC 2.7.2.2) in some organisms
	putative Aminomethyltransferase	2.1.2.10	Nide1059	Saccharopolyspora erythraea (A4FLG1: 367/363, 5-366, 32%)	Stigmatella aurantiaca (Q09DI0: 358/363, 10-353, 37%)	
ald	Alanine dehydrogenase	1.4.1.1	Nide1244	Oceanobacillus iheyensis (Q8CX61: 376/367, 1-363, 54%)	Anaeromyxobacter dehalogenans (B8J6K8: 370/367, 1-364, 61%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
nadE	NAD(+) synthase (glutamine-hydrolyzing)	6.3.5.1	Nide1358	Thermotoga maritima (Q9X0Y0: 576/589, 1-573, 48%)	Persephonella marina (C0QUN3: 574/589, 1-572, 53%)	can use both glutamine or ammonia as a nitrogen source
glnA	Glutamine synthetase	6.3.1.2	Nide1363	Anabaena sp. PCC 7120 (P00964: 474/469, 5-474, 62%)	bacterium Ellin514 (B9XCQ2: 470/469, 5-470, 68%)	
cynS	Cyanate hydratase	4.2.1.104	Nide1365	Thiobacillus denitrificans (Q3SHJ2: 147/146, 1-147, 40%)	Bordetella petrii (A9HZN1: 148/146, 2-148, 39%)	
tadA	tRNA-specific adenosine deaminase	3.5.4.-	Nide1745	Bacillus subtilis (P21335: 161/162, 4-157, 53%)	Staphylococcus carnosus (B9DKU7: 159/163, 4-150, 61%)	
lpd/ gcvL	Dihydrolipoyl dehydrogenase, E3 component of pyruvate and 2-oxoglutarate dehydrogenase complexes	1.8.1.4	Nide2727	Pseudomonas fluorescens (P14218: 478/473, 7-475, 42%)	Moorella thermoacetica (Q2RHM5: 459/473, 5-458, 49%)	also acts as L protein of glycine cleavage system
gcvH	Glycine cleavage system, H protein		Nide2728	Bacillus subtilis (O32174: 127/128, 3-121, 54%)	Bacillus pumilus (B4AG69: 127/128, 4-121, 53%)	
dcd	2'-deoxycytidine 5'-triphosphate deaminase	3.5.4.13	Nide3575	Sulfurihydrogenibium sp. YO3AOP1 (B2V937: 180/383, 20-151, 32%)	alpha proteobacterium BAL199 (A8TKV4: 382/383, 11-374, 49%)	
	putative Carbon-nitrogen hydrolase	3.5.-.-	Nide3616	Schizosaccharomyces pombe (O59829: 272/260, 41-244, 25%)	Pelobacter carbinolicus (Q3A713: 262/260, 9-254, 61%)	
pncA	Nicotine deamidase	3.5.1.19	Nide3776	Escherichia coli (P21369: 213/193, 1-200, 42%)	Methylococcus capsulatus (Q603P2: 198/193, 10-189, 62%)	
Energy metabolism						
	Complex I					
nuoN	NADH-quinone oxidoreductase, membrane subunit N	1.6.99.5	Nide0225	Synechocystis sp. PCC 6803 (P72714: 521/498, 16-512, 40%)	Thermosinus carboxydivorans Nor1 (A1HPU1: 471/498, 35-466, 47%)	
nuoM	NADH-quinone oxidoreductase, membrane subunit M	1.6.99.5	Nide0226	Paracoccus denitrificans (P29925: 513/551, 5-504, 30%)	Chloroflexus aurantiacus (A9WED0: 503/551, 13-499, 41%)	
nuoM	NADH-quinone oxidoreductase, membrane subunit M	1.6.99.5	Nide0227	Rickettsia conorii (Q92G96: 493/519, 5-490, 37%)	Candidatus Kuenenia stuttgartiensis (Q1PWH2: 525/519, 2-490, 43%)	
nuoL	NADH-quinone oxidoreductase, membrane subunit L	1.6.99.5	Nide0228	Paracoccus denitrificans (P29924: 703/667, 3-696, 40%)	Candidatus Kuenenia stuttgartiensis (Q1PWH4: 642/667, 4-639, 51%)	
nuoK	NADH-quinone oxidoreductase, membrane subunit K	1.6.99.5	Nide0229	Paulinella chromatophora (B1X495: 103/100, 3-103, 49%)	Candidatus Kuenenia stuttgartiensis (Q1PWH5: 100/100, 1-100, 53%)	
nuoJ	NADH-quinone oxidoreductase, membrane subunit J	1.6.99.5	Nide0230	Chlorokybus atmophyticus (Q19V56: 214/174, 10-185, 33%)	Leptospirillum sp. Group II '5-way CG' (B6AQU6: 169/174, 1-163, 37%)	
nuoI	NADH-quinone oxidoreductase, subunit I	1.6.99.5	Nide0231	Polaromonas sp. JS666 (Q127Y0: 165/202, 1-150, 39%)	Leptospirillum rubarum (A3ERI9: 186/202, 4-484, 45%)	protein ~50aa longer than most
nuoG	putative NADH-quinone oxidoreductase, subunit G	1.6.99.5	Nide0232	Bacillus subtilis (Q795Y4: 980/902, 2-941, 27%)	Carboxydothermus hydrogenoformans (Q3AE47: 893/902, 2-893, 35%)	
nuoD	NADH-quinone oxidoreductase, subunit D	1.6.99.5	Nide0233	Geobacter metallireducens (Q39QB0: 390/415, 4-390, 55%)	Leptospirillum rubarum (A3ERJ4: 423/415, 34-423, 66%)	
nuoC	NADH-quinone oxidoreductase, subunit C	1.6.99.5	Nide0234	Geobacter sulfurreducens (Q74GA6: 162/165, 10-151, 46%)	Leptospirillum rubarum (A3ERJ5: 184/165, 1-160, 59%)	
nuoB	NADH-quinone oxidoreductase, subunit B	1.6.99.5	Nide0235	Herpetosiphon aurantiacus (A9B4Z5: 179/182, 1-156, 60%)	Leptospirillum rubarum (A3ERJ6: 178/182, 5-160, 75%)	
nuoA	NADH-quinone oxidoreductase, membrane subunit A	1.6.99.5	Nide0236	Nitrosomonas europaea (Q82TU3: 122/127, 1-122, 52%)	Beggiatoa sp. PS (A7C2N8: 118/127, 2-118, 51%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
nuoN	NADH-quinone oxidoreductase, membrane subunit N	1.6.99.5	Nide0602	Rhizobium meliloti (P56911: 479/491, 7-476, 47%)	Leptospirillum rubarum (A3ERI4: 481/491, 1-465, 47%)	
nuoM	NADH-quinone oxidoreductase, membrane subunit M	1.6.99.5	Nide0603	Rhodobacter capsulatus (P50974: 512/522, 5-500, 51%)	delta proteobacterium MLMS-1 (Q1NQL7: 507/522, 10-489, 64%)	
nuoL	NADH-quinone oxidoreductase, membrane subunit L	1.6.99.5	Nide0604	Neisseria meningitidis serogroup A (Q9JX92: 674/632, 6-665, 45%)	Leptospirillum sp. Group II '5-way CG' (B6AQU8: 639/632, 3-363, 51%)	
nuoK	NADH-quinone oxidoreductase, membrane subunit K	1.6.99.5	Nide0605	Paracoccus denitrificans (P29923: 101/101, 2-100, 51%)	Leptospirillum rubarum (A3ERI7: 100/101, 2-99, 66%)	
nuoJ	NADH-quinone oxidoreductase, membrane subunit J	1.6.99.5	Nide0606	Paulinella chromatophora (B1X496: 205/206, 10-170, 40%)	Leptospirillum sp. Group II '5-way CG' (B6AQU6: 169/206, 1-163, 55%)	
nuoI	NADH-quinone oxidoreductase, subunit I	1.6.99.5	Nide0607	Polaromonas sp. JS666 (Q127Y0: 165/191, 6-149, 43%)	Leptospirillum rubarum (A3ERI9: 186/191, 7-147, 54%)	
nuoH	NADH-quinone oxidoreductase, membrane subunit H	1.6.99.5	Nide0608	Geobacter sulfurreducens (Q74GA1: 348/355, 17-342, 57%)	Leptospirillum rubarum (A3ERJ0: 337/355, 5-337, 63%)	
nuoG	NADH-quinone oxidoreductase, subunit G	1.6.99.5	Nide0609	Rhizobium meliloti (P56914: 853/889, 2-817, 35%)	Carboxydothermus hydrogenoformans (Q3AE47: 893/889, 4-893, 35%)	
nuoF	NADH-quinone oxidoreductase, subunit F	1.6.99.5	Nide0611	Rhizobium meliloti (P56913: 421/435, 2-413, 57%)	Sinorhizobium medicae (A6UFK2: 421/435, 2-413, 57%)	
nuoE	NADH-quinone oxidoreductase, subunit E	1.6.99.5	Nide0612	Rattus norvegicus (P19234: 248/178, 62-208, 52%)	Acidithiobacillus ferrooxidans (B5EN67: 163/178, 1-156, 53%)	
nuoCD	NADH-quinone oxidoreductase, subunits C and D	1.6.99.5	Nide0613	Aquifex aeolicus (O67335: 586/583, 51-586, 44%)	Leptospirillum sp. Group II '5-way CG' (B6AQU0: 423/583, 27-423, 70%)	
nuoB	NADH-quinone oxidoreductase, subunit B	1.6.99.5	Nide0614	Herpetosiphon aurantiacus (A9B4Z5: 179/159, 1-156, 65%)	Leptospirillum rubarum (A3ERJ6: 178/159, 22-160, 78%)	
nuoA	NADH-quinone oxidoreductase, membrane subunit A	1.6.99.5	Nide0615	Acidovorax avenae subsp. citrulli (A1TLL6: 119/123, 6-119, 52%)	Leptospirillum rubarum (A3ERJ7: 123/123, 1-123, 57%)	
nuoF	NADH dehydrogenase I, subunit F (fragment)	1.6.99.5	Nide1769	Synechocystis sp. PCC 6803 (Q55429: 681/555, 10-450, 33%)	Halothiobacillus neapolitanus c2 (C0H3D1: 559/555, 7-537, 39%)	
nuoF	NADH dehydrogenase I, subunit F	1.6.99.5	Nide3560	Rhizobium meliloti (P56913: 421/425, 18-420, 47%)	Leptospirillum rubarum (A3EW61: 453/425, 26-441, 47%)	
nuoL	NADH dehydrogenase I, subunit L	1.6.99.5	Nide4386	Zygnema circumcarinatum (Q32RH9: 702/564, 7-400, 35%)	Nitrobacter winogradskyi Nb-255 (Q3SR40: 566/564, 9-550, 38%)	
nuoM	NADH dehydrogenase I, subunit M	1.6.99.5	Nide4387	Synechococcus sp. JA-3-3Ab (Q2JWW3: 526/435, 200-487, 27%)	Nitrosococcus oceani (Q3JC26: 493/435, 229-483, 27%)	
	Other NADH oxidoreductases					
	putative Ferredoxin-NAD(+) reductase	1.18.1.3	Nide0018	Pseudomonas putida (P23101: 336/246, 106-333, 29%)	Lutiella nitroferrum 2002 (B9Z3R7: 239/246, 2-239, 52%)	
ndh	NADH dehydrogenase II	1.6.99.3	Nide0026	Bacillus subtilis (P80861: 392/439, 5-372, 30%)	Acidobacteria bacterium Ellin345 (Q1INM3: 444/439, 8-431, 51%)	
fre	FMN reductase	1.5.1.29	Nide0722	Pseudomonas sp. F600 (P19734: 353/235, 102-335, 35%)	Fervidobacterium nodosum (A7HKQ5: 369/235, 128-368, 36%)	
fprA	putative Ferredoxin-NADP(+) reductase	1.18.1.2	Nide2455	Mycobacterium leprae (O32886: 456/434, 27-451, 25%)	Acidobacterium capsulatum (C1F5X8: 478/434, 38-468, 40%)	
wrbA	NADH:Quinone oxidoreductase, type IV	1.6.5.2	Nide3258	Rhizobium meliloti (Q92Y27: 212/210, 4-207, 36%)	Syntrophus aciditrophicus (Q2LWH9: 160/210, 1-159, 41%)	
	putative Ferredoxin reductase	1.18.1.-	Nide4021	Pseudomonas sp. CF600 (P19734: 353/232, 111-275, 37%)	Thauera sp. MZ1T (C4KBD2: 351/232, 112-274, 41%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	Complex II					
sdhA	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	1.3.99.1	Nide0310	Rickettsia bellii (Q1RHB9: 596/551, 13-580, 49%)	Magnetococcus sp. MC-1 (A0L4R3: 568/551, 7-549, 55%)	
	Ferredoxin		Nide0838	Methanopyrus kandleri (Q8TXF7: 760/236, 376-438, 33%)	Leptospirillum sp. Group II '5-way CG' (B6AQJ9: 228/236, 11-225, 47%)	possible sdhB
sdhA or nadB	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit or L-aspartate oxidase		Nide0839	Methanocaldococcus jannaschii (Q60356: 539/533, 3-388, 42%)	Geobacter metallireducens (Q39RJ5: 531/533, 12-526, 54%)	could link L-aspartate degradation to TCA if not sdhA
sdhB	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	1.3.99.1	Nide2517	Haemophilus influenzae (P44893: 256/324, 11-236, 39%)	Magnetococcus sp. MC-1 (A0L4R2: 328/324, 38-313, 43%)	
sdhE	Succinate dehydrogenase, subunit C	1.3.99.1	Nide2527	Methanobacterium thermoautotrophicum (O27907: 302/300, 1-295, 34%)	Gloeobacter violaceus (Q7NCN3: 298/300, 5-296, 49%)	
	Complex III					
qcrA	Cytochrome *bc1* complex, iron-sulfur subunit	1.10.2.2	Nide0898	Synechococcus sp. PR-6 (P26292: 180/312, 108-165, 44%)	Leptospirillum sp. Group II '5-way CG' (B6AQZ9: 309/312, 21-308, 26%)	contains N-terminal PRC-barrel
qcrB	Cytochrome *bc1* complex, cytochrome b subunit	1.10.2.2	Nide0899	Chlorobium limicola f.sp. thiosulfatophilum (Q59297: 428/371, 109-417, 39%)	Leptospirillum rubarum (A3ER63: 458/371, 9-354, 46%)	
qcrC	putative Cytochrome *bc1* complex, cytochrome c subunit	1.10.2.2	Nide3886	Pavlova lutherii (P00107: 83/281, 2-81, 32%)	Ralstonia metallidurans (Q1LCC0: 272/281, 33-271, 49%)	
qcrB	Cytochrome *bc1* complex, fused cytochrome b/c subunit	1.10.2.2	Nide3889	Bacillus thermodenitrificans (Q45658: 224/441, 1-223, 48%)	Leptospirillum rubarum (A3ER63: 458/441, 8-450, 46%)	fused with cytochrome *c*
qcrA	putative Cytochrome *bc1* complex, iron-sulfur subunit	1.10.2.2	Nide3890	Chlorobium limicola f.sp. thiosulfatophilum (Q46136: 181/155, 58-178, 31%)	Solibacter usitatus (Q01T85: 173/155, 31-162, 41%)	
	Terminal oxidases					
	putative cytochrome *bd*-like oxidase		Nide0896	Escherichia coli O6 (P0ABK0: 522/567, 14-247, 21%)	Candidatus Kuenenia stuttgartiensis (Q1PZE5: 699/567, 97-387, 45%)	
	putative cyt. *bd*-like cytochrome *c* oxidase		Nide0901	Escherichia coli K12 (P0ABJ9: 522/626, 14-196, 23%)	Candidatus Kuenenia stuttgartiensis (Q1PZE5: 699/626, 78-490, 49%)	contains putative Cu-and heme binding sites
cydB	Cytochrome *bd* oxidase, subunit II	1.10.3.-	Nide2609	Escherichia coli O157:H7 (P0ABK4: 379/340, 10-359, 30%)	Solibacter usitatus (Q01RW6: 342/340, 3-342, 43%)	
cydA	Cytochrome *bd* oxidase, subunit I	1.10.3.-	Nide2610	Bacillus subtilis (P94364: 468/447, 8-442, 35%)	Solibacter usitatus (Q01RW7: 441/447, 3-440, 59%)	
	putative cytochrome *bd*-like oxidase		Nide3296	Haemophilus influenzae (P45021: 521/639, 19-156, 23%)	Candidatus Kuenenia stuttgartiensis (Q1PZE5: 699/639, 17-570, 38%)	
	putative cytochrome *bd*-like oxidase		Nide3303	Haemophilus influenzae (P45021: 521/621, 8-235, 24%)	Candidatus Kuenenia stuttgartiensis (Q1PZE5: 699/621, 70-696, 35%)	
	Complex V					
yidC	Inner-membrane protein insertion factor OxaA		Nide0360	Syntrophus aciditrophicus (Q2LSF9: 544/580, 1-537, 38%)	Geobacter sp. M21 (B3JYU8: 536/580, 1-529, 37%)	membrane insertion factor for F_1F_0-complex
atpH	putative ATP synthase F1, delta subunit	3.6.3.14	Nide0369	Ochrobactrum anthropi (A6WXW8: 186/179, 9-182, 32%)	Roseovarius sp. 217 (A3VZT3: 218/179, 39-211, 32%)	
atpA	ATP synthase F1, alpha subunit	3.6.3.14	Nide0370	Geobacter uraniireducens (A5G9D6: 502/505, 1-499, 71%)	Desulfuromonas acetoxidans (Q1JZG6: 502/505, 1-497, 71%)	
atpG	ATP synthase F1, gamma subunit	3.6.3.14	Nide0371	Carboxydothermus hydrogenoformans (Q3A945: 282/295, 1-281, 47%)	Thermodesulfovibrio yellowstonii (B5YI23: 295/295, 1-291, 50%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
atpD	ATP synthase F1, beta subunit	3.6.3.14	Nide0372	Anaeromyxobacter sp. Fw109-5 (A7HIX7: 481/480, 8-478, 76%)	Desulfuromonas acetoxidans (Q1JZG8: 468/480, 1-466, 76%)	
atpC	ATP synthase F1, epsilon subunit	3.6.3.14	Nide0373	Syntrophobacter fumaroxidans (A0LLF7: 134/140, 1-133, 51%)	Desulfovibrio salexigens (B7DDB2: 142/140, 1-138, 49%)	
atpF	ATP synthase F0, subunit B	3.6.3.14	Nide3732	Rubrobacter xylanophilus (Q1AVH5: 174/170, 12-164, 35%)	Rhodothermus marinus (C1ZTS9: 171/170, 16-169, 38%)	
atpE	ATP synthase F0, subunit C	3.6.3.14	Nide3733	Rhodospirillum rubrum (P15014: 75/76, 1-72, 51%)	Leptospirillum rubarum (A3EQ54: 76/76, 1-76, 82%)	
atpB	ATP synthase F0, subunit A	3.6.3.14	Nide3734	Rhodospirillum rubrum (P15012: 241/249, 3-241, 48%)	Leptospirillum rubarum (A3EQ55: 244/249, 4-243, 48%)	
atpI	putative ATP synthase F0, subunit I		Nide3735	Rhodospirillum rubrum (P15011: 123/94, 41-95, 46%)	Rhizobium loti (Q986D5: 135/94, 59-115, 48%)	guides assembly of ATPase enzyme complex
	Other proton pumps					
hppA	Pyrophosphate-energized proton pump	3.6.1.1	Nide3266	Xanthomonas campestris pv. campestris (Q8P5M6: 675/684, 10-675, 62%)	Thiobacillus denitrificans (Q3SIS2: 679/684, 7-679, 64%)	
	Cytochromes c					
pchC	4-cresol dehydrogenase (hydroxylating), cytochrome c subunit		Nide0020	Pseudomonas putida (P09787: 113/113, 1-108, 42%)	Nitrosospira multiformis (Q2Y8Z1: 108/113, 1-107, 55%)	
ccpA	Cytochrome c551 peroxidase		Nide0057	Methylobacillus flagellatus (Q50426: 33/357, 39-321, 41%)	Solibacter usitatus (Q01SD6: 336/357, 20-336, 51%)	
cyp	Cytochrome P450	1.14.14.1	Nide0186	Mycobacterium tuberculosis (P77900: 461/460, 42-456, 33%)	Nitrosospira multiformis (Q2Y711: 448/460, 12-448, 48%)	
	putative diheme cytochrome c		Nide0558	Rhizobium sp. NGR234 (P55493: 596/592, 30-595, 29%)	Sorangium cellulosum (A9F9W9: 546/592, 35-545, 37%)	
	putative monoheme cytochrome c		Nide0698	Pongo abelii (Q5RCR9: 314/205, 31-83, 36%)	Methylacidiphilum infernorum (B3DZU2: 211/205, 9-200, 48%)	
	putative diheme cytochrome c		Nide0818	Agrobacterium tumefaciens (Q8UJ37: 192/257, 76-178, 32%)	Candidatus Kuenenia stuttgartiensis (Q1PZE8: 535/257, 51-267, 42%)	
	putative monoheme cytochrome c		Nide0895	Pan troglodytes (A5A6M6: 637/168, 118-144, 63%)	Nitrosococcus oceani AFC27 (B6BZR3: 201/168, 63-149, 33%)	
	putative monoheme cytochrome c		Nide0897	Anabaena sp. PCC 7120 (Q8YW74: 633/268, 552-622, 40%)	Chthoniobacter flavus Ellin428 (B4CVC7: 919/268, 606-700, 36%)	
	putative monoheme cytochrome c		Nide0902	Methylobacillus flagellatus (Q50426: 333/206, 6-76, 34%)	Candidatus Kuenenia stuttgartiensis (Q1PZE6: 316/206, 6-166, 28%)	
	putative monoheme cytochrome c		Nide0903	Gluconobacter oxydans (Q47945: 478/241, 330-410, 30%)	Candidatus Kuenenia stuttgartiensis (Q1PZE7: 256/241, 14-256, 40%)	
	putative diheme cytochrome c		Nide0905	Bacillus pseudofirmus (Q04441: 342/308, 251-339, 30%)	Leptospira biflexa serovar Patoc (B0SP11: 234/308, 168-234, 46%)	
	putative monoheme cytochrome c		Nide1114	Roseobacter denitrificans (Q16CP0: 642/259, 180-370, 22%)	Geobacter sp. FRC-32 (B9M8A7: 245/259, 32-244, 40%)	
	putative monoheme cytochrome c		Nide1159	Hyphomonas neptunium (Q0C577: 701/200, 425-540, 27%)	Verrucomicrobiae bacterium DG1235 (B5JNZ9: 198/200, 37-197, 42%)	
	putative monoheme cytochrome c		Nide1251	Brassica oleracea (P62773: 111/130, 8-68, 34%)	Fulvimarina pelagi HTCC2506 (Q0G019: 137/130, 46-137, 59%)	
	putative diheme cytochrome c		Nide2637	Mus musculus (P07903: 298/80, 44-90, 34%)	delta proteobacterium MLMS-1 (Q1NSX9: 100/80, 17-92, 35%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	putative multiheme cytochrome c		Nide3221	Parabacteroides distasonis (A6L890: 494/424, 162-303, 28%)	Candidatus Kuenenia stuttgartiensis (Q1PX28: 390/424, 40-354, 32%)	
	putative multiheme cytochrome c		Nide3268	Erwinia tasmaniensis (B2VDA7: 522/412, 329-436, 29%)	Photobacterium profundum (Q6LRI9: 496/412, 91-454, 24%)	
	putative multiheme cytochrome c		Nide3269	Homo sapiens (Q53ET0: 693/435, 244-425, 23%)	Dethiobacter alkaliphilus AHT 1 (C0GJ19: 413/435, 19-325, 23%)	
	putative diheme cytochrome c		Nide3293	Thauera selenatis (Q9S1G7: 239/594, 120-237, 26%)	Candidatus Kuenenia stuttgartiensis (Q1PZE8: 535/594, 57-522, 31%)	
	putative monoheme cytochrome c		Nide3294	Gluconobacter oxydans (Q47945: 478/345, 332-412, 32%)	Candidatus Kuenenia stuttgartiensis (Q1PZE7: 256/345, 43-250, 30%)	
	putative monoheme cytochrome c		Nide3295	Geobacter sulfurreducens (Q749A8: 217/439, 7-117, 26%)	Candidatus Kuenenia stuttgartiensis (Q1PZE6: 316/439, 52-236, 30%)	
	putative monoheme cytochrome c		Nide3336	Synechococcus sp. JA-2-3B'a(2-13) (Q2JMC1: 340/156, 72-171, 29%)	Sorangium cellulosum (A9FK99: 890/156, 46-123, 36%)	
	putative multiheme cytochrome c		Nide3339	Wolinella succinogenes (Q9S1E6: 177/234, 1-158, 27%)	Carboxydothermus hydrogenoformans (Q3AEH4: 454/234, 5-218, 29%)	
	putative monoheme cytochrome c		Nide3350	Xenopus tropicalis (Q5XGC7: 153/150, 16-64, 39%)	Leptospirillum rubarum (A3EU12: 185/150, 8-150, 50%)	
	putative monoheme cytochrome c		Nide3450	Homo sapiens (A6NHN0: 477/118, 269-308, 45%)	Roseiflexus castenholzii (A7NPI4: 690/118, 57-158, 35%)	
	putative diheme cytochrome c		Nide3512	Paracoccus denitrificans (Q51658: 387/691, 270-373, 38%)	Sorangium cellulosum (A9FK28: 721/691, 133-707, 44%)	
	putative diheme cytochrome c		Nide3650	Rhodobacter sphaeroides (P81238: 129/271, 1-67, 39%)	Persephonella marina (C0QU35: 253/271, 17-248, 24%)	
	putative cytochrome c552		Nide3775	Monoraphidium braunii (Q09099: 89/126, 3-80, 34%)	Vibrio sp. MED222 (A3Y1M5: 121/126, 16-117, 38%)	
	putative monoheme cytochrome c		Nide3818	Vibrio fischeri (B5FDB6: 268/145, 146-189, 36%)	Sulfurihydrogenibium yellowstonense SS-5 (C4FLG5: 278/145, 24-152, 33%)	
	putative cytochrome c553		Nide3864	Synechocystis sp. PCC 6803 (P46445: 120/120, 10-112, 25%)	Thermomicrobium roseum (B9L2D6: 273/120, 89-172, 32%)	
	putative multiheme cytochrome c		Nide3868	Shewanella oneidensis (P83223: 596/240, 23-116, 26%)	Leptothrix cholodnii (B1Y4G6: 288/240, 12-153, 32%)	
	putative diheme cytochrome c		Nide3870	Halobacterium salinarium (P57715: 401/128, 309-363, 38%)	Geobacter sp. M21 (B3JWQ9: 201/128, 76-142, 33%)	
	putative monoheme cytochrome c		Nide3872	Homo sapiens (A8MTZ7: 301/158, 140-224, 23%)	Mariprofundus ferrooxydans PV-1 (Q0F295: 145/158, 32-144, 30%)	contains Ankyrin repeats
	putative monoheme cytochrome c		Nide3879	Thiocapsa roseopersicina (P86052: 192/186, 91-186, 32%)	Nitrosococcus oceani AFC27 (B6BZR3: 201/186, 44-192, 50%)	
	putative multiheme cytochrome c		Nide3882	Pectobacterium cypripedii (O34215: 441/470, 249-413, 25%)	Solibacter usitatus (Q021B0: 606/470, 201-593, 27%)	
	putative diheme cytochrome c		Nide3883	Euglena gracilis (P00119: 87/211, 2-78, 32%)	Deinococcus radiodurans (Q9RXH0: 340/211, 269-319, 51%)	
qcrC	putative Cytochrome *bc1* complex, cytochrome c subunit	1.10.2.2	Nide3886	Pavlova lutherii (P00107: 83/281, 2-81, 32%)	Ralstonia metallidurans (Q1LCC0: 272/281, 33-271, 49%)	
	putative cytochrome c55X		Nide3888	Pseudomonas stutzeri (P24039: 113/116, 15-103, 30%)	Methylibium petroleiphilum (A2SIC0: 653/114, 121-220, 39%)	
qcrB	Cytochrome *bc1* complex, fused cytochrome b/c subunit	1.10.2.2	Nide3889	Bacillus thermodenitrificans (Q45658: 224/441, 1-223, 48%)	Leptospirillum rubarum (A3ER63: 458/441, 8-450, 46%)	fused with cytochrome *c*

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	putative monoheme cytochrome c		Nide3892	Mus musculus (Q8CG79: 1128/156, 347-442, 27%)	Nitrobacter hamburgensis (Q1QM08: 158/156, 1-149, 49%)	
	putative cytochrome c553		Nide3893	Desulfovibrio vulgaris (P04032: 103/121, 4-103, 34%)	Thiobacillus denitrificans (Q3SGY3: 171/121, 61-154, 38%)	
	putative monoheme cytochrome c		Nide3923	Haemophilus influenzae (P45069: 421/158, 85-217, 27%)	Nitrobacter sp. Nb-311A (A3X2M8: 158/158, 6-152, 43%)	
	putative monoheme cytochrome c		Nide4057	Pseudomonas stutzeri (P00101: 104/148, 10-93, 33%)	Leptospirillum rubarum (A3EUN6: 184/148, 8-146, 41%)	
ccpA	Cytochrome c peroxidase		Nide4060	Methylobacterium extorquens (Q49128: 353/356, 52-353, 41%)	Leptospirillum rubarum (A3EU40: 338/356, 32-327, 54%)	
	putative monoheme cytochrome c		Nide4247	Paracoccus denitrificans (Q51702: 103/115, 13-97, 31%)	Pseudomonas fluorescens (C3K886: 429/115, 335-426, 40%)	
	putative cytochrome c55x		Nide4373	Pseudomonas aeruginosa (Q51479: 119/215, 13-104, 40%)	Leptothrix cholodnii (B1XXP9: 124/215, 44-122, 46%)	
Carbon metabolism						
	Carbonate uptake					
cah	Carbonic anhydrase	4.2.1.1	Nide1449	Erwinia carotovora (O52538: 244/262, 23-242, 39%)	Microcoleus chthonoplastes PCC 7420 (B4W0E2: 255/262, 32-250, 44%)	
cam	Carbonic anhydrase	4.2.1.1	Nide2799	Methanosarcina thermophila (P40881: 247/228, 51-241, 57%)	Syntrophus aciditrophicus (Q8PSJ1: 204/228, 2-196, 55%)	
cam	Carbonic anhydrase	4.2.1.1	Nide3172	Methanosarcina thermophila (P40881: 247/230, 45-241, 54%)	Syntrophus aciditrophicus (Q2LUP7: 204/230, 2-197, 57%)	
	RubisCO-like proteins					
rlp	Ribulose-1,5-bisphosphate carboxylase, large subunit	4.1.1.39	Nide1881	Rhodopseudomonas palustris (Q6N7T7: 368/389, 5-345, 37%)	Beggiatoa sp. PS (A7BYT9: 372/389, 4-351, 40%)	form IV RubisCO-like protein
	Oxidative and reductive carboxylate cycle					
sdhA	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	1.3.99.1	Nide0310	Rickettsia bellii (Q1RHB9: 596/551, 13-580, 49%)	Magnetococcus sp. MC-1 (A0L4R3: 568/551, 7-549, 55%)	
ppsA	putative Phosphoenolpyruvate synthase	2.7.9.2	Nide0497	Bacillus subtilis (O34309: 866/876, 12-349, 34%)	Desulfitobacterium hafniense (B8FPE4: 891/876, 18-887, 30%)	
forB	2-oxoglutarate:ferredoxin oxidoreductase, beta subunit	1.2.7.3	Nide0823	Methanocaldococcus jannaschii (Q57714: 298/299, 16-231, 31%)	Leptospirillum rubarum (A3EQL1: 288/299, 1-276, 72%)	OGOR
forC	2-oxoglutarate:ferredoxin oxidoreductase, gamma subunit	1.2.7.3	Nide0824	Pyrococcus horikoshii (O58411: 185/237, 4-164, 35%)	Leptospirillum rubarum (A3EQL2: 232/237, 6-223, 67%)	OGOR
forE	2-oxoglutarate:ferredoxin oxidoreductase, epsilon subunit	1.2.7.3	Nide0825	Aquifex aeolicus (O67251: 79/107, 1-78, 41%)	Leptospirillum sp. Group II '5-way CG' (B6ANA8: 101/107, 1-73, 57%)	OGOR
forD	2-oxoglutarate:ferredoxin oxidoreductase, delta subunit	1.2.7.3	Nide0826	Haemophilus influenzae (P45354: 928/254, 359-402, 36%)	Hydrogenivirga sp. 128-5-R1-1 (A8UW23: 239/254, 3-195, 62%)	OGOR
forA	2-oxoglutarate:ferredoxin oxidoreductase, alpha subunit	1.2.7.3	Nide0827	Methanocaldococcus jannaschii (Q57715: 386/448, 10-375, 29%)	Leptospirillum sp. Group II '5-way CG' (B6ANA5: 412/448, 25-404, 67%)	OGOR
aclB	ATP-citrate lyase, beta subunit	2.3.3.8	Nide0834	Ovis aries (Q2TCH3: 1101/399, 31-419, 33%)	Pelodictyon phaeoclathratiforme (B4S9V9: 398/399, 1-398, 61%)	indicative for rTCA
aclA	ATP citrate lyase, alpha subunit	2.3.3.8	Nide0835	Caenorhabditis elegans (P53585: 1106/606, 492-1088, 38%)	Chlorobium tepidum (Q8KDG2: 610/606, 1-610, 67%)	indicative for rTCA

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
acnA	Aconitate hydratase	4.2.1.3	Nide0836	Gracilaria verrucosa (P49609: 779/748, 43-773, 58%)	Chloroherpeton thalassium (B3QXI7: 755/748, 6-749, 61%)	
idh	Isocitrate dehydrogenase (NADP(+))	1.1.1.42	Nide0837	Azotobacter vinelandii (P16100: 741/743, 5-741, 72%)	Geobacter uraniireducens (A5G3L3: 743/743, 1-743, 73%)	
sdhA or nadB	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit or L-aspartate oxidase		Nide0839	Methanocaldococcus jannaschii (Q60356: 539/533, 3-388, 42%)	Geobacter metallireducens (Q39RJ5: 531/533, 12-526, 54%)	could link L-aspartate degradation to TCA if not sdhA
sucC	Succinyl-CoA synthetase, beta subunit	6.2.1.5	Nide0840	Parvibaculum lavamentivorans (A7HT39: 389/392, 1-389, 60%)	alpha proteobacterium BAL199 (A8TIM8: 389/392, 1-389, 58%)	
sucD	Succinyl-CoA synthetase, NAD(P)-binding, alpha subunit	6.2.1.5	Nide0841	Coxiella burnetii (P53591: 294/290, 1-290, 67%)	Sulfurihydrogenibium azorense (C1DU38: 293/290, 1-290, 71%)	
porE	Pyruvate:ferredoxin oxidoreductase, epsilon subunit	1.2.7.1	Nide0968	Aquifex aeolicus (O67232: 78/92, 1-71, 46%)	Hydrogenivirga sp. 128-5-R1-1 (A8UW27: 74/92, 1-70, 48%)	POR
porC	Pyruvate:ferredoxin oxidoreductase, gamma subunit	1.2.7.1	Nide0969	Pyrococcus kodakaraensis (Q5JIK2: 185/235, 2-179, 35%)	Leptospirillum rubarum (A3EQL8: 232/235, 4-224, 69%)	POR
porB	Pyruvate:ferredoxin oxidoreductase, beta subunit	1.2.7.1	Nide0970	Methanobacterium thermoautotrophicum (O27771: 288/300, 13-274, 34%)	Leptospirillum sp. Group II '5-way CG' (B6ANB2: 296/300, 1-293, 78%)	POR
porA	Pyruvate:ferredoxin oxidoreductase, alpha subunit	1.2.7.1	Nide0971	Methanobacterium thermoautotrophicum (P56810: 383/403, 7-382, 32%)	Leptospirillum rubarum (B6ANB1: 406/403, 12-400, 70%)	POR
porD	Pyruvate:ferredoxin oxidoreductase, delta subunit	1.2.7.1	Nide0972	Gallus gallus (Q5F3P8: 2008/211, 1669-1732, 28%)	Leptospirillum rubarum (A3EQD4: 193/211, 1-191, 65%)	POR
pycB	Pyruvate carboxylase, subunit B	6.4.1.1	Nide1204	Methanocaldococcus jannaschii (Q58628: 567/591, 21-564, 49%)	Mariprofundus ferrooxydans PV-1 (Q0EZ39: 617/591, 27-617, 56%)	oTCA
pycA	Pyruvate carboxylase, subunit A	6.4.1.1	Nide1205	Methanocaldococcus jannaschii (Q58626: 501/472, 1-441, 57%)	Thermodesulfovibrio yellowstonii (B5YH39: 471/472, 1-468, 61%)	oTCA
porB-N	partial 2-oxoacid:ferredoxin oxidoreductase beta subunit (fragment N-terminal)	1.2.7.3	Nide1461	Caenorhabditis elegans (Q18801: 399/64, 343-384, 32%)	Leptospirillum rubarum (A3EQL7: 296/64, 1-52, 62%)	
porB-C	partial 2-oxoacid:ferredoxin oxidoreductase beta subunit (fragment C-terminal)	1.2.7.3	Nide1464	Methanobacterium thermoautotrophicum (O27771: 288/240, 34-277, 32%)	Leptospirillum rubarum (A3EQL7: 296/240, 59-293, 80%)	
forB	2-oxoacid:ferredoxin oxidoreductase beta subunit	1.2.7.3	Nide1465	Methanocaldococcus jannaschii (Q57714: 298/297, 16-231, 31%)	Leptospirillum rubarum (A3EQL1: 288/297, 1-276, 71%)	
gltA	Citrate synthase	2.3.3.1	Nide2451	Synechocystis sp. PCC 6803 (Q59977: 397/377, 11-383, 59%)	Cyanothece sp. PCC 7822 (B4AVT4: 389/377, 6-379, 61%)	oTCA
fumC	Fumarate hydratase, class II	4.2.1.2	Nide2454	Halobacterium salinarium (Q9HQ29: 470/484, 1-463, 56%)	Pelobacter carbinolicus (Q3A7R0: 468/484, 2-464, 64%)	
sdhB	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	1.3.99.1	Nide2517	Haemophilus influenzae (P44893: 256/324, 11-236, 39%)	Magnetococcus sp. MC-1 (A0L4R2: 328/324, 38-313, 43%)	
sdhE	Succinate dehydrogenase/fumarate reductase, subunit C	1.3.99.1	Nide2527	Methanobacterium thermoautotrophicum (O27907: 302/300, 1-295, 34%)	Gloeobacter violaceus (Q7NCN3: 298/300, 5-296, 49%)	
lpd	Dihydrolipoyl dehydrogenase, E3 component of Pyruvate and 2-oxoglutarate dehydrogenase complexes	1.8.1.4	Nide2727	Pseudomonas fluorescens (P14218: 478/473, 7-475, 42%)	Moorella thermoacetica (Q2RHM5: 459/473, 5-458, 49%)	oTCA
mdh	Malate dehydrogenase	1.1.1.37	Nide3562	Chloroherpeton thalassium (B3QSH8: 310/313, 2-309, 61%)	Gemmatimonas aurantiaca (C1A902: 309/313, 4-307, 61%)	
pdhA	Pyruvate dehydrogenase E1 component, alpha subunit (acetyl-transferring)	1.2.4.1	Nide3852	Rhizobium meliloti (Q9R9N5: 348/324, 31-346, 42%)	Desulfotalea psychrophila (Q6ALF0: 335/324, 21-330, 65%)	oTCA

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
pdhB	Pyruvate dehydrogenase E1 component, beta subunit (Transketolase)	1.2.4.1	Nide3853	Arabidopsis thaliana (Q38799: 363/324, 36-357, 47%)	Polaromonas sp. JS666 (Q12FH3: 330/324, 5-327, 76%)	oTCA
pdhC	putative Dihydrolipoamide acetyltransferase (E2) component of pyruvate dehydrogenase complex	2.3.1.12	Nide3854	Mus musculus (Q8BMF4: 642/390, 219-637, 33%)	Thauera sp. MZ1T (C4ZNL1: 404/390, 1-404, 53%)	oTCA
pdhC	Dihydrolipoamide acetyltransferase (E2) component of pyruvate dehydrogenase complex	2.3.1.12	Nide3951	Rhizobium meliloti (Q9R9N3: 447/400, 1-446, 41%)	Rhodothermus marinus DSM 4252 (C1ZRZ9: 441/400, 1-440, 46%)	oTCA
pdhB	Pyruvate dehydrogenase E1 component, beta subunit (Transketolase)	1.2.4.1	Nide3952	Rickettsia felis (Q4UKQ7: 326/325, 1-322, 58%)	Gemmatimonas aurantiaca (C1A6D1: 326/325, 3-322, 61%)	oTCA
pdhA	Pyruvate dehydrogenase E1 component, alpha subunit (acetyl-transferring)	1.2.4.1	Nide3953	Rhizobium meliloti (Q9R9N5: 348/325, 28-348, 48%)	Myxococcus xanthus (Q1D8Y8: 389/325, 23-339, 57%)	oTCA
lpd	putative Dihydrolipoyl dehydrogenase, E3 component of pyruvate and 2-oxoglutarate dehydrogenase complexes	1.8.1.4	Nide4341	Staphylococcus epidermidis (P0A0E4: 547/453, 90-526, 33%)	bacterium Ellin514 (B9XE79: 489/453, 24-470, 47%)	oTCA
	Embden-Meyerhof-Parnas pathway					
glk	Glucokinase	2.7.1.2	Nide0354	Synechocystis sp. PCC 6803 (Q55855: 355/361, 9-354, 43%)	Gloeobacter violaceus (Q7NLF5: 327/361, 1-324, 46%)	
pgi	putative Glucose-6-phosphate isomerase	5.3.1.9	Nide0356	Carboxydothermus hydrogenoformans (Q3AFH3: 464/567, 60-440, 29%)	Acidobacteria bacterium Ellin345 (Q1IMT9: 958/567, 393-958, 45%)	modular protein
gpmA	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	5.4.2.1	Nide0478	Nitrobacter hamburgensis X14 (Q1QRT7: 207/201, 6-201, 63%)	Leptospirillum sp. Group II '5-way CG' (B6AKS2: 223/201, 16-200, 65%)	
tpiA	Triosephosphate isomerase	5.3.1.1	Nide1105	Carboxydothermus hydrogenoformans (Q3AFD0: 251/259, 1-251, 51%)	Thermosinus carboxydivorans Nor1 (A1HSR1: 250/259, 1-247, 52%)	
pgk	Phosphoglycerate kinase	2.7.2.3	Nide1106	Geobacter uraniireducens (A5G382: 399/399, 4-393, 64%)	Thermodesulfovibrio yellowstonii (B5YI30: 404/399, 10-401, 75%)	
gapA	Glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	Nide1107	Bacillus stearothermophilus (P00362: 335/335, 1-334, 65%)	Thermosinus carboxydivorans Nor1 (A1HSR3: 334/335, 1-333, 66%)	
fbpV	Fructose-1,6-bisphosphatase, class V	3.1.3.11	Nide2031	Methanocaldococcus jannaschii (Q57747: 389/370, 1-366, 59%)	Roseiflexus castenholzii (A7NNE8: 379/370, 2-366, 67%)	archaeal type enzyme
pgm	Phosphoglucomutase	5.4.2.2	Nide2083	Acetobacter xylinus (P38569: 555/550, 3-549, 64%)	Stigmatella aurantiaca DW4/3-1 (Q08SD1: 545/550, 3-545, 68%)	
pykF	Pyruvate kinase	2.7.1.40	Nide2807	Nicotiana tabacum (Q40546: 562/478, 91-562, 41%)	gamma proteobacterium HTCC2207 (Q1YVJ3: 469/478, 1-465, 63%)	
eno	Enolase (Phosphopyruvate hydratase)	4.2.1.11	Nide2913	Geobacter sulfurreducens (Q74AR6: 428/428, 1-421, 70%)	Geobacillus sp. Y412MC52 (C3J7A1: 430/428, 1-421, 70%)	
fbaB	Fructose-bisphosphate aldolase class I	4.1.2.13	Nide2980	Escherichia coli K12 (P0A991: 350/307, 52-346, 34%)	Methylobacterium radiotolerans (B1M1T6: 307/307, 4-306, 71%)	
fbp	Fructose-1-6-bisphosphatase	3.1.3.11	Nide2981	Gloeobacter violaceus (Q7NGN9: 348/332, 13-339, 55%)	Candidatus Kuenenia stuttgartiensis (Q1PYS0: 338/332, 7-338, 54%)	
manB/ pgm	bifunctional Phosphoglucomutase / Phosphomannomutase	5.4.2.2	Nide2983	Pseudomonas putida (Q88C93: 463/466, 12-452, 45%)	Solibacter usitatus (Q01SH3: 454/466, 6-448, 53%)	
	putative Aldose 1-epimerase	5.1.3.3	Nide3181	Escherichia coli K12 (P32139: 308/317, 64-283, 33%)	Frankia alni (Q0RN20: 311/317, 10-304, 36%)	
pfkA	6-phosphofructokinase	2.7.1.11	Nide3182	Streptomyces coelicolor (O08333: 342/421, 3-321, 33%)	Stigmatella aurantiaca DW4/3-1 (Q08ST9: 476/421, 51-458, 57%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
apgM	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	5.4.2.1	Nide4112	Geobacter sulfurreducens (Q74C57: 399/416, 1-399, 40%)	Geobacter metallireducens (Q39VR2: 399/416, 1-399, 39%)	
pykA	Pyruvate kinase	2.7.1.40	Nide4286	Bacillus stearothermophilus (Q02499: 587/483, 3-475, 46%)	Roseiflexus sp. RS-1 (A5UT73: 478/483, 6-475, 48%)	
pfkA	putative 6-phosphofructokinase	2.7.1.11	Nide4287	Haemophilus ducreyi (Q8GNC1: 324/779, 3-277, 32%)	Psychromonas ingrahamii (A1SSI0: 320/779, 2-313, 33%)	modular protein
	Pentose phosphate pathway					
rbsK	Ribokinase	2.7.1.15	Nide0315	Escherichia coli O157:H7 (P0A9J7: 309/308, 7-302, 43%)	Enterobacter sakazakii (A7MMV8: 309/308, 7-308, 45%)	
rpiA	Ribose-5-phosphate isomerase A	5.3.1.6	Nide0353	Nitrobacter hamburgensis X14 (Q1QN95: 232/238, 3-228, 52%)	Nitrobacter sp. Nb-311A (A3WWE4: 232/238, 3-225, 52%)	
pgl	6-phosphogluconolactonase	3.1.1.31	Nide0355	Anabaena sp. PCC 7120 (P46016: 240/253, 10-239, 43%)	Sphaerobacter thermophilus (C4CNK6: 252/253, 11-246, 47%)	
pgi	putative Glucose-6-phosphate isomerase	5.3.1.9	Nide0356	Carboxydothermus hydrogenoformans (Q3AFH3: 464/567, 60-440, 29%)	Acidobacteria bacterium Ellin345 (Q1IMT9: 958/567, 393-958, 45%)	modular protein
rpe	D-ribulose-5-phosphate 3-epimerase	5.1.3.1	Nide0397	Bacillus subtilis (O34557: 217/234, 2-214, 60%)	Desulfuromonas acetoxidans (Q1K3H3: 217/217, 2-213, 65%)	
zwf	Glucose-6-phosphate 1-dehydrogenase	1.1.1.49	Nide0468	Synechocystis sp. PCC 6803 (P73411: 509/509, 13-506, 48%)	Synechococcus sp. PCC 7335 (B4WRF8: 557/509, 46-555, 58%)	
talC	Transaldolase, putative Fructose-6-phosphate aldolase	2.2.1.2	Nide0493	Geobacter lovleyi (B3E1K9: 214/215, 1-214, 61%)	Haliangium ochraceum (C1UMI2: 214/215, 1-210, 63%)	
prsA	Ribose-phosphate pyrophosphokinase	2.7.6.1	Nide0781	Desulfotalea psychrophila (Q6AJL7: 313/313, 5-313, 60%)	Leptospirillum rubarum (A3EVV2: 314/313, 2-314, 69%)	
deoC	Deoxyribose-phosphate aldolase	4.1.2.4	Nide1710	Thermoanaerobacter sp. X514 (B0K709: 223/229, 2-217, 52%)	Geobacillus sp. Y412MC52 (C3IXI2: 223/229, 4-214, 53%)	
deoB	Phosphopentomutase	5.4.2.7	Nide1711	Moorella thermoacetica (Q2RID0: 389/384, 3-387, 40%)	Desulforudis audaxviator (B1I498: 397/384, 6-389, 40%)	
fbpV	Fructose-1,6-bisphosphatase, class V	3.1.3.11	Nide2031	Methanocaldococcus jannaschii (Q57747: 389/370, 1-366, 59%)	Roseiflexus castenholzii (A7NNE8: 379/370, 2-366, 67%)	archaeal type enzyme
pgm	Phosphoglucomutase	5.4.2.2	Nide2083	Acetobacter xylinus (P38569: 555/550, 3-549, 64%)	Stigmatella aurantiaca DW4/3-1 (Q08SD1: 545/550, 3-545, 68%)	
zwf	Glucose-6-phosphate dehydrogenase	1.1.1.49	Nide2203	Synechocystis sp. PCC 6803 (P73411: 509/507, 20-509, 48%)	Herpetosiphon aurantiacus (A9B4T6: 508/507, 14-508, 53%)	
gnd	6-phosphogluconate dehydrogenase (decarboxylating)	1.1.1.44	Nide2204	Bacillus subtilis (P54448: 297/297, 1-296, 41%)	Roseiflexus sp. RS-1 (A5V174: 299/297, 1-297, 60%)	
tktA	Transketolase	2.2.1.1	Nide2536	Mus musculus (Q9D4D4: 627/628, 12-625, 54%)	bacterium Ellin514 (B9XJW4: 612/628, 2-612, 59%)	
fbaB	Fructose-bisphosphate aldolase class I	4.1.2.13	Nide2980	Escherichia coli K12 (P0A991: 350/307, 52-346, 34%)	Methylobacterium radiotolerans (B1M1T6: 307/307, 4-306, 71%)	
fbp	Fructose-1-6-bisphosphatase	3.1.3.11	Nide2981	Gloeobacter violaceus (Q7NGN9: 348/332, 13-339, 55%)	Candidatus Kuenenia stuttgartiensis (Q1PYS0: 338/332, 7-338, 54%)	
manB/ pgm	bifunctional Phosphoglucomutase/ Phosphomannomutase	5.4.2.2	Nide2983	Pseudomonas putida (Q88C93: 463/466, 12-452, 45%)	Solibacter usitatus (Q01SH3: 454/466, 6-448, 53%)	
pfkA	6-phosphofructokinase	2.7.1.11	Nide3182	Streptomyces coelicolor (O08333: 342/421, 3-321, 33%)	Stigmatella aurantiaca DW4/3-1 (Q08ST9: 476/421, 51-458, 57%)	
xfp	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase	4.1.2.9, 4.1.2.22	Nide3515	Gloeobacter violaceus (Q7NLX2: 793/791, 12-791, 76%)	Desulfomicrobium baculatum (C1T5J4: 797/791, 6-782, 78%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
gntK	Gluconate kinase	2.7.1.12	Nide3654	Mus musculus (Q8R0J8: 184/125, 47-158, 42%)	Trichodesmium erythraeum (Q111A8: 170/125, 43-156, 50%)	fragment
pfkA	putative 6-phosphofructokinase	2.7.1.11	Nide4287	Haemophilus ducreyi (Q8GNC1: 324/779, 3-277, 32%)	Psychromonas ingrahamii (A1SSI0: 320/779, 2-313, 33%)	modular protein
Use of organic substrates						
	Use of acetate, formate, pyruvate					
pycB	Pyruvate carboxylase, subunit B	6.4.1.1	Nide1204	Methanocaldococcus jannaschii (Q58628: 567/591, 21-564, 49%)	Mariprofundus ferrooxydans PV-1 (Q0EZ39: 617/591, 27-617, 56%)	
pycA	Pyruvate carboxylase, subunit A	6.4.1.1	Nide1205	Methanocaldococcus jannaschii (Q58626: 501/472, 1-441, 57%)	Thermodesulfovibrio yellowstonii (B5YH39: 471/472, 1-468, 61%)	
acyP	Acylphosphatase	3.6.1.7	Nide1245	Thermus thermophilus (Q72L64: 88/116, 7-81, 51%)	Meiothermus ruber (C1XJV3: 88/116, 3-81, 51%)	
ppdK	Pyruvate, phosphate dikinase	2.7.9.1	Nide1634	Clostridium symbiosum (P22983: 874/949, 3-871, 56%)	Thermodesulfovibrio yellowstonii (B5YL47: 906/949, 23-900, 60%)	
acsA	Acetyl-CoA synthetase	6.2.1.1	Nide1643	Synechocystis sp. PCC 6803 (Q55404: 653/629, 1-643, 56%)	Methanospirillum hungatei (Q2FLA8: 629/629, 1-629, 59%)	
	putative Pyruvate decarboxylase	4.1.1.1	Nide1669	Azospirillum brasilense (P51852: 545/550, 3-534, 35%)	Planctomyces limnophilus (C1ZAI4: 601/550, 60-601, 51%)	
fdsG	Formate dehydrogenase, gamma subunit	1.2.1.2	Nide1930	Pseudomonas aeruginosa (Q9I0J8: 166/149, 37-165, 40%)	Methylococcus capsulatus (Q608U6: 159/149, 9-155, 39%)	
fdsB	Formate dehydrogenase, beta subunit	1.2.1.2	Nide1931	Rhizobium meliloti (P56913: 421/498, 18-409, 44%)	Burkholderia vietnamiensis (A4JCG1: 525/498, 3-508, 51%)	might belong to formate hydrogenlyase complex
fdsA	Formate dehydrogenase, alpha subunit	1.2.1.2	Nide1932	Bacillus subtilis (O34720: 985/908, 7-944, 34%)	Burkholderia multivorans CGD1 (B9BAE2: 983/908, 48-952, 48%)	
maeA	Malate dehydrogenase (oxaloacetate-decarboxylating)	1.1.1.38	Nide2442	Bacillus stearothermophilus (P16468: 478/480, 11-467, 55%)	Rubrobacter xylanophilus (Q1AX02: 481/480, 5-467, 60%)	
lpd	Dihydrolipoyl dehydrogenase, E3 component of pyruvate and 2-oxoglutarate dehydrogenase complexes	1.8.1.4	Nide2727	Pseudomonas fluorescens (P14218: 478/473, 7-475, 42%)	Moorella thermoacetica (Q2RHM5: 459/473, 5-458, 49%)	
	putative NAD-dependent alcohol dehydrogenase	1.1.1.1	Nide2776	Sulfolobus acidocaldarius (Q4J781: 344/343, 1-344, 33%)	Anaeromyxobacter dehalogenans (B8JFI5: 342/343, 1-341, 52%)	
atoB	Acetyl-CoA acetyltransferase	2.3.1.9	Nide3116	Arabidopsis thaliana (Q8S4Y1: 403/394, 15-401, 53%)	Plesiocystis pacifica SIR-1 (A6GI51: 395/394, 1-394, 58%)	
maeA	Malate dehydrogenase (oxaloacetate-decarboxylating)	1.1.1.38	Nide3846	Bacillus subtilis (O34962: 410/447, 5-398, 49%)	Dethiobacter alkaliphilus AHT 1 (C0GJW5: 446/447, 1-445, 56%)	
acsA	Acetyl-CoA synthetase	6.2.1.1	Nide3851	Bacillus subtilis (P39062: 572/586, 14-571, 49%)	Geobacter uraniireducens (A5GEE9: 584/586, 1-582, 65%)	
pdhA	Pyruvate dehydrogenase E1 component, alpha subunit (acetyl-transferring)	1.2.4.1	Nide3852	Rhizobium meliloti (Q9R9N5: 348/324, 31-346, 42%)	Desulfotalea psychrophila (Q6ALF0: 335/324, 21-330, 65%)	
pdhB	Pyruvate dehydrogenase E1 component, beta subunit (Transketolase)	1.2.4.1	Nide3853	Arabidopsis thaliana (Q38799: 363/324, 36-357, 47%)	Polaromonas sp. JS666 (Q12FH3: 330/324, 5-327, 76%)	
pdhC	putative Dihydrolipoamide acetyltransferase (E2) component of pyruvate dehydrogenase complex	2.3.1.12	Nide3854	Mus musculus (Q8BMF4: 642/390, 219-637, 33%)	Thauera sp. MZ1T (C4ZNL1: 404/390, 1-404, 53%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
pdhC	Dihydrolipoamide acetyltransferase (E2) component of pyruvate dehydrogenase complex	2.3.1.12	Nide3951	Rhizobium meliloti (Q9R9N3: 447/400, 1-446, 41%)	Rhodothermus marinus (C1ZRZ9: 441/400, 1-440, 46%)	
pdhB	Pyruvate dehydrogenase E1 component, beta subunit (Transketolase)	1.2.4.1	Nide3952	Rickettsia felis (Q4UKQ7: 326/325, 1-322, 58%)	Gemmatimonas aurantiaca (C1A6D1: 326/325, 3-322, 61%)	
pdhA	Pyruvate dehydrogenase E1 component, alpha subunit (acetyl-transferring)	1.2.4.1	Nide3953	Rhizobium meliloti (Q9R9N5: 348/325, 28-348, 48%)	Myxococcus xanthus (Q1D8Y8: 389/325, 23-339, 57%)	
	Aldehyde dehydrogenase	1.2.1.3	Nide4062	Arabidopsis thaliana (Q9SYG7: 508/518, 9-508, 52%)	Nitrosococcus oceani AFC27 (B6BXB5: 513/518, 2-503, 69%)	
adhC	Alcohol dehydrogenase, NADP-dependent	1.1.1.2	Nide4251	Mycobacterium bovis (P0A4X1: 346/348, 1-346, 54%)	Myxococcus xanthus (Q1CVY7: 349/348, 1-347, 64%)	
hyfI	putative Hydrogenase, small subunit	1.12.7.2	Nide4260	Methanocaldococcus jannaschii (Q57936: 148/173, 8-131, 47%)	Uncultured methanogenic archaeon RC-I (Q0W2C0: 247/173, 111-232, 65%)	inactive or unusual hydrogenase, might belong to formate hydrogenlyase complex
hyfG	putative Hydrogenase, large subunit	1.12.7.2	Nide4261	Escherichia coli K12 (P16431: 569/530, 37-536, 34%)	Uncultured methanogenic archaeon RC-I (Q0W2B9: 524/530, 3-524, 44%)	
hyfF	Hydrogenase, membrane subunit	1.12.7.2	Nide4262	Escherichia coli K12 (P77437: 526/510, 9-495, 36%)	Uncultured methanogenic archaeon RC-I (Q0W2B8: 487/510, 6-482, 43%)	
hyfE	putative Hydrogenase, membrane subunit	1.12.7.2	Nide4263	Mycobacterium tuberculosis (P64681: 220/215, 7-220, 35%)	Uncultured methanogenic archaeon RC-I (Q0W2B7: 220/215, 6-220, 49%)	
hyfC	putative Hydrogenase, membrane subunit	1.12.7.2	Nide4264	Escherichia coli K12 (P77858: 315/318, 11-313, 35%)	Candidatus Kuenenia stuttgartiensis (Q1PZL7: 312/318, 5-311, 54%)	
hyfB	Hydrogenase, membrane subunit	1.12.7.2	Nide4265	Escherichia coli K12 (P23482: 672/675, 6-668, 38%)	Opitutus terrae (B1ZSE3: 683/675, 8-682, 50%)	
lpd	putative Dihydrolipoyl dehydrogenase, E3 component of pyruvate and 2-oxoglutarate dehydrogenase complexes	1.8.1.4	Nide4341	Staphylococcus epidermidis (P0A0E4: 547/453, 90-526, 33%)	bacterium Ellin514 (B9XE79: 489/453, 24-470, 47%)	
	Branched chain amino acid degradation					
ilvE	Branched-chain amino acid aminotransferase	2.6.1.42	Nide1103	Pseudomonas aeruginosa (O86428: 307/304, 10-306, 56%)	Thermodesulfovibrio yellowstonii (B5YL62: 304/304, 1-304, 61%)	
mmsB	putative 3-hydroxyisobutyrate dehydrogenase	1.1.1.31	Nide1138	Drosophila melanogaster (Q9V8M5: 324/299, 28-321, 29%)	bacterium Ellin514 (B9XQ05: 293/299, 4-292, 65%)	
lpd	Dihydrolipoyl dehydrogenase, E3 component of pyruvate and 2-oxoglutarate dehydrogenase complexes	1.8.1.4	Nide2727	Pseudomonas fluorescens (P14218: 478/473, 7-475, 42%)	Moorella thermoacetica (Q2RHM5: 459/473, 5-458, 49%)	
	putative Acyl-CoA dehydrogenase	1.3.99.-	Nide3100	Mycobacterium bovis (P63430: 650/639, 65-633, 37%)	Saccharopolyspora erythraea (A4FK37: 641/639, 20-623, 38%)	
paaF	putative Enoyl-CoA hydratase	4.2.1.17	Nide3102	Bacillus subtilis (P94549: 258/262, 4-255, 46%)	Brevibacillus brevis (C0Z9H6: 257/262, 13-257, 53%)	
paaC	putative 3-hydroxyacyl-CoA dehydrogenase	1.1.1.35	Nide3103	Serratia proteamaculans (A8GH86: 715/407, 312-707, 39%)	Campylobacterales bacterium GD 1 (B6BLQ4: 705/407, 316-693, 39%)	
fadA	3-ketoacyl-CoA thiolase	2.3.1.16	Nide3104	Clostridium acetobutylicum (P45359: 392/396, 1-391, 47%)	Dethiobacter alkaliphilus AHT 1 (C0GEG4: 392/396, 1-392, 51%)	
	putative 2-oxoisovalerate dehydrogenase, alpha subunit (TPP-binding module)	1.2.4.4	Nide3113	Ralstonia eutropha (P27745: 333/333, 19-330, 36%)	Candidatus Kuenenia stuttgartiensis (Q1Q668: 325/333, 2-323, 64%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	putative 2-oxoisovalerate dehydrogenase, beta subunit (Transketolase)	1.2.4.4	Nide3114	Bacillus subtilis (P37941: 327/330, 4-327, 54%)	Candidatus Kuenenia stuttgartiensis (Q1Q665: 344/330, 23-334, 67%)	
	putative branched-chain alpha-keto acid dehydrogenase, dihydrolipoamide acyltransferase (E2) component	2.3.1.168	Nide3115	Bacillus subtilis (P37942: 424/385, 6-418, 40%)	Thermobaculum terrenum (C0UUD5: 420/385, 3-415, 48%)	
atoB	Acetyl-CoA acetyltransferase	2.3.1.9	Nide3116	Arabidopsis thaliana (Q8S4Y1: 403/394, 15-401, 53%)	Plesiocystis pacifica SIR-1 (A6GI51: 395/394, 1-394, 58%)	
scpAa	Methylmalonyl-CoA mutase	5.4.99.2	Nide3118	Caenorhabditis elegans (Q23381: 744/531, 78-567, 46%)	Roseiflexus sp. RS-1 (A5UPU7: 557/531, 29-557, 61%)	N-terminal fragment
scpAb	Methylmalonyl-CoA mutase	5.4.99.2	Nide3120	Escherichia coli K12 (P27253: 714/135, 586-708, 37%)	Candidatus Chloracidobacterium thermophilum (A8DJV5: 138/135, 5-127, 67%)	C-terminal fragment
pccB	Propionyl-CoA carboxylase	6.4.1.3	Nide3121	Mus musculus (Q3ULD5: 563/537, 31-563, 61%)	Thermomonospora curvata (C2AHZ8: 534/537, 5-534, 64%)	
paaF	putative Enoyl-CoA hydratase	4.2.1.17	Nide3122	Mycobacterium leprae (O07137: 257/282, 2-254, 33%)	Ralstonia solanacearum IPO1609 (B5SEC1: 264/282, 1-260, 46%)	
mvaB	Hydroxymethylglutaryl-CoA lyase	4.1.3.4	Nide3123	Mus musculus (Q8JZS7: 343/315, 48-335, 47%)	Legionella pneumophila (A5ICY7: 302/315, 7-269, 54%)	
fadE	putative Acyl-CoA dehydrogenase	1.3.99.3	Nide3129	Mycobacterium tuberculosis (P63429: 650/636, 17-612, 38%)	Psychromonas ingrahamii (A1SXV7: 633/636, 31-629, 39%)	
vorB	2-ketoisovalerate ferredoxin reductase, beta subunit	1.2.7.7	Nide3873	Methanocaldococcus jannaschii (Q57714: 298/312, 3-290, 43%)	Desulfotalea psychrophila (Q6ANZ4: 308/312, 9-303, 56%)	VOR
vorA	2-ketoisovalerate ferredoxin reductase, alpha subunit	1.2.7.7	Nide3874	Methanocaldococcus jannaschii (Q57715: 389/412, 8-377, 44%)	Pelobacter carbinolicus (Q3A569: 409/412, 3-409, 59%)	VOR
vorCD	2-ketoisovalerate ferredoxin reductase, fused gamma and delta subunit	1.2.7.7	Nide3875	Pyrococcus furiosus (Q51799: 185/305, 1-182, 41%)	Candidatus Kuenenia stuttgartiensis (Q1PVD2: 304/305, 9-300, 41%)	VOR
	Aldehyde dehydrogenase, NAD-dependent	1.2.1.-	Nide3974	Rhizobium sp. NGR234 (Q53197: 512/431, 90-510, 40%)	Acidobacteria bacterium Ellin345 (Q1IRG7: 481/431, 53-479, 56%)	
	Aldehyde dehydrogenase	1.2.1.3	Nide4062	Arabidopsis thaliana (Q9SYG7: 508/518, 9-508, 52%)	Nitrosococcus oceani AFC27 (B6BXB5: 513/518, 2-503, 69%)	
lpd	putative Dihydrolipoyl dehydrogenase, E3 component of 2-oxoacid oxidoreductase complexes	1.8.1.4	Nide4341	Staphylococcus epidermidis (P0A0E4: 547/453, 90-526, 33%)	bacterium Ellin514 (B9XE79: 489/453, 24-470, 47%)	
Uptake and storage						
	Phosphorous storage					
ppk2	putative Polyphosphate kinase 2	2.7.4.-	Nide0147	Plectonema boryanum (Q04605: 136/277, 4-115, 51%)	Roseiflexus castenholzii (A7NLF6: 290/277, 1-245, 57%)	
ppx	putative Exopolyphosphatase	3.6.1.11	Nide1410	Fusobacterium nucleatum subsp. nucleatum (Q8RF47: 664/316, 355-649, 33%)	Nocardia farcinica (Q5YQ33: 315/316, 6-315, 40%)	
ppx	putative Exopolyphosphatase	3.6.1.11	Nide2415	Pseudomonas aeruginosa (Q9ZN70: 506/517, 15-429, 31%)	Rhodothermus marinus (C1ZSU4: 534/517, 17-524, 41%)	
ppk	Polyphosphate kinase	2.7.4.1	Nide3408	Pseudomonas aeruginosa (Q9S646: 690/712, 7-682, 59%)	Nitrosococcus oceani AFC27 (B6BZP1: 690/709, 5-687, 63%)	
ppk2	Polyphosphate kinase 2	2.7.4.-	Nide3880	Plectonema boryanum (Q04605: 136/327, 7-97, 41%)	Polaribacter sp. MED152 (A2TWW9: 283/327, 4-279, 63%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	Glycogen metabolism					
pgmB	Beta-phosphoglucomutase hydrolase	5.4.2.6	Nide0042	Mycobacterium leprae (Q49741: 261/258, 24-260, 50%)	Stigmatella aurantiaca DW4/3-1 (Q092D5: 531/258, 10-255, 54%)	
otsA	Alpha,alpha-trehalose-phosphate synthase (UDP-forming)	2.4.1.15	Nide0144	Mycobacterium avium (A0QAK7: 492/473, 18-483, 38%)	Rhodoferax ferrireducens (Q21WG6: 743/743, 1-739, 62%)	
otsB	Trehalose phosphatase	3.1.3.12	Nide0145	Mycobacterium tuberculosis (A5U846: 391/257, 124-340, 35%)	Geobacter lovleyi (B3EBD0: 273/257, 1-252, 38%)	
glk	Glucokinase	2.7.1.2	Nide0354	Synechocystis sp. PCC 6803 (Q55855: 355/361, 9-354, 43%)	Gloeobacter violaceus (Q7NLF5: 327/361, 1-324, 46%)	
pgi	putative Glucose-6-phosphate isomerase	5.3.1.9	Nide0356	Carboxydothermus hydrogenoformans (Q3AFH3: 464/567, 60-440, 29%)	Acidobacteria bacterium Ellin345 (Q1IMT9: 958/567, 393-958, 45%)	modular protein
malQ	putative Glycogen debranching enzyme	2.4.1.25	Nide1132	Canis familiaris (Q2PQH8: 1533/661, 1084-1532, 26%)	Solibacter usitatus (Q022H9: 649/661, 8-648, 46%)	archaeal type
malQ	4-alpha-glucanotransferase	2.4.1.25	Nide1254	Escherichia coli K12 (P15977: 694/753, 10-691, 37%)	Nitrococcus mobilis Nb-231 (A4BM92: 1711/753, 3-734, 41%)	
glgP	Glycogen phosphorylase	2.4.1.1	Nide1289	Aquifex aeolicus (O66932: 692/719, 10-685, 43%)	Roseiflexus sp. RS-1 (A5V0X2: 719/719, 21-713, 54%)	
pgmB	putative Beta-phosphoglucomutase	5.4.2.6	Nide1522	Rhodobacter capsulatus (O33513: 227/235, 2-217, 30%)	Pelobacter propionicus (A1APU1: 227/235, 5-222, 40%)	
glgC	Glucose-1-phosphate adenylyltransferase	2.7.7.27	Nide1796	Rhodopirellula baltica (Q7UXF5: 446/412, 22-436, 54%)	Blastopirellula marina (A4A2T6: 420/412, 1-411, 57%)	
otsA	Alpha,alpha-trehalose-phosphate synthase (UDP-forming)	2.4.1.15	Nide1892	Schizosaccharomyces pombe (P40387: 513/508, 1-473, 32%)	Pelobacter propionicus (A1AU43: 747/508, 259-730, 42%)	
pgm	Phosphoglucomutase	5.4.2.2	Nide2083	Acetobacter xylinus (P38569: 555/550, 3-549, 64%)	Stigmatella aurantiaca DW4/3-1 (Q08SD1: 545/550, 3-545, 68%)	
	UDP-glucose/GDP-mannose dehydrogenase family protein	1.1.1.-	Nide2831	Staphylococcus aureus (P39861: 424/431, 1-424, 59%)	Geobacillus sp. WCH70 (C5D949: 426/431, 1-426, 66%)	
udg	UDP-glucose 6-dehydrogenase	1.1.1.22	Nide2887	Bacillus subtilis (O32271: 461/448, 7-430, 39%)	Archaeoglobus fulgidus (O29659: 465/448, 48-465, 46%)	
	NAD-dependent epimerase/dehydratase		Nide2888	Danio rerio (Q6GMI9: 418/377, 87-391, 37%)	Trichodesmium erythraeum (Q112T4: 347/377, 1-339, 51%)	
rfbF	Glucose-1-phosphate cytidylyltransferase	2.7.7.33	Nide2890	Salmonella typhimurium (P26396: 257/256, 1-256, 54%)	Candidatus Kuenenia stuttgartiensis (Q1Q6Y3: 277/256, 21-277, 70%)	
glgA	Glycogen synthase	2.4.1.21	Nide2906	Acidobacteria bacterium Ellin345 (Q1ILA0: 484/496, 7-481, 45%)	Syntrophobacter fumaroxidans (A0LK44: 488/496, 10-482, 49%)	
manB/ pgm	bifunctional Phosphoglucomutase / Phosphomannomutase	5.4.2.2	Nide2983	Pseudomonas putida (Q88C93: 463/466, 12-452, 45%)	Solibacter usitatus (Q01SH3: 454/466, 6-448, 53%)	
galU	UTP--glucose-1-phosphate uridylyltransferase	2.7.7.9	Nide3045	Bacillus subtilis (Q05852: 292/313, 4-284, 56%)	Leptospirillum sp. Group II '5-way CG' (B6AMK5: 297/313, 4-290, 62%)	
	putative Amylo-alpha-1,6-glucosidase	3.2.1.33	Nide3191	Bombyx mori (P32358: 579/721, 184-254, 37%)	Nitrosococcus oceani AFC27 (B6C283: 723/721, 3-722, 57%)	
glgB	1,4-alpha-glucan branching enzyme	2.4.1.18	Nide3464	Aquifex aeolicus (O66936: 630/656, 5-627, 64%)	Sphaerobacter thermophilus (C4CIM1: 656/656, 19-649, 70%)	
glgB	putative 1,4-alpha-glucan branching enzyme	2.4.1.18	Nide3768	Mycobacterium paratuberculosis (Q73X75: 731/636, 133-726, 28%)	bacterium Ellin514 (B9XCV6: 639/636, 12-635, 52%)	
glgP	Alpha-glucan phosphorylase	2.4.1.1	Nide3903	Mycobacterium tuberculosis (Q10639: 863/573, 113-724, 37%)	Opitutus terrae (B1ZPX7: 566/573, 2-564, 55%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
udg	UDP-glucose 6-dehydrogenase	1.1.1.22	Nide4145	Rhizobium meliloti (O54068: 437/439, 1-436, 53%)	Thermodesulfovibrio yellowstonii (B5YKR5: 434/439, 1-433, 60%)	
	Iron uptake and storage					
bfrB	Bacterioferritin		Nide1021	Azotobacter vinelandii (P22759: 156/158, 1-154, 47%)	Anaeromyxobacter sp. Fw109-5 (A7HCJ8: 156/158, 1-156, 54%)	
bfrA	Bacterioferritin		Nide1022	Azotobacter vinelandii (P22759: 156/159, 1-154, 40%)	Nitrococcus mobilis Nb-231 (A4BL46: 154/159, 1-154, 44%)	
	putative Multi-domain non-ribosomal peptide synthetase		Nide1742	Brevibacillus parabrevis (Q70LM4: 5085/3120, 5-2096, 40%)	Sorangium cellulosum (A9G1U1: 3445/3120, 1795-3315, 44%)	
	Cyclic peptide transporter	3.6.3.-	Nide1743	Pseudomonas syringae pv. syringae (P33951: 566/565, 30-554, 35%)	bacterium Ellin514 (B9XGC9: 541/565, 1-522, 44%)	
	putative Multi-domain non-ribosomal peptide synthetase		Nide2150	Brevibacillus parabrevis (Q70LM5: 7756/1907, 3091-4545, 36%)	Sorangium cellulosum (A9FNK4: 1791/1907, 14-1606, 47%)	
	putative Non-ribosomal peptide synthetase		Nide2151	Brevibacillus parabrevis (Q70LM5: 7756/214, 905-1043, 42%)	Nostoc punctiforme (B2J0Z1: 1769/214, 1581-1721, 47%)	fragment
	putative Polyketide synthase		Nide2152	Streptomyces antibioticus (Q07017: 3519/1517, 27-1580, 32%)	bacterium Ellin514 (B9XIK8: 1911/1517, 30-1409, 46%)	
	Cyclic peptide transporter		Nide2153	Pseudomonas syringae pv. syringae (P33951: 566/565, 14-554, 34%)	Sorangium cellulosum (A9FNJ8: 598/565, 1-568, 55%)	
dat	Diaminobutyrate-2-oxoglutarate transaminase	2.6.1.76	Nide2154	Rhizobium meliloti (Q9Z3R2: 470/468, 31-461, 58%)	Anabaena variabilis (Q3M984: 493/468, 44-484, 68%)	
	putative Penicillin amidase	3.5.1.11	Nide2155	Pseudomonas sp. SE83 (P15558: 774/795, 14-755, 31%)	Sorangium cellulosum (A9FNJ4: 795/795, 6-795, 50%)	
	MbtH-like protein		Nide2156	Mycobacterium bovis (P59965: 71/81, 8-70, 41%)	Myxococcus xanthus (Q1D7Q2: 71/81, 3-68, 83%)	
	Thioesterase, type II	3.1.2.-	Nide2157	Brevibacillus parabrevis (Q70LM8: 263/253, 25-256, 43%)	Cyanothece sp. PCC 7424 (B7K7V0: 1337/253, 1094-1325, 50%)	
syrP	putative Pyoverdine biosynthesis regulatory protein SyrP		Nide2158	Arabidopsis thaliana (Q9LIG0: 330/344, 25-330, 31%)	Sorangium cellulosum (A9FNJ0: 393/344, 82-389, 66%)	Clavaminate synthase-like
	putative Multi-domain non-ribosomal peptide synthetase		Nide2159	Brevibacillus parabrevis (Q70LM5: 7756/2178, 5186-7264, 36%)	Sorangium cellulosum (A9FNI5: 4722/2178, 1538-2591, 54%)	
	putative Multi-domain non-ribosomal peptide synthetase		Nide2160	Brevibacillus parabrevis (Q70LM5: 7756/2623, 2603-5177, 42%)	Sorangium cellulosum (A9FNI0: 3015/2623, 1128-2653, 52%)	
	putative Multi-domain non-ribosomal peptide synthetase		Nide2161	Brevibacillus parabrevis (Q70LM5: 7756/2999, 2585-4079, 43%)	Sorangium cellulosum (A9FNI0: 3015/2999, 4-2997, 55%)	
	putative Multi-domain non-ribosomal peptide synthetase		Nide2162	Brevibacillus parabrevis (Q70LM4: 5085/2392, 37-2094, 39%)	Myxococcus xanthus (Q1D6A2: 5741/2392, 1756-3169, 46%)	
	putative Multidrug efflux transporter		Nide2163	Haemophilus influenzae (Q57124: 1032/1038, 7-1012, 33%)	Methylococcus capsulatus (Q605Z1: 1024/1038, 1-1007, 51%)	
	putative Multidrug efflux transporter, membrane fusion protein		Nide2164	Haemophilus influenzae (Q57500: 382/359, 5-360, 29%)	Azoarcus sp. BH72 (A1KA96: 365/359, 3-351, 42%)	
tonB	putative Protein TonB		Nide2166	Homo sapiens (Q96L91: 3160/343, 2539-2690, 32%)	Human herpesvirus 3 (Q0Q9E6: 2958/343, 2434-2637, 35%)	
exbD	Biopolymer transport protein ExbD		Nide2167	Aquifex aeolicus (O67694: 132/124, 10-129, 45%)	Nitratiruptor sp. SB155-2 (A6Q4X1: 131/124, 14-131, 42%)	
exbB	Biopolymer transport protein ExbB		Nide2168	Aquifex aeolicus (O67637: 148/141, 2-141, 52%)	Hydrogenivirga sp. 128-5-R1-1 (A8URJ9: 141/141, 1-140, 60%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	putative Prolyl oligopeptidase	3.4.-.-	Nide2169	Caenorhabditis elegans (P34422: 740/695, 313-655, 31%)	uncultured bacterium (Q6WLC7: 676/695, 43-661, 42%)	
	Ferrichrome iron receptor		Nide2171	Haemophilus influenzae (P45220: 181/45, 58-96, 36%)	Anabaena sp. PCC 7120 (Q8YTX5: 851/45, 212-251, 68%)	fragment
fhuA	Ferrichrome iron receptor		Nide2173	Escherichia coli K12 (P06971: 747/855, 58-747, 30%)	Anabaena sp. PCC 7120 (Q8YV06: 863/855, 176-863, 41%)	
fecR	Protein FecR, ferric citrate sensor		Nide2174	Escherichia coli K12 (P23485: 317/310, 28-304, 32%)	Rhodopseudomonas palustris (B3QC13: 321/310, 27-308, 41%)	
fecI	putative RNA polymerase sigma factor FecI (sigma-19)		Nide2175	Escherichia coli K12 (P23484: 173/172, 11-164, 32%)	Xanthobacter autotrophicus (A7IJ24: 171/172, 1-169, 41%)	
fur	Ferric uptake regulation protein Fur		Nide2176	Campylobacter jejuni (P0C631: 157/172, 12-150, 34%)	Geobacter metallireducens (Q39SV4: 143/172, 13-143, 44%)	
fecA	putative Iron(III) dicitrate transport protein FecA		Nide2178	Escherichia coli K12 (P13036: 774/864, 35-763, 26%)	Caulobacter crescentus (B8GXM9: 730/864, 61-730, 38%)	
fecR	putative Protein FecR, ferric citrate sensor		Nide2179	Escherichia coli K12 (P23485: 317/339, 12-315, 27%)	Pseudomonas entomophila (Q1IGG4: 325/339, 15-322, 34%)	
fecI	putative RNA polymerase sigma factor FecI (sigma-19)		Nide2180	Escherichia coli K12 (P23484: 173/201, 2-164, 30%)	Xanthobacter autotrophicus (A7IJ24: 171/201, 19-169, 41%)	
	putative Ferric iron uptake protein		Nide2183	Escherichia coli K12 (P06971: 747/857, 62-721, 28%)	Gloeobacter violaceus (Q7NEA6: 730/857, 17-730, 40%)	
fecR	Protein FecR, ferric citrate sensor		Nide2185	Escherichia coli K12 (P23485: 317/329, 7-312, 30%)	Azorhizobium caulinodans (A8HWH5: 315/329, 1-314, 40%)	
fecI	putative RNA polymerase sigma factor FecI (sigma-19)		Nide2188	Escherichia coli K12 (P23484: 173/173, 1-164, 31%)	Pseudomonas fluorescens (Q4KBX1: 230/173, 68-222, 51%)	
tonB	putative Protein TonB		Nide2189	Saccharomyces cerevisiae (P08640: 1367/325, 657-800, 29%)	Methylobacillus flagellatus (Q1GYW4: 246/325, 29-244, 26%)	
	protein of unknown function, Ferritin-like		Nide2585	Polaromonas sp. JS666 (Q125Q0: 236/194, 107-223, 27%)	Toxoplasma gondii ME49 (B6KNT4: 1840/194, 744-844, 30%)	
	putative Bacterioferritin		Nide3169	Listeria welshimeri serovar 6b (A0AHX3: 785/165, 430-558, 29%)	Synechococcus sp. WH 5701 (A3YVS7: 154/165, 10-152, 76%)	
	putative Bacterioferritin-associated ferredoxin		Nide3257	Plectonema boryanum (Q00241: 205/95, 54-111, 29%)	Marinobacter sp. ELB17 (A3JE77: 68/95, 1-49, 45%)	
Stress response and defence						
	Various resistance mechanisms					
merR	Mercuric resistance operon regulatory protein		Nide0064	Pseudomonas sp SB3 (P69413: 144/135, 1-133, 70%)	Burkholderia multivorans (A9ADW6: 135/135, 1-135, 99%)	
merT	Mercury ion transport protein		Nide0065	Serratia marcescens (P13112: 116/123, 2-116, 71%)	Burkholderia multivorans (A9ADW7: 123/123, 1-123, 100%)	
merP	Periplasmic mercury ion binding protein		Nide0066	Enterobacter cloacae (P0A218: 91/94, 1-89, 72%)	Burkholderia multivorans (A9ADW8: 94/94, 1-94, 99%)	
merA	putative Mercuric reductase	1.16.1.1	Nide1146	Staphylococcus epidermidis (P0A0E4: 547/517, 84-526, 34%)	Rhodopirellula baltica (Q7UEQ0: 507/517, 4-503, 59%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
cynS	Cyanate hydratase	4.2.1.104	Nide1365	Thiobacillus denitrificans (Q3SHJ2: 147/146, 1-147, 40%)	Bordetella petrii (A9HZN1: 148/146, 2-148, 39%)	
cld	Chlorite dismutase	1.13.11.49	Nide1387	Geobacillus kaustophilus (Q5KUD5: 248/264, 25-213, 26%)	Pseudomonas stutzeri (B1AAM4: 282/264, 44-282, 43%)	
arsC	putative Arsenate reductase (glutaredoxin)	1.20.4.1	Nide2447	Escherichia coli K12 (P76569: 119/116, 4-118, 38%)	Nitrosomonas europaea (Q82U35: 115/116, 4-113, 47%)	
cld	Chlorite dismutase	1.13.11.49	Nide3081	Bacillus weihenstephanensis (A9VSI0: 247/235, 24-221, 32%)	Rubrobacter xylanophilus (Q1AY13: 240/235, 10-240, 51%)	
	putative Methyltransferase, UbiE-family	2.1.1.-	Nide3699	Rattus norvegicus (Q8VHT6: 369/395, 50-328, 31%)	Sorangium cellulosum (A9G3P6: 389/395, 8-389, 42%)	possible Arsenite methyltransferase
arsR	putative Transcriptional regulator, ArsR family		Nide3700	Methanocaldococcus jannaschii (Q58721: 89/121, 6-67, 47%)	Chthoniobacter flavus Ellin428 (B4CZ22: 106/121, 11-101, 40%)	
	Predicted Fe-S oxidoreductase		Nide3701	Bacillus subtilis (O31423: 410/346, 69-241, 27%)	Pelotomaculum thermopropionicum (A5D616: 349/346, 15-343, 61%)	
	putative Inorganic phosphate transporter		Nide3702	Pyrococcus kodakaraensis (Q5JHX4: 406/398, 10-402, 28%)	Sorangium cellulosum (A9G3P1: 374/398, 7-374, 42%)	might function as arsenite/arsenate transporter
	conserved exported protein of unknown function		Nide3703	Mycobacterium tuberculosis (Q79FW4: 645/491, 256-441, 29%)	Hydrogenivirga sp. 128-5-R1-1 (A8V0N1: 437/491, 20-434, 36%)	putative outer membrane porin
aoxA	Arsenite oxidase, small subunit	1.20.98.1	Nide3704	Herminiimonas arsenicoxydans (Q8GGJ7: 173/174, 6-173, 43%)	Achromobacter sp. SY8 (A5A3H6: 176/174, 8-176, 48%)	
aoxB	Arsenite oxidase, large subunit	1.20.98.1	Nide3705	Herminiimonas arsenicoxydans (Q8GGJ6: 826/820, 6-817, 47%)	Chloroflexus aurantiacus (A9WJY7: 836/820, 3-831, 55%)	
arsC	Arsenate reductase	1.20.4.-, 3.1.3.48	Nide3707	Bacillus halodurans (Q9K8K8: 139/138, 3-139, 47%)	Acidobacteria bacterium Ellin345 (Q1IME2: 150/138, 12-145, 62%)	also dephosphorylates tyrosine phosphorylated proteins
arsB	putative Arsenite resistance protein ArsB		Nide3708	Synechocystis sp. PCC 6803 (P74311: 383/353, 16-365, 45%)	Solibacter usitatus (Q01Y95: 354/353, 6-348, 73%)	
arsM	putative Arsenite S-adenosylmethyltransferase	2.1.1.137	Nide3709	Mus musculus (Q91WU5: 376/274, 6-235, 32%)	Solibacter usitatus (Q01Y94: 281/274, 1-271, 67%)	
arsR	Arsenical resistance operon repressor		Nide3710	Methanocaldococcus jannaschii (Q58721: 89/111, 6-89, 45%)	Gemmatimonas aurantiaca (C1A7V1: 107/111, 4-106, 50%)	
	ROS protection					
ccpA	Cytochrome c551 peroxidase	1.11.1.5	Nide0057	Methylobacillus flagellatus (Q50426: 333/357, 39-321, 41%)	Solibacter usitatus (Q01SD6: 336/357, 20-336, 51%)	
mntC	Manganese transport system, permease component		Nide0383	Bacillus halodurans (Q9KD29: 292/273, 10-273, 41%)	Halothermothrix orenii (B8CZ63: 277/273, 1-266, 53%)	
mntB	Manganese transport system, ATPase component	3.6.3.-	Nide0384	Bacillus subtilis (O34338: 250/259, 4-248, 38%)	Sphaerobacter thermophilus (C4CJR3: 275/259, 22-272, 43%)	
mntA	Manganese transport system, periplasmic binding component		Nide0385	Treponema pallidum (O83077: 316/271, 38-307, 30%)	Sphaerobacter thermophilus (C4CJR2: 312/271, 50-311, 35%)	
bfrB	Bacterioferritin		Nide1021	Azotobacter vinelandii (P22759: 156/158, 1-154, 47%)	Anaeromyxobacter sp. Fw109-5 (A7HCJ8: 156/158, 1-156, 54%)	
bfrA	Bacterioferritin		Nide1022	Azotobacter vinelandii (P22759: 156/159, 1-154, 40%)	Nitrococcus mobilis Nb-231 (A4BL46: 154/159, 1-154, 44%)	
trxB	Thioredoxin-disulfide reductase	1.8.1.9	Nide1127	Mycobacterium smegmatis (O30973: 311/304, 7-310, 60%)	Rhodothermus marinus (C1ZU61: 334/304, 20-323, 64%)	
tpx	putative Thiol peroxidase (Peroxiredoxin)	1.11.1.-	Nide1202	Bacillus halodurans (Q9K813: 166/228, 3-163, 41%)	Geobacillus sp. WCH70 (C5D685: 166/228, 4-163, 47%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
btuE	Glutathione peroxidase	1.11.1.9	Nide1235	Synechocystis sp. PCC 6803 (P74250: 169/165, 8-169, 51%)	Solibacter usitatus (Q01QA9: 180/165, 14-180, 56%)	
trxA	Thioredoxin		Nide1582	Pisum sativum (P48384: 172/115, 74-172, 55%)	Leptospirillum rubarum (A3EWE7: 110/115, 1-108, 58%)	
bcp	Peroxiredoxin	1.11.1.15	Nide1596	Coxiella burnetii (Q83CY8: 151/155, 3-151, 58%)	Thermoanaerobacter tengcongensis (Q8R9N1: 157/155, 2-150, 64%)	
	Thioredoxin-like protein		Nide1753	Geobacillus kaustophilus (Q5KXL9: 174/191, 32-148, 27%)	Thermosynechococcus elongatus (Q8DI04: 193/191, 5-191, 68%)	
	probable Peroxiredoxin	1.11.1.15	Nide1875	Dictyostelium discoideum (Q54SE2: 241/211, 33-241, 60%)	bacterium Ellin514 (B9XB85: 211/211, 1-211, 78%)	
	Thioredoxin-like protein		Nide1884	Bacillus subtilis (P39598: 200/190, 6-194, 24%)	Bdellovibrio bacteriovorus (Q6MMV4: 188/190, 7-185, 40%)	
	Methionine sulfoxide reductase	1.8.4.11	Nide1889	Leptospira interrogans serogroup Icterohaemorrhagiae serovar copenhageni (Q72NN2: 132/139, 5-129, 54%)	Roseobacter sp. AzwK-3b (A6FQV8: 158/139, 32-158, 64%)	
	putative Thioredoxin		Nide2084	Methanocaldococcus jannaschii (Q57755: 85/88, 6-85, 29%)	Methanococcoides burtonii (Q12ZK6: 80/88, 1-79, 46%)	
trxA	Thioredoxin		Nide2202	Rickettsia felis (Q4UNK3: 105/104, 1-102, 37%)	Actinobacillus pleuropneumoniae serotype 5b (A3N184: 105/104, 1-85, 49%)	
	possible Alkylhydroperoxidase AhpD		Nide2793	Paracoccus denitrificans (P08304: 111/116, 26-102, 48%)	Beijerinckia indica subsp. indica (B2IJN4: 101/116, 1-97, 67%)	
	putative Bacterioferritin		Nide3169	Listeria welshimeri serovar 6b (A0AHX3: 785/165, 430-558, 29%)	Synechococcus sp. WH 5701 (A3YVS7: 154/165, 10-152, 76%)	
	Glutaredoxin		Nide3573	Mus musculus (Q8BWM0: 384/81, 99-180, 33%)	Halomicrobium mukohataei (C1VIK9: 84/81, 4-80, 48%)	
	Glutaredoxin		Nide3582	Synechocystis sp. PCC 6803 (P73056: 107/108, 9-107, 36%)	Synechococcus sp. JA-3-3Ab (Q2JWT2: 113/108, 1-106, 42%)	
	Thioredoxin-like protein		Nide3865	Photorhabdus luminescens subsp. laumondii (Q7MZX2: 575/400, 102-386, 23%)	Methylococcus capsulatus (Q603L8: 405/400, 18-404, 49%)	
trxA	Thioredoxin	1.8.1.8	Nide3877	Corynebacterium nephridii (P52228: 145/144, 3-139, 43%)	Desulfovibrio desulfuricans (Q30VT7: 146/144, 1-144, 66%)	
	Peroxiredoxin	1.11.1.15	Nide4033	Mycobacterium bovis (P65689: 153/156, 2-151, 37%)	Leptospirillum rubarum (A3ERT1: 153/156, 1-152, 65%)	
	putative Thioredoxin		Nide4034	Chlamydia pneumoniae (Q9Z7P5: 102/110, 2-101, 30%)	Delftia acidovorans (A9BMI9: 110/110, 26-109, 36%)	
ccpA	Cytochrome c peroxidase	1.11.1.5	Nide4060	Methylobacterium extorquens (Q49128: 353/356, 52-353, 41%)	Leptospirillum rubarum (A3EU40: 338/356, 32-327, 54%)	
	Carotenoid biosynthesis					
ispA	Geranyltranstransferase	2.5.1.1, 2.5.1.10	Nide0701	Hevea brasiliensis (Q94ID7: 370/326, 78-370, 47%)	Geobacter bemidjiensis (B5EI12: 295/326, 1-295, 56%)	
ispE	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	2.7.1.148	Nide0780	Geobacter metallireducens (Q39RQ7: 281/304, 3-263, 39%)	Desulfococcus oleovorans (A8ZXZ2: 285/304, 1-267, 38%)	
ispH	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	1.17.1.2	Nide0859	Bradyrhizobium japonicum (Q89QW7: 308/320, 1-305, 61%)	Rhodopseudomonas palustris (Q218A1: 314/320, 7-311, 61%)	
dxs	1-deoxy-D-xylulose-5-phosphate synthase	2.2.1.7	Nide0869	Geobacter sulfurreducens (Q74FC3: 637/648, 3-628, 55%)	Desulfotomaculum reducens (A4J3G0: 635/648, 6-620, 56%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
ispG	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	1.17.7.1	Nide0870	Geobacter sulfurreducens (Q74D60: 353/401, 2-352, 59%)	Geobacter metallireducens (Q39VY7: 355/401, 3-352, 60%)	
sqs	Squalene synthase	2.5.1.21	Nide1238	Candida glabrata (Q9HGZ6: 443/319, 64-327, 26%)	Nitrosospira multiformis (Q2YAG1: 380/319, 50-368, 68%)	
dxr	1-deoxy-D-xylulose-5-phosphate reductoisomerase	1.1.1.267	Nide1511	Chlorobium phaeobacteroides (B3EK13: 382/386, 1-382, 57%)	Pelobacter carbinolicus (Q3A3A1: 394/386, 4-386, 57%)	
uppS	Undecaprenyl pyrophosphate synthase	2.5.1.31	Nide1513	Heliobacillus mobilis (Q8GDY3: 260/261, 15-258, 53%)	Heliobacterium modesticaldum (B0THE2: 260/261, 10-258, 53%)	
pds	putative Pytoene desaturase	1.14.99.-	Nide1519	Synechococcus elongatus (P26294: 474/436, 3-455, 25%)	Prosthecochloris vibrioformis (A4SFI6: 453/436, 3-444, 25%)	
crtB	Phytoene synthase	2.5.1.32	Nide1520	Spirulina platensis (O07333: 309/303, 12-308, 35%)	Methylobacillus flagellatus (Q1H1A0: 278/303, 4-278, 44%)	
ispF	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	4.6.1.12	Nide2976	Desulfococcus oleovorans (A9A0H0: 168/160, 1-154, 55%)	delta proteobacterium MLMS-1 (Q1NQ03: 400/160, 242-396, 55%)	
ispD	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	2.7.7.60	Nide2977	Syntrophus aciditrophicus (Q2LUS9: 234/251, 3-227, 49%)	Geobacter sp. M21 (B3JVB6: 231/251, 3-230, 46%)	
	Polyprenyl-diphosphate synthase	2.5.1.-	Nide3385	Shigella flexneri (P0AD58: 323/334, 12-323, 42%)	Desulfuromonas acetoxidans (Q1K0W2: 322/334, 1-322, 48%)	
	CRISPR-associated genes					
cas3	CRISPR-associated helicase Cas3		Nide1541	Escherichia coli K12 (P38036: 888/909, 9-807, 30%)	Methanococcoides burtonii (Q12YB2: 921/909, 2-916, 39%)	
cse1	putative CRISPR-associated protein Cse1		Nide1542	Escherichia coli K12 (Q46901: 502/520, 1-283, 22%)	Methanococcoides burtonii (Q12YB1: 528/520, 4-522, 47%)	
cse2	putative CRISPR-associated protein Cse2		Nide1544	Chlamydia muridarum (Q9PKK3: 804/181, 646-709, 33%)	Methanosphaerula palustris (B8GIV5: 178/181, 26-175, 38%)	
cse4	CRISPR-associated protein Cse4		Nide1545	Escherichia coli K12 (Q46899: 363/398, 1-313, 31%)	Pelobacter carbinolicus (Q3A5Z5: 373/398, 1-370, 49%)	
cas5e	CRISPR-associated protein Cas5e		Nide1546	Escherichia coli K12 (Q46898: 224/266, 1-173, 36%)	Methanococcoides burtonii (Q12YA8: 244/266, 1-238, 44%)	
cse3	CRISPR-associated protein Cse3		Nide1547	Escherichia coli K12 (Q46897: 199/232, 15-194, 30%)	Methanosphaerula palustris (B8GIV2: 225/232, 3-223, 42%)	
cas1	CRISPR-associated protein Cas1		Nide1548	Escherichia coli K12 (Q46896: 305/306, 5-304, 68%)	Leptospirillum rubarum (A3EQA2: 306/306, 1-306, 77%)	
cas2	CRISPR-associated protein Cas2		Nide1549	Escherichia coli K12 (P45956: 94/104, 3-93, 67%)	Leptospirillum rubarum (A3EQ99: 102/104, 1-102, 76%)	
Cofactor metabolism						
	Cobalamin and heme biosynthesis					
cobC	putative Alpha-ribazole phosphatase cobC	3.1.3.73	Nide0515	Salmonella typhimurium (P39701: 202/202, 5-196, 30%)	Moorella thermoacetica (Q2RJH0: 214/202, 6-205, 36%)	
hemE	Uroporphyrinogen decarboxylase	4.1.1.37	Nide1215	Sorangium cellulosum (A9FZW7: 349/341, 1-338, 64%)	Sphaerobacter thermophilus (C4CNU5: 348/341, 9-346, 66%)	
hemH	Ferrochelatase	4.99.1.1	Nide1216	Herpetosiphon aurantiacus (A9B546: 305/316, 6-302, 45%)	Sphaerobacter thermophilus (C4CNU7: 336/316, 21-309, 47%)	
hemY	Protoporphyrinogen oxidase	1.3.3.4	Nide1217	Bacillus subtilis (P32397: 470/484, 6-468, 35%)	Sphaerobacter thermophilus (C4CNU8: 481/484, 6-466, 46%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
hemN	putative Oxygen-independent coproporphyrinogen III oxidase	1.3.99.22	Nide1853	Mycobacterium tuberculosis (P71756: 375/392, 7-374, 37%)	Thermobaculum terrenum (C0UWN2: 389/392, 17-387, 37%)	
cbiX	Sirohydrochlorin cobaltochelatase	4.99.1.3	Nide2091	Bacillus megaterium (O87690: 306/393, 9-288, 33%)	Sorangium cellulosum (A9GQ78: 395/393, 5-395, 59%)	
cbiD	putative Cobalt-precorrin-6A synthase (deacetylating)	2.1.1.-	Nide2095	Leptospira interrogans serogroup Icterohaemorrhagiae serovar copenhageni (P61986: 367/368, 4-355, 59%)	Sorangium cellulosum (A9GQ64: 381/368, 6-363, 70%)	
cobH	Precorrin-8X methylmutase	5.4.1.2	Nide2096	Methanocaldococcus jannaschii (Q58340: 210/215, 32-209, 44%)	Sorangium cellulosum (A9GQ61: 249/215, 13-255, 77%)	
cbiET	probable cobalt-precorrin-6Y C(5,15)-methyltransferase (decarboxylating)	2.1.1.132	Nide2097	Pseudomonas denitrificans (P21921: 413/418, 16-412, 36%)	Sorangium cellulosum (A9GQ58: 442/418, 1-411, 73%)	
cbiL	Cobalt-precorrin-2 C(20)-methyltransferase	2.1.1.151	Nide2098	Salmonella typhimurium (Q05593: 237/254, 3-232, 36%)	Sorangium cellulosum (A9GQ55: 266/254, 5-249, 66%)	
cbiG	Cobalt-precorrin 5A hydrolase		Nide2099	Salmonella typhimurium (Q05631: 351/383, 8-334, 33%)	Sorangium cellulosum (A9GMA5: 379/383, 1-375, 68%)	presence of CbiG indicates anaerobic biosynthesis pathway
cobJ	Precorrin-3B C(17)-methyltransferase	2.1.1.131	Nide2100	Salmonella typhimurium (Q05590: 241/277, 1-240, 52%)	Sorangium cellulosum (A9GMA2: 281/277, 5-279, 80%)	
cobM	Precorrin-4 C(11)-methyltransferase	2.1.1.133	Nide2102	Pseudomonas aeruginosa (Q9HZP9: 250/263, 1-249, 55%)	Sorangium cellulosum (A9GM51: 264/263, 1-263, 73%)	
cobA/ hemD	Uroporphyrinogen-III C-methyltransferase and Uroporphyrinogen-III synthase	4.2.1.75, 2.1.1.107	Nide2103	Pseudomonas syringae pv. syringae (Q4ZRL6: 464/489, 213-460, 51%)	Desulfovibrio desulfuricans (Q30XG4: 503/489, 3-502, 36%)	modular protein
bluB	Cob(II)yrinic acid a,c-diamide reductase	1.16.8.1	Nide2104	Rhodobacter capsulatus (Q52685: 207/249, 8-201, 35%)	Nitrosopumilus maritimus (A9A2V5: 236/219, 7-219, 51%)	
cbiA	Cobyrinic acid a,c-diamide synthase	6.3.1.-	Nide2105	Leptospira interrogans serogroup Icterohaemorrhagiae serovar copenhageni (Q75FQ8: 545/458, 5-451, 53%)	Sorangium cellulosum (A9GM45: 500/458, 5-493,61 %)	
cbiP	Cobyric acid synthase	6.3.5.10	Nide2659	Syntrophobacter fumaroxidans (A0LJ24: 514/523, 4-511, 46%)	Geobacter metallireducens (Q39YE6: 797/523, 293-792, 49%)	
cbiZ	putative Adenosylcobinamide amidohydrolase	3.5.1.90	Nide2660	Methanocaldococcus jannaschii (Q59008: 255/242, 31-249, 29%)	Pelotomaculum thermopropionicum (A5D257: 247/242, 22-226, 37%)	
cobD	Threonine-phosphate decarboxylase	4.1.1.81	Nide2661	Salmonella typhimurium (P97084: 364/369, 8-359, 36%)	Geobacter sulfurreducens (Q748L2: 361/369, 7-357, 44%)	
cbiB	Cobalamin biosynthesis protein CbiB	6.3.1.10	Nide2662	Citrobacter koseri (A8AEP3: 319/318, 8-316, 46%)	Symbiobacterium thermophilum (Q67N31: 312/318, 13-309, 52%)	
cobS	Cobalamin synthase	2.7.8.26	Nide2663	Syntrophobacter fumaroxidans (A0LLI5: 248/254, 29-241, 39%)	Geobacter lovleyi (B3E976: 251/254, 2-240, 40%)	
cobT	Nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase	2.4.2.21	Nide2664	Geobacter sulfurreducens (Q748J3: 352/350, 4-351, 63%)	Geobacter sp. M21 (B3JVD3: 349/350, 4-348, 62%)	
cobU	Bifunctional adenosylcobalamin biosynthesis protein CobU	2.7.7.62, 2.7.1.156	Nide2665	Pseudomonas denitrificans (P29931: 174/177, 2-173, 43%)	Geobacter sulfurreducens (Q748J2: 172/177, 2-172, 47%)	
cobO	Cob(I)yrinic acid a,c-diamide adenosyltransferase	2.5.1.17	Nide2666	Citrobacter freundii (P45515: 176/211, 3-166, 43%)	Sorangium cellulosum (A9GX46: 237/211, 10-230, 46%)	
bluB	putative Cob(II)yrinic acid a,c-diamide reductase	1.16.8.1	Nide2667	Rhodobacter capsulatus (Q52685: 207/271, 3-203, 38%)	Azoarcus sp. EbN1 (Q5P2Q4: 250/271, 14-223, 60%)	
btuR	Cob(I)yrinic acid a,c-diamide adenosyltransferase	2.5.1.17	Nide2669	Pseudomonas aeruginosa (Q9I472: 203/199, 11-203, 58%)	Acaryochloris marina (B0C1C8: 214/199, 20-214, 59%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
hemB	Delta-aminolevulinic acid dehydratase	4.2.1.24	Nide2741	Aquifex aeolicus (O67876: 330/322, 6-325, 59%)	Rubrobacter xylanophilus (Q1AUK5: 326/327, 1-321, 64%)	
hemD	Porphyrin biosynthesis protein HemD	4.2.1.75, 2.1.1.107	Nide2743	Clostridium josui (Q59294: 504/514, 6-496, 39%)	Pelotomaculum thermopropionicum (A5D3L6: 511/514, 6-503, 55%)	
hemC	Porphobilinogen deaminase	2.5.1.61	Nide2744	Geobacter sulfurreducens (Q747I1: 318/294, 19-306, 60%)	Desulfovibrio piger (B6WTE9: 311/294, 18-309, 59%)	
hemA	Glutamyl-tRNA reductase	1.2.1.70	Nide2745	Geobacter lovleyi (B3E2H8: 434/468, 1-420, 51%)	Desulfuromonas acetoxidans (Q1JVU1: 434/468, 1-420, 49%)	
	putative Precorrin-2 dehydrogenase	1.3.1.76	Nide2764	Neisseria meningitidis serogroup C (A9LZ77: 480/217, 4-185, 32%)	Geobacter uraniireducens (A5GCW5: 249/217, 31-206, 36%)	
gltX	Glutamyl-tRNA synthetase	6.1.1.17	Nide4019	Geobacter sulfurreducens (Q74DU6: 466/471, 1-459, 58%)	Geobacter sp. FRC-32 (B9M2R8: 468/471, 1-462, 56%)	
hemD	putative Uroporphyrinogen III synthase HemD	4.2.1.75	Nide4176	Methanobacterium thermoautotrophicum (O26268: 253/280, 10-228, 24%)	Blastopirellula marina (A3ZTC1: 645/280, 9-282, 43%)	
hemL	Glutamate-1-semialdehyde-2,1-aminomutase	5.4.3.8	Nide4293	Geobacter uraniireducens (A5G9C0: 427/427, 1-427, 65%)	Desulfotomaculum reducens (A4J6H0: 432/427, 6-427, 62%)	
	Biotin biosynthesis					
bioB	Biotin synthase	2.8.1.6	Nide1095	Thermosynechococcus elongatus (Q8DL38: 360/344, 31-351, 55%)	Alicyclobacillus acidocaldarius LAA1 (B7DQ13: 333/344, 9-329, 57%)	
birA	putative Biotin-(acetyl-CoA-carboxylase) ligase	6.3.4.15	Nide1731	Bacillus subtilis (P42975: 325/290, 67-308, 36%)	Anaeromyxobacter sp. Fw109-5 (A7HAE0: 329/290, 68-310, 40%)	
bioD	Dethiobiotin synthetase	6.3.3.3	Nide2471	Anaeromyxobacter dehalogenans (B8J7V5: 223/254, 3-180, 48%)	Myxococcus xanthus (Q1DCV7: 227/254, 10-209, 46%)	
bioA	Adenosylmethionine-8-amino-7-oxononanoate transaminase	2.6.1.62	Nide2550	Bacillus subtilis (P53555: 448/455, 12-444, 54%)	Candidatus Chloracidobacterium thermophilum (A8DJL1: 461/455, 15-459, 59%)	
bioF	8-amino-7-oxononanoate synthase	2.3.1.47	Nide3447	Geobacter sulfurreducens (Q749W3: 391/386, 9-391, 49%)	Candidatus Kuenenia stuttgartiensis (Q1Q6F5: 391/386, 10-383, 48%)	
bioH	putative Carboxylesterase BioH	3.1.1.1	Nide3554	Acinetobacter sp. ADP1 (P00632: 267/267, 30-264, 30%)	Geobacter sulfurreducens (Q747V8: 266/267, 7-263, 42%)	carries out reaction of 6.2.1.14
	Riboflavin biosynthesis					
ribBA	Riboflavin biosynthesis protein RibBA	3.5.4.25, 4.1.99.12	Nide1383	Syntrophomonas wolfei subsp. wolfei (Q0AXM5: 398/402, 2-395, 63%)	Geobacter sp. FRC-32 (B9M1E5: 400/402, 6-400, 67%)	
ribH	6,7-dimethyl-8-ribityllumazine synthase	2.5.1.9	Nide1384	Geobacter lovleyi (B3E758: 155/180, 3-155, 57%)	Sulfurihydrogenibium azorense (B3U4H4: 154/180, 1-151, 56%)	Riboflavin synthase, beta subunit
ribD	fused Diaminohydroxyphosphoribosylamino-pyrimidine deaminase / 5-amino-6-(5-phosphoribosylamino)uracil reductase	1.1.1.193, 3.5.4.26	Nide1635	Aquifex aeolicus (O66534: 356/365, 10-351, 44%)	Pelobacter carbinolicus (Q3A4L6: 371/365, 9-367, 49%)	
ribE	Riboflavin synthase, alpha subunit	2.5.1.9	Nide1641	Actinobacillus pleuropneumoniae (P50854: 215/218, 1-214, 44%)	Pelotomaculum thermopropionicum (A5D1C9: 216/218, 1-216, 50%)	
	putative Acid phosphatase, class B	3.1.3.2	Nide2215	Haemophilus influenzae (P26093: 274/240, 47-248, 30%)	Rhodothermus marinus (C1ZSU9: 264/240, 36-258, 47%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
cobT	Nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase	2.4.2.21	Nide2664	Geobacter sulfurreducens (Q748J3: 352/350, 4-351, 63%)	Geobacter sp. M21 (B3JVD3: 349/350, 4-348, 62%)	
ribF	Riboflavin biosynthesis protein RibF	2.7.1.26, 2.7.7.2	Nide2740	Escherichia coli O6 (P0AG41: 313/321, 1-304, 38%)	Syntrophobacter fumaroxidans (A0LN53: 312/321, 1-308, 44%)	
ppa	Inorganic pyrophosphatase	3.6.1.1	Nide3643	Chlamydophila caviae (Q821T4: 216/227, 11-209, 55%)	Stigmatella aurantiaca DW4/3-1 (Q09BP6: 222/227, 11-206, 58%)	can also have function of EC 3.1.3.1
	putative Acid phosphatase	3.1.3.2	Nide3744	Penicillium chrysogenum (P37274: 421/291, 75-301, 29%)	Catenulispora acidiphila (C1QQC5: 292/291, 19-292, 49%)	
	Thiamine biosynthesis					
apbE	putative Thiamine biosynthesis lipoprotein ApbE		Nide1029	Treponema pallidum (O83774: 362/365, 34-355, 34%)	Sorangium cellulosum (A9FZX7: 381/365, 55-373, 41%)	
thi4	putative Thiazole biosynthesis enzyme		Nide1062	Haloquadratum walsbyi (Q18KP1: 307/266, 7-290, 42%)	uncultured marine crenarchaeote HF4000_APKG3K8 (B3T7X2: 278/266, 15-268, 47%)	
thiC	Thiamine biosynthesis protein		Nide1063	Dechloromonas aromatica (Q478T2: 638/638, 21-635, 77%)	Mariprofundus ferrooxydans PV-1 (Q0F0G9: 622/638, 21-619, 76%)	
thiI	putative Thiamine biosynthesis protein ThiI		Nide1125	Thermus thermophilus (Q72HP1: 406/392, 10-387, 46%)	Thermus aquaticus Y51MC23 (B7A8B1: 406/392, 9-386, 45%)	
thiD	Hydroxymethylpyrimidine / phosphomethylpyrimidine kinase	2.7.1.49, 2.7.4.7	Nide1195	Haemophilus influenzae (P44697: 269/264, 4-266, 49%)	Rhodothermus marinus (C1ZNM9: 285/264, 10-266, 61%)	
tenA	putative Thiaminase	3.5.99.2	Nide1200	Bacillus subtilis (P25052: 236/222, 1-218, 35%)	Thermobaculum terrenum (C0UTR9: 230/222, 10-230, 49%)	
thiF	Adenylyltransferase ThiF	2.7.7.-	Nide1820	Sus scrofa (A5GFZ6: 455/275, 54-307, 50%)	Leptospirillum rubarum (A3ERN6: 270/275, 1-267, 63%)	
thiS	putative Thiamine biosynthesis protein ThiS		Nide1822	Magnetococcus sp. MC-1 (A0L5E8: 326/66, 1-67, 30%)	Clostridium kluyveri (B9E2K3: 71/66, 1-67, 45%)	
	Nucleoside-triphosphatase	3.6.1.15	Nide1872	Thermodesulfovibrio yellowstonii (B5YHP2: 204/204, 1-192, 50%)	Desulfotomaculum acetoxidans (C1TEM3: 202/204, 4-199, 51%)	
thiE	Thiamine-phosphate pyrophosphorylase	2.5.1.3	Nide2470	Moorella thermoacetica (Q2RGI8: 210/204, 18-204, 48%)	Chloroflexus aggregans (B8G7I9: 221/204, 28-209, 47%)	
	putative L-2-hydroxyglutarate oxidase	1.1.99.2	Nide3018	Nematostella vectensis (A7SMW7: 456/400, 38-448, 36%)	Thermodesulfovibrio yellowstonii (B5YIC0: 394/400, 2-385, 44%)	could function as thiO, EC 1.4.3.19
thiL	Thiamine-monophosphate kinase	2.7.4.16	Nide3047	Buchnera aphidicola subsp. Acyrthosiphon pisum (P57532: 323/349, 4-305, 36%)	Pelobacter carbinolicus (Q3A8K1: 329/349, 8-327, 42%)	
thiE	Thiamine-phosphate pyrophosphorylase	2.5.1.3	Nide3059	Geobacter sulfurreducens (P61411: 213/208, 1-206, 47%)	Geobacter metallireducens (Q39RH1: 213/208, 1-208, 49%)	
thiG	Thiazole biosynthesis protein ThiG		Nide3060	Thermodesulfovibrio yellowstonii (B5YJ73: 257/259, 1-256, 79%)	Geobacter sp. FRC-32 (B9M5W4: 259/259, 4-258, 76%)	
thiS	Thiamine biosynthesis protein ThiS		Nide3061	Magnetospirillum magneticum (Q2W3R9: 324/68, 5-67, 44%)	Pelobacter propionicus (A1AUP6: 66/68, 1-66, 48%)	
	putative Thiamine biosynthesis protein ThiS		Nide3395	Trypanosoma brucei brucei (P16355: 1744/66, 785-835, 31%)	Thermoanaerobacter sp. X514 (B0K6H9: 64/66, 9-64, 45%)	
iscS	Cysteine desulfurase	2.8.1.7	Nide4024	Anaeromyxobacter sp. Fw109-5 (A7H804: 404/405, 1-401, 62%)	Acidobacterium capsulatum (C1F1Z6: 428/405, 25-428, 66%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
yjeF	conserved protein of unknown function, putative Carbohydrate kinase		Nide4119	Escherichia coli K12 (P31806: 515/516, 41-490, 36%)	Moorella thermoacetica (Q2RGI2: 530/516, 3-512, 46%)	related to Hydroxyethylthiazole kinase, could function as EC 2.7.1.50
	Folate biosynthesis					
folK	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase	2.7.6.3	Nide0502	Schizosaccharomyces pombe (Q4LB35: 686/187, 251-403, 41%)	Thermus thermophilus (Q72HU7: 159/187, 2-155, 51%)	
	putative Dihydroneopterin aldolase and SAM-dependent methyltransferase	4.1.2.25	Nide1532	Streptococcus pyogenes serotype M1 (P0C0G5: 119/309, 3-119, 32%)	Candidatus Kuenenia stuttgartiensis (Q1PY54: 222/309, 53-204, 37%)	modular protein
folC	Bifunctional protein FolC: Folylpolyglutamate synthase and Dihydrofolate synthase	6.3.2.17, 6.3.2.12	Nide1579	Bacillus subtilis (Q05865: 430/431, 5-430, 35%)	Desulfococcus oleovorans (A8ZTL3: 434/431, 7-425, 43%)	
	Dihydrofolate reductases superfamily enzyme		Nide2810	Bacillus subtilis (P45862: 174/184, 5-164, 31%)	Geobacter bemidjiensis (B5E9R1: 187/184, 1-181, 68%)	could function as EC 1.5.1.3
folP	Dihydropteroate synthase	2.5.1.15	Nide2736	Haemophilus influenzae (P43776: 275/269, 18-271, 48%)	Halorhodospira halophila (A1WXX1: 286/269, 25-278, 56%)	
folE	GTP cyclohydrolase I	3.5.4.16	Nide3555	Chlorobium tepidum (Q8KEA8: 224/211, 7-221, 56%)	Solibacter usitatus (Q028J6: 194/211, 6-191, 61%)	
ptpS	6-pyruvoyl-tetrahydropterin synthase	4.2.3.12	Nide3556	Rattus norvegicus (P27213: 144/149, 10-143, 49%)	Geobacillus kaustophilus (Q5KZY1: 134/149, 1-129, 53%)	
pabB	Aminodeoxychorismate synthase, component I	2.6.1.85	Nide3736	Bacillus subtilis (P28820: 470/477, 24-461, 43%)	Moorella thermoacetica (Q2RGN9: 466/477, 4-465, 47%)	
pabC	putative Aminodeoxychorismate lyase	4.1.3.38	Nide3737	Methanocaldococcus jannaschii (Q58414: 288/293, 1-281, 39%)	Blastopirellula marina (A3ZV31: 286/293, 6-280, 46%)	
ptpS	putative 6-pyruvoyl-tetrahydropterin synthase	4.2.3.12	Nide4038	Rattus norvegicus (P27213: 144/250, 13-136, 44%)	Roseiflexus sp. RS-1 (A5UR67: 275/250, 2-268, 40%)	modular protein , contains 2 pts domains
	Coenzyme A biosynthesis					
panC	Pantoate--beta-alanine ligase	6.3.2.1	Nide0501	Heliobacterium modesticaldum (B0TBP5: 281/285, 1-278, 51%)	Sphaerobacter thermophilus (C4CPL0: 299/285, 20-294, 54%)	
panB	3-methyl-2-oxobutanoate hydroxymethyltransferase	2.1.2.11	Nide0812	Pelobacter propionicus (A1AUV8: 269/262, 8-269, 56%)	Thermodesulfovibrio yellowstonii (B5YIG6: 264/262, 4-264, 59%)	
ilvI	Acetolactate synthase, large subunit	2.2.1.6	Nide0997	Haemophilus influenzae (P45261: 573/592, 3-570, 52%)	Thermodesulfovibrio yellowstonii (B5YFP1: 576/592, 3-574, 61%)	
ilvH	Acetolactate synthase, small subunit	2.2.1.6	Nide0998	Aquifex aeolicus (O67703: 192/172, 27-183, 55%)	Desulfuromonas acetoxidans (Q1K337: 165/172, 1-158, 68%)	
ilvC	Ketol-acid reductoisomerase	1.1.1.86	Nide0999	Geobacter sulfurreducens (Q74BW9: 338/337, 1-338, 75%)	Leptospirillum sp. Group II '5-way CG' (B6AN25: 339/337, 4-339, 71%)	
ilvE	Branched-chain amino acid aminotransferase	2.6.1.42	Nide1103	Pseudomonas aeruginosa (O86428: 307/304, 10-306, 56%)	Thermodesulfovibrio yellowstonii (B5YL62: 304/304, 1-304, 61%)	
coaE	Dephospho-CoA kinase	2.7.1.24	Nide1128	Geobacter metallireducens (Q39R83: 201/200, 1-176, 52%)	Heliobacterium modesticaldum (B0TES0: 200/200, 1-192, 48%)	
coaX	Type III pantothenate kinase	2.7.1.33	Nide1730	Pelotomaculum thermopropionicum (A5D5P9: 256/256, 1-253, 55%)	Desulforudis audaxviator (B1I1G9: 262/256, 7-258, 55%)	
panE	putative 2-dehydropantoate 2-reductase (modular protein)	1.1.1.169	Nide1738	Anabaena sp. PCC 7120 (Q8YX96: 319/4554, 4-313, 28%)	alpha proteobacterium BAL199 (A8TJ04: 302/454, 2-291, 33%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
ilvB	putative Acetolactate synthase, large subunit	2.2.1.6	Nide2332	Bacillus subtilis (P37251: 574/612, 31-548, 32%)	Magnetospirillum magneticum (Q2W979: 613/612, 1-608, 71%)	
coaBC	Coenzyme A biosynthesis bifunctional protein CoaBC	4.1.1.36, 6.3.2.5	Nide2464	Mycobacterium tuberculosis (P67733: 418/419, 10-415, 45%)	Desulfuromonas acetoxidans (Q1K0Z1: 405/419, 6-400, 52%)	
coaD	Phosphopantetheine adenylyltransferase	2.7.7.3	Nide2566	Thermoanaerobacter tengcongensis (Q8R9U9: 160/162, 1-158, 56%)	Azotobacter vinelandii (C1DIB2: 159/162, 1-159, 54%)	
panD	Aspartate 1-decarboxylase	4.1.1.11	Nide2919	Ralstonia solanacearum (Q8XVU6: 120/121, 1-114, 60%)	Leptospirillum rubarum (A3EU20: 121/121, 1-114, 66%)	
ilvD	Dihydroxy-acid dehydratase	4.2.1.9	Nide3406	Acidobacteria bacterium Ellin345 (Q1ILZ0: 573/557, 8-559, 66%)	Acidobacterium capsulatum (C1F6Z8: 562/557, 16-561, 65%)	
Chemotaxis and motility						
	Flagellum biosynthesis					
fleQ	Sigma-54 dependent response regulator FleQ		Nide2280	Escherichia coli K12 (Q06065: 461/471, 5-458, 44%)	Syntrophus aciditrophicus (Q2LSK2: 472/471, 12-465, 53%)	
fleS	putative Sensor histidine kinase FleS		Nide2283	Escherichia coli K12 (Q06067: 608/429, 215-599, 29%)	Syntrophus aciditrophicus (Q2LRY6: 415/429, 2-411, 33%)	
fleR	Sigma-54 dependent response regulator FleR		Nide2284	Escherichia coli K12 (Q06065: 461/472, 7-454, 46%)	Syntrophus aciditrophicus (Q2LRZ0: 473/472, 8-465, 49%)	
flgB	Flagellar basal-body rod protein FlgB		Nide2285	Yersinia enterocolitica (Q56893: 137/135, 6-137, 33%)	Geobacter sp. M21 (B3JZF7: 139/135, 6-136, 41%)	
flgC	Flagellar basal-body rod protein FlgC		Nide2286	Borrelia burgdorferi (Q57466: 152/149, 1-149, 44%)	Geobacter metallireducens (Q39QZ6: 147/149, 1-145, 49%)	
fliE	Flagellar hook-basal body complex protein FliE		Nide2287	Geobacter lovleyi (B3EBH2: 99/103, 29-99, 45%)	Leptospirillum sp. Group II '5-way CG' (B6AMC8: 109/103, 11-107, 43%)	
fliF	Flagellar M-ring protein FliF		Nide2288	Salmonella typhimurium (P15928: 560/518, 18-558, 36%)	Geobacter uraniireducens (A5G988: 525/518, 19-523, 43%)	
fliG	Flagellar motor switch protein FliG		Nide2289	Thermotoga maritima (Q9WY63: 335/331, 4-334, 40%)	Thioalkalivibrio sp. HL-EbGR7 (B8GT34: 338/331, 7-338, 41%)	
fliH	putative Flagellar biosynthesis protein FliH with response regulator receiver domain		Nide2290	Rhizobium meliloti (P13632: 460/349, 3-134, 35%)	Roseovarius nubinhibens ISM (A3SQ96: 446/349, 1-150, 37%)	modular protein
fliI	ATPase FliI	3.6.3.14	Nide2291	Bacillus subtilis (P23445: 438/435, 6-435, 52%)	Pelobacter carbinolicus (Q3A5B9: 433/435, 8-428, 59%)	
fliJ	putative Flagellar biosynthesis protein FliJ		Nide2292	Sarcoptes scabiei (Q9BMM8: 876/146, 527-641, 25%)	Geobacter sulfurreducens (Q74G35: 146/146, 15-143, 30%)	
FliK	putative Flagellar hook-length control protein FliK		Nide2295	Saccharomyces cerevisiae (P08640: 1367/524, 326-678, 22%)	Rhodopseudomonas palustris (Q131Q7: 526/524, 8-501, 24%)	
flgD	Flagellar hook capping protein FlgD		Nide2296	Bacillus subtilis (P23455: 140/101, 11-79, 43%)	Heliobacterium modesticaldum (B0THA9: 188/101, 16-84, 55%)	fragment
flgE	Flagellar hook protein FlgE		Nide2297	Salmonella typhimurium (P0A1J1: 403/407, 2-403, 38%)	Thermodesulfovibrio yellowstonii (B5YIG4: 434/407, 1-434, 42%)	
fliL	Flagellar basal body-associated protein FliL		Nide2298	Aquifex aeolicus (O67712: 161/178, 2-159, 30%)	Geobacter bemidjiensis (B5EEN4: 174/178, 2-174, 40%)	
fliM	Flagellar motor switch protein FliM		Nide2299	Pseudomonas aeruginosa (Q51465: 323/323, 6-315, 43%)	Syntrophus aciditrophicus (Q2LT03: 333/323, 1-322, 43%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
fliN	Flagellar motor switch protein FliN		Nide2300	Pseudomonas aeruginosa (Q51466: 157/124, 34-156, 51%)	Desulfomicrobium baculatum (C1SZL7: 188/124, 104-188, 73%)	
fliO	putative Flagellar biosynthetic protein FliO, export component		Nide2301	Pseudomonas aeruginosa (Q51467: 150/122, 44-148, 31%)	Syntrophus aciditrophicus (Q2LT01: 130/122, 4-109, 25%)	
fliP	Flagellar biosynthetic protein FliP, export component		Nide2302	Bacillus subtilis (P35528: 221/266, 20-221, 56%)	Geobacter uraniireducens (A5G974: 251/266, 4-250, 57%)	
fliQ	Flagellar biosynthetic protein FliQ, export component		Nide2303	Escherichia coli O6 (P0AC08: 89/89, 1-88, 41%)	Thermosinus carboxydivorans Nor1 (A1HN23: 89/89, 1-88, 52%)	
fliR	Flagellar biosynthetic protein FliR, export component		Nide2304	Bacillus subtilis (P35537: 259/263, 8-257, 31%)	Desulfotalea psychrophila (Q3V7H4: 260/263, 4-258, 36%)	
flhB	Flagellar biosynthetic protein FlhB, export component		Nide2305	Bacillus subtilis (P35538: 360/360, 14-359, 41%)	Syntrophus aciditrophicus (Q2LT07: 354/360, 1-352, 44%)	
flhA	Flagellar biosynthetic protein FlhA, export component		Nide2306	Escherichia coli K12 (P76298: 692/702, 31-689, 51%)	Syntrophus aciditrophicus (Q2LT06: 691/702, 12-691, 56%)	
flhF	putative Flagellar biosynthetic protein FlhF, GTP binding		Nide2307	Bacillus subtilis (Q01960: 366/427, 1-352, 29%)	Geobacter sulfurreducens (Q3V8C7: 451/427, 1-447, 32%)	
fleN	Flagellar number regulator FleN		Nide2308	Bacillus subtilis (P40742: 298/288, 30-271, 33%)	Pelobacter propionicus (A1AUL8: 308/288, 23-302, 49%)	
fliA	Flagellar-specific RNA polymerase sigma 28 factor		Nide2309	Streptomyces coelicolor (P17211: 280/247, 47-275, 41%)	Bdellovibrio bacteriovorus (Q6MI54: 261/247, 16-249, 45%)	
flgF	Flagellar basal-body rod protein FlgF		Nide2311	Buchnera aphidicola subsp. Schizaphis graminum (Q8K9K4: 260/255, 10-259, 33%)	Geobacter sp. FRC-32 (B9M0F5: 244/255, 1-244, 43%)	
flgG	Flagellar basal-body rod protein FlgG		Nide2312	Salmonella typhi (P0A1J4: 260/263, 1-255, 53%)	Geobacter sp. FRC-32 (B9M0F6: 262/263, 1-262, 56%)	
flgA	putative Flagella basal body P-ring formation protein FlgA		Nide2313	Agrobacterium tumefaciens (Q44339: 162/265, 39-159, 28%)	Magnetococcus sp. MC-1 (A0LC96: 236/265, 63-231, 34%)	
flgH	Flagellar L-ring protein FlgH		Nide2314	Geobacter lovleyi (B3EAV2: 224/238, 3-224, 41%)	Geobacter sp. M21 (B3JV79: 229/238, 14-229, 43%)	
flgI	Flagellar P-ring protein FlgI		Nide2315	Desulfotalea psychrophila (Q6AJR7: 390/436, 26-390, 47%)	Pelobacter propionicus (A1AUM4: 364/436, 5-364, 49%)	
flgJ	putative Peptidoglycan hydrolase FlgJ		Nide2316	Escherichia coli K12 (P75942: 313/109, 22-90, 30%)	Desulfovibrio desulfuricans (B8J2Y0: 560/109, 48-105, 50%)	fragment
flgM	Anti-sigma 28 factor FlgM		Nide2317	Bacillus subtilis (P39809: 88/101, 11-82, 28%)	Desulfotomaculum acetoxidans (C1TAV4: 97/101, 1-94, 34%)	
flgN	putative Flagellar chaperone FlgN		Nide2318	Emericella nidulans (Q5B797: 487/128, 380-483, 26%)	Variovorax paradoxus S110 (C5CXV9: 148/128, 15-144, 30%)	
flgK	First flagellar hook-filament junction protein FlgK		Nide2319	Bacillus subtilis (P39810: 507/464, 14-499, 29%)	Geobacter uraniireducens (A5G8X4: 478/464, 3-476, 35%)	
flgL	putative Second flagellar hook-filament junction protein FlgL		Nide2320	Bacillus subtilis (P96501: 298/296, 6-298, 29%)	Geobacter metallireducens (Q39YJ2: 294/296, 1-294, 34%)	
csrA	Carbon storage regulator		Nide2321	Geobacter uraniireducens (A5G8X2: 79/81, 1-67, 54%)	Providencia alcalifaciens (B6XBP0: 61/81, 1-58, 62%)	can also regulate swarming
fliW	Flagellar assembly factor FliW		Nide2322	Natranaerobius thermophilus (B2A827: 151/158, 1-141, 40%)	Desulfatibacillum alkenivorans (B8FK12: 148/158, 3-135, 41%)	
fliC	Flagellin, class B		Nide2339	Bacillus halodurans (Q05203: 272/275, 1-272, 45%)	Geobacter sulfurreducens (Q748G4: 276/275, 1-276, 54%)	
flaG	Flagellar protein FlaG		Nide2340	Vibrio parahaemolyticus (Q56704: 144/119, 32-141, 28%)	Geobacter metallireducens (Q39YG8: 124/119, 12-124, 35%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
fliD	putative Flagellar capping protein FliD		Nide2341	Escherichia coli K12 (P24216: 468/463, 3-457, 29%)	Thauera sp. MZ1T (C4KAR2: 447/463, 1-438, 32%)	
fliS	Flagellin-specific chaperone FliS		Nide2342	Pseudomonas aeruginosa (Q9I4N6: 126/130, 6-125, 34%)	Geobacter metallireducens (Q39YI6: 140/130, 6-119, 38%)	
motA	Flagellar motor protein MotA		Nide2356	Helicobacter pylori (P65410: 257/256, 1-252, 37%)	Anaeromyxobacter dehalogenans (Q2IQS4: 261/256, 1-258, 54%)	
motB	Flagellar motor protein MotB		Nide2357	Bacillus subtilis (P28612: 261/259, 6-257, 35%)	Dechloromonas aromatica (Q47I30: 257/259, 4-242, 45%)	
	Chemotaxis					
cheR	Chemotaxis protein methyltransferase CheR	2.1.1.80	Nide2347	Vibrio cholerae (Q9KS61: 288/287, 6-281, 35%)	Thermodesulfovibrio yellowstonii (B5YIY3: 283/287, 12-282, 48%)	
cheY	Chemotaxis regulator CheY		Nide2348	Escherichia coli O6 (Q8FGP6: 129/120, 6-125, 68%)	Hydrogenivirga sp. 128-5-R1-1 (A8UVU3: 127/120, 8-127, 66%)	
cheZ	putative Chemotaxis regulator CheZ		Nide2349	Nide2349 (P0A9I0: 214/216, 21-214, 25%)	Hydrogenobaculum sp. Y04AAS1 (B4U6X3: 186/216, 6-185, 27%)	
cheA	Chemotaxis protein histidine kinase CheA	2.7.13.3	Nide2350	Thermotoga maritima (Q56310: 671/630, 256-671, 45%)	Leptospirillum sp. Group II '5-way CG' (B6AMA5: 646/630, 7-633, 53%)	
cheV	Chemotaxis signal transducer CheV		Nide2351	Bacillus subtilis (P37599: 303/327, 12-301, 31%)	Methylophaga thiooxidans (C0N2Q8: 318/327, 1-313, 46%)	
cheB	Chemotaxis methylesterase CheB		Nide2352	Desulfovibrio vulgaris (P62637: 367/365, 2-363, 47%)	Thermodesulfovibrio yellowstonii (B5YIX8: 351/365, 2-351, 53%)	
mcp	putative Methyl-accepting chemotaxis protein		Nide2353	Vibrio cholerae (P15492: 548/561, 205-548, 35%)	Geobacter lovleyi (B3EA57: 533/561, 122-533, 42%)	
cheW	Chemotaxis signal transducer CheW		Nide2354	Escherichia coli O157:H7 (P0A966: 167/1172, 18-157, 38%)	Desulfovibrio magneticus RS-1 (C4XS46: 158/172, 6-151, 53%)	
mcp	putative Methyl-accepting chemotaxis protein		Nide2359	Bacillus subtilis (P39214: 661/371, 356-620, 25%)	Denitrovibrio acetiphilus (C1SKD0: 400/371, 60-400, 31%)	
cheY	Chemotaxis regulator CheY		Nide2360	Helicobacter pylori (P71403: 124/123, 3-120, 42%)	Geobacter metallireducens (Q39SX1: 128/123, 8-127, 68%)	
cheA	Chemotaxis protein histidine kinase CheA	2.7.13.3	Nide2361	Salmonella typhimurium (P09384: 671/715, 1-662, 42%)	Bermanella marisrubri (Q1N2W7: 704/715, 1-689, 44%)	
cheW	Chemotaxis signal transduction protein CheW		Nide2362	Shigella flexneri (P0A967: 167/169, 16-164, 48%)	Halorhodospira halophila (A1WZ18: 205/169, 32-181, 55%)	
cheM	Globin-coupled methyl-accepting chemotaxis protein		Nide2363	Salmonella typhimurium (P02941: 553/772, 230-540, 52%)	Rhodopseudomonas palustris (Q07SD7: 910/772, 381-906, 51%)	modular protein
cheR	Chemotaxis protein methyltransferase CheR	2.1.1.80	Nide2365	Pseudomonas aeruginosa (Q9I6V7: 280/272, 14-277, 43%)	Candidatus Kuenenia stuttgartiensis (Q1PXN4: 287/272, 1-277, 47%)	
cheD	Chemoreceptor glutamine deamidase CheD	3.5.1.44	Nide2366	Halorhodospira halophila (A1WZ13: 216/205, 20-216, 52%)	Bermanella marisrubri (Q1N2W3: 210/205, 11-199, 47%)	
cheB	Chemotaxis methylesterase CheB	3.1.1.61	Nide2367	Nitrosomonas europaea (Q820K0: 358/349, 1-351, 60%)	Candidatus Kuenenia stuttgartiensis (Q1PXP9: 346/349, 4-343, 62%)	
	Histidine kinases					
	putative Histidine kinase with N-terminal NAD-binding region		Nide0166	Bacillus subtilis (P13799: 385/363, 176-384, 37%)	Thermosinus carboxydivorans Nor1 (A1HPR4: 377/373, 175-376, 41%)	
	putative Hybrid histidine kinase	2.7.13.3	Nide0222	Pseudomonas syringae pv. phaseolicola (Q48IV1: 534/545, 139-528, 37%)	Acidobacteria bacterium Ellin345 (Q1IM85: 1132/545, 745-1126, 55%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	putative Sensor histidine kinase	2.7.13.3	Nide0224	Anabaena sp. PCC 7120 (Q9LCC2: 765/499, 491-746, 37%)	bacterium Ellin514 (B9XNB7: 468/499, 13-457, 40%)	
	putative Sensor histidine kinase	2.7.13.3	Nide0241	Agrobacterium tumefaciens (P18540: 833/532, 451-695, 32%)	Cyanothece sp. PCC 7425 (B8HRN4: 1428/532, 783-4301, 31%)	
cztS	Heavy metal sensor histidine kinase	2.7.13.3	Nide0304	Ralstonia metallidurans (Q44007: 476/476, 10-454, 28%)	Geobacter metallireducens (Q39Q80: 470/467, 6-466, 44%)	
	putative Histidine kinase	2.7.13.3	Nide0542	Escherichia coli K12 (P14377: 465/910, 232-455, 33%)	Pelobacter carbinolicus (Q3A202: 1154/910, 399-1146, 25%)	
	putative Histidine kinase	2.7.13.3	Nide0543	Caulobacter crescentus (P37894: 842/771, 542-828, 33%)	Chloroflexus aggregans (B8G4L8: 875/771, 351-858, 32%)	
	putative Histidine kinase	2.7.13.3	Nide0557	Shigella flexneri (P0AEC4: 778/949, 276-645, 34%)	Acaryochloris marina (B0C3R5: 551/949, 105-547, 33%)	
	putative Sensor histidine kinase		Nide0578	Bradyrhizobium japonicum (P15939: 889/1332, 112-661, 27%)	Methylobacterium populi (B1ZBV9: 1002/1332, 10-806, 31%)	
	putative Histidine kinase	2.7.13.3	Nide0584	Dictyostelium discoideum (Q54YZ9: 2062/954, 1336-1848, 41%)	Desulfatibacillum alkenivorans (B8FJD5: 1104/954, 198-986, 41%)	
	putative Histidine kinase	2.7.13.3	Nide0585	Dictyostelium discoideum (Q54YZ9: 2062/1610, 1336-1848, 43%)	Chthoniobacter flavus Ellin428 (B4CU67: 1248/1610, 47-1124, 36%)	
	putative Histidine kinase	2.7.13.3	Nide0686	Pseudomonas syringae pv. tomato (Q881J7: 534/922, 140-513, 32%)	Chthoniobacter flavus Ellin428 (B4D7A5: 733/922, 305-732, 49%)	
	putative Histidine kinase	2.7.13.3	Nide0694	Escherichia coli K12 (P14377: 465/610, 236-455, 30%)	Geobacter uraniireducens (A5G4Y5: 544/610, 313-543, 37%)	
	putative Sensor protein PilS	2.7.13.3	Nide0778	Escherichia coli K12 (Q06067: 608/619, 239-599, 33%)	Anaeromyxobacter dehalogenans (B8JCY8: 525/619, 3-525, 37%)	
	putative Hybrid histidine kinase	2.7.13.3	Nide0804	Dictyostelium discoideum (Q54YZ9: 2062/415, 1320-1582, 46%)	Microcoleus chthonoplastes PCC 7420 (B4W4M9: 684/415, 30-395, 44%)	
	putative Hybrid histidine kinase	2.7.13.3	Nide0805	Dictyostelium discoideum (Q86CZ2: 1213/1210, 794-1203, 37%)	Desulfomicrobium baculatum (C1T1Z8: 1131/1210, 467-1125, 38%)	
	putative Histidine kinase, containing PAS domain S-box	2.7.13.3	Nide0807	Escherichia coli K12 (P0AEC5: 918/578, 243-519, 49%)	Rubrobacter xylanophilus (Q1AWJ6: 1069/578, 175-576, 45%)	
	putative Hybrid histidine kinase	2.7.13.3	Nide0815	Dictyostelium discoideum (Q54YZ9: 2062/1431, 1326-1848, 43%)	Chthoniobacter flavus Ellin428 (B4CU67: 1248/1431, 187-1242, 35%)	
	putative Hybrid histidine kinase	2.7.13.3	Nide0816	Shigella flexneri (P59342: 918/945, 103-892, 32%)	Thauera sp. MZ1T (C4KD25: 921/945, 147-918, 40%)	
	putative Phosphate regulon sensor histidine kinase PhoR	2.7.13.3	Nide0857	Bacillus subtilis (P23545: 579/609, 58-578, 33%)	Geobacter uraniireducens (A5GEB2: 592/619, 1-591, 38%)	
	putative Sensor histidine kinase		Nide0936	Brevibacillus brevis (P54663: 386/511, 131-385, 32%)	bacterium Ellin514 (B9XSI9: 656/511, 406-642, 51%)	
	putative Histidine kinase, contains GAF domain	2.7.13.3	Nide0964	Escherichia coli K12 (P14377: 465/1002, 226-457, 39%)	Myxococcus xanthus (Q1D732: 747/1002, 146-707, 31%)	
	putative Hybrid histidine kinase	2.7.13.3	Nide1046	Dictyostelium discoideum (Q54YZ9: 2062/1340, 1328-1847, 42%)	Candidatus Kuenenia stuttgartiensis (Q1Q3E7: 967/1340, 297-966, 44%)	
	putative Sensor histidine kinase	2.7.13.3	Nide1109	Bacillus subtilis (P23545: 579/768, 248-571, 26%)	Geobacillus sp. Y412MC52 (C3J5Z3: 470/768, 64-462, 28%)	
	putative Histidine kinase	2.7.13.3	Nide1119	Escherichia coli K12 (P14377: 465/645, 213-452, 34%)	Anaeromyxobacter sp. K (B4UB40: 1527/645, 1265-1507, 40%)	
	putative Histidine kinase		Nide1163	Brevibacillus brevis (P54663: 386/351, 128-382, 32%)	Burkholderia vietnamiensis (A4JL87: 598/351, 322-580, 39%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	Sensor histidine kinase	2.7.13.3	Nide1222	Azorhizobium caulinodans (Q04850: 771/765, 29-715, 29%)	Geobacter uraniireducens (A5GA24: 746/765, 3-741, 41%)	
	Histidine kinase	2.7.13.3	Nide1227	Dictyostelium discoideum (Q54YZ9: 2062/589, 1336-1847, 40%)	Heliobacterium modesticaldum (B0TG16: 909/589, 366-898, 52%)	
	putative Histidine kinase	2.7.13.3	Nide1255	Escherichia coli K12 (Q06067: 608/383, 375-599, 31%)	Myxococcus xanthus (Q1D4I6: 702/383, 455-690, 38%)	
	putative Hybrid sensor histidine kinase		Nide1491	Brevibacillus brevis (P54663: 386/998, 182-382, 33%)	bacterium Ellin514 (B9XSJ3: 795/998, 127-769, 34%)	
	putative Sensor histidine kinase with MCP region	2.7.13.3	Nide1493	Anabaena sp. PCC 7120 (Q9LCC2: 765/590, 484-761, 56%)	uncultured archaeon GZfos26B2 (Q64C59: 1172/590, 917-1163, 62%)	
	putative Sensor histidine kinase CusS	2.7.13.3	Nide1563	Burkholderia pseudomallei (O31396: 464/481, 4-460, 30%)	Geobacter metallireducens (Q39Q80: 470/481, 16-466, 32%)	
	putative Histidine kinase, contains GAF domain	2.7.13.3	Nide1586	Escherichia coli K12 (P14377: 465/405, 236-455, 39%)	Myxococcus xanthus (Q1D732: 747/405, 289-692, 30%)	
	putative Sensory histidine kinase with methyltransferase region		Nide1587	Bradyrhizobium japonicum (P15939: 889/1396, 490-809, 22%)	Desulfuromonas acetoxidans (Q1JWS1: 745/1396, 29-335, 29%)	
	putative Sensor histidine kinase	2.7.13.3	Nide1772	Escherichia coli (strain K12 (P14377: 465/645, 226-456, 32%)	Leptospirillum sp. Group II '5-way CG' (B6ARP9: 660/645, 40-656, 47%)	
	putative Sensor histidine kinase	2.7.13.3	Nide1908	Dictyostelium discoideum (Q54YZ9: 2062/1013, 1335-1848, 39%)	Chthoniobacter flavus Ellin428 (B4CU67: 1248/1013, 198-1114, 38%)	
	putative Histidine kinase, contains GAF domain	2.7.13.3	Nide2021	Synechococcus elongatus (Q06904: 387/438, 141-380, 28%)	Sorangium cellulosum (A9FXI9: 692/438, 288-692, 26%)	
	putative Histidine kinase with protein phosphatase region		Nide2119	Guillardia theta (O78428: 254/563, 16-133, 41%)	Cellvibrio japonicus (B3PEZ3: 592/563, 30-586, 43%)	
	putative Sensor histidine kinase	2.7.13.3	Nide2265	Bacillus subtilis (Q45614: 611/490, 368-598, 36%)	Syntrophobacter fumaroxidans (A0LKM4: 488/490, 1-477, 33%)	
	putative Sensor histidine kinase FleS	2.7.13.3	Nide2283	Escherichia coli K12 (Q06067: 608/429, 215-599, 29%)	Syntrophus aciditrophicus (Q2LRY6: 415/429, 2-411, 33%)	
	putative Hybrid histidine kinase	2.7.13.3	Nide2323	Dictyostelium discoideum (Q54YZ9: 2062/950, 1302-1847, 38%)	Chthoniobacter flavus Ellin428 (B4CU67: 1248/950, 441-1242, 39%)	
cheA	Chemotaxis protein histidine kinase CheA	2.7.13.3	Nide2350	Thermotoga maritima (Q56310: 671/630, 256-671, 45%)	Leptospirillum sp. Group II '5-way CG' (B6AMA5: 646/630, 7-633, 53%)	
cheA	Chemotaxis protein histidine kinase CheA	2.7.13.3	Nide2361	Salmonella typhimurium (P09384: 671/715, 1-662, 42%)	Bermanella marisrubri (Q1N2W7: 704/715, 1-689, 44%)	
	putative Histidine kinase, contains GAF domain	2.7.13.3	Nide2629	Salmonella typhi (Q8Z332: 465/686, 238-459, 29%)	Anabaena sp. PCC 7120 (Q8YQQ8: 1550/686, 489-699, 38%)	
	putative Chemotaxis protein CheA modulated with response regulator receiver region	2.7.13.3	Nide2654	Thermotoga maritima (Q56310: 671/1129, 4-668, 32%)	Meiothermus silvanus (C1XT07: 925/1129, 1-921, 34%)	modular protein
	putative Sensor histidine kinase	2.7.13.3	Nide3050	Anabaena sp. PCC 7120 (Q9R6X3: 751/915, 515-747, 34%)	Methanosarcina acetivorans (Q8TRB0: 1456/915, 801-1441, 32%)	
	putative Hybrid sensor histidine kinase	2.7.13.3	Nide3052	Pseudomonas syringae pv. syringae (Q4ZSY3: 534/809, 139-527, 41%)	Syntrophobacter fumaroxidans (A0LMY3: 733/809, 349-727, 51%)	
	putative Sensor histidine kinase		Nide3092	Brevibacillus brevis (P54663: 386/1051, 122-381, 32%)	Ralstonia eutropha (Q0KA85: 534/1051, 48-529, 36%)	
	putative Sensor histidine kinase	2.7.13.3	Nide3162	Synechocystis sp. PCC 6803 (Q55168: 748/528, 500-745, 32%)	Geobacter sulfurreducens (Q74A24: 583/528, 264-576, 36%)	

Gene	Product	EC no.	CDS[a]	Best BLAST hit in SwissProt database[b]	Best BLAST hit in TrEMBL database[c]	Notes
	putative Sensor histidine kinase	2.7.13.3	Nide3244	Escherichia coli K12 (P14377: 465/651, 231-456, 37%)	Syntrophobacter fumaroxidans (A0LKN7: 551/651, 169-549, 32%)	
	putative Histidine kinase	2.7.13.3	Nide3246	Bacillus subtilis (P13799: 385/380, 102-382, 29%)	Burkholderia cenocepacia (B1K614: 598/380, 269-584, 34%)	
	putative Histidine kinase	2.7.13.3	Nide3284	Escherichia coli K12 (Q06067: 608/1058, 245-600, 33%)	Dehalococcoides sp. CBDB1 (Q3ZWK0: 1062/1058, 426-1046, 28%)	
	putative Sensor histidine kinase		Nide3415	Rhizobium sp. NGR234 (P55552: 827/590, 175-402, 32%)	Methanosarcina mazei (Q8Q0G8: 1584/590, 1075-1569, 37%)	
	putative Sensor histidine kinase	2.7.13.3	Nide3462	Escherichia coli K12 (Q06067: 608/924, 253-604, 29%)	Cyanothece sp. PCC 7425 (B8HVE3: 1676/924, 297-769, 34%)	
	putative Histidine kinase	2.7.13.3	Nide3521	Rhizobium leguminosarum (P10047: 622/677, 361-619, 35%)	Solibacter usitatus (Q02C12: 532/677, 59-516, 31%)	
	putative Hybrid sensor histidine kinase	2.7.13.3	Nide3570	Pseudomonas syringae pv. tomato (Q881J7: 534/1033, 135-529, 39%)	Acidobacteria bacterium Ellin345 (Q1IM85: 1132/1033, 745-1126, 59%)	
	putative Histidine kinase	2.7.13.3	Nide3586	Escherichia coli K12 (P14377: 465/349, 209-455, 31%)	Chloroflexus aurantiacus (A9W9X5: 797/349, 419-658, 39%)	
	putative Histidine kinase	2.7.13.3	Nide3587	Escherichia coli K12 (Q06067: 608/830, 223-604, 27%)	Candidatus Kuenenia stuttgartiensis (Q1PXQ2: 769/830, 309-767, 26%)	
	putative Hybrid sensor histidine kinase	2.7.13.3	Nide3593	Pseudomonas syringae pv. phaseolicola (Q48IV1: 534/786, 47-527, 33%)	Solibacter usitatus (Q01WP9: 1210/786, 709-1208, 47%)	
	Hybrid sensor histidine kinase	2.7.13.3	Nide3753	Pseudomonas syringae pv. phaseolicola (Q48IV1: 534/504, 34-527, 32%)	Acidobacteria bacterium Ellin345 (Q1IM85: 1132/504, 711-1126, 48%)	
	putative Sensor histidine kinase	2.7.13.3	Nide3941	Bacillus subtilis (P16497: 606/892, 143-604, 31%)	Chthoniobacter flavus Ellin428 (B4D3J9: 944/892, 21-799, 31%)	
	putative Hybrid histidine kinase	2.7.13.3	Nide3963	Pseudomonas syringae pv. tomato (Q881J7: 534/620, 139-528, 30%)	Solibacter usitatus (Q01W39: 505/620, 71-495, 37%)	
	putative Histidine kinase		Nide3983	Staphylococcus haemolyticus (Q4L8Q7: 344/389, 124-344, 32%)	Burkholderia pseudomallei S13 (B1HHH6: 575/389, 295-559, 41%)	
	putative Hybrid sensor histidine kinase	2.7.13.3	Nide4240	Pseudomonas syringae pv. phaseolicola (Q48IV1: 534/1314, 51-528, 35%)	Solibacter usitatus (Q01WP9: 1210/1314, 172-1209, 31%)	
	putative Sensor histidine kinase	2.7.13.3	Nide4272	Escherichia coli K12 (Q06067: 608/465, 255-608, 31%)	Rhodopirellula baltica (Q7UF63: 459/465, 13-428, 31%)	

[a] *Ca*. N. defluvii protein identifier

[b] Organism with highest scoring BLAST hit to *Ca*. N. defluvii protein in SwissProt database. In parentheses: SwissProt accession number of best hit: length of SwissProt entry / length of *Ca*. N. defluvii protein, alignment positions on SwissProt entry, amino acid identity.

[c] Organism with highest scoring BLAST hit to *Ca*. N. defluvii protein in TrEMBL database. In parentheses: TrEMBL accession number of best hit: length of TrEMBL entry / length of *Ca*. N. defluvii protein, alignment positions on TrEMBL entry, amino acid identity.

References

Aamand, J., T. Ahl and E. Spieck (1996). Monoclonal antibodies recognizing nitrite oxidoreductase of *Nitrobacter hamburgensis*, *N. winogradskyi*, and *N. vulgaris*. *Appl Environ Microbiol* **62**(7): 2352-2355.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic local alignment search tool. *J Mol Biol* **215**(3): 403-410.

Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones and A. J. Weightman (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**(12): 7724-7736.

Ashida, H., Y. Saito, C. Kojima, K. Kobayashi, N. Ogasawara and A. Yokota (2003). A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science* **302**(5643): 286-290.

Blamey, J. M. and M. W. Adams (1993). Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochim Biophys Acta* **1161**(1): 19-27.

Blasco, F., B. Guigliarelli, A. Magalon, M. Asso, G. Giordano and R. A. Rothery (2001). The coordination and function of the redox centres of the membrane-bound nitrate reductases. *Cell Mol Life Sci* **58**(2): 179-193.

Bocs, S., S. Cruveiller, D. Vallenet, G. Nuel and C. Medigue (2003). AMIGene: Annotation of MIcrobial Genes. *Nucl Acids Res* **31**(13): 3723-3726.

Calzolai, L., C. M. Gorst, Z. H. Zhao, Q. Teng, M. W. Adams and G. N. La Mar (1995). ^{1}H NMR investigation of the electronic and molecular structure of the four-iron cluster ferredoxin from the hyperthermophile *Pyrococcus furiosus*. Identification of Asp14 as a cluster ligand in each of the four redox states. *Biochemistry* **34**(36): 11373-11384.

Carrondo, M. A. (2003). Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J* **22**(9): 1959-1568.

Caspi, R., H. Foerster, C. A. Fulcher, R. Hopkinson, J. Ingraham, P. Kaipa, M. Krummenacker, S. Paley, J. Pick, S. Y. Rhee, C. Tissier, P. Zhang and P. D. Karp (2006). MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucl Acids Res* **34**(Database issue): D511-516.

Daims, H., J. L. Nielsen, P. H. Nielsen, K. H. Schleifer and M. Wagner (2001). In situ characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl. Environ. Microbiol.* **67**(11): 5273-5284.

Daims, H., K. Stoecker and M. Wagner (2005). Fluorescence *in situ* hybridisation for the detection of prokaryotes. *In* A. M. Osborn and C. J. Smith (ed.), Molecular Microbial Ecology. Abingdon, U.K., Bios-Garland**:** 213-239.

Daims, H., S. Lücker and M. Wagner (2006). *daime*, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* **8**(2): 200-213.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Edwards, U., T. Rogall, H. Blocker, M. Emde and E. C. Bottger (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucl Acids Res* **17**(19): 7843-7853.

Ehrich, S., D. Behrens, E. Lebedeva, W. Ludwig and E. Bock (1995). A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Arch. Microbiol.* **164**: 16-23.

Frickey, T. and A. N. Lupas (2004). PhyloGenie: automated phylome generation and analysis. *Nucleic Acids Res.* **32**(17): 5231-5238.

Glöckner, F. O., R. Amann, A. Alfreider, J. Pernthaler, R. Psenner, K.-H. Trebesius and K.-H. Schleifer (1996). An in situ hybridization protocol for detection and identification of planctonic bacteria. *Syst Appl Microbiol* **19**(3): 403-406.

Goltsman, D. S., V. J. Denef, S. W. Singer, N. C. VerBerkmoes, M. Lefsrud, R. S. Mueller, G. J. Dick, C. L. Sun, K. E. Wheeler, A. Zemla, B. J. Baker, L. Hauser, M. Land, M. B. Shah, M. P. Thelen, R. L. Hettich and J. F. Banfield (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**(13): 4599-4615.

Gomes, C. M., R. S. Lemos, M. Teixeira, A. Kletzin, H. Huber, K. O. Stetter, G. Schäfer and S. Anemüller (1999). The unusual iron sulfur composition of the *Acidianus ambivalens* succinate dehydrogenase complex. *Biochim Biophys Acta* **1411**(1): 134-141.

Guindon, S. and O. Gascuel (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**(5): 696-704.

Haft, D. H., B. J. Loftus, D. L. Richardson, F. Yang, J. A. Eisen, I. T. Paulsen and O. White (2001). TIGRFAMs: A protein family resource for the functional identification of proteins. *Nucl Acids Res* **29**(1): 41-43.

Hanson, T. E. and F. R. Tabita (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**(8): 4397-4402.

Hatzenpichler, R., E. V. Lebedeva, E. Spieck, K. Stoecker, A. Richter, H. Daims and M. Wagner (2008). A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc Natl Acad Sci USA* **105**(6): 2134-2139.

Holland, H. D. (2006). The oxygenation of the atmosphere and oceans. *Phil Trans R Soc B* **361**(1470): 903-915.

Hughes, N. J., C. L. Clayton, P. A. Chalk and D. J. Kelly (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**(5): 1119-1128.

Ikeda, T., T. Ochiai, S. Morita, A. Nishiyama, E. Yamada, H. Arai, M. Ishii and Y. Igarashi (2006). Anabolic five subunit-type pyruvate:ferredoxin oxidoreductase from Hydrogenobacter thermophilus TK-6. *Biochem. Biophys. Res. Commun.* **340**(1): 76-82.

Jeanjean, R., E. Talla, A. Latifi, M. Havaux, A. Janicki and C. C. Zhang (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**(10): 2574-2585.

Jormakka, M., D. Richardson, B. Byrne and S. Iwata (2004). Architecture of NarGH reveals a structural classification of Mo-bisMGD enzymes. *Structure* **12**(1): 95-104.

Kanehisa, M. and S. Goto (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucl Acids Res* **28**(1): 27-30.

Kisker, C., H. Schindelin, D. Baas, J. Retey, R. U. Meckenstock and P. M. Kroneck (1998). A structural comparison of molybdenum cofactor-containing enzymes. *FEMS Microbiol Rev* **22**(5): 503-521.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones and M. A. Marra (2009). Circos: An information aesthetic for comparative genomics. *Genome Res* **19**(9): 1639-1645.

Kuo, C. H. and H. Ochman (2009). Inferring clocks when lacking rocks: The variable rates of molecular evolution in bacteria. *Biol Direct* **4**: 35.

Lancaster, C. R. (2002). Succinate:quinone oxidoreductases: An overview. *Biochim Biophys Acta* **1553**(1-2): 1-6.

Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin and N. R. Pace (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA* **82**(20): 6955-6959.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2948.

Lemos, R. S., A. S. Fernandes, M. M. Pereira, C. M. Gomes and M. Teixeira (2002). Quinol:fumarate oxidoreductases and succinate:quinone oxidoreductases: Phylogenetic relationships, metal centres and membrane attachment. *Biochim Biophys Acta* **1553**(1-2): 158-170.

Levican, G., J. A. Ugalde, N. Ehrenfeld, A. Maass and P. Parada (2008). Comparative genomic analysis of carbon and nitrogen assimilation mechanisms in three indigenous bioleaching bacteria: Predictions and validations. *BMC Genomics* **9**: 581.

Ludwig, W. (2004). ARB: A software environment for sequence data. *Nucl Acids Res* **32**: 1363-1371.

Lueders, T., M. Manefield and M. W. Friedrich (2004). Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environ Microbiol* **6**(1): 73-78.

Maixner, F., D. R. Noguera, B. Anneser, K. Stoecker, G. Wegl, M. Wagner and H. Daims (2006). Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. *Environ. Microbiol.* **8**(8): 1487-1495.

Maixner, F., M. Wagner, S. Lücker, E. Pelletier, S. Schmitz-Esser, K. Hace, E. Spieck, R. Konrat, D. Le Paslier and H. Daims (2008). Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium *'Candidatus* Nitrospira defluvii'. *Environ. Microbiol.* **10**(11): 3043-3056.

Martinez-Espinosa, R. M., E. J. Dridge, M. J. Bonete, J. N. Butt, C. S. Butler, F. Sargent and D. J. Richardson (2007). Look on the positive side! The orientation, identification and bioenergetics of "Archaeal" membrane-bound nitrate reductases. *FEMS Microbiol Lett* **276**(2): 129-139.

McDevitt, C. A., P. Hugenholtz, G. R. Hanson and A. G. McEwan (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**(6): 1575-1587.

Mook, W. G., J. C. Bommerson and W. H. Staverman (1974). Carbon isotope fractionation between dissolved bicarbonate and gaseous carbon-dioxide. *Earth Planet Sci Lett* **22**(2): 169-176.

Ochman, H. and A. C. Wilson (1987). Evolution in bacteria - evidence for a universal substitution rate in cellular genomes. *J Mol Evol* **26**(1-2): 74-86.

Paiva, S., F. Devaux, S. Barbosa, C. Jacq and M. Casal (2004). Ady2p is essential for the acetate permease activity in the yeast *Saccharomyces cerevisiae*. *Yeast* **21**(3): 201-210.

Pereira, M. M., M. Santana and M. Teixeira (2001). A novel scenario for the evolution of haem-copper oxygen reductases. *Biochim Biophys Acta* **1505**(2-3): 185-208.

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies and F. O. Glöckner (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl Acids Res* **35**(21): 7188-7196.

Quail, M. A., P. Jordan, J. M. Grogan, J. N. Butt, M. Lutz, A. J. Thomson, S. C. Andrews and J. R. Guest (1996). Spectroscopic and voltammetric characterisation of the bacterioferritin-associated ferredoxin of *Escherichia coli*. *Biochem Biophys Res Comm* **229**(2): 635-642.

Reichard, K. and M. Kaufmann (2003). EPPS: Mining the COG database by an extended phylogenetic patterns search. *Bioinformatics* **19**(6): 784-785.

Rohmer, M., P. Bouvier-Navé and G. Ourisson (1984). Distribution of hopanoid triterpenes in prokaryotes. *J Gen Microbiol* **130**: 1137-1150.

Rowe, J. J., T. Ubbink-Kok, D. Molenaar, W. N. Konings and A. J. Driessen (1994). NarK is a nitrite-extrusion systeminvolved in anaerobic nitrate respiration by *Escherichia coli*. *Mol Microbiol* **12**(4): 579-586.

Sayers, E. W., T. Barrett, D. A. Benson, E. Bolton, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. Dicuccio, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, Z. Lu, T. L. Madden, T. Madej, D. R. Maglott, A. Marchler-Bauer, V. Miller, I. Mizrachi, J. Ostell, A. Panchenko, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, D. Slotta, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, Y. Wang, W. John Wilbur, E. Yaschenko and J. Ye (2010). Database resources of the National Center for Biotechnology Information. *Nucl Acids Res* **38**(Database issue): D5-16.

Schmidt, H. A., K. Strimmer, M. Vingron and A. von Haeseler (2002). TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**(3): 502-504.

Schouten, S., W. C. M. K. Breteler, P. Blokker, N. Schogt, W. I. C. Rijpstra, K. Grice, M. Baas and J. S. S. Damste (1998). Biosynthetic effects on the stable carbon isotopic compositions of algal lipids: Implications for deciphering the carbon isotopic biomarker record. *Geochim Cosmochim Acta* **62**(8): 1397-1406.

Serres, M. H. and M. Riley (2000). MultiFun, a multifunctional classification scheme for *Escherichia coli* K-12 gene products. *Microb Comp Genomics* **5**(4): 205-222.

Spieck, E., C. Hartwig, I. McCormack, F. Maixner, M. Wagner, A. Lipski and H. Daims (2006). Selective enrichment and molecular characterization of a previously uncultured *Nitrospira*-like bacterium from activated sludge. *Environ. Microbiol.* **8**(3): 405-415.

Strous, M., E. Pelletier, S. Mangenot, T. Rattei, A. Lehner, M. W. Taylor, M. Horn, H. Daims, D. Bartol-Mavel, P. Wincker, V. Barbe, N. Fonknechten, D. Vallenet, B. Segurens, C. Schenowitz-Truong, C. Medigue, A. Collingro, B. Snel, B. E. Dutilh, H. J. Op den Camp, C. van der Drift, I. Cirpus, K. T. van de Pas-Schoonen, H. R. Harhangi, L. van Niftrik, M. Schmid, J. Keltjens, J. van de Vossenberg, B. Kartal, H. Meier, D. Frishman, M. A. Huynen, H. W. Mewes, J. Weissenbach, M. S. Jetten, M. Wagner and D. Le Paslier (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**(7085): 790-794.

Trieber, C. A., R. A. Rothery and J. H. Weiner (1996). Engineering a novel iron-sulfur cluster into the catalytic subunit of *Escherichia coli* dimethyl-sulfoxide reductase. *J Biol Chem* **271**(9): 4620-4626.

Tsugawa, H., H. Suzuki, I. Nakagawa, T. Nishizawa, Y. Saito, M. Suematsu and T. Hibi (2008). Alpha-ketoglutarate oxidoreductase, an essential salvage enzyme of energy metabolism, in coccoid form of *Helicobacter pylori*. *Biochem Biophys Res Comm* **376**(1): 46-51.

Vallenet, D., L. Labarre, Z. Rouy, V. Barbe, S. Bocs, S. Cruveiller, A. Lajus, G. Pascal, C. Scarpelli and C. Medigue (2006). MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* **34**(1): 53-65.

Vignais, P. M. and B. Billoud (2007). Occurrence, classification, and biological function of hydrogenases: An overview. *Chem Rev* **107**(10): 4206-4272.

Watson, S. W., E. Bock, F. W. Valois, J. B. Waterbury and U. Schlosser (1986). *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. *Arch. Microbiol.* **144**: 1-7.

Widdel, F. and F. Bak (1992). Gram-negative mesophilic sulfate-reducing bacteria. *In* A. Balows, H. G. Trüper, M. Dworking, W. Harder and K.-H. Schleifer (ed.), The Prokaryotes. New York, Springer**:** 3352-3378.

Yun, N. R., M. Yamamoto, H. Arai, M. Ishii and Y. Igarashi (2002). A novel five-subunit-type 2-oxoglutalate:ferredoxin oxidoreductases from *Hydrogenobacter thermophilus* TK-6. *Biochem Biophys Res Comm* **292**(1): 280-286.

Acknowledgments

Eine Arbeit wie diese kann niemals von einer Person alleine bewältigt werden. Daher möchte ich mich an dieser Stelle ganz herzlich bei allen am Entstehungsprozess beteiligten Personen bedanken.

Zunächst gilt meine ausgesprochene Dankbarkeit natürlich meinen Betreuern, Holger Daims und Michael Wagner, die mich über all die Jahre fachlich sowie finanziell großartig unterstützt haben. Ganz besonders, da sie mich nicht nur in der Arbeit stets dazu motivierten vorwärts zu streben, sondern weil sie mir auch in schwierigen Zeiten starken Rückhalt gegeben haben. Ferner natürlich ein herzlichstes Dankeschön an Christian, dem besten TA den man sich vorstellen kann, ohne den die Kleinigkeiten, die ein Labor am Laufen halten sicher lange nicht so gut funktioniert hätten. Ein besonderer Dank auch an das DoME-Anfangsteam, Alexander, Assi, Eva, Frank, Kilian, Matthias, Stephan, Susanne und Tobi, ohne wäre vor allem auch mein Start in Wien mit Sicherheit schwerer gewesen. Nicht zu vergessen natürlich alle andern DoME Mitglieder, ohne euch wäre meine Doktorarbeit nicht dieselbe gewesen.

Aber die letzten Jahre bestanden nicht nur aus Arbeit. Auch das sollte hier erwähnt werden, denn Freizeit ist ein wichtiger Gegenpol, der zum Gelingen einer erfolgreichen Dissertation beiträgt. Zuallererst möchte ich mich an dieser Stelle bei zwei weiteren Domies bedanken. Zum einen bei Doris dafür, dass sie mir eine großartige Stütze war als es darauf ankam und in unschönen Zeiten restlos für mich da war, und zum anderen bei Elena, da sie mir in den letzten Jahren eine sehr enge Freundin geworden ist, die mich in der Arbeit wie auch privat sehr stark geprägt hat. Dann noch ein ganz dickes Dankeschön an meine Mädels, Bettina, Birgit, Ingrid, Sigrid und Michi, die mich mit unzähligen Kaffeepausen und Feierabendbierchen durch den Alltag begleiteten und mit denen gemeinsam ich auch die Probleme und Motivationstiefs, die während einer Doktorarbeit auftreten, erfolgreich bewältigen konnte.

Ein Danke von Herzen geht auch nach München an all meine Freunde aus Studien- und Schulzeiten. Ihr habt immer zu mir gehalten, auch wenn die Kommunikation aus der Ferne mal wieder etwas spärlicher ausfiel. Und zu guter Letzt natürlich ein ganz dickes Dankeschön an meine Eltern, Dieter und Martha, sowie meine Geschwister Tobi, Io und Nina. Niemand hat so rückhaltlos über all die Jahre an mich geglaubt. Ihr habt mich schon allein mit eurem Stolz auf mich zum Durchhalten bewegt.

Danke euch allen!

Curriculum vitae

Personal Information

Name	Sebastian Lücker
Date of birth	6 February 1975
Place of birth	Munich, Germany
Citizenship	German

Education

Aug 2004 – present	Doctoral studies of Life Sciences Ecology at the University of Vienna, Vienna, Austria. PhD theses title "**Exploring the ecology and genomics of nitrite-oxidizing bacteria**," supervised by Univ.-Ass. Dipl.-Biol. Dr. Holger Daims and Univ.-Prof. Dipl. Biol. Dr. Michael Wagner, Department of Microbial Ecology.
Jul 2004	Diploma award (Dipl. Biol.) with grade "sehr gut bestanden." Exam subjects: Technical Biology, Microbiology, Neurobiology, and Biophysics. Diploma thesis at the Department of Microbial Ecology, University of Vienna. Thesis title "**Entwicklung und Evaluierung 16S rRNS-gerichteter Oligonukleotidsonden zum spezifischen Nachweis deltaproteobakterieller Sulfatreduzierer mittels Fluoreszenz *in situ* Hybridisierung und Erweiterung des *dsrAB*-Sequenzdatensatzes**," supervised by Dipl.-Biol. Dr. Alexander Loy and Univ.-Prof. Mag. Dr. Michael Wagner, Department of Microbial Ecology, University of Vienna, and Prof. Dr. rer. nat. Karl-Heinz Schleifer, Department of Microbiology, Technical University of Munich.
Oct 1998 – Jul 2004	Course of studies: Biology, major Technical Biology, at the Technical University of Munich, Munich, Germany. Areas of specialization: Technical Biology, Microbiology, Neurobiology, and Biophysics.
Jun 1996	Secondary-school diploma (Abitur), Lion-Feuchtwanger-Gymnasium, Munich, Germany.

Graduate Research Assistance

May 2010 – present	WWTF project "Multiphasic comparative analysis of key nitrite-oxidizing bacteria in wastewater treatment plants."
Apr 2008 – Apr 2010	FWF project "MICDIF – Linking microbial diversity and ecosystem functions across scales and interfaces."
Aug2004 – Mar 2008	WWTF project "Genomics and Ecology of Novel Uncultured Nitrite-Oxidizing Bacteria in Natural and Engineered Surroundings."

Undergraduate Research Assistance

May 2003 – Jul 2004	Bavarian Research Foundation project „Entwicklung eines universellen DNA-Chips zum Nachweis von Bakterien in komplexem Probenmaterial."

Teaching

Jul 2004, Feb 2006, 2007, 2008, 2010

Tutor of the „International FISH Course" at the Department of Microbial Ecology, University of Vienna.

Jul 2004, 2005, 2009, Feb 2005, 2006

Tutor of the course (Projektpraktikum) „Fluoreszenz-in-situ-Hybridisierung (FISH) - Identifizierung nicht-kultivierter Mikroorganismen" at the Department of Microbial Ecology, University of Vienna

Publications

Lebedeva, E. V., S. Off, S. Zumbrägel, M. Kruse, A. Shagzhina, **S. Lücker**, F. Maixner, A. Lipski, H. Daims and E. Spieck (in preparation). Isolation and characterization of a moderately thermophilic nitrite-oxidizing bacterium from a geothermal spring.

Lücker, S., M. Wagner, F. Maixner, E. Pelletier, H. Koch, B. Vacherie, T. Rattei, J. S. S. Damsté, E. Spieck, D. Le Paslier and H. Daims (2010). A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci USA* 107**(30)**: 13479-13484.

Bayer, C., N. R. Heindl, C. Rinke, **S. Lücker**, J. A. Ott and S. Bulgheresi (2009). Molecular characterization of the symbionts associated with marine nematodes of the genus *Robbea*. *Environ Microbiol Reports* 1**(2)**: 136-144.

Webster, N. S., M. W. Taylor, F. Behnam, **S. Lücker**, T. Rattei, S. Whalan, M. Horn and M. Wagner (2009). Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* 12**(8)**: 2070–2082.

Maixner, F., M. Wagner, **S. Lücker**, E. Pelletier, S. Schmitz-Esser, K. Hace, E. Spieck, R. Konrat, D. Le Paslier and H. Daims (2008). Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium *"Candidatus* Nitrospira defluvii". *Environ Microb* 10**(11)**: 3043-3056.

Lücker, S., D. Steger, K. U. Kjeldsen, B. J. MacGregor, M. Wagner and A. Loy (2007). Improved 16S rRNA-targeted probe set for analysis of sulfate-reducing bacteria by fluorescence in situ hybridization. *J Microbiol Meth* 69**(3)**: 523-528.

Daims, H., **S. Lücker** and M. Wagner (2006). daime, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* 8**(2)**: 200-213.

Daims, H., F. Maixner, **S. Lücker**, K. Stoecker, K. Hace and M. Wagner (2006). Ecophysiology and niche differentiation of *Nitrospira*-like bacteria, the key nitrite oxidizers in wastewater treatment plants. *Water Sci Tech* 54**(1)**: 21-27.

Loy, A., C. Schulz, **S. Lücker**, A. Schopfer-Wendels, K. Stoecker, C. Baranyi, A. Lehner and M. Wagner (2005). 16S rRNA Gene-Based Oligonucleotide Microarray for Environmental Monitoring of the Betaproteobacterial Order "Rhodocyclales". *Appl Environ Microbiol* 71**(3):** 1373-1386.

Zverlov, V., M. Klein, **S. Lücker**, M. W. Friedrich, J. Kellermann, D. A. Stahl, A. Loy and M. Wagner (2005). Lateral gene transfer of dissimilatory (bi)sulfite reductase revisited. *J Bacteriol* 187**(6)**: 2203-2208.

Book Chapter

Daims, H., **S. Lücker**, D. Le Paslier and M. Wagner (2010). Diversity, Environmental Genomics, and Ecophysiology of Nitrite-Oxidizing Bacteria. *In*: Nitrification, B. B. Ward, D. J. Arp and M. G. Klotz (ed.). Washington, DC, ASM Press: in press.

Oral Presentations

Sebastian Lücker, Michael Wagner, and Holger Daims (2010). A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. IECB Wissenschaftstag, Vienna, Austria, April 2010.

Sebastian Lücker (2009). Introducing FISH. Hands-On Cryo Electron Microscopy and Tissue in-situ Hybridization course, Vienna, November 2009

Sebastian Lücker, Michael Wagner, and Holger Daims (2009). The complete genome of "*Candidatus* Nitrospira defluvii". 7th International meeting on Novel Techniques in Microbial Ecology (INTIME), Lackenhof, Austria, August 2009.

Sebastian Lücker, Michael Wagner, and Holger Daims (2007). The complete genome of the uncultured nitrite oxidizer "*Candidatus* Nitrospira defluvii" as basis for comparative genomics

with other *Nitrospira*-like bacteria. 6th International meeting on Novel Techniques in Microbial Ecology (INTIME), Aarhus, Denmark, October 2007.

Sebastian Lücker, Karin Hace, Frank Maixner, Bettina Anneser, Gertrude Wegl, Kilian Stoecker, Christian Baranyi, Michael Wagner, and Holger Daims (2005). From phylogeny to ecophysiology: Niche differentiation of nitrite-oxidizing *Nitrospira*-like bacteria. 5th International meeting on Novel Techniques in Microbial Ecology (INTIME), Lunz am See, Austria, July 2005.

Sebastian Lücker, Alexander Loy, and Michael Wagner (2004). 16S rRNA-based oligonucleotide microarray for the detection of all recognized and yet uncultured members of the order *Rhodocyclales*. IECB Wissenschaftstag, Vienna, Austria, April 2004.

Poster Presentations

Sebastian Lücker, Christiane Dorninger, Eva Spieck, Michael Wagner, and Holger Daims (2010). Novel nitrite oxidizers related to "*Candidatus* Nitrotoga arctica" frequently appear in wastewater treatment systems. 13th International Symposium on Microbial Ecology (ISME13), Seattle, WA, USA, August 2010

Holger Daims**, Sebastian Lücker**, Eva Spieck, Thomas Rattei, Denis Le Paslier, and Michael Wagner (2010). Genomics and functional analyses reveal unexpected features and suggest a complex evolution of nitrite-oxidizing bacteria. 13th International Symposium on Microbial Ecology (ISME13), Seattle, WA, USA, August 2010

Sebastian Lücker, Eric Pelletier, Frank Maixner, Karin Hace, Eva Spieck, Denis Le Paslier, Michael Wagner, and Holger Daims (2008). The complete genome of the uncultured nitrite oxidizer "*Candidatus* Nitrospira defluvii" as basis for comparative genomics with other *Nitrospira*-like bacteria. VAAM Jahrestagung, Frankfurt, Germany, March 2008.

Niels R Heindl, Christoph Bayer, **Sebastian Lücker**, Jörg A Ott, Silvia Bulgheresi (2008). Detection of free-living forms of marine nematode symbionts. 12th International Symposium on Microbial Ecology (ISME12), Cairns, Australia, August 2008.

Sebastian Lücker, Eric Pelletier, Frank Maixner, Karin Hace, Eva Spieck, Denis Le Paslier, Michael Wagner, and Holger Daims (2007). The complete genome of the uncultured nitrite oxidizer "*Candidatus* Nitrospira defluvii" as basis for comparative genomics with other *Nitrospira*-like bacteria. ProkaGENOMICS 2007, Göttingen, Germany, October 2007.

Sebastian Lücker, Karin Hace, Frank Maixner, Christian Baranyi, Kilian Stoecker, Michael Wagner and Holger Daims (2006). Molecular screening of chemically diverse habitats reveals a hitherto unexplored biodiversity of nitrite-oxidizing *Nitrospira*-like bacteria. 11th International Symposium on Microbial Ecology (ISME 11), Vienna, Austria, August 2006.

Sebastian Lücker, Frank Maixner, Bettina Anneser, Gertrude Wegl, Kilian Stoecker, Michael Wagner and Holger Daims (2005). Responses of closely related nitrite-oxidizing *Nitrospira*-like

bacteria to shifts of nitrite concentration: an example of niche differentiation in the microbial world. VAAM Jahrestagung, Göttingen, Germany, September 2005.

Karin Hace, **Sebastian Lücker**, Frank Maixner, Christian Baranyi, Kilian Stoecker, Michael Wagner and Holger Daims (2005). Molecular screening of chemically diverse habitats reveals a hitherto unexplored biodiversity of nitriteoxidizing *Nitrospira*-like bacteria. VAAM Jahrestagung, Göttingen, Germany, September 2005.

Sebastian Lücker, Alexander Loy, and Michael Wagner (2004). 16S rRNA-based oligonucleotide microarray for the detection of all recognized and yet uncultured members of the betaproteobacterial order *Rhodocyclales*. Minisymposium: Microbial ecology: from ecosystem functioning to biotechnology application, Vienna, Austria, March 2004