

## **Phylogenetics and environmental distribution of nitric oxide forming nitrite reductases reveals their distinct functional and ecological roles**

**Authors:** Grace Pold, Germán Bonilla-Rosso, Aurélien Saghaï, Marc Strous, Christopher M. Jones, Sara Hallin

### **Supplementary Materials and Methods**

#### ***Naming clades:***

Major clades in each tree were first delineated based on groupings identified in previous literature [1,2], reserving clade 1 in each protein for the “canonical clade”, and letters were used within each larger clade to delineate sub clades. Primarily structural features and secondarily taxonomy and ecological traits were used to refine clade delineation. Clades were defined as deeply as possible on the phylogeny within these constraints, leading to statistically supported clades of variable phylogenetic depth. In some instances this meant generating new clades from previously proposed clades that were not well-supported in the present phylogeny. Thus, clades with similar naming number-letter hierarchy level do not necessarily originate at similar depths on the phylogeny, and we refer to these well-supported groups of sequences in the phylogeny as clades independent of number-letter hierarchy.

#### **Exclusion of halophilic archaea and *Pyrobaculum* NirS**

Functional NirS has also been demonstrated in *Pyrobaculum* species (Thermoproteota [3,4]). Despite carrying all the conserved motifs, we excluded these *Pyrobaculum* sequences from our phylogeny because the haem d<sub>1</sub> and *cyt<sub>c</sub>* domains are encoded in opposite directions in

the genome, which led to an unresolvable long branch upon re-orienting and concatenating them.

Halobacteriota NirS-like proteins were included as an outgroup and excluded from *nirS* counts in the metagenome survey because they lack the first two characteristic motifs corresponding to the *cyt<sub>c</sub>* domain. Using the previously described search for genes encoding enzymes involved in NirS assembly combined with genome viewer in NCBI to look for potential alternative heme assembly proteins [3] and *cyt<sub>c</sub>* domain-containing motifs, we confirmed the absence of the evidence that these proteins are *cyt<sub>cd</sub>* NirS. An additional reason for excluding this clade from *nirS* gene fragments counts is that absence of the *cyt<sub>c</sub>* domain led to strange behaviour of the search and place algorithm. An excess of *nirS* reads were placed in this gapped region of the alignment and subsequently annotated as haloarchaeal, despite being derived from habitats such as forest soils where Halobacteriota are rare, and where Halobacteriota *nirK* reads were below detection. Furthermore, BLAST of a subset (n=20) of the reads annotated as haloarchaeal *nirS* but mapped in their entirety to this 75aa gap in the beginning of the alignment against the UniProt database indicated that none of them most closely matched archaea; instead, most (18; 90%) mapped as non-nitrite reductase bacterial cytochrome C or cytochrome C oxidase, often with at least 60% identity (13; 65%). By contrast, short reads from this N-terminal region which GraftM placed in the canonical *nirS* portion of the tree correctly mapped to proteobacterial nitrite reductase or cytochrome C.

### ***Structural features and nitrite reductase helper genes***

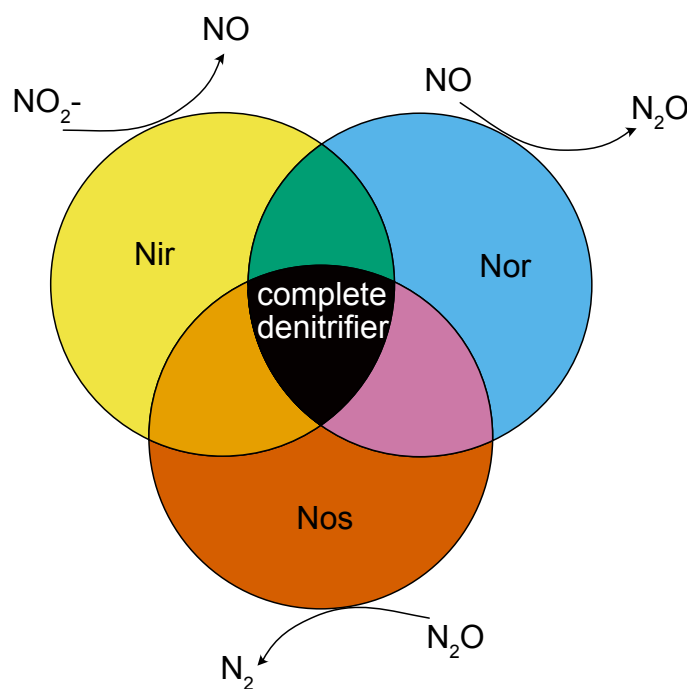
We searched for *nirF*, *nirN*, *nirJ*, *nirE*, *nirB* and *nirT* in assemblies carrying *nirS*, and *nirV* in assemblies carrying *nirK*. The seed alignments for NirJ (TIGR04051, TIGR04055, TIGR04054) and NirE (cd11642) were derived from the NCBI's Conserved Domains Database. Seed alignments for NirF and NirN were derived from the original NirS search,

and were readily differentiated from NirS using a phylogenetic approach. The HMMs for NirB (UniProt P24037), NirT (UniProt P24038) and NirV (NCBI AAK08123.1) were generated using protein BLAST searches with the aforementioned reference sequences against NCBI's ClusteredNR database [5]. We stochastically selected a subset of the 1000 top hits to be aligned and exported using the multiple alignment function accessible from the BLAST outputs, and then checked if the sequences were aligned at the important ligands and catalytic residues in ARB [6]. We searched for the structural features TAT, lipobox, and Sec-type signal peptides using the online version of SignalP 6.0 [7] and transmembrane domains using DeepTMHMM [8], both with default settings. Clade-specific insertions and deletions[9], and cytochrome C (-CX<sub>2</sub>CHX<sub>50</sub>M-) and cupredoxin (CX<sub>2-4</sub>HX<sub>2-4</sub>M) motifs in the C and N termini of the proteins were identified in ARB [6].

### **Inferring redox traits from protein-coding gene composition**

We looked for nitric oxide reductases within the heme-copper oxidase (HCO) superfamily using the same method used for identifying NirK and NirS, but used the database generated by Murali and colleagues as a starting seed for our hmmsearch [10,11]. Genes were categorized into classes of *nor* also following Murali *et al.* [10,11]. We identified nitrous oxide reductases (*nosZ*) similarly, but used the reference database from Graf *et al.* 2014 as our seed; hydrogenases associated with hydrogen oxidation (group 1, 2a) and reduction (group 4 b,c,e) using the alignment in [12]; nitrite oxidation using the alignment and structural information provided by [13]; ammonium (*amoA*) and methane (*pmoA*) oxidation against the database provided by [14]; and sulfur compound oxidation (*sox*, *ox-apr*, *ox-dsr*, *sat*, *shdr*, or *sor*) and reduction (APS sulfate reduction: *aprAB*, *qmoABC*; sulfite reduction: *dsrABCDNTMK*, *mccABCD*, *AsrABC*; tetrathionate reduction: *TtrABC*) using HMMs and gene neighborhood information provided in [15]. The capacity for anammox (*hzxA*, *hzsA*),

methane oxidation (*mmoB/dmpM, pmoABC*), iron (*fmnA/dmkA/fmnB/pplA/ndh2/eetAB/dmkB, mtrABC*), selenate (*ygfKMN*), nitrate (*napAB, narGH*) and dissimilatory arsenate reduction (*arrA*), and iron (*cyc123; foxABCXYZ*), nitrite (*nxrAB*) and manganese oxidation (*mnxG*) were predicted using HMMs and gene neighborhood information from FeGenie and Lithogenie [16]. *nxA/napA* and *amoA/pmoA/hcoA* were further checked by assessing the alignment and constructing a phylogeny using FastTree, and additional taxonomy-based verification. The capacity for reductive dehalogenation was evaluated using HMMs built from the Reductive Dehalogenase Database [17]; sequences lying within known reductive dehalogenase diversity in a combined FastTree phylogeny were kept. The complete denitrification trait was defined as having *nir*, *nor*, and *nosZ* genes as depicted below.



**Supplementary methods figure 1:** complete denitrification refers to the potential to transform nitrite into dinitrogen via NO (completed by Nir),  $\text{N}_2\text{O}$  (Nor) and finally  $\text{N}_2$  (Nos). Incomplete denitrification in organisms encoding NirK and/or NirS refers to the presence of Nir+Nor (green), Nir+Nos (orange), or the presence of just Nir (yellow).

## References:

1. Wei W, Isobe K, Nishizawa T *et al.* Higher diversity and abundance of denitrifying microorganisms in environments than considered previously. *ISME J* 2015;**9**:1954–65.
2. Bonilla-Rosso G, Wittorf L, Jones CM *et al.* Design and evaluation of primers targeting genes encoding NO-forming nitrite reductases: implications for ecological inference of denitrifying communities. *Sci Rep* 2016;**6**:39208.
3. Storbeck S, Rolfes S, Raux-Deery E *et al.* A Novel Pathway for the Biosynthesis of Heme in Archaea: Genome-Based Bioinformatic Predictions and Experimental Evidence. *Archaea* 2010;**2010**:e175050.
4. Feinberg LF, Holden JF. Characterization of dissimilatory Fe(III) versus NO<sub>3</sub>- reduction in the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *J Bacteriol* 2006;**188**:525–31.
5. Staff N. New ClusteredNR database: faster searches and more informative BLAST results. *NCBI Insights* 2022. [https://ncbiinsights.ncbi.nlm.nih.gov/2022/05/02/clusterednr\\_1/](https://ncbiinsights.ncbi.nlm.nih.gov/2022/05/02/clusterednr_1/)
6. Westram R, Bader K, Preusse E *et al.* ARB: a software environment for sequence data. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*. Wiley, 399–406.
7. Teufel F, Almagro Armenteros JJ, Johansen AR *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 2022;**40**:1023–5.
8. Hallgren J, Tsigos KD, Pedersen MD *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. 2022:2022.04.08.487609.
9. Decleyre H, Heylen F, Bjorn T *et al.* Highly diverse nirK genes comprise two major clades that harbour ammonium-producing denitrifiers. *BMC Genomics* 2016;**17**:155.
10. Murali R, Pace LA, Sanford RA *et al.* Diversity and evolution of nitric oxide reduction. 2021:2021.10.15.464467.
11. Murali R, Hemp J, Gennis RB. Evolution of quinol oxidation within the heme-copper oxidoreductase superfamily. *Biochim Biophys Acta BBA - Bioenerg* 2022:148907.
12. Greening C, Biswas A, Carere CR *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H<sub>2</sub> is a widely utilised energy source for microbial growth and survival. *ISME J* 2016;**10**:761–77.
13. Chicano TM, Dietrich L, de Almeida NM *et al.* Structural and functional characterization of the intracellular filament-forming nitrite oxidoreductase multiprotein complex. *Nat Microbiol* 2021;**6**:1129–39.
14. Diamond S, Lavy A, Crits-Christoph A *et al.* Soils and sediments host Thermoplasmata archaea encoding novel copper membrane monooxygenases (CuMMOs). *ISME J* 2022;**16**:1348–62.

15. Tanabe TS, Dahl C. HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies. *Mol Ecol Resour* 2022;**22**:2758–74.
16. Garber AI, Neilson KH, Okamoto A *et al.* FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front Microbiol* 2020;**11**, DOI: doi.org/10.3389/fmicb.2020.00037.
17. Molenda O, Jácome LAP, Cao X *et al.* Insights into origins and function of the unexplored majority of the reductive dehalogenase gene family as a result of genome assembly and ortholog group classification. *Environ Sci Process Impacts* 2020;**22**:663–78.



1 10 20 30 40 50 60

GCA\_000337795.1\_Haloferax\_denitrificans ANRIAADPRDIPAPITRTSPETVSVELETLEQVAEVEPGVTFYMTFNNGQVPGPFIRTRV  
GCA\_013407385.1\_Nitrosarchaeum\_sp.\_AC2 . . . . . ADSGRKVVFEFNLTGESVTLPIMGKTFNAMTFSGQVVPGPFILRVTQ  
GCA\_000380525.1\_Thioalkalivibrio\_sp.\_AKL8 . . . . . SQDAEPDHFVFRIDLDEGVPIGQGVVYDGFIDGKLPGLTIRVTE  
GCA\_000478095.1\_Corynebacterium\_sp.\_KPL1824 REGFKAHDAASLAP. . . . . AANTTTHEETWMSMTEEIVEVAPGVKQMWLNFNGQAPGPTLRGKV  
GCA\_000775045.1\_Burkholderia\_pseudomallei . . . . . ALQP. . . . . LDAARDQAIRLDTHTTVIRIAPGIAFAAWTFNGNQVPGPTVHVKV  
GCA\_001458695.1\_Nitrospira\_inopinata . . . . . SAGAKVHDVTFATATESEIVIDGGTKYKAWTFNGQMPGPFVVRVTQ  
GCA\_001723765.1\_Scalindua\_rubra . . . . . GNIPKTVRVTLTAQEITATLDSNVSFRYFTFNFRVPGPFIRVME  
GCA\_003097635.1\_Flavobacterium\_psychrotolerans MEAELTAPPLVPKPIGSRDATKLIIVMEVKEMESELADGVKTYTWTFFGGSVPGSFIIRTRV  
GCA\_003135435.1\_Nitrospira\_sp. IKAVVTTAPNVPPPIITRTEPANVLELEAHEWVGNLSDDNKYKFWGFGGTVPGPFIRVMV  
GCA\_003335105.1\_Rhodanobacter\_denitrificans IHAVLTSPHPVPPPIHRNYPKVVIVELEVVEKEMPISEGVSYTFWTFGGTVPGPFIRVRO  
GCA\_003495885.1\_Neisseria\_meningitidis IDAVTTHAPEVPPAIDRDYPAKVRVKMERTVEKTMMEDGVYRYWTFDGDVPGPFIRVRE  
GCA\_0013267085.1\_Pseudomonas\_stutzeri GKDLTMDPSKVGEVPGKREPKNLTIDLRTTEEGEGLSDGSSYRFWTFNGTVPGPFMRIRE  
GCA\_900492205.1\_Rhizobium\_naphthalenivorans VKVDLVKPPFVHTQKAEGGPKVVEFTMTIEEKKLIVDDGTEVHAMTFFDGSVPGPFLMVVHE  
GCA\_905339145.1\_Anaerolineae\_bacterium . QNPALVGLPIQLRPGGATPRTVTLNLKAAEEVIAEVAPMNF . . . . . WTFFNGAVPGPFMLRVME  
Globobulimina\_pacifica . . . . . LPRRHSVPVTHLITLTSVEVVAELEEGHTFEFFTYNATVPGPFILRVRE  
GCA\_000004985.1\_Naegleria\_gruberi VEAILTKAPGIPPPIDRDYVQLLVNLDTTIEVKPILSKQYKYPYWTFNGTVPGPFIRARV  
GCA\_014843455.1\_2\_Fusarium\_oxysporum EAAVLTSAPNVPPPIITRKHPLVQLQVALATETKLAQLTSQYKYEQWTFNGTVPGPFIRARV  
GCA\_000204585.1\_4\_outgroup1 . . . . . NTIRDFTIIAEDNKIETSPGVFYNVWTFNGTVPGPFIRATE  
GCA\_003113895.1\_outgroup2 . . . . . KQENTIREFQLTAGTTSIIQLNSAVSYNIWDLNGRITPGPFLRAKQ  
consensus>80 . . . . . y . . . . . f . g . v . P . G . p . . . . . r . . . . .

70 80 90 100 110 120

GCA\_000337795.1\_Haloferax\_denitrificans GDTVDLIIIRNNHEDNSMVEHNVDFHACRPPGGGAEATNVAPGEEERQLRFKVTYYPGARFYHC  
GCA\_013407385.1\_Nitrosarchaeum\_sp.\_AC2 GDVVKMTLTIIPAGEVTCGNGNDMHSASQMSA . GNFESVNPGETSQYCYIAEAAGVFHYHC  
GCA\_000380525.1\_Thioalkalivibrio\_sp.\_AKL8 GNIIVRMIEISNSGD . . VMHSASIAAYTQTSKHVGHILPQGTQKSIYFRATTPGVFHYHC  
GCA\_000478095.1\_Corynebacterium\_sp.\_KPL1824 GDKFKITIKNNEGS . . MAHSIDFHAAGEVSPDENMKSIIQPEEELTYEFTANRAGIWFHYHC  
GCA\_000775045.1\_Burkholderia\_pseudomallei GDRVRLSMTNRSDEPMMHSMDFFHAMVSPTDKYRSIAPGQTMHLEFTFNYPGVFHYRC  
GCA\_001458695.1\_Nitrospira\_inopinata GDTVNFITLIGHKDNAFFHSMDFFAAELDFLKNYKTVGPGGETHFKFSFVAKKPGVFHYC  
GCA\_001723765.1\_Scalindua\_rubra GDETLVETLVNPKTNTETHTVDFHAIKFRGGATRMMAVPPGQSRRSFQITRPGLYHYC  
GCA\_003097635.1\_Flavobacterium\_psychrotolerans GDEVEFHLRNHPDNKLSHNIDLHAVTQGGGAASSLVAPAGHEKVFNFKTLNPGLYHYC  
GCA\_003135435.1\_Nitrospira\_sp. GDTVEINLKNDKNSKESHNIDFHAVNPGGGAAMLNTEPQGQESKLRFKALNAGLYHYC  
GCA\_003335105.1\_Rhodanobacter\_denitrificans GDTVEFHLKNAPDSKMEHNIDLHGVTPGGGAASSFTAPGHESQFTFKALNQGIYHYC  
GCA\_003495885.1\_Neisseria\_meningitidis GDTVEVEFNSNPSSSTVFEHNVDFHAAATQGGGAATFTAPGRTSTFSFKALQPGLYHYC  
GCA\_013267085.1\_Pseudomonas\_stutzeri GDTVTLNLTNELDSNHIEHSIDLHAVTQGGGAAVTQAAPGQTRSFYFKALQPGLYHYC  
GCA\_900492205.1\_Rhizobium\_naphthalenivorans GDYVELTTLINPETNTLQHNIDFHSSTALGGGALITVNPGEKTIILRFKATKAGVFHYC  
GCA\_905339145.1\_Anaerolineae\_bacterium GDTVVIINLNSDSKNTRAYAIDMPELNPRTVVTNLLMPGGETATLTFSAAKSGAYYYG  
Globobulimina\_pacifica GDWIDLTFINPNTSLHFEHSVDFFSMTLDGGAASLRINPGERARTVWQAIIPGMFYHC  
GCA\_000004985.1\_Naegleria\_gruberi GDVVMQVNYLNLDETGMANHNIDFHAVTQGGGAEMLLAEKDEEKTGFKLLTSGLFYHC  
GCA\_014843455.1\_2\_Fusarium\_oxysporum GDVVELTLTKNDPAGNEHNIDFHAFTPGGGAAVTTVEENESKTAGFKLLYPLGYHYC  
GCA\_000204585.1\_4\_outgroup1 GDLVRIHFINNGS . . . . . KHHTIHFFSIIHAEMDGVFEIVGAGGQFTYEFVAGPVGVHYC  
GCA\_003113895.1\_outgroup2 GDRIRVLFNLNAG . . . . . HSHLHFFSIVHPAEMDGI RPI SN G SATIYEEDAEPYGVHYC  
consensus>80 G# . v . . . . . n . . . . . h . . . . . d . h . . . . . g . . . . . f . . . . . G . f . Y . h . . . . .

130 140 150 160 170 180

GCA\_000337795.1\_Haloferax\_denitrificans ANVDYHISACGMFGIILVEPEEGLPEVDHHEFYLGQHELYYTNKGKQKGGHHEFDTRMAMEDP  
GCA\_013407385.1\_Nitrosarchaeum\_sp.\_AC2 VKMDQHVLSGMYGLTIVDPIIDGYNADALEFTLQYNQLYL . . . . . TPEGNYDAGKMFQHN  
GCA\_000380525.1\_Thioalkalivibrio\_sp.\_AKL8 GGIPMIVMFGQYGMIVVEPDRDPYKEDLKLMLQHELYA . . . . . SGKEAVEGEDA  
GCA\_000478095.1\_Corynebacterium\_sp.\_KPL1824 APMSIHIANGMAGNVIIDPDP.LKDVDLAEYFIANDVFL . . . . . GEEKTGADARVADGEF  
GCA\_000775045.1\_Burkholderia\_pseudomallei PMVLIHIASGMYGVVVVAPP RDGYPRADREYVIVQSEFYTKPDGTDALHVLDDGERLRRKAP  
GCA\_001458695.1\_Nitrospira\_inopinata SPMIQHVARGMFGAIIVDPKDVWPKADREFVLLVQSELWK . . . . . NPDNVQAMFDRKW  
GCA\_001723765.1\_Scalindua\_rubra NGVLIHIANNCMYGLLILVEPRD.LDPNLKEFYVMQGEFHI . . . . . NDEVPGNMDHEKGLREDP  
GCA\_003097635.1\_Flavobacterium\_psychrotolerans APVGMHIANCMYGLLILVEPEGGLPAVDKIEFYIMQGEFYTKGKGDDQLQAFDMDKAVKEQP  
GCA\_003135435.1\_Nitrospira\_sp. PSIPMHIANCMYGLLILVEPVGGLKKVDKIEFYIVQSEFYTKDGKKGDTLEFSFENGLAEHP  
GCA\_003335105.1\_Rhodanobacter\_denitrificans APVGMHIANCMYGLLILVEPEGLSPVDHEEYVMQGEFYTTGKREKGGHPFDMEKAIDEHP  
GCA\_003495885.1\_Neisseria\_meningitidis APVGMHIANCMYGLLILVEPEKGLPKVDKIEFYIVQGEFYTKGKGAQGLQPFDDMDKAI AEQP  
GCA\_0013267085.1\_Pseudomonas\_stutzeri PMVACHIITNCMYGLLILVEPEGLLAPVDHEFYVMQGEFYVTASPGERGLHEFSLDMLLRET  
GCA\_900492205.1\_Rhizobium\_naphthalenivorans PGVAVHVTSGMNGAMMVLPRDGLKVVYDYVVEGEFYVVPKDTAGDAYQDVLQVMRLTLP  
GCA\_905339145.1\_Anaerolineae\_bacterium AERTTHAAHGMFGALLLVEPLGGLPLMQKEFYIGRSEWYLGSTLKMSSFDDLEQKILAEQP  
Globobulimina\_pacifica GWASIHIAKGMFGAIIVEPYNGLPYSMDVYIIGQSEIYMHYPTPNLHNTFFDIKERYEMS  
GCA\_000004985.1\_Naegleria\_gruberi GPVPSHISNGMYGLMLVQPEEGLPKVDKIEFYVMQSEFYCEPSDDPKLMEHSYANGLDEKP  
GCA\_014843455.1\_2\_Fusarium\_oxysporum APVPMHIANCMYGLMLVQPEEEDLPVDKIEFYVMQSEFYHEPPRRSDTVEFSYPNGLREEP  
GCA\_000204585.1\_4\_outgroup1 MPLLEHHSISGLYGMFIVDPIKIPRP.QADEEMVVLNGLFD . . . . . D . . . . . FDTEN  
GCA\_003113895.1\_outgroup2 EPVTHHIAKGLYGMFIIDPPT.ARPPADEMVLVMG . . . . . GYD . . . . . VNDSDSHN  
consensus>80 . . . . . Hi . . . . . GmyG . . . . . !eP . . . . . e . . . . . e . y . . . . .

190 200 210 220 230 240

GCA\_000337795.1\_Haloferax\_denitrificans TYVLMNGEKYAITYAE MNVKTGETARIFYGVGGPNLNFSSFHPIGSDVWDEVWEQ GALASEP  
GCA\_013407385.1\_Nitrosarchaeum\_sp.\_AC2 TATVVNGMQFGYVSQQLLFVENDQH VRLFVENQG . NEPVFFFHIVG EILDR . VTQGN . . RVQ  
GCA\_000380525.1\_Thioalkalivibrio\_sp.\_AKL8 SYTAFNGQLFRYVENPIIPVRPGDYVRMYFLNVGNLSTFHIVGIVWDYVYVQGH . PEAL  
GCA\_000478095.1\_Corynebacterium\_sp.\_KPL1824 DLMANFYRPSQYDVPDIKAKVGDTVRLFVILNVGPDQPLSFHVVG EIQFDTAYKEGAYIKDE  
GCA\_000775045.1\_Burkholderia\_pseudomallei TYTVFNGRYRNGMVTQPLIAKPGERVRLYILNAGGSDTSSFHVVG EIFDRWLDGN . PDNQ  
GCA\_001458695.1\_Nitrospira\_inopinata DHTVFNGLIFKYHGEPLEVKVGERVRIYFVNA GPNFSAHPPIAEIWDAVYESGN . PSNK  
GCA\_001723765.1\_Scalindua\_rubra DYVVFNGRVNALSYP LRSITIGDNVYIYFGNAGPNKISSFHIVG EIFDKVWREGLVSP  
GCA\_003097635.1\_Flavobacterium\_psychrotolerans DYVVFNGKTGSLVSKAI TAKVGETVRLFVINGGPNLVSSFHIVG EIFDNVRIEPEG . LISMP  
GCA\_003135435.1\_Nitrospira\_sp. SHVVFNGMAGALIKNPLKAKVGDTRMFFYINAGPNLHVSAWHIIG EIFDNVRIEPEG . GALISMP  
GCA\_003335105.1\_Rhodanobacter\_denitrificans TYVLFNGREGSITDNAL TAKTDQKVRLFVNGGPNLVSSFHIVG EIFDKVQPEGG . TVA  
GCA\_003495885.1\_Neisseria\_meningitidis EYVVFNGHVGAIAADNAL KAKAGETVRMYVNGGPNLVSSFHIVG EIFDKVYVEGG . KLI  
GCA\_013267085.1\_Pseudomonas\_stutzeri QFMVFNGALDALTTHQMTV KAGDSVRIFFGVGGPNLISSFHIVG EIFDRVYDQGSLSAP  
GCA\_900492205.1\_Rhizobium\_naphthalenivorans SHIVFNGLAVGALTENAL KAEVGDRLIVHSQANRDRTRFHILGHG D YVWATGKFNHP  
GCA\_905339145.1\_Anaerolineae\_bacterium DFFTFNGHTQALQTNII VVDQYDQVRLFVVASGPNKGSDFYITGDIIDFK . VYTGHLCKVN  
Globobulimina\_pacifica DVVMMNGAPFALLHNPIITTVGSI VRAFIVSGGPNHLSSFHILIGHDFDRVWLHGDFANS  
GCA\_000004985.1\_Naegleria\_gruberi TYVVFNGREGSLIDTPLLANTGERIRFYFGNAGPNLSSSFHIVG EIFDKVREGLVSP  
GCA\_014843455.1\_2\_Fusarium\_oxysporum NVIAANTSESALTDKPLKANTGDSVRIFFGNAGPNLTSAFHILIGHFLFKVYRDGDVISPP  
GCA\_000204585.1\_4\_outgroup1 DFYAANTIPFYYQHHPIQINTNELIRIYVNMGEFDPINFHLHGNLYQY . FPTGT . SLVP  
GCA\_003113895.1\_outgroup2 NFYAFNGLPHYMHNP IRTYKDQLIRLYVLNIEYDPAVTFHLHANF . FD . VYRYGM . SMT  
consensus>80 . . . . . Ng . . . . . d . vr . f . . . . . g . n . . . . . fh . ig . . . . . d . . . . . G . . . . .

250 260 270 280 290

GCA\_000337795.1\_Haloferax\_denitrificans MRYVQTTPLVLPGSACVA TMSFPVVEGDFKLV DHALSRVARGALAIITAE GPEDTD  
GCA\_013407385.1\_Nitrosarchaeum\_sp.\_AC2 SAATETWLLGGSQGMIVDLVFDEH GAYAAVNH DYAAYTGAATVFA . . . . .  
GCA\_000380525.1\_Thioalkalivibrio\_sp.\_AKL8 WPGGQTVTTPGPSDSWVIEFRIPERGVYTMLDHGVGATSRGAI GLLDARADADTP  
GCA\_000478095.1\_Corynebacterium\_sp.\_KPL1824 KTGSQAFD LPAQGGFVEMTFNEHGTYSFVNHIMTNAEKQDHGK FV . . . . .  
GCA\_000775045.1\_Burkholderia\_pseudomallei LRGMQTVL LGSSGSAIVAFVPEAGAYVMVDHGFANASQAGVIVIDAGTHEEST  
GCA\_001458695.1\_Nitrospira\_inopinata FIGVQTYVVGPGSAA TFDVIADHEGAYPVVTHSLTGALRGAI AVV . . . . . NPK  
GCA\_001723765.1\_Scalindua\_rubra FRYVQTTAVPAGGAAVHLKSEEA GSYLLIDHSVFRIAKGAMGHLVSG . . . . .  
GCA\_003097635.1\_Flavobacterium\_psychrotolerans NHNVQTTLIPSGGAAIVEFKVDVHGTLLILVDHSIFRANKGALGLMKVEGPEDKT  
GCA\_003135435.1\_Nitrospira\_sp. LQNIQTTVVPAGGSSMAEFKVEVHGTYLNV DHSIFRIAKGAVGLLKVVEGAEQPD  
GCA\_003335105.1\_Rhodanobacter\_denitrificans QHNVQTTIPSGGAAIVEFKTDVHGTYLVDHSIFRANKGALGLLKVVEGAENKA  
GCA\_003495885.1\_Neisseria\_meningitidis NENVQSTIVPAGGAAIVEFKTDI HGSYTLVDHSIFRANKGALGQLVEGAEENKA  
GCA\_013267085.1\_Pseudomonas\_stutzeri LTDVQTTLVPAGGATMVEFVADYH GKYTLDV DHALSRAEKGLAGVLTVTDEGDSS  
GCA\_900492205.1\_Rhizobium\_naphthalenivorans ETDQETWFIPGGAAGAA Y YTFLOHGTIYAYVHNLI EAELGAAAHFAVTGDWDDD  
GCA\_905339145.1\_Anaerolineae\_bacterium . . . . . ARTAFVAPGSAVFDWRDLVHEGKYITGLDHALYRMAKVG . . . . .  
Globobulimina\_pacifica QQNIQTTVPPGGCAVADWDALHMEGAYFTLVDHALTRAVG . . . . .  
GCA\_000004985.1\_Naegleria\_gruberi ERNVQVTQVPPGGATMI EFEAIVEGTYSFV DHAIFRIEKGAVGFLKIAGAPRPD  
GCA\_014843455.1\_2\_Fusarium\_oxysporum . . . . .  
GCA\_000204585.1\_4\_outgroup1 STYTDMITLSQTERGIM E FQYQYH GKYL F H A H K V E F S E K G W V G L F L V K . . . . .  
GCA\_003113895.1\_outgroup2 SEKTDVITMGVAERH ILEFAFRYHEGKYMFH HODAI AENGMGQFEV . . . . .  
consensus>80 . . . . . q . . . . . g . y . . . . . n . . . . . g . . . . .

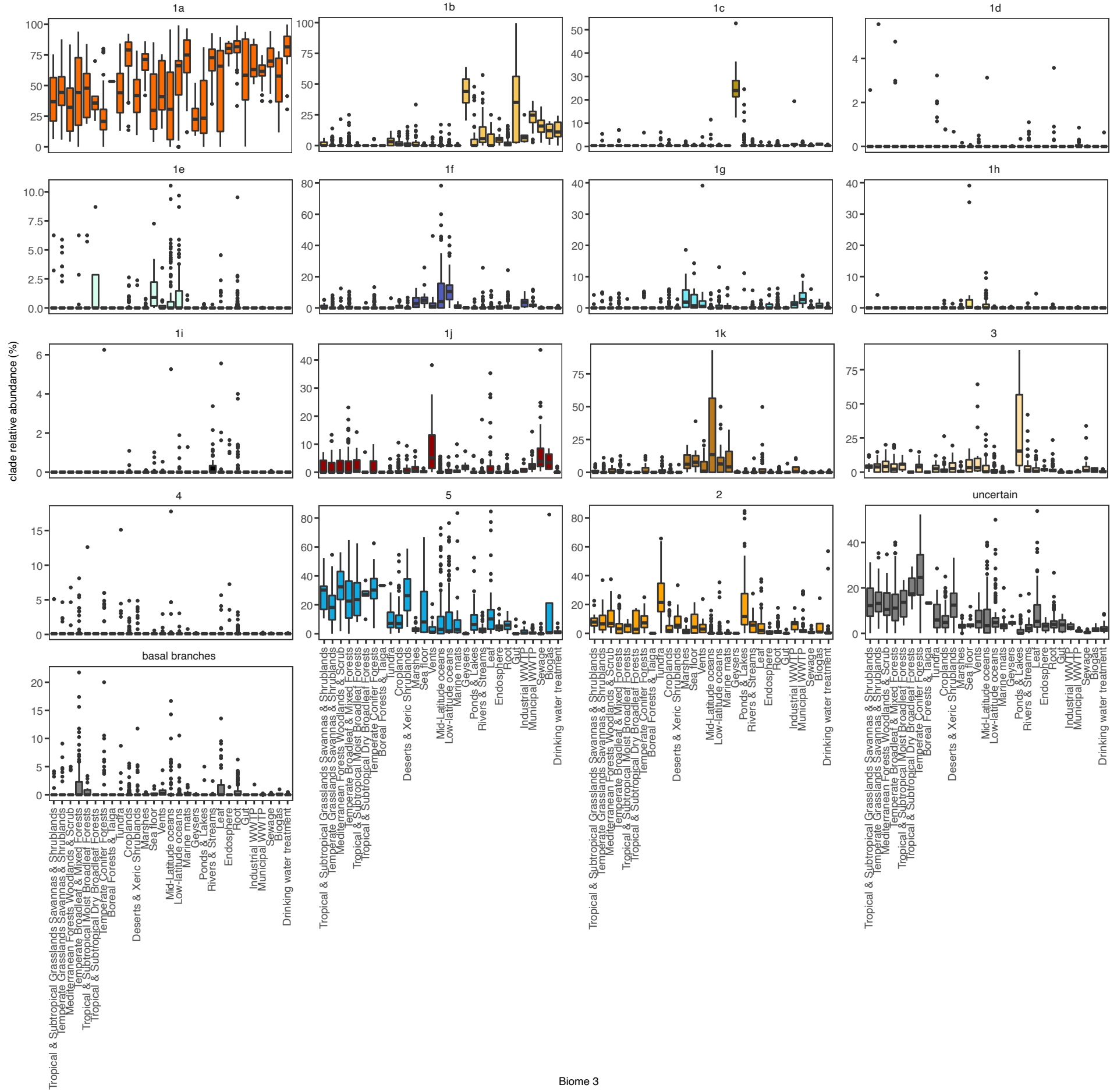
**Fig. S1 Representative NirS alignment.** Yellow boxes represent >70% conserved residues and red boxes represent identical residues. Numbered red boxes mark characteristic conserved motifs (I, III, IV, V, VII and IV distinguish NirS from NirN and NirF) and blue boxes mark functionally important residues, and asterisks beneath alignment mark specific residues where nirS can be differentiated from NirF and NirN. Notably, motif I on the *cyt<sub>c</sub>* domain was generally characterised by -Ca**G**CHg- in NirS, motif VI by -LHD- (vs. -PyD- in NirF and -LDD- in NirN), and motif VII at the *cyt<sub>d1</sub>* domain is identified by the universal absence of a proline in the third position in NirN and NirF (i.e. -PHpGpG- vs. PH!PgeG). Clade 1c: *Hydrogenophilales sp.*; clade 3: *Sulfurimonas parvalvinellae*; clade 4: *Calidifontibacillus erzurumensis*; clade 1d: *Brocadia sp. WS118*; clade 1e: *c. Magnetaquicoccus inordinatus*; clade 1a: *Stutzerimonas stutzeri*; clade 1f: *Deltaproteobacteria bact. GWA2\_43\_19*; clade 5: *Levilinea saccharolytica*; clade 1h: *Scalindua brodae*; clade 1j: *Thermus oshimai*; clade 1b: *Thauera linaloolentis*; clade 2: *Methylomicrobium album*; clade 6: *candidatus Methylomirabilis oxyfera*; clade 1g: *Roseiflexus castenholzii*; clade 1k: *Colwellia psychrerythraea*; halophilic archaea NirS-like: *Natrinema pellirubrum*. The Fig. was prepared in ESPscript/ENDscript.



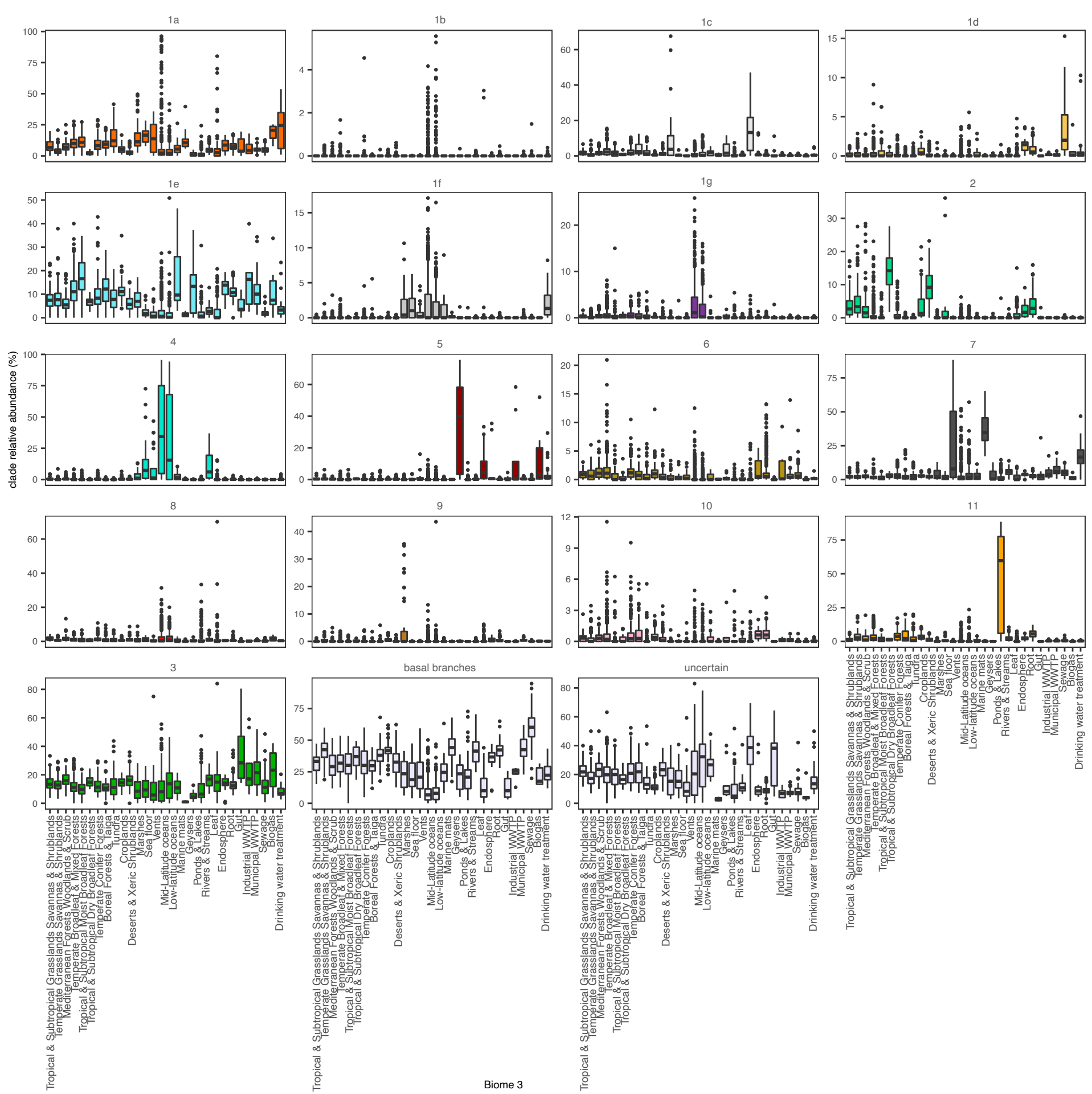




**Fig. S2 Representative NirK alignment.** Yellow boxes represent >70% conserved residues and red boxes represent identical residues. Green boxes mark characteristic conserved motifs differentiating NirK from other MCOs (**HniDfH-** (or **-HniSfH-** in clade 5; [91]) rather than **HtiHFH-** in motif I; **YHCap-** instead of **-YHCHvm-** in motif II; **-trpHvIG-** instead of **-insfHlHG-** in motif IV; and **-GiYAydh-** instead of **GkYmFHAH-** in motif V), and blue boxes mark functionally important residues. From top to bottom, the sequences represent: clade 1c: *Haloferax denitrificans*; clade 4: *Nitrosoarchaeum sp.*; clade 5: *Thioalkalivibrio thiocyanoxidans*; clade 4: *Nitrososphaera viennensis*; Clade 7: *Nitrospira nitrificans*; clade 1g: *Fusarium oxysporum*; cladeless anammox: *Scalindua rubra*; clade 8: *Paenibacillus amylolyticus*; clade 3: *Propionibacterium australiense*; clade 11: *Burkholderia mallei*; clade 1a: *Nitrosomonas oligotropha*; clade 9: *Rhodanobacter glycinis*. The Fig. was prepared in ESPscript/ENDscript.



**Fig. S3 Relative abundance of *nirS* clades in metagenomes from various biomes.** Basal branches indicate the best placement for a read occurred in regions of the phylogeny not belonging to a specific clade. Unknown indicates best placements was distributed across multiple clades. Remaining clades follow Fig. 2. Boxplots show median and quartiles, and whiskers show 95 percentiles, and values outside 95 percentiles are shown as points. Please note differences in y-axis scale across plots.



**Fig. S4 Relative abundance of *nirK* clades in metagenomes from various biomes.** Basal branches indicate the best placement for a read occurred in regions of the phylogeny not belonging to a specific clade. Unknown indicates best placements was distributed across multiple clades. Remaining clades follow Fig. 1. Boxplots show median and quartiles, and whiskers show 95 percentiles, and values outside 95 percentiles are shown as points. Please note differences in y-axis scale across plots.