Novel nitrite reductase domain structure suggests a chimeric denitrification repertoire in Phylum Chloroflexi

Sarah Schwartz[1], Lily Momper[1], L. Thiberio Rangel[1], Cara Magnabosco[2], Jan Amend[3], and Gregory Fournier[1]

[1]Massachusetts Institute of Technology
[2]ETH Zurich
[3]University of Southern California

June 11, 2021

Abstract

Denitrification plays a central role in the global nitrogen cycle, reducing and removing nitrogen from marine and terrestrial ecosystems. The flux of nitrogen species through this pathway has a widespread impact, affecting ecological carrying capacity, agriculture, and climate. Nitrite reductase (Nir) and nitric oxide reductase (NOR) are the two central enzymes in this pathway. Here we present a previously unreported Nir domain architecture in members of Phylum Chloroflexi. Phylogenetic analyses of protein domains within Nir indicate that an ancestral horizontal transfer and fusion event produced this chimeric domain architecture. We also identify an expanded genomic diversity of a rarely reported nitric oxide reductase subtype, eNOR. Together, these results suggest a greater diversity of denitrification enzyme arrangements exist than have been previously reported.

RESEARCH PAPER

TITLE: Novel nitrite reductase domain structure suggests a chimeric denitrification repertoire in Phylum Chloroflexi

SHORT TITLE: Novel Denitrification Architecture in Chloroflexi

Sarah L. Schwartz[1,2*], Lily M. Momper[2,3], L. Thiberio Rangel[2], Cara Magnabosco[4], Jan P. Amend[5,6], and Gregory P. Fournier[2]

1. Microbiology Graduate Program, Massachusetts Institute of Technology
2. Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology
3. Exponent, Inc., Pasadena, CA
4. Department of Earth Sciences, ETH Zurich
5. Department of Earth Sciences, University of Southern California
6. Department of Biological Sciences, University of Southern California

*Correspondence: sschwart@mit.edu, +1 (415) 497-1747

SUMMARY

Denitrification plays a central role in the global nitrogen cycle, reducing and removing nitrogen from marine and terrestrial ecosystems. The flux of nitrogen species through this pathway has widespread impact, affecting ecological carrying capacity, agriculture, and climate. Nitrite reductase (Nir) and nitric oxide reductase (NOR) are the two central enzymes in this pathway. Here we present a previously unreported Nir domain architecture in members of Phylum Chloroflexi. Phylogenetic analyses of protein domains within Nir

indicate that an ancestral horizontal transfer and fusion event produced this chimeric domain architecture. We also identify an expanded genomic diversity of a rarely reported nitric oxide reductase subtype, eNOR. Together, these results suggest a greater diversity of denitrification enzyme arrangements exist than have been previously reported.

KEYWORDS

Denitrification, phylogeny, Chloroflexi, nitrite reductases, nitric-oxide reductase, cytochromes

INTRODUCTION

Microbial denitrification is a key pathway in global nitrogen cycling and has been studied extensively for its role in fixed nitrogen loss and as a source of potent greenhouse gases (Zumft, 1997; Decleyre *et al.* , 2016). Diverse bacteria are capable of denitrification, often facultatively using nitrate or nitrite as an alternative electron acceptor in oxygen-limited zones. Several diverse microorganisms have the genomic capacity to perform complete denitrification (Fig. 1), reducing nitrate to dinitrogen gas (Philippot, 2002; Canfield, Glazer and Falkowski, 2010).

Figure 1 link

Denitrification has been widely reported in various taxa, and is especially diverse among proteobacterial groups (Zumft, 1997; Philippot, 2002); the utility of the pathway is underscored by the diversity of key constituent enzymes. The canonical denitrification enzyme is nitrite reductase, Nir*,* which reduces nitrite to nitric oxide. Nir functionality is found in two distinct enzymes—the copper-based nitrite reductase NirK, and the cytochrome-type reductase NirS (Braker*et al.* , 2000; Priemé, Braker and Tiedje, 2002; Decleyre *et al.* , 2016).

The next step in the pathway—the reduction of nitric oxide to nitrous oxide—is catalyzed by nitric oxide reductases (NORs). Most bacterial NORs are homologous and closely related to one another, and to oxygen reductases in the heme-copper oxygen reductase superfamily (Hemp and Gennis, 2008). The most widely studied NOR enzymes are cNOR and qNOR (Hendriks *et al.* , 2000; Hemp and Gennis, 2008; Graf, Jones and Hallin, 2014), distinguished by their cytochrome or quinol electron donors, respectively. Rarer, alternative NOR enzymes, including sNOR, gNOR and eNOR, have been more recently identified and characterized in limited members of the Proteobacteria, Firmicutes, Archaea and Chloroflexi (Stein *et al.* , 2007; Hemp and Gennis, 2008; Sievert*et al.* , 2008; Hemp *et al.* , 2015).

Many bacteria contain only one or a partial subset of the four denitrification genes. Such organisms may perform partial denitrification, while others may use one of these enzymes for non-denitrifying functions (Hendriks *et al.* , 2000; Sanford*et al.* , 2012; Graf, Jones and Hallin, 2014; Roco *et al.* , 2017). In partial denitrifers, the co-occurrence of denitrification pathway genes appears to vary across different taxa and environments (Graf, Jones and Hallin, 2014). Some of this variation may be constrained by the chemistry of certain intermediates. For example, nitric oxide (NO), the product of NirS and NirK, is highly cytotoxic. Both Nir types are periplasmic, and so cells require a means of effluxing or detoxifying nitric oxide before it accumulates to lethal levels. Denitrifiers are thought to immediately reduce nitric oxide to nitrous oxide to avoid injury, using membrane-bound NOR enzymes (Hendriks *et al.* , 2000). Perhaps for this reason, it is rare to find genomes that contain *nir* but not *nor* , while organisms showing the inverse—the presence of a *nor* gene but not a*nir* gene—are far more common (Hendriks *et al.* , 2000; Graf, Jones and Hallin, 2014). While cNORs are only found in denitrifying microbes, other types of NOR—for example, quinol-dependent qNOR—are found in non-denitrifiers and can presumably detoxify environmental nitric oxides (Hendriks *et al.* , 2000). Beyond NORs, alternative pathways to nitric oxide detoxification are possible, including alternative enzymes such as cytochrome c oxidase (Blomberg and Ädelroth, 2018) or oxidoreductase (Gardner, Helmick and Gardner, 2002), flavorubredoxin (Gardner, Helmick and Gardner, 2002), or flavohemoglobins (Sánchez *et al.* , 2011).

While denitrification has been most widely studied and observed in Proteobacteria, the process has also been identified in other phyla, including Chloroflexi. Chloroflexi are ecologically and physiologically diverse,

and often key players in oxygen-, nutrient-, and light-limited environments, including anaerobic sludge and subsurface sediments (Hug*et al.* , 2013; Ward *et al.* , 2018). Previous surveys have indicated that, within Chloroflexi, members of Classes Chloroflexia and Thermomicrobia may have the capacity for nitrite reduction via the copper-type NirK (Wei *et al.* , 2015; Decleyre *et al.* , 2016). Recent studies indicate that certain Chloroflexi—especially members of Order Anaerolineales—may possess *nirS* instead of*nirK* (Hemp *et al.* , 2015; Ward, McGlynn and Fischer, 2018), and may also harbor a divergent variant of *nor* previously reported in members of Archaea (Hemp and Gennis, 2008; Hemp *et al.* , 2015). These findings suggest that the evolution and/or biochemistry of denitrification may be unusual for this subset of bacteria, and informative for a broader understanding of microbial denitrification metabolisms and their origin.

RESULTS

To investigate divergent denitrification genes in Chloroflexi, we performed a comprehensive analysis of denitrification homologs in over 100 recently sequenced Chloroflexi metagenome-assembled genomes, as well as previously available genomes and metagenomes from the National Center for Biotechnology Information's (NCBI) protein databases.

Apparent chimeric fusion in Chloroflexi NirS

Domain analysis of the Anaerolineales-type nitrite reductase open reading frame (ORF) from SURF MAG 4indicated three putative functional regions of interest: one cytochrome-type NirS domain and two cytochrome C superfamily domains (Fig. 2, Supplementary Datafile S1).

Figure 2 link

The first cytochrome domain (C1) in the Anaerolineales-type ORF was identified by NCBI's Conserved Domains Database (Marchler-Bauer *et al.* , 2015; Lu *et al.* , 2020) as a cytochrome c551/552 (NCBI COG4654, e-value = 6.4×10^{-4}). The second cytochrome domain (C2) was predicted with high specificity as a cytochrome C mono- and diheme variant (NCBI COG2010, e-value=5.31×10^{-9}). C2 included a region predicted as a cbb3-type cytochrome c oxidase subunit III (pfam 13442, e-value = 4.06×10^{-7}); such subunits frequently contain two cytochromes (Bertini, Cavallaro and Rosato, 2006).

Though the C2 and NirS functional domains frequently co-occur in nitrite reductases, the inclusion of C1 in the ORF appears extremely rare and limited to Chloroflexi. A lineage-specific fusion of multiple gene domains could explain this novel C1-C2-NirS arrangement. Different evolutionary histories among the domain subunits within Chloroflexi would provide evidence for an ancestral horizontal acquisition and fusion event. To determine if the different domains of the enzyme have different ancestry, maximum-likelihood domain trees were independently reconstructed (See Methods) for C1, C2, and the NirS-specific domain.

Domain phylogenies indicate a similar overall topology for C2 (Fig. 3a) and the NirS domains (Fig 3b). These trees both recover a clade of Chloroflexi closely related to a polyphyletic group containing diverse NirS sequences from Aquificae, Bacteroidetes, Epsilonproteobacteria, and Spirochaetia. In both cases, this polyphyletic group is sister to a large grouping of proteobacterial sequences from Alpha-, Beta-, and Gammaproteobacteria. The relative placement of Chloroflexi sequences varies slightly between both domain trees: For C2, Chloroflexi sequences fall within the polyphyletic group comprising Aquificae, Bacteroidetes, Epsilonproteobacteria, and Spirochaetia, with the combined group placing sister to the large proteobacterial group; for NirS, Chloroflexi are sister to the polyphyletic grouping.

Figure 3a link

Figure 3b link

C2 and NirS domain trees reconstructed exclusively from ORFs containing both domains inverts the relative placement of Chloroflexi, suggesting that sampling and phylogenetic noise are likely responsible for the observed differences in these phylogenies (Fig. S3, Fig. S4). Additionally, there are notable differences in placement among subclades within the Proteobacteria, suggesting that patterns unrelated to Chloroflexi evolution may be polarizing the relative placements of groups in the tree. This lack of robustness caused by

alternative sampling, combined with poor support values within the polyphyletic clade or between this clade and the Proteobacteria, suggest that the differences in tree topology may be artefactual, and not reflective of gene reticulation events. There seems to be no reason to reject the null hypothesis that the C2 and NirS domains have the same evolutionary origins within Chloroflexi.

The inferred phylogeny for C1 shows a much different evolutionary history than the other two domains (Fig. 4). In contrast to the domain trees for C2 or NirS, the C1 tree shows sequences from Nitrospirae and Nitrospinae grouping together within a large clade of Chloroflexi C1 domains. Additional Chloroflexi sequences group with a small number of more distantly related Proteobacteria. However, the placement and taxonomic representation of Proteobacteria in the C1 tree is different from that seen in the other domain trees.

Figure 4 link

The majority of ORFs represented in the C1 domain tree contain the C1 domain homolog either as a free cytochrome, or as one of multiple cytochrome-type or cytochrome superfamily domains. In rare or isolated cases, C1 homologs co-occur in ORFs with membrane or structural protein domains (Table S1). The ORF containing the C1 homolog was annotated as a nitric oxide reductase in several members of the Nitrospirae and one Geobacteraceae genome (Table S1); domain analysis of these genes yielded limited additional data, but representative sequences showed detectable sequence similarity to *Pseudomonas norC* genes.

The occurrence of C1 domain homologs within predicted nitrite reductase genes is restricted to the Chloroflexi. The majority of these C1-containing *nir* genes have the cytochrome-type NirS domain; however, a small number of Chloroflexi MAGs contain an ORF pairing the C1 cytochrome with the copper-type NirK domain instead. This NirK ORF also included an N-terminal cupredoxin/plastocyanin domain. As this fusion is only apparent within a small number of MAGs, which are identical across the length of the analyzed ORF, this may represent an assembly artifact. However, several of the *nir* genes that contained a C1 homolog and a cytochrome-type NirS (not copper-type NirK) also contained cupredoxins or other copper-containing domains (Fig. S1).

The distinct phylogeny and taxonomic distribution of C1, as compared with C2 and NirS domains, strongly suggest that the C1-C2-NirS domain structure observed in Chloroflexi is the result of fusion of the C1 domain with a horizontally-acquired Proteobacterial *nir* gene. Though topology and extant taxon sampling of these gene trees does not allow us to reliably infer the donor lineage, the C2-NirS architecture—or similar arrangements of functional domains—is widespread among members of the Alpha-, Beta-, and Gammaproteobacteria. Additionally, it appears that Chloroflexi may have been the source for an independent transfer of the free C1 domain into Nitrospirae and Nitrospinae.

C1-C2-nirS Domain Architecture is Unique to Chloroflexi

Though putative homologs exist independently for the constituent C1 and C2-NirS regions, respectively, these hits reflect different cytochrome or cytochrome-type nitrite reductases (largely in Proteobacteria, Nitrospirae, and Nitrospinae). The full C1-C2-NirS architecture appears unique to Chloroflexi, and is not observed in other groups. Querying NCBI's non-redundant environmental database (env-nr) with the full ORF from SURF MAG 42 did not identify additional examples of the full gene construct. While several hits were identified that reflected putative homology to the joint C2-NirS domains, none of these included the C1 domain as well (Supplementary Datafile S2). An independent search of the env-nr database using the C1 domain as a query returned few overall hits. While some of these putative C1 homologs were identified in ORFs containing additional cytochrome-type enzyme superfamily domains or subunits, none co-occurred with NirS or NirK domains. These data suggest that there is little to no missing diversity of the Chloroflexi-type chimeric nitrite reductase in existing metagenomes.

Attempts to visualize the full enzyme structure using homology modeling (Bienert *et al.* , 2017; Waterhouse *et al.* , 2018) were unsuccessful; structural models were only able to predict a close match for the conserved C2-NirS region of the putative gene (Supplementary Datafile S3). Efforts to independently model the C1

structure could not recover predicted QMEAN scores above -4.50 (Benkert *et al* 2011). The poor scores may reflect the relatively short length of the cytochrome coding region. However, the Chloroflexi *nirS* gene sequence does retain several conserved residues present in the crystal structure of *Pseudomonas aeruginosa* NirS. In *P. aeruginosa*NirS, His51 and Met88 coordinate heme c; His182 coordinates heme d1; and His327 and His369 are believed to stabilize the active site nitrite anion (Rinaldo *et al.* , 2011; Maia and Moura, 2014). Corresponding residues are conserved within the C2 (His65, Met125) and NirS alignments (His46, His239, His300) for the Chloroflexi NirS ORF; interestingly, the residue corresponding to His327 (His239) is not universally conserved, though it is conserved among Chloroflexi with the novel NirS architecture (Supplementary Datafile S1).

Expansion of eNOR diversity

Notably, the majority of genomes with the unique C1-C2-NirS structure do not appear to contain a nitric oxide reductase gene (*nor*). Though the absence of the *nor* gene in genomes with *nirS* is not unprecedented, previous genomic surveys suggest it is relatively uncommon, and the toxicity of the product of Nir (nitric oxide, NO) makes this absence counterintuitive (Hendriks *et al.* , 2000; Graf, Jones and Hallin, 2014). However, analysis of the SURF MAG 42 metagenome—the original assembled genome in which the novel*nirS* ORF was observed—did reveal the presence of an unusual*nor* homolog. Previous studies have identified the established cNOR and qNOR family enzymes (which contain cytochrome c or quinols as electron donors, respectively) in Chloroflexi, as well as a broad distribution of Proteobacteria (Hendriks *et al.* , 2000; Zumft, 2005; Hemp and Gennis, 2008). However, the predicted NOR in SURF MAG 42 included an active site glutamine substitution characteristic of eNOR (Hemp and Gennis, 2008; Hemp *et al.* , 2015) (Table S4). eNOR has been previously described in Archaea (Hemp and Gennis, 2008) and at least one isolated Anaerolineales bacterium (Hemp *et al.* , 2015).

Phylogenetic analysis indicates the presence of *eNOR* in an expanded diversity of genomes (Fig. 5). Previous studies have described eNOR in *Natronomonas* ; these data indicate a cluster of*eNOR* genes throughout other Halobacteria as well. Additional putative *eNOR* genes appear in multiple members of Anaerolineales, as well as other Chloroflexi, and in many members of the Alpha-, Beta-, Gamma-, and Deltaproteobacteria. Many of these putative *eNOR* hits appear to have been misannotated or mislabeled as cytochrome c oxidase genes, likely because of the structural similarity of the heme-copper oxidase subunits (Hemp and Gennis, CDD).

Figure 5 link

DISCUSSION

The phylogenetic analyses of *nirS* and *eNOR* ORFs in Chloroflexi suggest that subsurface ecosystems may harbor an under-described diversity of denitrification enzymes, which may reflect adaptations to the unique challenges of nutrient cycling within these environments. More broadly, a deeper understanding of the ecological extent of microbial denitrification has important implications for basic and applied microbial ecology. The reduction of fixed nitrogen species plays a crucial role in global nitrogen cycling and is also an essential component of smaller-scale systems, such as those associated with agricultural or waste treatment (Butterbach-Bahl and Dannenmann, 2011; Lu, Chandran and Stensel, 2014). The discovery and characterization of novel variants of genes such as *nirS* and *eNOR* may therefore pave the way for future biotechnological applications.

Although the C2 and NirS domains do not have identical evolutionary histories or distributions, the taxonomic representation of these groups is very similar, and the presence of the paired C2-NirS domains in cytochrome-type nitrite reductases appears broadly throughout the Proteobacteria. In contrast, the taxonomic distribution and phylogeny of the C1 domain tree is strikingly different than that of the other domains in the nitrite reductase ORF. Combined with the apparent absence of a full C1-C2-NirS ORF in any taxonomic group other than Chloroflexi, these data suggest that the C1 cytochrome was likely incorporated into*nirS* in a gene fusion event within Chloroflexi, following HGT. As there is no evidence of the C2-NirS ORF in Chloroflexi without the fused C1 domain present, the fusion probably occurred very soon after the

acquisition of the C2-NirS region and may be necessary for the function of the gene in Chloroflexi.

Interestingly, putative homologs of C1 cytochrome domains were found in some Chloroflexi genomes in ORFs containing *nirK* , not *nirS*(Fig. 4, Fig. S1). Though NirS and NirK are functionally equivalent, the two enzymes do not show a shared evolutionary origin, and are often—though not always—mutually exclusive among known denitrifier genomes (Jones *et al.* , 2008; Graf, Jones and Hallin, 2014). Unlike the cytochrome-containing NirS, NirK is a copper-type enzyme. The co-occurrence of cytochrome *c* domains in ORFs with the copper-type *nirK* has been identified in rare instances in Proteobacteria, and noted as surprising, given the cupredoxin-like fold of the NirK enzyme (Bertini, Cavallaro and Rosato, 2006). Similarly surprising is the inverse relationship revealed in the C1 domain tree: Several Chloroflexi ORFs contain a cupredoxin or similar copper-containing domain N-terminal to the C1-C2-NirS architecture (Fig. 4, Fig. S1). The co-occurrence of C1 with both cytochrome- and copper-dependent Nir domains suggests a general evolutionary trend within Chloroflexi to incorporate this cytochrome into denitrification ORFs. This distribution pattern raises the possibility that the C1-type cytochrome may serve an important but generalized role in nitrite reduction—regardless of the evolutionary history or genetic profile of the nitrite reduction domain itself.

The apparent absence of a *nor* homolog in the majority of genomes with the C1-*nirS* fusion is unexpected. Beyond providing downstream redox capacity, nitric oxide reductase provides an efficient means of reducing and detoxifying nitric oxide, the highly cytotoxic product of NirS. It is not unprecedented for bacterial genomes to harbor a *nir* gene without a *nor* gene, particularly for organisms with *nirK* (Heylen *et al.* , 2007; Graf, Jones and Hallin, 2014). This *nir-nor* mismatch is much rarer for putative denitrifiers with *nirS* , representing fewer than 4% of genomes in a recent survey—but a small number of surveyed bacteria do, interestingly, appear to harbor *nirS* without also harboring*cNOR* or *qNOR* (Heylen *et al.* , 2007; Graf, Jones and Ha llin, 2014). To our knowledge, however, *eNOR* has not been included in such analyses of the genomic correlation between nitrite reductases and nitric oxide reductases. The phylogenetic evidence for diverse *eNOR* homologs suggests likely undocumented or underexplored diversity for divergent nitric oxide reductases. Diversity and function of cytochrome-dependent (cNOR) and quinol-dependent nitric oxide reductases (qNOR) are fairly well-established. However, divergent enzymes such as eNOR and sNOR are less-extensively documented and may not be accurately distinguished from broader oxygen reductase superfamily members in genomic or metagenomic analyses.

Cytochrome *c* proteins function as electron transfer proteins in anaerobic respiration and are often fused to redox enzymes to allow electron passage (Bertini, Cavallaro and Rosato, 2006). It is not surprising, therefore, to find cytochrome c-containing subunits in frame with nitrite reductase. NirS itself is cytochrome-dependent (Bertini, Cavallaro and Rosato, 2006). However, the unusual addition of the upstream cytochrome domain (C1) may reflect additional redox requirement or capacity. It is also possible that the inclusion of this construct could be linked to the conspicuous absence of nitric oxide reductase enzymes in several metagenome-assembled genomes containing a NirS ORF with the C1 fusion. Nitric oxide reduction can be cytochrome-dependent; the well-studied cNOR nitric oxide reductases contain a membrane-anchored cytochrome *c* (Hemp and Gennis, 2008). Further, the C1 domain tree recovers ORFs in the Nitrospirae that contain C1 homologs and are annotated as nitric oxide reductases, with detectable similarity to Proteobacteria nitric oxide reductase subunits. It is therefore possible that the inclusion of a C1 domain in *nir* genes within genomes lacking *eNOR* reflects some generalized NOR-like role in detoxification of the cytotoxic product of NirS. Additionally, while the presence of NirS suggests an active denitrification pathway, and the NirS domain tree reflects the homology between this domain and NirS from known denitrifying groups, the possibility remains that this group of Chloroflexi do not perform denitrification, and instead use this gene product for a different metabolic function, potentially enabled or constrained by the C1 domain. Experimental validation would be necessary to determine if the novel Chloroflexi-associated NirS performs differently than canonical NirS *in vivo* ; this work, therefore, suggests a promising direction for future investigation.

The divergent denitrification enzymes described above may or may not reflect different metabolic strategies *in situ* . But the identification of both a novel *nirS* ORF and an expanded diversity of eNOR enzymes suggests that the existing understanding of denitrification may underestimate the genetic diversity and ecological

distribution of constituent enzymes. This may be especially true in deep subsurface biomes, such as those from which several Chloroflexi analyzed in this study were isolated. These systems have garnered increasing attention in recent years; extensive evidence supports the existence of dynamic, diverse microbial subsurface ecosystems with the metabolic potential to influence global biogeochemical cycles (Hug *et al.* , 2013; Osburn *et al.* , 2014, 2019; Momper *et al.* , 2017). Chloroflexi are frequently cited as well-represented members of deep sediment and aquifer systems, where they play key roles in carbon cycling dynamics (Hug *et al.* , 2013; Momper *et al.* , 2017; Kadnikov *et al.* , 2020). But Chloroflexi are known to also harbor diverse nitrogen metabolisms (Hemp *et al.* , 2015; Denef *et al.* , 2016; Spieck *et al.* , 2020), and previous studies have linked subsurface Chloroflexi to denitrification pathway genes such as nitrous oxide reductase (*nos*) (Sanford *et al.* , 2012; Hug*et al.* , 2016; Momper *et al.* , 2017). The role of Chloroflexi in subsurface nitrogen cycling—as well as the scope of subsurface microbial nitrogen dynamics at large—requires further investigation.

EXPERIMENTAL PROCEDURES

Genome sampling and assembly

Collection of all fluid samples and total genomic DNA extractions from those fluids, as well as corresponding physical and geochemical data have been described previously (Lau *et al.* , 2014, 2016; Osburn*et al.* , 2014; Magnabosco *et al.* , 2016; Heard *et al.* , 2017; Momper *et al.* , 2017). All MAGs from North America and Africa were reconstructed according to the methods used in (Momper*et al.* , 2017). MAG identifiers and sources are listed in Table S2. Completeness was calculated using the composite values from five widely accepted core essential gene metrics. Duplicate copies of any of these single copy marker genes was interpreted as a measure of contamination (Creevey *et al.* , 2011; Dupont *et al.* , 2012; Wu and Scott, 2012; Campbell *et al.* , 2013; Alneberg *et al.* , 2014). Individual genomes were then submitted for gene calling and annotations through the DOE Joint Genome Institute IMG-ER (Integrated Microbial Genomes expert review) pipeline (Markowitz *et al.* , 2008; Huntemann *et al.* , 2015). For quality control purposes, the genes flanking every denitrification gene presented in this study were individually searched on the NCBI RefSeq database using the BLASTp algorithm, confirming that top hits for all flanking genes were also to Chloroflexi. This step ensured that the nitrogen transforming genes of interest presented here were not simply on scaffolds that were incorrectly binned into a putative Chloroflexi genome.

Genetic database construction and sequence sampling

Sequences for *nirS* and *eNOR* genes from SURF MAG 42 (Table S1) were used as queries to BLAST (Camacho *et al.* , 2009) three genomic repositories:

1. Genome databases constructed for 21 Chloroflexi genomes assembled from deep-subsurface MAG data (Jungbluth, Amend and Rappé, 2017; Momper*et al.* , 2017) (Table S1).
2. Genome databases constructed for 86 genomes from recent MAG assembled sludge bioreactor genomes (Parks *et al.* , 2017) (Table S3)
3. The full NCBI non-redundant protein database (as of 25 September, 2019)(Agarwala *et al.* , 2018)

Additionally, putative environmental homologs were evaluated using protein sequence data from SURF MAG 42 to query NCBI's non-redundant environmental metagenomic sequence database (env-nr, as of June 2020)(Agarwala *et al.* , 2018) (Supplementary Datafile S2) .

Hits from all databases (Table S4) were combined and assessed for quality; hits with E [?] $1x10^{-10}$ were included for initial analyses. To capture diversity while limiting imprecision and biased sampling of over-represented groups (e.g., Proteobacteria), hits were subsampled to the genus level, with the exception of members of the Chloroflexi (to fully capture the taxonomic distribution of the novel gene variant). One additional, divergent multispecies hit was allowed per genus. The genus-level filter was also removed for C1, where non-Chloroflexi hits were severely limited (see below). Duplicate sequences (from strains with multiple genome entries or in multiple databases surveyed) were removed.

Sequence alignment

Putative homologous protein sequences were aligned with MAFFT, using auto-parameterization (Nakamura *et al.* , 2018), and visualized in Jalview (Waterhouse *et al.* , 2009). Alignments were manually curated; partial sequences with substantial missing regions or anomalous insertions in conserved regions of the protein were removed to avoid confounding phylogenetic analyses and evolutionary model selection (Table S3). Protein sequence alignments were trimmed to the length of individual domains identified by NCBI's conserved domains database (Supplementary Datafile S1). Each domain was then re-aligned.

eNOR

A preliminary alignment for the *eNOR* gene showed a poorly conserved region near the C-terminal end of the ORF (see Table S4); to improve accuracy and avoid misalignment, this region was manually removed, and the remaining sequences were realigned prior to tree construction. Two sequences (Actinobacteria bacterium RBG_16_68_12, OFW73639.1, and *Thermus* WP_015717644.1) with missing N-terminal regions and three sequences (Chloroflexi bacterium, RME47896.1; Rhodocyclaceae bacterium UTPRO2, OQY7467.1; and *Rhodothermus profundi* , WP_072715415.1) with missing C-terminal regions were included in the final alignment; the placement of these sequences is therefore based upon fewer alignment sites than other taxa. All retain key active site residues and show no clear evidence of long branch attraction artifacts in the tree.

C1

An initial alignment for the C1 domain showed a poorly conserved N-terminal region (see Table S4). To improve accuracy, this region was manually removed, and the remaining sequences realigned prior to tree construction.

NirS

Because the *nirS* domain had a C-terminal placement in the ORF across hits, C-terminal sites extending beyond the identified*nirS* domain were included in the trimmed alignment.

Rooting and outgroup identification

eNOR

Ingroup *eNOR* sequences were identified by the presence of a conserved Gln residue in alignment position 333. This site distinguishes eNOR not only from other nitric oxide reductases, but also from members of the oxygen reductase superfamily, which have a conserved Tyr in this site that plays a role in cofactor crosslinking. (Hemp and Gennis, 2008) (Table S5). Outgroup sequences (oxygen reductase superfamily or other divergent nitric oxide reductases) were subsampled to a single taxon representative per major subgroup observed in a preliminary tree (Supplementary Datafile S4). Retained outgroup sequences CCQ74688.1, WP_100277903.1, WP_097280063.1, WP_089728124.1, RLC59399.1, and WP_083704903.1 are annotated as uncharacterized domains. Remaining sequences were re-aligned before tree construction, and manually rooted on the branch leading to the outgroup.

C1

Due to a paucity of initial hits (17 total genera), the genus-level filter was removed for all phyla to increase resolution of the domain phylogeny. A preliminary tree (Fig. S2) was expanded to identify outgroup sequences by including hits with E [?] 10^{-10}. The resulting tree was rooted using minimal ancestor deviation (MAD) rooting (Tria, Landan and Dagan, 2017).

C2 and NirS

Sequences were rooted using minimal ancestor deviation (MAD) rooting (Tria, Landan and Dagan, 2017).

Tree construction

Maximum-likelihood trees were constructed using IQ-Tree (Nguyen *et al.* , 2015), under the optimal model defined by the ModelFinder (-MFP) command (Kalyaanamoorthy *et al.* , 2017) (Table S6). Ultrafast bootstraps and approximate likelihood ratio tests were performed using IQ-Tree's ultrafast bootstrap and

Sh-aLRT parameters (Minh, Nguyen and Von Haeseler, 2013; Hoang *et al.* , 2018). Full treefiles, supports, and expanded accession data for all sequences are provided in Supplementary Datafile S5.

Scripts and Jupyter Notebook files (Kluyver *et al.* , 2016) used for automating alignment or treefile analysis, curation, and visualization are available at https://github.com/slschwartz/fournierlab-scripts.

Gene and enzyme structural analysis :

FIND (Murali *et al.* , 2019) was used to identify structural features and conserved denitrification pathway genes in deep subsurface genomes. Putative domains within denitrification gene ORFs were identified and compared across genomes using NCBI's Conserved Domains Database (CDD) (Marchler-Bauer *et al.* , 2015; Lu *et al.* , 2020) and EMBL Interpro (Mitchell *et al.* , 2019).

Existing enzyme structures for canonical denitrification genes were downloaded from the RCSB Protein Data Bank (PDB) (Berman *et al.* , 2000). Anaerolineales-type enzyme structures were predicted using SWISS-MODEL (Waterhouse *et al.* , 2018). All enzyme structures were visualized and analyzed in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.)

ACKNOWLEDGEMENTS

The authors thank Ranjani Murali for advice on domain and residue identification and analysis. This work was supported by a National Defense Science and Engineering Graduate Fellowship to SLS, and Simons Collaboration on Origins of Life award #339603 to GPF. The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Experimental design and planning (SLS, LMM, GPF, JPA), production and analysis of data (SLS, LMM, LTR, CM), preparation of the manuscript (SLS, GPF, LMM).

GRAPHICAL ABSTRACT

Hosted file

`image1.emf` available at https://authorea.com/users/419181/articles/525743-novel-nitrite-reductase-domain-structure-suggests-a-chimeric-denitrification-repertoire-in-phylum-chloroflexi

ABBREVIATED SUMMARY

Nitrite reductase (Nir) and nitric oxide reductase (NOR) are the two central enzymes in denitrification, a key process in the global nitrogen cycle. This study identifies a novel *nir* domain architecture and expanded diversity in a rarely reported nitric oxide reductase variant (*eNOR*) in members of the bacterial phylum Chloroflexi.

REFERENCES

Agarwala, R. *et al.* (2018) 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Research* . doi: 10.1093/nar/gkx1095.

Alneberg, J. *et al.* (2014) 'Binning metagenomic contigs by coverage and composition', *Nature Methods* . doi: 10.1038/nmeth.3103.

Berman, H. M. *et al.* (2000) 'The Protein Data Bank',*Nucleic Acids Research* . doi: 10.1093/nar/28.1.235.

Bertini, I., Cavallaro, G. and Rosato, A. (2006) 'Cytochrome c: Occurrence and functions', *Chemical Reviews* , pp. 90–115. doi: 10.1021/cr050241v.

Bienert, S. *et al.* (2017) 'The SWISS-MODEL Repository-new features and functionality', *Nucleic Acids Research* . doi: 10.1093/nar/gkw1132.

Blomberg, M. R. A. and Adelroth, P. (2018) 'Mechanisms for enzymatic reduction of nitric oxide to nitrous oxide - A comparison between nitric oxide reductase and cytochrome c oxidase', *Biochimica et Biophysica Acta - Bioenergetics* . Elsevier, 1859(11), pp. 1223–1234. doi: 10.1016/j.bbabio.2018.09.368.

Braker, G. *et al.* (2000) 'Nitrite reductase genes (nirK and nirS) as functional markers to investigate diversity of denitrifying bacteria in pacific northwest marine sediment communities', *Applied and Environmental Microbiology* , 66(5), pp. 2096–2104. doi: 10.1128/AEM.66.5.2096-2104.2000.

Butterbach-Bahl, K. and Dannenmann, M. (2011) 'Denitrification and associated soil N2O emissions due to agricultural activities in a changing climate', *Current Opinion in Environmental Sustainability* . Elsevier B.V., 3(5), pp. 389–395. doi: 10.1016/j.cosust.2011.08.004.

Camacho, C. *et al.* (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics* . doi: 10.1186/1471-2105-10-421.

Campbell, J. H. *et al.* (2013) 'UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota', *Proceedings of the National Academy of Sciences of the United States of America* . doi: 10.1073/pnas.1303090110.

Canfield, D. E., Glazer, A. N. and Falkowski, P. G. (2010) 'The evolution and future of earth's nitrogen cycle', *Science* , 330(6001), pp. 192–196. doi: 10.1126/science.1186120.

Creevey, C. J. *et al.* (2011) 'Universally distributed single-copy genes indicate a constant rate of horizontal transfer', *PLoS ONE* , 6(8). doi: 10.1371/journal.pone.0022099.

Decleyre, H. *et al.* (2016) 'Erratum to: Highly diverse nirK genes comprise two major clades that harbour ammonium-producing denitrifiers',*BMC Genomics* . BMC Genomics, 17(1), p. 12864. doi: 10.1186/s12864-016-2812-1.

Denef, V. J. *et al.* (2016) 'Chloroflexi CL500-11 populations that predominate deep-lake hypolimnion bacterioplankton rely on nitrogen-rich dissolved organic matter metabolism and C1 compound oxidation',*Applied and Environmental Microbiology* , 82(5), pp. 1423–1432. doi: 10.1128/AEM.03014-15.

Dupont, C. L. *et al.* (2012) 'Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage', *ISME Journal* . Nature Publishing Group, 6(6), pp. 1186–1199. doi: 10.1038/ismej.2011.189.

Gardner, A. M., Helmick, R. A. and Gardner, P. R. (2002) 'Flavorubredoxin, an inducible catalyst for nitric oxide reduction and detoxification in Escherichia coli', *Journal of Biological Chemistry* , 277(10), pp. 8172–8177. doi: 10.1074/jbc.M110471200.

Graf, D. R. H., Jones, C. M. and Hallin, S. (2014) 'Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N2O emissions',*PLoS ONE* , 9(12), pp. 1–20. doi: 10.1371/journal.pone.0114118.

Heard, A. W. *et al.* (2017) 'Origins and ages of fracture fluids in the South African Crust', in *Agufm* , pp. H11A-1153.

Hemp, J. *et al.* (2015) 'Draft genome sequence of Ardenticatena maritima 110S, a thermophilic nitrate- and iron-reducing member of the Chloroflexi Class Ardenticatenia', *Genome Announcements* , 3(6), p. 2920. doi: 10.1128/genomeA.01347-15.

Hemp, J. and Gennis, R. B. (2008) 'Diversity of the Heme-Copper superfamily in archaea: Insights from genomics and structural modeling',*Results and Problems in Cell Differentiation* , 45(January), pp. 1–31. doi: 10.1007/400_2007_046.

Hendriks, J. *et al.* (2000) 'Nitric oxide reductases in bacteria',*Biochimica et Biophysica Acta - Bioenergetics* , 1459(2–3), pp. 266–273. doi: 10.1016/S0005-2728(00)00161-4.

Heylen, K. *et al.* (2007) 'Nitric oxide reductase (norB) gene sequence analysis reveals discrepancies with nitrite reductase (nir) gene phylogeny in cultivated denitrifiers', *Environmental Microbiology* , 9(4), pp. 1072–1077. doi: 10.1111/j.1462-2920.2006.01194.x.

Hoang, D. T. *et al.* (2018) 'UFBoot2: Improving the ultrafast bootstrap approximation', *Molecular Biology and Evolution* . doi: 10.1093/molbev/msx281.

Hug, L. A. *et al.* (2013) 'Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling', *Microbiome* , 1(1), pp. 1–17. doi: 10.1186/2049-2618-1-22.

Hug, L. A. *et al.* (2016) 'Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages', *Environmental Microbiology* , 18(1), pp. 159–173. doi: 10.1111/1462-2920.12930.

Huntemann, M. *et al.* (2015) 'The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4)',*Standards in Genomic Sciences* . doi: 10.1186/s40793-015-0077-y.

Jones, C. M. *et al.* (2008) 'Phylogenetic analysis of nitrite, nitric oxide, and nitrous oxide respiratory enzymes reveal a complex evolutionary history for denitrification', *Molecular Biology and Evolution* , 25(9), pp. 1955–1966. doi: 10.1093/molbev/msn146.

Jungbluth, S. P., Amend, J. P. and Rappe, M. S. (2017) 'Corrigendum: Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids', *Scientific data* , 4, p. 170080. doi: 10.1038/sdata.2017.80.

Kadnikov, V. V. *et al.* (2020) 'Microbial Life in the Deep Subsurface Aquifer Illuminated by Metagenomics', *Frontiers in Microbiology* , 11(September), pp. 1–15. doi: 10.3389/fmicb.2020.572252.

Kalyaanamoorthy, S. *et al.* (2017) 'ModelFinder: Fast model selection for accurate phylogenetic estimates', *Nature Methods* . doi: 10.1038/nmeth.4285.

Kluyver, T. *et al.* (2016) 'Jupyter Notebooks—a publishing format for reproducible computational workflows', *Positioning and Power in Academic Publishing: Players, Agents and Agendas - Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016* , pp. 87–90. doi: 10.3233/978-1-61499-649-1-87.

Lau, M. C. Y. *et al.* (2014) 'of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships', *Frontiers in Microbiology* , 5(OCT), pp. 1–17. doi: 10.3389/fmicb.2014.00531.

Lau, M. C. Y. *et al.* (2016) 'An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers', *Proceedings of the National Academy of Sciences* , 113(49), pp. E7927–E7936. doi: 10.1073/pnas.1612244113.

Lu, H., Chandran, K. and Stensel, D. (2014) 'Microbial ecology of denitrification in biological wastewater treatment', *Water Research* . Elsevier Ltd, 64, pp. 237–254. doi: 10.1016/j.watres.2014.06.042.

Lu, S. *et al.* (2020) 'CDD/SPARCLE: The conserved domain database in 2020', *Nucleic Acids Research* . doi: 10.1093/nar/gkz991.

Magnabosco, C. *et al.* (2016) 'A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust', *ISME Journal* . Nature Publishing Group, 10(3), pp. 730–741. doi: 10.1038/ismej.2015.150.

Maia, L. B. and Moura, J. J. G. (2014) 'How biology handles nitrite',*Chemical Reviews* , 114(10), pp. 5273–5357. doi: 10.1021/cr400518y.

Marchler-Bauer, A. *et al.* (2015) 'CDD: NCBI's conserved domain database', *Nucleic Acids Research* . doi: 10.1093/nar/gku1221.

Markowitz, V. M. *et al.* (2008) 'IMG/M: A data management and analysis system for metagenomes', *Nucleic Acids Research* . doi: 10.1093/nar/gkm869.

Minh, B. Q., Nguyen, M. A. T. and Von Haeseler, A. (2013) 'Ultrafast approximation for phylogenetic bootstrap', *Molecular Biology and Evolution* . doi: 10.1093/molbev/mst024.

Mitchell, A. L. *et al.* (2019) 'InterPro in 2019: Improving coverage, classification and access to protein sequence annotations',*Nucleic Acids Research* . doi: 10.1093/nar/gky1100.

Momper, L. *et al.* (2017) 'Major phylum-level differences between porefluid and host rock bacterial communities in the terrestrial deep subsurface', *Environmental Microbiology Reports* , 9(5), pp. 501–511. doi: 10.1111/1758-2229.12563.

Murali, R. *et al.* (2019) 'FIND: Identifying Functionally and Structurally Important Features in Protein Sequences with Deep Neural Networks', *bioRxiv* . doi: 10.1101/592808.

Nakamura, T. *et al.* (2018) 'Parallelization of MAFFT for large-scale multiple sequence alignments', *Bioinformatics* . doi: 10.1093/bioinformatics/bty121.

Nguyen, L. T. *et al.* (2015) 'IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies',*Molecular Biology and Evolution* . doi: 10.1093/molbev/msu300.

Osburn, M. R. *et al.* (2014) 'Chemolithotrophy in the continental deep subsurface: Sanford underground research facility (SURF), USA',*Frontiers in Microbiology* , 5(NOV), pp. 1–14. doi: 10.3389/fmicb.2014.00610.

Osburn, M. R. *et al.* (2019) 'Establishment of the Deep Mine Microbial Observatory (DeMMO), South Dakota, USA, a Geochemically Stable Portal Into the Deep Subsurface', *Frontiers in Earth Science* , 7(July), pp. 1–17. doi: 10.3389/feart.2019.00196.

Parks, D. H. *et al.* (2017) 'Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life',*Nature Microbiology* . Springer US, 2(11), pp. 1533–1542. doi: 10.1038/s41564-017-0012-7.

Philippot, L. (2002) 'Philippot 2002 Denitrifying genes in bacterial and Archaeal genomes', 1577, pp. 355–376. doi: 10.1016/S0167-4781(02)00420-7.

Prieme, A., Braker, G. and Tiedje, J. M. (2002) 'Diversity of nitrite reductase (nirK and nirS) gene fragments in forested upland and wetland soils', *Applied and Environmental Microbiology* , 68(4), pp. 1893–1900. doi: 10.1128/AEM.68.4.1893-1900.2002.

Rinaldo, S. *et al.* (2011) 'The catalytic mechanism of Pseudomonas aeruginosa cd1 nitrite reductase', *Biochemical Society Transactions* , 39(1), pp. 195–200. doi: 10.1042/BST0390195.

Roco, C. A. *et al.* (2017) 'Modularity of nitrogen-oxide reducing soil bacteria: linking phenotype to genotype', *Environmental Microbiology* , 19(6), pp. 2507–2519. doi: 10.1111/1462-2920.13250.

Sanchez, C. *et al.* (2011) 'Nitric oxide detoxification in the rhizobia-legume symbiosis', *Biochemical Society Transactions* , 39(1), pp. 184–188. doi: 10.1042/BST0390184.

Sanford, R. A. *et al.* (2012) 'Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils',*Proceedings of the National Academy of Sciences* , 109(48), pp. 19709–19714. doi: 10.1073/pnas.1211238109.

Sievert, S. M. *et al.* (2008) 'Genome of the epsilonproteobacterial chemolithoautotroph Sulfurimonas denitrificans',*Applied and Environmental Microbiology* , 74(4), pp. 1145–1156. doi: 10.1128/AEM.01844-07.

Spieck, E. *et al.* (2020) 'Extremophilic nitrite-oxidizing Chloroflexi from Yellowstone hot springs', *ISME Journal* . Springer US, 14(2), pp. 364–379. doi: 10.1038/s41396-019-0530-9.

Stein, L. Y. *et al.* (2007) 'Whole-genome analysis of the ammonia-oxidizing bacterium, Nitrosomonas eutropha C91: Implications for niche adaptation', *Environmental Microbiology* , 9(12), pp. 2993–3007. doi: 10.1111/j.1462-2920.2007.01409.x.

'The PyMOL Molecular Graphics System' (no date). Schrodinger, LLC.

Tria, F. D. K., Landan, G. and Dagan, T. (2017) 'Phylogenetic rooting using minimal ancestor deviation', *Nature ecology & evolution* . doi: 10.1038/s41559-017-0193.

Ward, L. M. *et al.* (2018) 'Evolution of phototrophy in the Chloroflexi phylum driven by horizontal gene transfer', *Frontiers in Microbiology* . doi: 10.3389/fmicb.2018.00260.

Ward, L. M., McGlynn, S. E. and Fischer, W. W. (2018) 'Draft genome sequence of a divergent anaerobic member of the Chloroflexi class Ardenticatenia from a sulfidic hot spring', *Genome Announcements* , 6(25), pp. 1–2. doi: 10.1128/genomeA.00571-18.

Waterhouse, A. *et al.* (2018) 'SWISS-MODEL: Homology modelling of protein structures and complexes', *Nucleic Acids Research* . doi: 10.1093/nar/gky427.

Waterhouse, A. M. *et al.* (2009) 'Jalview Version 2-A multiple sequence alignment editor and analysis workbench',*Bioinformatics* . doi: 10.1093/bioinformatics/btp033.

Wei, W. *et al.* (2015) 'Higher diversity and abundance of denitrifying microorganisms in environments than considered previously',*ISME Journal* . Nature Publishing Group, 9(9), pp. 1954–1965. doi: 10.1038/ismej.2015.9.

Wu, M. and Scott, A. J. (2012) 'Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2', *Bioinformatics* . doi: 10.1093/bioinformatics/bts079.

Zumft, W. G. (1997) 'Cell biology and molecular basis of denitrification.', *Microbiology and molecular biology reviews : MMBR* , 61(4), pp. 533–616. doi: http://dx.doi.org/10.1016/j.sbspro.2014.08.122.

Zumft, W. G. (2005) 'Nitric oxide reductases of prokaryotes with emphasis on the respiratory, heme-copper oxidase type', *Journal of Inorganic Biochemistry* , 99(1), pp. 194–215. doi: 10.1016/j.jinorgbio.2004.09.024.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in figshare at http://doi.org/10.6084/m9.figshare.14515554.

Scripts used to automate analyses are available at https://github.com/slschwartz/fournierlab-scripts.

FIGURE LEGENDS

Fig. 1. Denitrification

Complete denitrification reduces nitrate to dinitrogen gas. The nitrate reductase gene (*nar*) is widely distributed in non-denitrifying organisms; nitrite reductase (Nir) is considered the canonical first enzyme of denitrification (Graf, Jones and Hallin, 2014), followed by nitric oxide reductase (NOR) and nitrous oxide reductase (Nos).

Figure 2. Open reading frame domain map

Conserved domain analysis of SURF MAG 42 Chloroflexi nitrite reductase (GenBank RJP53747.1) indicates the presence of two distinct cytochrome superfamily domains and a C-terminal nitrite reductase domain.

Figure 3a. Phylogenetic tree for C2 domain

Phylogenetic analysis of C2 domain homologs places the Chloroflexi in a diverse clade including Epsilonproteobacteria, Aquificae, Bacteroidetes, and Planctomycetes; this clade is sister to a broad radiation of Alpha-, Beta-, and Gammaproteobacteria. Support values for selected bipartitions are labeled (aLRT/bb). Support

for other nodes is indicated with the following color scheme: Strong support with both values [?] 90 (black); weak support with both values [?] 50 (white); intermediate support with one or both values between 50 and 90 (grey); conflicting support, with one value [?] 50 and the other [?] 90 (grey).

Figure 3b. Phylogenetic tree for NirS domain

Phylogenetic analysis of NirS (cd-1 type) domain homologs places the Chloroflexi within a clade dominated by Alpha-, Beta-, Gammaproteobacteria, but also including members of the Epsilonproteobacteria, Aquificae, Bacteroidetes, and Planctomycetes. Statistical support values were displayed using a simplified color scheme for aLRT/rapid bootstrap values. Support values for selected bipartitions are labeled (aLRT/bb). Support for other nodes is indicated with the following color scheme: Strong support with both values [?] 90 (black); weak support with both values [?] 50 (white); intermediate support with one or both values between 50 and 90 (grey); conflicting support, with one value [?] 50 and the other [?] 90 (grey).

Figure 4. C1 domain tree

A phylogenetic tree for the C1 domain—with no genus-level filter and inclusion of more distant hits (See Methods)—indicates a limited taxonomic distribution of the domain. The largest group of sequences in Chloroflexi places sister to domains found in Nitrospirae, Nitrospinae, and Deltaproteobacteria. Within this clade, the branch along which C1 is inferred to have fused into nitrite reductase genes in Chloroflexi is labeled. C1 homologs that co-occur in ORFs with nitrite reductase are indicated with magenta diamonds (NirS) or yellow diamonds (NirK).

Support values for selected bipartitions are labeled (aLRT/bb). Support for other nodes is indicated with the following color scheme: Strong support with both values [?] 90 (black); weak support with both values [?] 50 (white); intermediate support with one or both values between 50 and 90 (grey); conflicting support, with one value [?] 50 and the other [?] 90 (grey).

Figure 5. *eNOR* domain tree

A phylogenetic tree of homologs to the nitric oxide reductase from SURF MAG 42 reveals an expanded diversity of putative *eNOR* homologs in not only Archaea and Chloroflexi, but also Proteobacteria and other diverse phyla. Putative *eNOR* sequences (red tips) have the characteristic Glu-323 in the alignment; outgroup sequences (blue tips) have Tyr-323 (oxygen reductase superfamily) or other substitutions. Support values for selected bipartitions are labeled (aLRT/bb). Support for other nodes is indicated with the following color scheme: Strong support with both values [?] 90 (black); weak support with both values [?] 50 (white); intermediate support with one or both values between 50 and 90 (grey); conflicting support, with one value [?] 50 and the other [?] 90 (grey).

FIGURES

Fig. 1

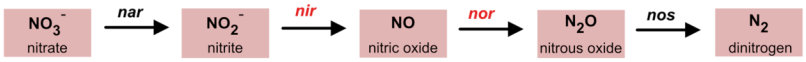

Fig. 2

Hosted file

image3.emf available at https://authorea.com/users/419181/articles/525743-novel-nitrite-reductase-domain-structure-suggests-a-chimeric-denitrification-repertoire-in-phylum-chloroflexi

Fig. 3a

Hosted file

image4.emf available at https://authorea.com/users/419181/articles/525743-novel-nitrite-reductase-domain-structure-suggests-a-chimeric-denitrification-repertoire-in-phylum-chloroflexi

Fig. 3b

Hosted file

image5.emf available at https://authorea.com/users/419181/articles/525743-novel-nitrite-reductase-domain-structure-suggests-a-chimeric-denitrification-repertoire-in-phylum-chloroflexi

Fig. 4

Hosted file

image6.emf available at https://authorea.com/users/419181/articles/525743-novel-nitrite-reductase-domain-structure-suggests-a-chimeric-denitrification-repertoire-in-phylum-chloroflexi

Hosted file

image7.emf available at https://authorea.com/users/419181/articles/525743-novel-nitrite-reductase-domain-structure-suggests-a-chimeric-denitrification-repertoire-in-phylum-chloroflexi

Fig. 5