

Phylogenetic Analysis of Nitrite, Nitric Oxide, and Nitrous Oxide Respiratory Enzymes Reveal a Complex Evolutionary History for Denitrification

Christopher M. Jones,* Blaž Stres,† Magnus Rosenquist,‡ and Sara Hallin*

*Department of Microbiology, Swedish Agricultural University, Uppsala, Sweden; †Zootechnical Department, Biotechnical Faculty, University of Ljubljana, Domžale, Slovenia; and ‡Department of Radiology, Oncology, and Clinical Immunology, Uppsala University Hospital, Uppsala, Sweden

Denitrification is a facultative respiratory pathway in which nitrite (NO_2^-), nitric oxide (NO), and nitrous oxide (N_2O) are successively reduced to nitrogen gas (N_2), effectively closing the nitrogen cycle. The ability to denitrify is widely dispersed among prokaryotes, and this polyphyletic distribution has raised the possibility of horizontal gene transfer (HGT) having a substantial role in the evolution of denitrification. Comparisons of 16S rRNA and denitrification gene phylogenies in recent studies support this possibility; however, these results remain speculative as they are based on visual comparisons of phylogenies from partial sequences. We reanalyzed publicly available *nirS*, *nirK*, *norB*, and *nosZ* partial sequences using Bayesian and maximum likelihood phylogenetic inference. Concomitant analysis of denitrification genes with 16S rRNA sequences from the same organisms showed substantial differences between the trees, which were supported by examining the posterior probability of monophyletic constraints at different taxonomic levels. Although these differences suggest HGT of denitrification genes, the presence of structural variants for *nirK*, *norB*, and *nosZ* makes it difficult to determine HGT from other evolutionary events. Additional analysis using phylogenetic networks and likelihood ratio tests of phylogenies based on full-length sequences retrieved from genomes also revealed significant differences in tree topologies among denitrification and 16S rRNA gene phylogenies, with the exception of the *nosZ* gene phylogeny within the data set of the *nirK*-harboring genomes. However, inspection of codon usage and G + C content plots from complete genomes gave no evidence for recent HGT. Instead, the close proximity of denitrification gene copies in the genomes of several denitrifying bacteria suggests duplication. Although HGT cannot be ruled out as a factor in the evolution of denitrification genes, our analysis suggests that other phenomena, such as gene duplication/divergence and lineage sorting, may have differently influenced the evolution of each denitrification gene.

Introduction

Denitrification, a microbial process in the nitrogen cycle, is a facultative respiratory pathway in which nitrate (NO_3^-), nitrite (NO_2^-), nitric oxide (NO), and nitrous oxide (N_2O), successively, are reduced to nitrogen gas (N_2). Denitrification has been the focus of numerous studies because it is a major cause of nitrogen loss in agriculture, contributes to production of the greenhouse gas N_2O , and is useful for removing excess nitrogen from water (Tiedje 1988)

Seven enzymes catalyzing the 4 reductive steps of the denitrification pathway are known (Zumft 1997; Philippot 2002). Nitrate reduction to NO_2^- , catalyzed by the reductases Nar or Nap, is the first step in the pathway, although it is not unique for denitrification. The defining step of denitrification is, instead, the reduction of soluble NO_2^- to gaseous NO, catalyzed by 1 of 2 nitrite reductases, NirK or NirS. The 2 proteins are different regarding structure and catalytic site, despite having identical functions. NirS contains a cytochrome *cd*₁ active site, whereas NirK is a member of the multi-copper oxidase metalloprotein family. NO is further reduced to N_2O by the nitric oxide reductases cNor or qNor, encoded by 2 variants of the same gene (*norB*). The 2 proteins differ in electron donor specificity as qNor derives electrons from a quinol pool, whereas cNor is associated with cytochrome *c* or blue copper proteins (Zumft 2005). The former is also found in nondenitrifying prokaryotes, where it plays a detoxifying role. A third Nor type has been found in *Bacillus azotoformans*, in which it was suggested to fulfill both bioenergetic and detoxifying

functions (Suharti et al. 2001). The final step in denitrification is the conversion of N_2O to N_2 . The only enzyme known to catalyze this reaction is nitrous oxide reductase, Nos. This step effectively closes the overall nitrogen cycle as N_2 gas may again be converted to soluble nitrogen oxides through the action of nitrogen-fixing and nitrifying organisms.

The ability to denitrify is observed in a range of genera among both bacteria and archaea, and the widespread distribution is likely due to a common ancestor that existed before the prokaryotes split into 2 domains (Knowles 1982; Zumft 1997). Closely related bacteria do not necessarily share the trait, and even when they do, they can have completely different denitrifying abilities. On the other hand, identical abilities can be found within distantly related bacteria (Clays-Josserand et al. 1999). Also, when looking at the distribution of denitrification genes, closely related genes can be found within distantly related organisms and vice versa (Philippot 2002). Altogether, this suggests that denitrification genes have been both lost and inherited during evolution.

Recent studies comparing the phylogenies of *nar*, *nir*, *nor*, and *nos* gene sequences in bacterial isolates showed that the functional gene phylogenies are more or less incongruent with that of 16S rRNA (Delorme et al. 2003; Gregory et al. 2003; Goregues et al. 2005; Heylen et al. 2006; Dandie et al. 2007; Heylen et al. 2007), and horizontal gene transfer (HGT) has often been proposed as a major cause behind the disagreement between gene phylogeny and 16S rRNA-based organism classification. This has led to criticism of the practice of using denitrification genes as molecular markers instead of the 16S rRNA genes for analysis of community structure and abundance of denitrifiers in the environment. It has also been proposed that *nir* and *nor* would be genetically linked among certain groups (Zumft 1997; Philippot 2002; Heylen et al.

Key words: denitrification, phylogeny, *nirK*, *nirS*, *norB*, *nosZ*.

E-mail: Chris.Jones@mikrob.slu.se.

Mol. Biol. Evol. 25(9):1955–1966. 2008

doi:10.1093/molbev/msn146

Advance Access publication July 8, 2008

2007). However, several of these proposals are inconclusive and claims remain largely speculative as testing of the underlying hypotheses has not been performed beyond visual comparison of bootstrap consensus trees generated with distance-based methods. Because most studies are based on analysis of partial sequences of polymerase chain reaction (PCR)-amplified genes, a potentially low signal-to-noise ratio may exist in these data sets, which can be problematic for distance-based phylogenetic inference when sequence divergence is high (Huelsenbeck and Hillis 1993).

We have revisited publicly available *nirK*, *nirS*, *norB*, and *nosZ* partial gene sequences from previous studies of denitrifying bacterial isolates, as well as full-length sequences from complete genomes, to estimate the extent to which denitrification gene phylogeny is determined by vertical inheritance, as compared with HGT. We used 2 likelihood-based phylogenetic methods, maximum likelihood and Bayesian inference, to avoid problems associated with short, and potentially divergent, sequence sets. The use of Bayesian analysis to approximate the posterior distribution of possible tree topologies allowed us to statistically determine if denitrification gene phylogeny deviates from organism classification, by comparing the probabilities of monophyletic groupings in denitrification and 16S rRNA data sets at several taxonomic levels. Because the 16S rRNA is the basis of bacterial and archaeal taxonomic classifications at the family level and above (Garrity et al. 2007), it served as a control to assess the validity of our analysis. Phylogenetic networks generated from full-length sequences retrieved from genomes were used to visualize the phylogenetic incongruence among genes in the denitrification pathway. Questions relating to genetic linkage and HGT were best answered by the location of denitrification genes in the chromosomes and examination of genome features, in conjunction with phylogenetic analysis of full-length sequences. To better understand the evolution of the denitrification pathway, we combined the use of phylogenetic network analyses, statistical comparison of functional gene tree topologies, and examination of genome features.

Materials and Methods

Data Sets

Nucleotide sequences for the *nirK*, *nirS*, *norB*, and *nosZ* genes, encoding enzymes in the denitrification pathway, and 16S rRNA genes were retrieved from the Fungene (<http://fungene.cme.msu.edu/>), TIGR (<http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi>), GenBank and Ribosomal Database Project (RDP) II (<http://rdp.cme.msu.edu>) sequence databases in May 2007. After excluding all sequences from environmental clones, we limited our data set to sequences originating from either genome sequencing projects or from cultivation studies in which strains were isolated from environmental samples, tested for complete or partial denitrification activity, and the 16S rRNA genes plus 1 or several of the functional genes listed above had been sequenced (supplementary table S1, Supplementary Material online). Sequences that contained multiple N

residues or stop codons in all reading frames were excluded. Functional genes that did not have definite annotation in published genomes were examined by comparison of conserved positions using PSI-Blast (Altschul et al. 1997) to confirm gene identity. The National Center for Biotechnology Information taxonomic classifications for all species used in this study were cross-checked using the RDP classifier (Wang et al. 2007) to ensure that classification errors or recent changes in classification were accounted for.

To compare the phylogeny of genes in the complete denitrification pathway, *nir*, *norB*, and *nosZ* sequences obtained from genome sequencing projects were used to assemble data sets comprised of concatenated alignments for all 3 genes, with each gene forming a single partition within the total data set. Multiple copies of 16S rRNA gene sequences from all genomes were combined into a separate data set. Because *nirS* and *nirK* were mutually exclusive within all genomes of denitrifying prokaryotes, functional and 16S rRNA gene data sets were divided into 2 subsets based on the presence of either *nirS* or *nirK* in the genomes. Multiple copies of functional genes were included when found.

Multiple Sequence Alignments and Phylogenetic Trees

Denitrification gene data sets were aligned with the Trans-align perl script (Bininda-Emonds 2005) using ClustalW (Thompson et al. 1994). The resulting alignments were inspected visually for error using the Se-AL alignment editor (<http://tree.bio.ed.ac.uk/software/seal/>) under amino acid settings. As the partial sequences were obtained from studies using different PCR primers, data sets were edited to include the same region for all sequences within each gene data set, and ambiguously aligned regions were manually selected and excluded from the analysis. The 16S rRNA gene sequences were aligned according to secondary structure using the RDP II hidden Markov model (HMM) alignment algorithm (Cole et al. 2003).

Phylogenetic trees of denitrification genes were constructed using maximum likelihood (ML) and Bayesian analysis of nucleotide and amino acid sequences. The general time reversible (GTR) + I + Γ model with 6 substitution categories was determined to be the most suitable model by Modeltest v3.6 (Posada and Crandall 1998) and was used for all subsequent nucleotide analyses. ML trees based on nucleotide sequences were inferred using PHYML v2.4.5 (Guindon and Gascuel 2003) with an estimated rate distribution shape parameter (α) and bootstrap consensus values calculated using 500 replicates. Bayesian inference of phylogeny was performed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003). Functional gene data sets were partitioned into first, second, and third codon position, and nucleotide substitution rate variation parameters were estimated for each partition with the covarion model parameter added to allow for rate variation across the tree (Galtier 2001). Amino acid alignments were initially analyzed in MrBayes using the model jumping option for amino acid analysis with estimated proportion of invariant sites and gamma distribution shape parameters. The model with the

highest posterior probability at stationary phase in MrBayes runs was used in subsequent ML analysis. The Whelan and Goldman (WAG) + I + Γ substitution model was used for *nirK*, *nirS*, and *nosZ* data sets, whereas *norB* was analyzed using the BLOSUM + I + Γ model. Amino acid ML trees were generated in PHYML, and bootstrap values were calculated using 500 replicates. For the 16S rRNA gene data sets, partitions for nucleotides in either loop or stem regions were created based on the RDP alignment structure guide, and the doublet model for paired nucleotide evolution (Schoniger and Von Haeseler 1994) was used for the stem partition as implemented in MrBayes. For the positions in the loop regions, the GTR + I + Γ model was used, as determined by the Modeltest output. MrBayes analyses were performed using 3 separate runs, consisting of 1 cold and 3 heated Markov Chain Monte Carlo (MCMC) chains with random starting trees. Chains were initially run for 1,000,000 generations with likelihood values sampled every 100 generations, and convergence was determined when the average deviation of split frequencies reached 0.01 (Huelsenbeck et al. 2002). The Tracer v1.4 program (<http://tree.bio.ed.ac.uk/software/tracer/>) was used to examine log-likelihood plots and MCMC summaries for all parameters. Runs that did not converge after 1,000,000 generations were increased until convergence was attained, with sampling frequencies set at every X/100 generation, where X equals the total number of generations. Consensus trees and Bayesian posterior probability values at nodes were calculated from the posterior distribution at stationary phase.

Probability of Taxonomic Constraints in Denitrification and 16S rRNA Genes

Monophyletic groupings of organisms were created at 5 different taxonomic levels, genus, family, order, class, and phylum, as determined by the Taxonomic Outline of Bacteria and Archaea (Garrity et al. 2007). These groupings were then used to create constraints for filtering trees sampled from the posterior distribution found when MCMC runs reached stationary phase using PAUP* version 4.b10 (Swofford 2000) for both the functional gene (amino acid and nucleotide) and matching 16S rRNA data sets. The posterior probability of a group being monophyletic was calculated as the proportion of trees sampled from the posterior distribution containing the monophyletic constraint.

Phylogenetic Networks and Comparison of Tree Topologies Using Likelihood Ratio Tests

Phylogenetic networks were generated from the concatenated DNA sequence alignments of *nir*, *norB*, and *nosZ* from genomes using parsimony splits implemented in Splitstree v4.8 (Huson 1998). Networks were also constructed from 16S rRNA gene sequences for comparison with the concatenated data sets. Initial analysis of 16S rRNA gene sequences, using uncorrected pairwise distances, showed minimal divergence of multiple copies within genomes (<0.01%). Therefore, a single 16S rRNA gene copy from each genome was used for generating networks and PHYML trees used in subsequent likelihood ratio tests.

Support for competing edges within the resulting networks was determined using 1,000 bootstrap replicates. Multiple functional gene copies were included in the analysis by creating a new partition for each additional gene copy, with gap sites for remaining genomes that did not possess additional copies.

The significance of incongruence between gene phylogenies was determined using the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999), implemented in the CONSEL program (Shimodaira and Hasegawa 2001). Denitrification and 16S rRNA gene data sets were analyzed using PHYML, under the same likelihood model parameters as described above. Site likelihood scores for each sequence, set under the different tree topologies, were calculated under the GTR + I + Γ model using PAUP. The resulting score files were analyzed in CONSEL to calculate the *P* values for rejection of null hypothesis, that is, that alternative tree topologies are not significantly worse than the inferred topology of the partition tested.

Genome Features

Complete genomes were inspected using either the Artemis genome annotation software (Rutherford et al. 2000) or the Jena Prokaryotic Genome Viewer (JPGV-<http://jpgv.fli-leibniz.de/cgi/index.pl>). In each genome, *nir*, *norB*, and *nosZ* genes were located and compared with the locations of regions that may have been subjected to HGT. Deviation from average G + C content and the score-based identification of genomic islands - hidden markov model algorithm (Waack et al. 2006) implemented in JPGV for detecting genomic islands were used to identify chromosomal regions that were likely to have undergone HGT.

Results

Sequence Data sets

The partial sequence data sets were predominately comprised of proteobacterial sequences, with a small proportion originating from more distantly related bacterial phyla or archaea (fig. 1). Full-length denitrification and 16S rRNA gene sequence sets from sequenced genomes had similar taxonomic distribution as the partial sequence data sets. Out of approximately 68 complete genomes in the database with either *nirS* or *nirK*, only 43 containing the complete denitrification pathway, that is, *nosZ* was included, were found at the time of sampling (May 2007). We observed several instances where the RDP classifier returned a different taxonomic designation than that listed in the original GenBank entry, with high statistical support (table 1). The majority of incongruent results were among closely related genera that may be difficult to resolve using 16S rRNA sequence information alone. However, several sequences were incongruent at the class level, such as *Pseudomonas* G-179 being classified as a *Rhizobium* and *Achromobacter cycloclastes* IAM-1013 classified as *Ensifer*. Although these 2 isolates contained references to the correct genus identification in the original GenBank file, several others did not and were discarded from the analysis as being unreliable.

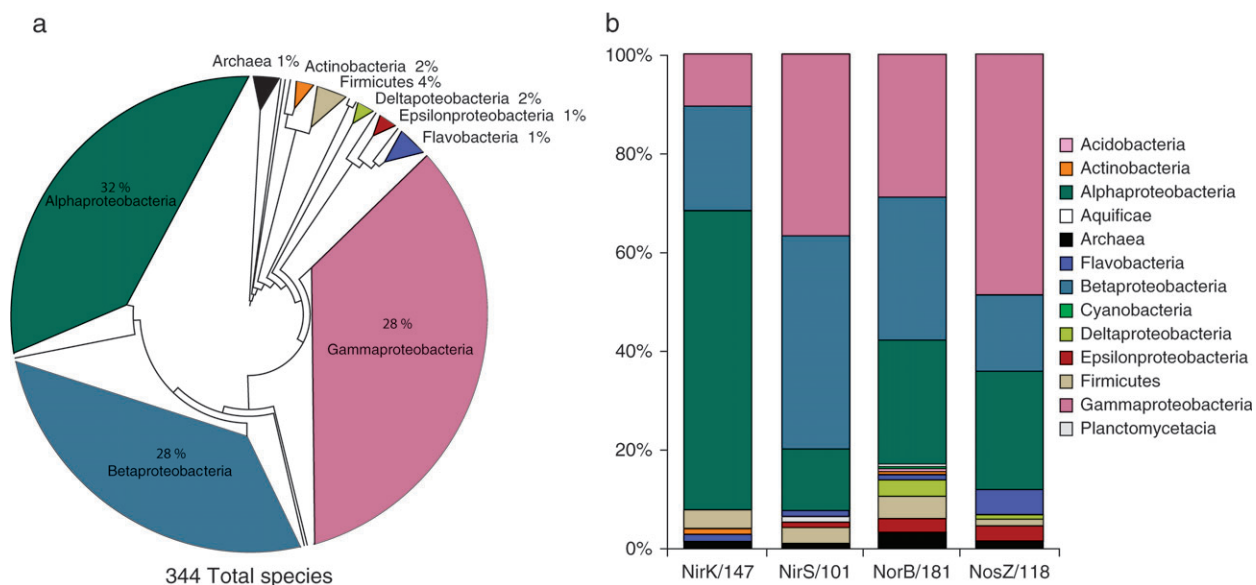


FIG. 1.—(a) Distribution of organisms sampled in all partial sequence alignments of functional denitrification genes. Phylogeny is based on PHYML analysis (GTR + I + Γ) of 16S rRNA sequences. (b) Distribution of organisms within each functional gene data set.

Structural Diversity of *nirK* Obscures HGT Events

Phylogenetic reconstruction of the *nirK* partial sequence data set, consisting of 147 genotypes, resulted in partially resolved phylogenies with low bootstrap and Bayesian probability values at basal nodes for both amino acid and nucleotide data (supplementary fig. S1, Supplementary Material online). High node support was, however, observed in several terminal nodes with distantly related organisms grouped together, such as the placement of *Staphylococcus*, *Chryseobacterium*, and *Enterococcus* species with proteobacteria. A well-supported clade within the tree corresponded to the grouping of a second NirK structural type (Ellis et al. 2007). Further inspection of the copper-binding domain within the partial sequence alignment revealed distinct amino acid motifs containing the copper-binding histidine residue for each type; TRPHL for group I *nirK* and SSFH(V/I/P) for group II *nirK*. The group I clade was dominated by proteobacterial sequences, whereas the group II clade was a more proportionate range of archaeal, Gram-positive, and proteobacterial sequences and contained no sequences obtained by PCR amplification using currently published primers. Monophyletic groupings with high posterior probabilities within the *nirK* tree were only seen among several distantly related genera (*Paracoccus*, *Neisseria*, *Bosea*, *Nitrobacter*, and *Achromobacter*). Monophyletic groupings at the family, order, class, and phylum levels were not supported by the data, with the exception of the 2 archaeal species.

Analysis of the matching 16S rRNA data showed that the Lactobacillales and Bacillales orders were each not monophyletic within the 16S rRNA tree, whereas monophyly of the Rhizobiales received a negligible probability (3%). Monophyly of the Neisseriales was also poorly supported (35%), most likely due to the separate positioning of *Chromobacterium*. Monophyly at the class and phylum levels was well supported (>95%) for all groups.

Vertical Inheritance Patterns Observed in *nirS* and *nosZ*, with Instances of HGT

The *nirS* phylogeny, based on 101 genotypes, was largely unresolved in the basal region of the nucleotide tree, and this was also reflected in low node support values in the tree based on amino acids (supplementary fig. S2, Supplementary Material online). Visual comparison between *nirS* and 16S rRNA gene phylogenies indicated a higher level of congruence with organism classification than that seen in the *nirK* data set, with possible instances of HGT events. This was supported by the observation that *nirS* sequences from archaeal (*Pyrobaculum calidifontis*) and early-diverging lineages within the bacteria (*Hydrogenobacter thermophilus* and *Kuenenia stuttgartensis*) grouped separately from the bulk of the proteobacteria. Additionally, substantial support was seen for monophyly of members within the Rhodobacterales (100%), and monophyly of the Burkholderiaceae family was also well supported (100%). In contrast, substantial probabilities for monophyletic constraints at the genus level were observed only for *Paracoccus*, *Alicyclophilus*, and *Cupriavidus* (100, 95, and 96%, respectively). The multiple *nirS* copies in *Magnetospirillum*, *Dechloromonas*, and *Thiobacillus* appeared in different clades throughout the tree, whereas *Colwellia psychrerythraea* and 3 *Pseudomonas* isolates (Gammaproteobacteria) clustered with the Betaproteobacteria. Two species of Alphaproteobacteria, *Azospirillum brasiliense* Sp 7 and *Magnetospirillum magneticum* AMB-1 copies A and C, grouped together within one of the Alphaproteobacteria clades, and another *nirS* copy from *M. magneticum* AMB-1 (copy B) clustered with *Thiobacillus denitrificans* copy B, close to *Sulfurimonas denitrificans* (Epsilonproteobacteria). Additionally, *Flavobacterium* U43 and *Bacillus* R22 grouped together with *Marinobacter* spp., with strong node support (bootstrap probability = 100% and bayesian posterior probability = 1.0) and short branch

Table 1
Organisms with Conflicting GenBank Identification and RDP Classifier Results at the Genus Level

Accession	GenBank	RDP Classifier	% Support
AM292630	<i>Achromobacter cycloclastes</i> IAM-1013	<i>Ensifer</i>	87
AF229862	<i>Acidovorax</i> 2FB7	<i>Diaphorobacter</i>	97
NC008782	<i>Acidovorax</i> JS42	<i>Diaphorobacter</i>	99
NC003304	<i>Agrobacterium tumefaciens</i> C58	<i>Rhizobium</i>	100
AJ412686	<i>Alcaligenes</i> I	<i>Pusillimonas</i>	81
AJ277706	<i>Alcaligenes</i> N	<i>Castellaniella</i>	100
AB046605	<i>Alcaligenes</i> STC1	<i>Bordetella</i>	59
AM084051	<i>Alicyclophilus</i> R-26814	<i>Diaphorobacter</i>	44
NC006933	<i>Brucella abortus</i> 9.941	<i>Ochrobactrum</i>	96
NC003318	<i>Brucella melitensis</i> 16M	<i>Ochrobactrum</i>	92
AM084067	<i>Brucella</i> R-26895	<i>Ochrobactrum</i>	92
NC004310	<i>Brucella suis</i> 1330	<i>Ochrobactrum</i>	98
AM084097	<i>Chryseobacterium</i> R-25053	<i>Cloacibacterium</i>	82
NC007298	<i>Dechloromonas aromatica</i> RCB	<i>Ferritribacterium</i>	100
AM084101	<i>Diaphorobacter</i> R-24615	<i>Alicyclophilus</i>	52
AF229872	<i>Ensifer</i> 2FB6	<i>Sinorhizobium</i>	81
AM083996	<i>Enterococcus</i> R-24626	<i>Isobaculum</i>	40
AAOC00000000	<i>Flavobacteria</i> HTCC 2170	<i>Maribacter</i>	99
AJ623288	<i>Flavobacterium</i> U43	<i>Bizionia</i>	94
CT573071	<i>Kuenenia stuttgartensis</i>	<i>Pirellula</i>	22
NC008576	<i>Magnetococcus</i> MC1	<i>Piscirickettsia</i>	43
NC006300	<i>Mannheimia succiniciproducens</i> MBEL55E	<i>Actinobacillus</i>	65
NC007426	<i>Natronomonas pharaonis</i> DSM 2160	<i>Halorhabdus</i>	63
DQ665838	<i>Nisaea denitrificans</i> DR 4121	<i>Inquilinus</i>	35
DQ665839	<i>N. denitrificans</i> PD 4118	<i>Inquilinus</i>	38
AF109171	<i>Pseudomonas</i> G-179	<i>Rhizobium</i>	100
AM084044	<i>Rhizobium</i> R-24654	<i>Shinella</i>	100
AM084043	<i>Rhizobium</i> R-24658	<i>Shinella</i>	100
AM084102	<i>Rhizobium</i> R-26820	<i>Agrobacterium</i>	55
AAYC00000000	<i>Roseobacter</i> SK-209-2-6	<i>Phaeobacter</i>	72
NC003047	<i>Sinorhizobium meliloti</i> 1021	<i>Ensifer</i>	100
DQ377753	<i>Sinorhizobium</i> PD12	<i>Ensifer</i>	75
AM084000	<i>Sinorhizobium</i> R-24605	<i>Rhizobium</i>	99
AM084031	<i>Sinorhizobium</i> R-25067	<i>Rhizobium</i>	100
AM084032	<i>Sinorhizobium</i> R-25078	<i>Rhizobium</i>	100
NC000911	<i>Synechocystis</i> PCC 6803	<i>Microcystis</i>	81

NOTE.—Support indicates percent bootstrap support for classifier designations.

lengths, whereas *Flavobacterium* isolate BH12.12 was inserted into a clade consisting largely of *Thauera* isolates, also with short branches.

As observed in the *nirK* data set, several isolates were assigned different genus classifications from the original GenBank entry when the matching 16S rRNA data set was analyzed using the RDP classifier. Interestingly, *S. denitrificans* (Epsilonproteobacteria) clustered with Flavobacteria species. Despite poor node support for this grouping, the monophyly of the proteobacteria in the 16S rRNA data set was not supported due to this grouping. Class and genus monophyly constraints received probabilities greater than 99%, although monophyly of Rhodocyclales order was not supported. All remaining order and family groupings received 99–100% support of monophyly based on current classification.

For the *nosZ* data set with 118 genotypes, trees inferred using either MrBayes or PHYML produced well-resolved phylogenies with high bootstrap or Bayesian support values in the basal region of the tree (supplementary fig. S3, Supplementary Material online). Strong support was observed for the monophyletic grouping of Bacteroidetes and Firmicutes phyla (>90%), whereas the Flavobacteria and Epsilonproteobacteria classes both received 100% probability for monophyletic groupings, the latter being

dominated by *Pseudomonas* isolates. At lower taxonomic levels, strong support was observed for 2 orders (Pseudomonadales and Flavobacteriales at 93% and 99%, respectively) and 6 different families. Only *Ensifer*, *Cupriavidus*, and *Shewanella* genera received low to zero probability for monophyly, whereas *Pseudomonas* received low support (27%) when analyzed using nucleotide sequences. The archaeal representatives appeared separately in distant regions of the tree, with *nosZ* from *Haloarcula marismortui* forming a single branch between the Flavobacteria and proteobacterial clades. A majority of *nosZ* genes from the Betaproteobacteria clustered in a well-supported clade, although *Dechloromonas aromatica* grouped with *Magnetospirillum* in the same clade as the Epsilonproteobacteria, and all *Achromobacter* isolates were inserted into the Alphaproteobacteria clade.

Inspection of the 16S rRNA trees in the *nosZ* data set showed no support for the monophyly of the proteobacteria. Several bacterial orders, including the Alteromonadales, Oceanospirillales, Rhizobiales, and Rhodocyclales, received low to zero support for monophyly. It should be noted here that Zumft and Kroneck (2007) pointed out that *A. cycloclastes* IAM-1013 has been incorrectly named and is actually a species of *Ensifer*. This is also seen in our 16S rRNA tree, and we therefore classified this species as

Ensifer. The other *Achromobacter* isolates, however, grouped accordingly within the Betaproteobacteria in the 16S rRNA tree, and monophyly of all groupings at the class level was well supported (93–100%).

norB Variants Have Different Evolutionary Histories

The *norB* data set consisted of 181 genotypes, and *cnorB* and *qnorB* sequences formed distinct clades in both PHYML and MrBayes trees, with high overall node support in the basal region of the tree (supplementary fig. S4, Supplementary Material online). As expected, the divergence of the 2 *norB* types led to no supported monophyletic groupings at family, order, or class taxonomic levels. Nevertheless, the Alphaproteobacteria only appeared in the *cnorB* clade, whereas the Beta-, Gamma-, and Epsilonproteobacteria had either *cnorB* or *qnorB* or copies of both types in the same organism. Monophyly of the archaea, which all had *qnorB*, was not recovered because *H. marismortui* did not cluster with the thermophilic archaeal species but instead formed an early-diverging clade with members of the Beta- and Gammaproteobacteria. Disagreement with the 16S rRNA gene phylogeny was much more pronounced in the *qnorB* compared with the *cnorB* clade, particularly within the Beta- and Gammaproteobacteria. Although the Alphaproteobacteria formed a single cluster within the *cnorB* clade, the insertion of *nor* genes from *Pseudomonas* and *Acidovorax* in well-supported clades resulted in no support for monophyly of the Alphaproteobacteria. In addition, species from the Aquificae and Flavobacteria groups formed an early-diverging cluster with several proteobacterial species. Concurrent analysis of the matching 16S rRNA data set showed the same trend as in other data sets, with only the Rhizobiales, Alteromonadales, and the Rhodocyclales orders having low to zero support for monophyly.

Analysis of Pathways in Genomes—Phylogenetic Networks and Likelihood Ratio Tests

Only genomes that contained *nir*, *norB*, and *nosZ* genes were analyzed to compare gene phylogeny in the pathway starting with nitrite reduction. We did not find any sequenced genome that contained both *nirS* and *nirK*. Thus, sequences for the 16S rRNA, *norB* and *nosZ* genes were divided into 2 data sets, corresponding to 23 and 20 genomes containing either *nirK* or *nirS*, respectively (supplementary table S1, Supplementary Material online). These 2 sets are referred to as *nirK* and *nirS* genomes throughout the text. Both sets were dominated by Alpha-, Beta-, and Gammaproteobacteria, with one archaeal representative in each data set. The *nirK* set also had members of the Firmicutes and cytophaga-flavobacterium-bacteroides (CFB) groups (*Geobacillus thermodenitrificans* and *Flavobacterium* sp., respectively), whereas the *nirS* set contained one Epsilonproteobacteria (*S. denitrificans*). Five genomes were found with multiple denitrification gene copies. Duplicate copies of *nirS* were found in *T. denitrificans* and *Dechloromonas aromaticum*, whereas *M. magneticum*

had 3 *nirS* genes. The *T. denitrificans* genome also contained duplicate *cnorB* copies, whereas one copy of both *cnorB* and *qnorB* were found in *Hahella chejuensis*. Finally, 2 copies of *nosZ* were found in the *S. denitrificans* genome.

Phylogenetic networks were constructed to explore conflicting tree topologies between denitrification and 16S rRNA gene phylogenies. Networks of concatenated *nirK*, *norB*, and *nosZ* sequence alignments of the *nirK* genomes showed that a majority of the genotypes clustered within their respective taxonomic classes (fig. 2*a* and *b*), whereas *nirS* genomes were less resolved, particularly among organisms with instances of multiple *nirS* and *norB* copies (fig. 2*c* and *d*). The Alphaproteobacteria clustered together with no interior edges in the networks of both the *nirK* and the *nirS* concatenated data sets. The exception was *M. magneticum* AMB-1, which grouped with *Dechloromonas* with interior edges (fig. 2*d*). In the *nirK* genomes, type II *nirK* occurred in the Betaproteobacteria and CFB classes, whereas the type I *nirK* was found in the Alpha- and Gammaproteobacteria, as well as in *H. marismortui*. Also, the *norB* types divided into 2 groups, such that *cnorB* was exclusively found within the Alphaproteobacteria and *qnorB* within the Betaproteobacteria. Both *nor* types were found within the Gammaproteobacteria and Flavobacteria, possibly giving rise to the interior edges observed in the network. The same pattern in the distribution of *norB* among different taxonomic classes was seen in the *nirS* genomes, where the Alphaproteobacteria only possessed *cnorB*. As expected, the 2 16S rRNA gene data sets showed tree-like configurations, although the Beta- and Gammaproteobacteria clades of the *nirS* genomes were unresolved (fig. 2*a* and *c*).

Despite the apparent similarity between gene phylogeny from the concatenated alignments and 16S rRNA phylogeny, PHYML bootstrap consensus trees, inferred from individual genes, were significantly different as determined by SH tests (data not shown). The only exception to this was the *nosZ* tree within the *nirK*-harboring genomes, which was not significantly different from the 16S rRNA gene phylogeny ($P = 0.14$).

Gene Location in Genomes

Examination of genome features revealed numerous insertion sequences in *Hahella*, *Dechloromonas*, *Magnetospirillum*, and *Thiobacillus*, whereas no indications of recent HGT events could be observed in the *Sulfurimonas* genome. Both *norB* variants in *H. chejuensis* were located outside of genomic islands, that is, annotated HGT regions (Jeong et al. 2005), as were both *cnorB* copies in *Thiobacillus* (Beller et al. 2006). All copies of *nirS* in *Magnetospirillum* (Matsunaga et al. 2005), *Dechloromonas*, and *Thiobacillus* were also outside of potentially transferred regions. In all 3 genomes, the *nirS* copies were positioned relatively close together (ca. 50 kb in *Dechloromonas*; 10 kb in *Magnetospirillum*; and 5 kb in *Thiobacillus*), with the exception of the third copy of *nirS* in *Magnetospirillum* (fig. 3*a*). The 2 *nosZ* copies in *Sulfurimonas* were spaced far from each other in the chromosome (ca. 500 kb), but both copies displayed approximately the same deviation from

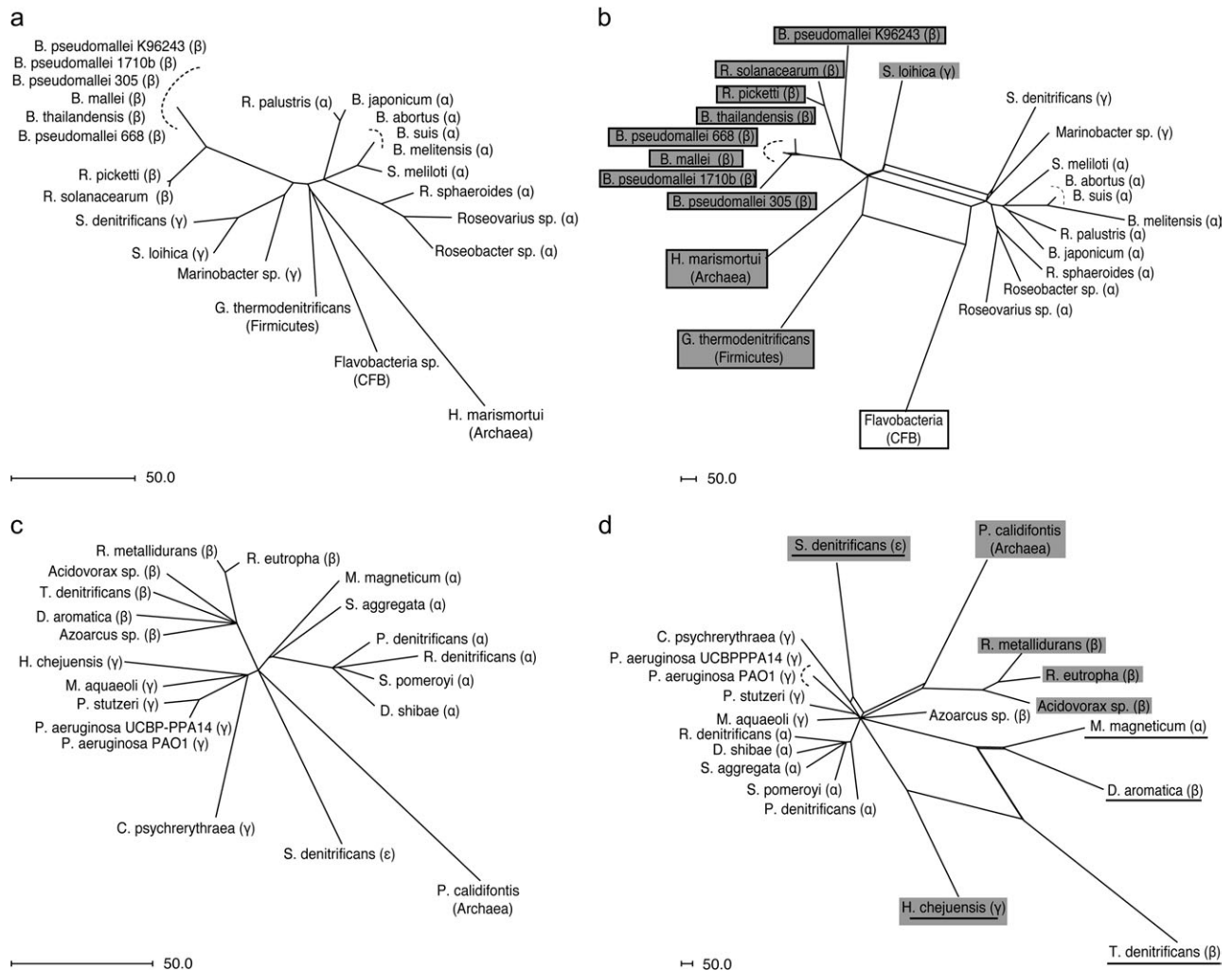


FIG. 2.—Phylogenetic networks based on parsimony splits analysis of full-length 16S rRNA and functional gene nucleotide data sets from genomes. Networks are presented as 95% confidence sets of bootstrap replicates ($n = 1000$). Species names are followed by class/phylum designations. (a) 16S rRNA gene network of genomes with *nirK*. (b) Concatenated *nirK*, *norB*, and *nosZ* data sets. Species with *qnorB* are shaded gray, whereas those with *cnorB* are unshaded. Boxed names indicate species with type II *nirK*. (c) 16S rRNA gene network of genomes with *nirS*. (d) Concatenated *nirS*, *norB*, and *nosZ* data sets. Species with *qnorB* are shaded gray; species with *cnorB* are unshaded. Underlined names indicate species with multiples copies of functional genes: *Sulfurimonas denitrificans*, duplicate *nosZ*; *Thiobacillus denitrificans*, duplicate *cnorB* and *nirS*; *Magnetospirillum magneticum*, 3 *nirS* copies; and *Haehella chejuensis*, single *cnorB* and *qnorB*. See supplementary table S1 (Supplementary Material online) for complete species names.

average G + C content. Additional analysis using the SIGI-HMM algorithm implemented in the JPGV server, which detects nonnative regions within the chromosome based on anomalous codon usage, indicated that all copies of denitrification genes in each genome were most likely native genes within each genome, despite the different phylogenetic signals observed among different copies.

The location of functional genes in the chromosomes indicated that phylogenetic relatedness of *nir* and *norB* genes between 2 species does not correspond to genetic linkage. For example, *Bradyrhizobium* and *Brucella* group relatively close together in the *nirK* and *norB* trees from genome sequences (fig. 4b). However, *nirK*, *norB*, *nosZ*, and all accessory genes required for the reduction of NO_3^- to N_2 are found within an ~50-Kbp region of the chromosome in all *Brucella* species, whereas spans greater

than 200 Kbp separate *nir*, *nor*, and *nos* genes in the *Bradyrhizobium* chromosome (fig. 4b). Similarly, *nirS*, *norB*, and accessory regulatory genes can be found within an 80-Kbp region in the *Dechloromonas* chromosome, whereas they are more than 400 Kbp apart in *Thiobacillus*, despite the clustering of both species in *nirS* (fig. 3a) and *norB* (data not shown) trees inferred using full-length sequences from *nirS* genomes.

Discussion

The analysis of individual genes in the first part of this study incorporates sequence data from denitrifiers, possessing the genes encoding all 3 reductive steps, and from organisms that may possess only the partial pathway.

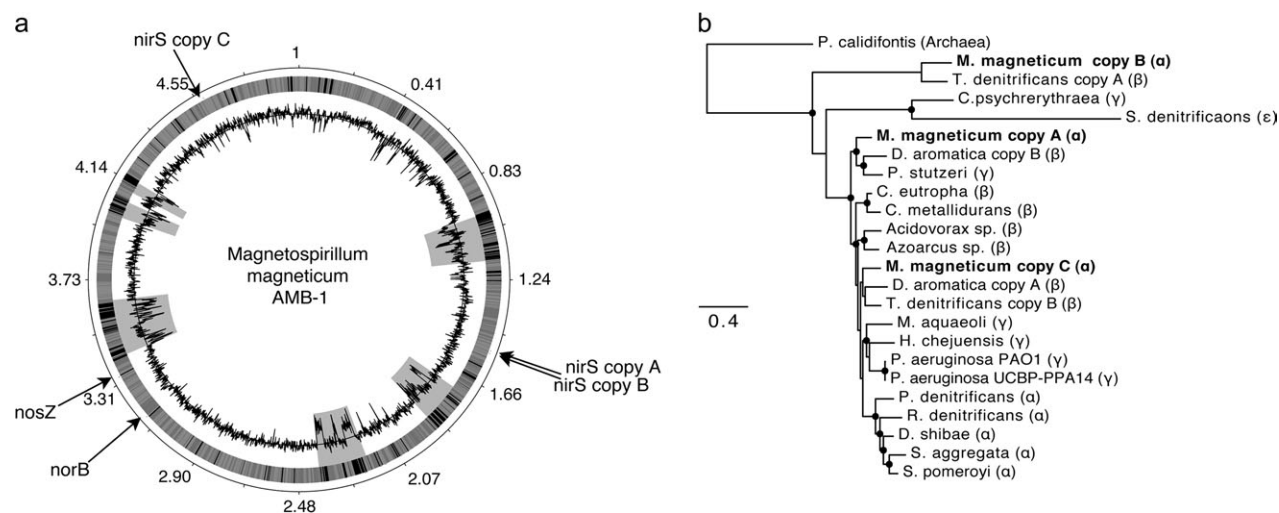


FIG. 3.—(a) Chromosome map of *Magnetospirillum magneticum* AMB-1 indicating positions of multiple *nirS* copies, *norB*, and *nosZ* in the genome. Inner circle, G + C content (1,000-bp window); second circle, codon usage based on SIGI-HMM output; and outer circle, distance from replication origin in mega base pair. Darker shading in coding sequence map denotes genes with nonnative codon usage. (b) Bootstrap PHYML tree of full-length *nirS* nucleotide sequences from genomes. Nodes with bootstrap probabilities >50% are indicated by symbols. *Magnetospirillum magneticum* *nirS* sequences are indicated in bold in the tree, and species names are followed by class/phylum designations. See supplementary table S1 (Supplementary Material online) for complete species names.

Although sequence data from genome sequencing projects are typically quite reliable, the use of partial sequences obtained from studies using PCR amplification requires extra vigilance as a large amount of variation may exist due to sequence quality and the identification of isolates. One issue that was encountered in this study is the potential for improper classification of bacterial isolates based on analysis of 16S rRNA. Most misclassifications are due to the inability to accurately resolve closely related organisms using the 16S rRNA alone, as is the case for the *Rhizobium*, *Sinorhizobium*, *Ensifer*, and *Agrobacterium* genera (Willems 2006), as well as *Ochrrobacterium* and *Brucella* (Moreno et al. 2002). This was reflected in our 16S rRNA trees as monophyly of these genera was often poorly supported or nonexistent. This, in addition to several instances where GenBank taxonomy was wildly different from RDP classification, underscores the difficulties encountered when using partial sequence data sets to infer complex evolutionary relationships.

Previous reports on the enzyme structure of NirK have indicated that 2 classes of the enzyme exist, with differing levels of species diversity between them (Boulanger and Murphy 2002). The structure of a group II NirK from *Hyphomicrobium* has been described in detail (Nojiri et al. 2007), and a recent survey of genome sequences has revealed widespread occurrence of this class across a diverse range of species (Ellis et al. 2007). The division between the 2 groups was detected in our analysis of partial sequences containing the type II copper-binding regions, and we observed that all *nirK* sequences amplified from bacterial isolates were group I *nirK* genes. This may explain the lower level of sequence diversity observed within the group I clade, as currently published primer combinations only target group I *nirK* (Throback et al. 2004). Although it is tempting to explain the distribution of *nirK* groups as

a function of habitat selection or differences in life cycles, we observed that both soil bacteria and pathogenic species clustered together, as well as mesophilic and extremophilic isolates. Additionally, no clear pattern is observed between the distribution of *nirK* types and organisms that have partial versus complete denitrification pathways as clades for both *nirK* types include organisms with or without *nosZ*. These results, in combination with recent reports on structural differences between the 2 groups, suggest that other evolutionary phenomenon, such as convergent evolution or lineage sorting of *nirK* structural variants, plays a role for the observed differences in *nirK* phylogeny and species taxonomy. This is further illustrated by previous research on the evolution of multi-copper oxidases, where the diverse range of functions found within this protein family could be partly attributed to evolution of duplicated and rearranged domains (Murphy et al. 1997; Nakamura et al. 2003). Although these observations do not rule out HGT as a cause of discordance between the phylogenies of *nirK* and 16S rRNA genes, the possibility of other events in the evolution of the 2 *nirK* types must also be taken into account.

Alternative possibilities for *nir* gene evolution are highlighted by recent studies demonstrating that both NirK and NirS may have additional or alternative enzymatic functions. One example is the ability of NirK to reduce toxic selenite to elemental selenium in a *Rhizobium* isolated from soils with high selenite content (Basaglia et al. 2007). Another example was that NirS in *Roseobacter denitrificans* and *Magnetospirillum* sp. have been assigned O₂ reductase and Fe(II) NO₂ oxidoreductase activities, respectively, in addition to dissimilatory nitrite reduction (Fukumori et al. 1997; Rinaldo and Cutruzzola 2007). Bayesian analysis of partial sequences and likelihood ratio tests of full-length genes revealed that differences in *nirS* and 16S rRNA

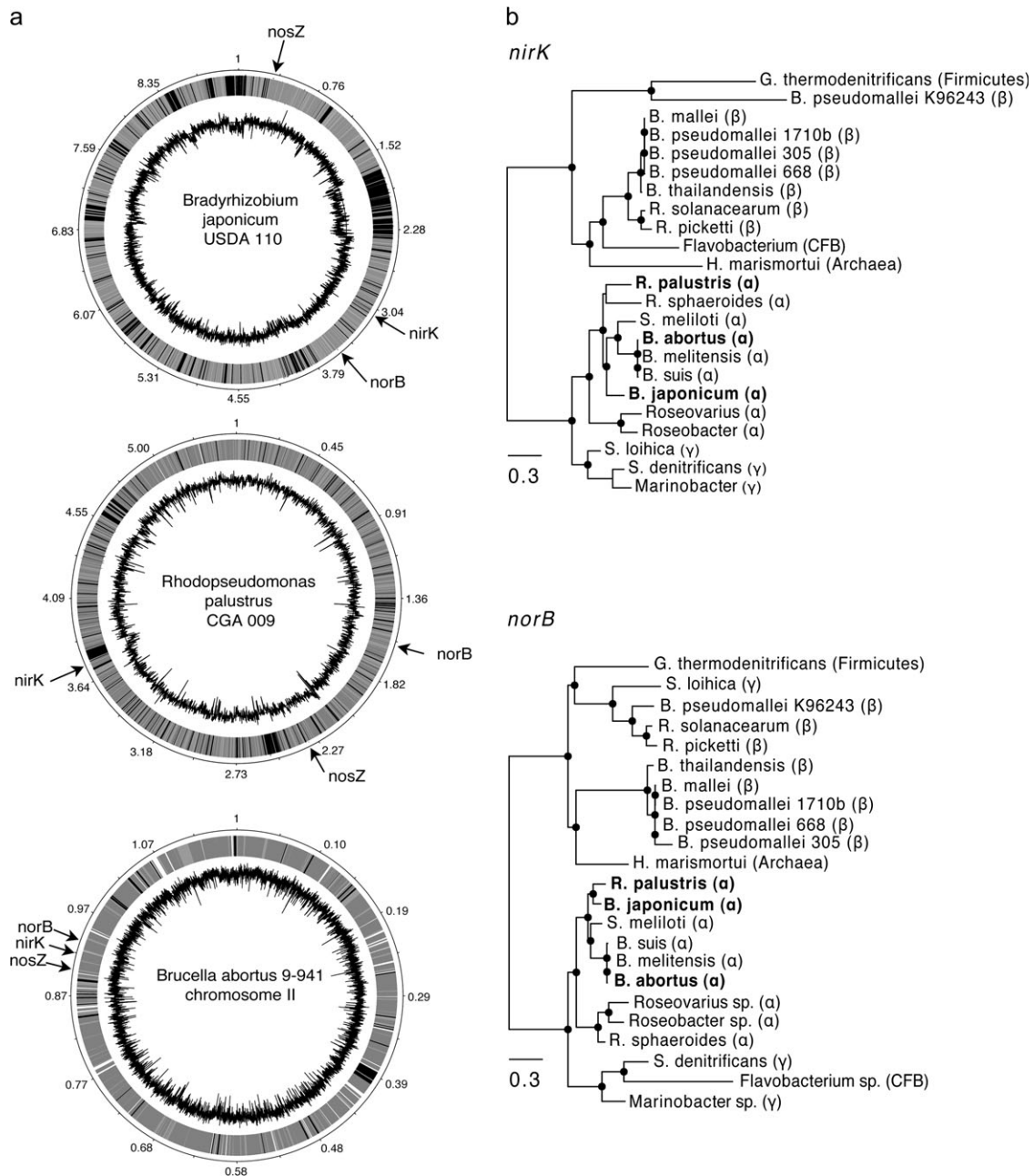


FIG. 4.—(a) Chromosome maps of *Bradyrhizobium japonicum* USDA 110, *Rhodopseudomonas palustris* CGA-009, and *Brucella abortus* 9-941 chromosome II comparing relative positions of *nirK*, *norB*, and *nosZ* genes, respectively, in genomes. Inner circle, G + C content (1,000-bp window; 100-bp window in *B. abortus*); second circle, codon usage based on SIGI-HMM output; outer circle, distance from replication origin in mega base pair. Darker shading in coding sequence map denotes genes with nonnative codon usage. (b) Bootstrap PHYML tree of full-length *nirK* and *norB* nucleotide sequences from *nirK*-harboring genomes. Nodes with bootstrap probabilities >50% are indicated by symbols. Sequences from genomes in (a) are indicated in bold in the tree, and species names are followed by class/phylum designations. See supplementary table S1 (Supplementary Material online) for complete species names.

phylogenies were indeed significant, although visual comparison between the *nirS* and the 16S rRNA gene phylogenies indicated a vertical inheritance pattern with instances of HGT events. Disagreement among the *nirS* genomes was mainly due to the different phylogenetic signals from multiple copies of *nirS* and *norB* genes, and it is appealing to conclude that copies grouping with distantly related species might have been horizontally transferred. However, analysis of genome features suggests that HGT events were either too

ancient to be detected using these methods or that other factors, such as selection pressure for alternative functions, may have lead to the discordant *nirS* groupings observed. The close proximity of duplicate *nirS* copies in the genomes of several organisms makes HGT less likely and duplication and divergence of copies a more plausible scenario. This could be the explanation to the 2 functional copies of *nirS* existing in a *Thauera* isolate, giving it a fitness advantage over isolates with a single copy (Etchebehere and Tiedje

2005). These 2 copies had different phylogenetic signals, similar to the multiple *nirS* copies observed in the *Magneto-spirillum*, *Thiobacillus*, and *Dechloromonas* genomes.

Previous phylogenetic analyses of *norB* have combined the 2 variants, *cnorB* and *qnorB* (Braker and Tiedje 2003; Heylen et al. 2007), and our analysis followed this convention for consistency with previous reports. Although these 2 proteins have been considered to be homologs rather than the same enzyme according to Zumft (2005), it has also been proposed that *qNor* resulted from a fusion of the *norC* and *norB* genes, where the *norC*-like region evolved from a heme c anchoring domain to a quinol-binding site (de Vries et al. 2007). Domains containing the copper-binding and heme-binding histidine residues were conserved within in the partial sequence alignment and flanked by several membrane-spanning alpha helices that also appeared to be relatively conserved between both *norB* types. Our database search resulted in *cnorB* being monophyletic for Alphaproteobacteria and no *qnorB* was found in any representative of this group. This was also observed in the network analysis, which contributed to the tree-like grouping of Alphaproteobacteria in both the *nirK* and *nirS* genome data sets. In agreement, Heylen et al. (2007) found that the 2 *nor* types were evenly distributed among the Betaproteobacteria, whereas the Alphaproteobacteria only had *cnorB*. In combination with the greater number of polyphyletic groupings observed within the proteobacteria in the *qnorB* variant, this indicates potential lineage sorting in addition to HGT events that occurred early in proteobacterial evolution. It is difficult to estimate which may have played the more dominant role, because it is not possible to distinguish ancient HGT events from lineage sorting using a phylogenetic approach, and we did not see evidence of potentially transferred *norB* genes in the genome features of species with multiple *norB* copies. Nevertheless, the greater extent of polyphyly in the *qnorB* clade, in combination with its apparent absence in the Alphaproteobacteria, indicates a different evolutionary trajectory than *cnorB*.

It has been hypothesized that *nosZ* is a highly likely candidate for HGT among denitrification genes, based on the evidence that *nosZ* has been found on plasmids, and is absent in some bacterial and archaeal species that harbor *nir* and *nor* genes (Zumft 1997). In some species, denitrification stops at N_2O , which would account for the nearly one-third of genomes found to contain *nir* and *nor*, but lacking *nos* in our database search. Additionally, the reduction of N_2O to N_2 , catalyzed by *nosZ*, contributes relatively little to the overall bioenergetic needs of denitrifying organisms. Thus, the loss of *nosZ* should not hinder endurance in anoxic conditions. Nevertheless, the *nosZ* partial sequence data set showed the greatest level of congruence with 16S rRNA-based hierarchical taxonomic classification, as noted in previous studies (Dandie et al. 2007; Zumft and Körner 2007). In fact, only a few groupings disagreed with taxonomic classification in our analysis. These rare exceptions were recently discussed by Zumft and Kroneck (2007) and could be consolidated by the NosDFYL phylogenies in the *nos* gene clusters. The polyphyly of the archaeal isolates is most likely attributed to the existence of different structural variants of the *nos* gene (*Z*, *H*, and *A*), distin-

guished by differences in signal peptide and C-terminal domains (Zumft and Körner 2007). *Pyrobaculum* sp. is thought to contain the A type *nos*, whereas *Haloarcula* possesses the Z type. Whether *Haloarcula* acquired its *nos* through lateral gene transfer is difficult to determine as genome features do not indicate HGT. However, genes encoding a transposase, recombinase, and a putative plasmid transfer protein were found to be located within close proximity (ca. 20 Kbp) to *nosZ* and its accessory genes in the chromosome. Zumft and Körner (2007) speculated that *nosZ* evolution, while subject to HGT events, may also be under selective pressure that limits inconsistencies with species taxonomy among distantly related groups. For example, internal selection pressure by protein interactions has been shown to govern the conservation of clusters of photosynthetic genes in cyanobacteria (Shi et al. 2005). One interesting pattern we observed among the genomes within our data set was that those with *nirS* also carried *nosZ*, with few exceptions (*Cupriavidus eutropha* JMP134), whereas this varied more among genomes with *nirK*. Because we observed several species where the *nirS* and *nosZ* genes are quite distant from each other on the chromosome, this cannot be reasonably explained by genetic linkage. Again, there may be other evolutionary mechanisms at work that selects for organisms with both *nirS* and *nosZ*, whereas those with *nirK* may be able to better utilize the truncated pathway under denitrifying conditions. This is particularly interesting when considering the higher level of congruence observed between the *nirS* and 16S rRNA phylogenies described above.

Looking in detail at the *Pseudomonas* genus, we observed that this clade in the *nosZ* and 16S rRNA trees showed significant differences in groupings of isolates similar to that described by Dandie et al. (2007). They found that *nosZ* and 16S rRNA gene phylogenies were incongruent within the genus, as isolates from similar species groups clustered together in the *nosZ* phylogeny, yet the overall grouping of species within each tree differed. In agreement, Delorme et al. (2003) found no similarity between 16S rRNA and *nosZ* phylogeny among a group of closely related fluorescent *Pseudomonas*. However, true evolutionary relationships among *Pseudomonas* species can be very difficult to ascertain, even with robust 16S rRNA sequence data. This again illustrates that questions relating to HGT can only be answered speculatively without additional information from whole genomes. Altogether, these results suggest that if HGT events have occurred in the evolution of *nosZ*, they most likely remain within more closely related organisms, although transfers between major bacterial clades have occurred much less frequently.

Coclustering of *nor* and *nir* genes inferred from trees was taxon specific, with no apparent relationship between taxonomic grouping. Congruence of *nir* and *norB* phylogenies has previously been taken as support for genetic linkage of these genes (Heylen et al. 2007), but the location of denitrification genes in the chromosomes, as observed in this study, indicated that phylogenetic relatedness of *nir* and *norB* genes between organisms does not correspond to genetic linkage. Zumft (2005) suggested independent evolutionary trajectories for *nir* and *nor* genes, which is supported by the high mobility of *nir* genes. This is further

underlined by the existence of genomes with *nor* present but not *nir*. Here, *Nor* might rather play a detoxifying role, and the lack of *nirK* or *nirS* seems inherent rather than the result of secondary evolutionary events (Zumft 2005). Nevertheless, even though not frequent, *nir* and *nor* are genetically linked in certain species, but this is likely taxon dependent. The *nir*–*nor* link in both *Brucella* and *Pyrobaculum* indicates an earlier evolution than in the archaea, although a very early HGT event cannot be excluded. Some bacteria even harbor gene clusters containing genes encoding all 4 reductive steps, for example, the Alphaproteobacteria *Sinorhizobium meliloti* and *Brucella*. This favors the hypothesis of a “denitrification island” that can be transferred horizontally (Zumft 1997).

The results from this study indicate that the evolution of the genes in the denitrification pathway may have several driving forces, including HGT, convergent evolution of different structural types, and lineage sorting. Another possibility for emergence of new genotypes is gene duplication and divergence events within the genome (Bergthorsson et al. 2007), as suggested by the presence of multiple copies of denitrification genes, each with different phylogenetic signals, in adjacent positions in some genomes. This, in combination with recent studies that have indicated alternative functions for the *nir* genes, raises new questions on how environmental factors influence the evolution of the denitrification genes. For example, what happens if a species requires the denitrification genes more often, due to higher frequency of anoxic conditions, than a closely related organism in a different environment? Recent evidence that gene expression levels may influence protein evolutionary rates independent of codon bias (Sallstrom et al. 2006), let us speculate that the less often used gene copy would be subject to a higher frequency of nonsynonymous substitutions than the one with higher expression levels, leading to a different phylogenetic signal over a number of generations. This could explain that denitrification gene phylotype diversity has been shown to be both higher (Delorme et al. 2003; Goregues et al. 2005) and lower (Heylen et al. 2006) than species diversity in bacterial strains isolated from different environments. Experiments aimed at examining these issues, using denitrification as a model facultative pathway, may provide further insight into the mechanism of molecular evolution in response to environmental change.

Supplementary Material

Supplementary table S1 and figures S1–S4 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We would like to thank Micke Tolleson and Rolf Bernander for helpful discussions, as well as Ann-Charlotte Berglund Sonnhammer at the UPPMAX high performance computing facility at Uppsala University, Sweden, for assistance with using UPPMAX resources for phylogenetic analysis. This work was supported by the Swedish

Research Council and the Formas financed Uppsala Microbiomics Center.

Literature Cited

- Altschul S, Madden T, Schaffer A, Zhang JH, Zhang Z, Miller W, Lipman D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Basaglia M, Toffanin A, Baldan E, Bottegai M, Shapleigh JP, Casella S. 2007. Selenite-reducing capacity of the copper-containing nitrite reductase of *Rhizobium sullae*. *FEMS Microbiol Lett.* 269:124–130.
- Beller HR, Chain PSG, Letain TE, Chakicherla A, Larimer FW, Richardson PM, Coleman MA, Wood AP, Kelly DP. 2006. The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium *Thiobacillus denitrificans*. *J Bacteriol.* 188:1473–1488.
- Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno’s dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci USA.* 104:17004–17009.
- Bininda-Emonds ORP. 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics.* 6:156.
- Boulanger MJ, Murphy MEP. 2002. Crystal structure of the soluble domain of the major anaerobically induced outer membrane protein (AniA) from pathogenic *Neisseria*: a new class of copper-containing nitrite reductases. *J Mol Biol.* 315:1111–1127.
- Braker G, Tiedje JM. 2003. Nitric oxide reductase (*norB*) genes from pure cultures and environmental samples. *Appl Environ Microbiol.* 69:3476–3483.
- Clays-Josserand A, Ghiglione JF, Philippot L, Lemanceau P, Lensi R. 1999. Effect of soil type and plant species on the fluorescent pseudomonads nitrate dissimilating community. *Plant Soil.* 209:275–282.
- Cole JR, Chai B, Marsh TL. 2003. The ribosomal database project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res.* 31:442–443.
- Dandie CE, Burton DL, Zebarth BJ, Trevors JT, Goyer C. 2007. Analysis of denitrification genes and comparison of *nosZ*, *cnorB* and 16S rDNA from culturable denitrifying bacteria in potato cropping systems. *Syst Appl Microbiol.* 30:128–138.
- Delorme S, Philippot L, Edel-Hermann V, Deulvot C, Mougel C, Lemanceau P. 2003. Comparative genetic diversity of the *narG*, *nosZ*, and 16S rRNA genes in fluorescent pseudomonads. *Appl Environ Microbiol.* 69:1004–1012.
- de Vries S, Pouvreau S, Pouvreau LAM. 2007. Nitric oxide reductase: structural variations and catalytic mechanism. In: Bothe H, Ferguson SJ, Newton WE, editors. *Biology of the nitrogen cycle*. Amsterdam (The Netherlands): Elsevier. p. 57–66.
- Ellis MJ, Grossmann JG, Eady RR, Hasnain SS. 2007. Genomic analysis reveals widespread occurrence of new classes of copper nitrite reductases. *J Biol Inorg Chem.* 12:1119–1127.
- Etchebehere C, Tiedje J. 2005. Presence of two different active *nirS* nitrite reductase genes in a denitrifying *Thauera* sp. from a high-nitrate-removal-rate reactor. *Appl Environ Microbiol.* 71:5642–5645.
- Fukumori Y, Oyanagi H, Yoshimatsu K, Noguchi Y, Fujiwara T. 1997. Enzymatic iron oxidation and reduction in magnetite synthesizing *Magnetospirillum magnetotacticum*. *J Phys IV.* 7:659–662.

- Galtier N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol Biol Evol.* 18:866–873.
- Garrity GM, Lilburn TG, Cole JR, Harrison SH, Euzeby J, Tindall BJ. 2007. The Taxonomic Outline of Bacteria and Archaea. TOBA release 7.7. East Lansing (MI): Michigan State University Board of Trustees.
- Goregues CM, Michotey VD, Bonin PC. 2005. Molecular, biochemical, and physiological approaches for understanding the ecology of denitrification. *Microbial Ecol.* 49:198–208.
- Gregory LG, Bond PL, Richardson DJ, Spiro S. 2003. Characterization of a nitrate-respiring bacterial community using the nitrate reductase gene (*narG*) as a functional marker. *Microbiology.* 149:229–237.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
- Heylen K, Gevers D, Vanparys B, Wittebolle L, Geets J, Boon N, De Vos P. 2006. The incidence of *nirS* and *nirK* and their genetic heterogeneity in cultivated denitrifiers. *Environ Microbiol.* 8:2012–2021.
- Heylen K, Vanparys B, Gevers D, Wittebolle L, Boon N, De Vos P. 2007. Nitric oxide reductase (*norB*) gene sequence analysis reveals discrepancies with nitrite reductase (*nir*) gene phylogeny in cultivated denitrifiers. *Environ Microbiol.* 9:1072–1077.
- Huelsenbeck JP, Hillis DM. 1993. Success of phylogenetic methods in the 4-taxon case. *Syst Biol.* 42:247–264.
- Huelsenbeck JP, Larget B, Miller RE, Ronquist F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst Biol.* 51:673–688.
- Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics.* 14:68–73.
- Jeong H, Yim JH, Lee C. 2005. Genomic blueprint of *Hahella chejuensis*, a marine microbe producing an algicidal agent. *Nucleic Acids Res.* 33:7066–7073.
- Knowles R. 1982. Denitrification. *Microbiol Rev.* 46:43–70.
- Matsunaga T, Okamura Y, Fukuda Y, Wahyudi AT, Murase Y, Takeyama H. 2005. Complete genome sequence of the facultative anaerobic magnetotactic bacterium *Magnetospirillum* sp. strain AMB-1. *DNA Res.* 12:157–166.
- Moreno E, Cloeckaert A, Moriyon I. 2002. *Brucella* evolution and taxonomy. *Vet Microbiol.* 90:209–227.
- Murphy MEP, Lindley PF, Adman ET. 1997. Structural comparison of cupredoxin domains: domain recycling to construct proteins with novel functions. *Protein Sci.* 6:761–770.
- Nakamura K, Kawabata T, Yura K, Go N. 2003. Novel types of two-domain multi-copper oxidases: possible missing links in the evolution. *FEBS Lett.* 553:239–244.
- Nojiri M, Xie Y, Inoue T, Yamamoto T, Matsumura H, Kataoka K, Deligeer, Yamaguchi K, Kai Y, Suzuki S. 2007. Structure and function of a hexameric copper-containing nitrite reductase. *Proc Natl Acad Sci USA.* 104:4315–4320.
- Philippot L. 2002. Denitrifying genes in bacterial and archaeal genomes. *Biochim Biophys Acta Gene Struct Expr.* 1577:355–376.
- Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 14:817–818.
- Rinaldo S, Cutruzzola F. 2007. Nitrite reductases in denitrification. In: Bothe H, Ferguson SJ, Newton WE, editors. *Biology of the nitrogen cycle.* Amsterdam (The Netherlands): Elsevier. p. 37–55.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 19:1572–1574.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics.* 16:944–945.
- Sallstrom B, Arnaut RA, Davids W, Bjelkmar P, Andersson SGE. 2006. Protein evolutionary rates correlate with expression independently of synonymous substitutions in *Helicobacter pylori*. *J Mol Evol.* 62:600–614.
- Schoniger M, Von Haeseler A. 1994. A stochastic model for the evolution of autocorrelated DNA sequences. *Mol Phylogenet Evol.* 3:240–247.
- Shi T, Bibby TS, Jiang L, Irwin AJ, Falkowski PG. 2005. Protein interactions limit the rate of evolution of photosynthetic genes in cyanobacteria. *Mol Biol Evol.* 22:2179–2189.
- Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol.* 16:1114–1116.
- Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics.* 17:1246–1247.
- Suharti, Strampraad MJF, Schroder I, de Vries S. 2001. A novel copper A containing menaquinol NO reductase from *Bacillus azotoformans*. *Biochemistry.* 40:2632–2639.
- Swofford DL. 2000. PAUP*. Phylogenetic inference using maximum parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
- Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Throbäck IN, Enwall K, Jarvis A, Hallin S. 2004. Reassessing PCR primers targeting *nirS*, *nirK* and *nosZ* genes for community surveys of denitrifying bacteria with DGGE. *FEMS Microbiol Ecol.* 49:401–417.
- Tiedje JM. 1988. Ecology of denitrification and dissimilatory nitrate reduction to ammonium. In: Zehnder AJB, editor. *Ecology of anaerobic microorganisms.* New York: Wiley. p. 179–243.
- Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics.* 7:142.
- Wang Q, Garrity G, Tiedje J, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 73:5261–5267.
- Willems A. 2006. The taxonomy of rhizobia: an overview. *Plant Soil.* 287:3–14.
- Zumft WG. 1997. Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev.* 61:533–616.
- Zumft WG. 2005. Nitric oxide reductases of prokaryotes with emphasis on the respiratory, heme-copper oxidase type. *J Inorg Biochem.* 99:194–215.
- Zumft WG, Körner H. 2007. Nitrous oxides reductases. In: Bothe H, Ferguson SJ, Newton WE, editors. *Biology of the nitrogen cycle.* Amsterdam (The Netherlands): Elsevier. p. 67–81.
- Zumft WG, Kroneck PMH. 2007. Respiratory transformation of nitrous oxide (N₂O) to dinitrogen by Bacteria and Archaea. *Adv Microb Physiol.* 52:107–227.

Jennifer Wernegreen, Associate Editor

Accepted June 25, 2008