

Environment

- Freshwater
- Soil
- Sediment
- Gut
- Oil
- Biofilm
- Marine
- Sludge

Tree scale: 0.1

Environment  
Clades

Phylum6

Phylum5

Phylum4-1

Phylum4

Phylum3

Phylum2.1

Phylum2.2

Lineage I

Endomicrobia

Phylum1.class5

Phylum2

Lineage IIc

Lineage IIa

Lineage IIb

Lineage IIa

Lineage VI

Lineage V

Lineage III

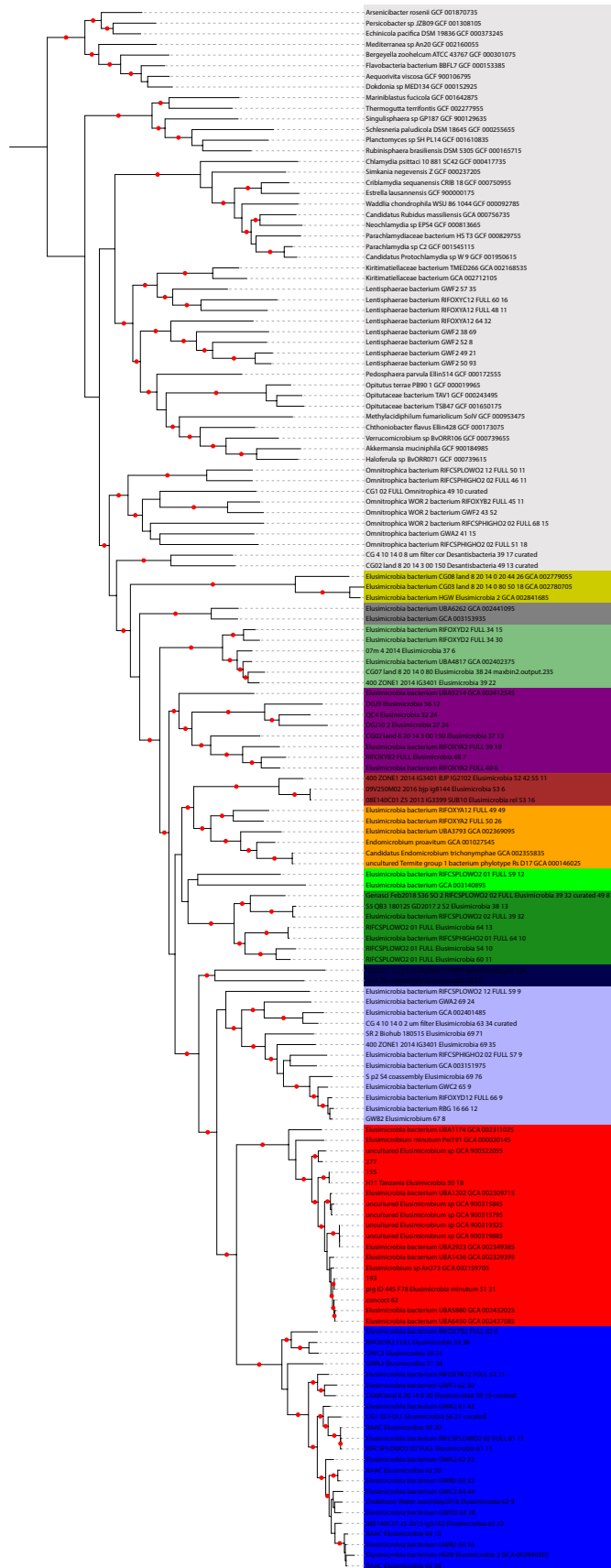
Elusimicrobiaceae

Lineage IV

Figure S1. Phylogenetic classification of 41 *Elusimicrobia*-related genomes based on their 16S rRNA gene sequences. Reference sequences from the SILVA database anchor the tree. The maximum likelihood phylogeny was inferred under the GTR model of evolution. For each *Elusimicrobia* genome, we indicated the lineage where the 16S sequence is nested in based on Herlemann *et al.* 2007; Geissinger *et al.* 2009 and Yarza *et al.* 2014. Scale bar indicates the average number of substitutions per site. The inner color bar indicates the environment of origin. Operational Taxonomic Unit (OTU) names colored in blue correspond to 16S from genome outgroups used in this study whereas OTUs colored in red correspond to the 16S from the *Elusimicrobia* genomes used in this study. OTUs colored in black were downloaded from the SILVA database.

Environments  
 New genomes  
 GTDB classification  
 16S rRNA classification  
 Final classification

Tree scale: 0.1 ⇐



*Bacteroidetes*

*Planctomycetes*

*Chlamydiae*

*Kiritimatiellaota*

*Lentisphaerae*

*Verrucomicrobia*

*Omnitrophica*

*Desantisbacteria*

*Elusimicrobia*-related lineage 2

*Elusimicrobia*-related lineage 1

Lineage IX

Lineage VIII

Lineage VII

Lineage I

*Endomicrobia*

Lineage IIa

Lineage IIc

Lineage VI

Lineage IV

Lineage V

Lineage III

*Elusimicrobiaceae*

Lineage V

## Environments

- Groundwater
- Animal-associated
- Sediment
- Palm oil effluent
- Peat
- Oil sand

## New genomes

- Genomes reported in this study
- Reference genomes

## GTDB classification (phylum-level)

- CG03
- UBA6262
- Elusimicrobiota

## 16S rRNA classification

- Lineage I
- Lineage IIa
- Lineage IIc
- Lineage III
- Lineage IV
- Lineage V
- Phylum1.class5
- Lineage VI
- Unknown
- Phylum3
- Phylum2.1
- Phylum2.2
- Phylum2

Figure S2. Placement of the 94 genomes related to the *Elusimicrobia* phylum. The maximum likelihood tree was constructed based on a concatenated alignment of 16 ribosomal proteins under an LG+gamma model of evolution. The inner color bar indicates the environment of origin. Black circles represent the four genomes described in previous studies whereas red circles represent the 30 new genomes added in this study. The remaining genomes come from the NCBI genome database. The two classifications based on 16S rRNA sequences and the GTDB are also indicated. Scale bars indicate the mean number of substitutions per site. The lineages of genomes are indicated by colored ranges and roman numbers. The bootstraps are indicated by red circles when  $\geq 85$ .

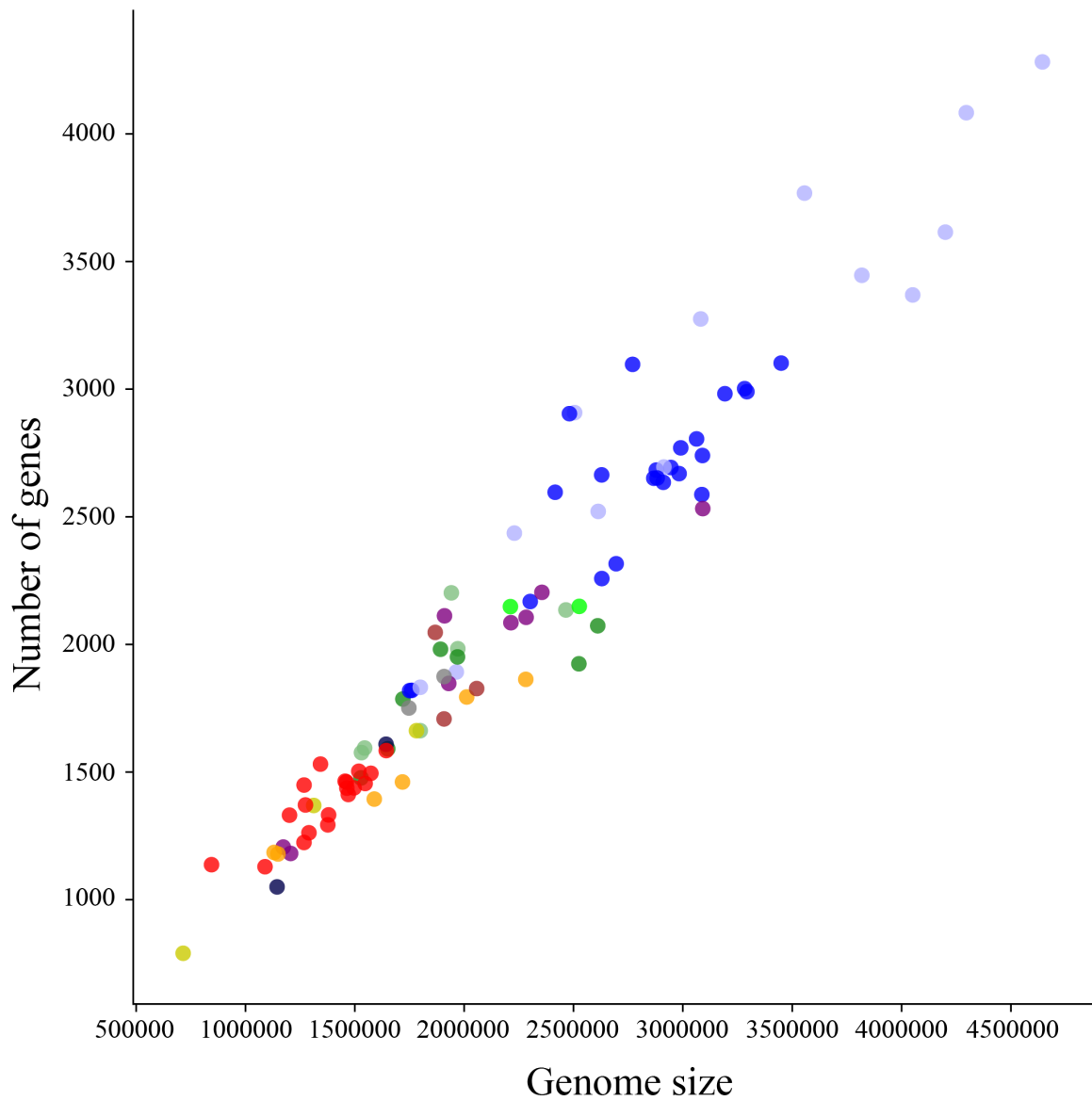


Figure S3. Number of genes and size (in base pairs) of the 94 *Elusimicrobia* genomes. Each genome is represented by a dot of the color of the lineage it has been classified to (see Figure 1 for color scheme).

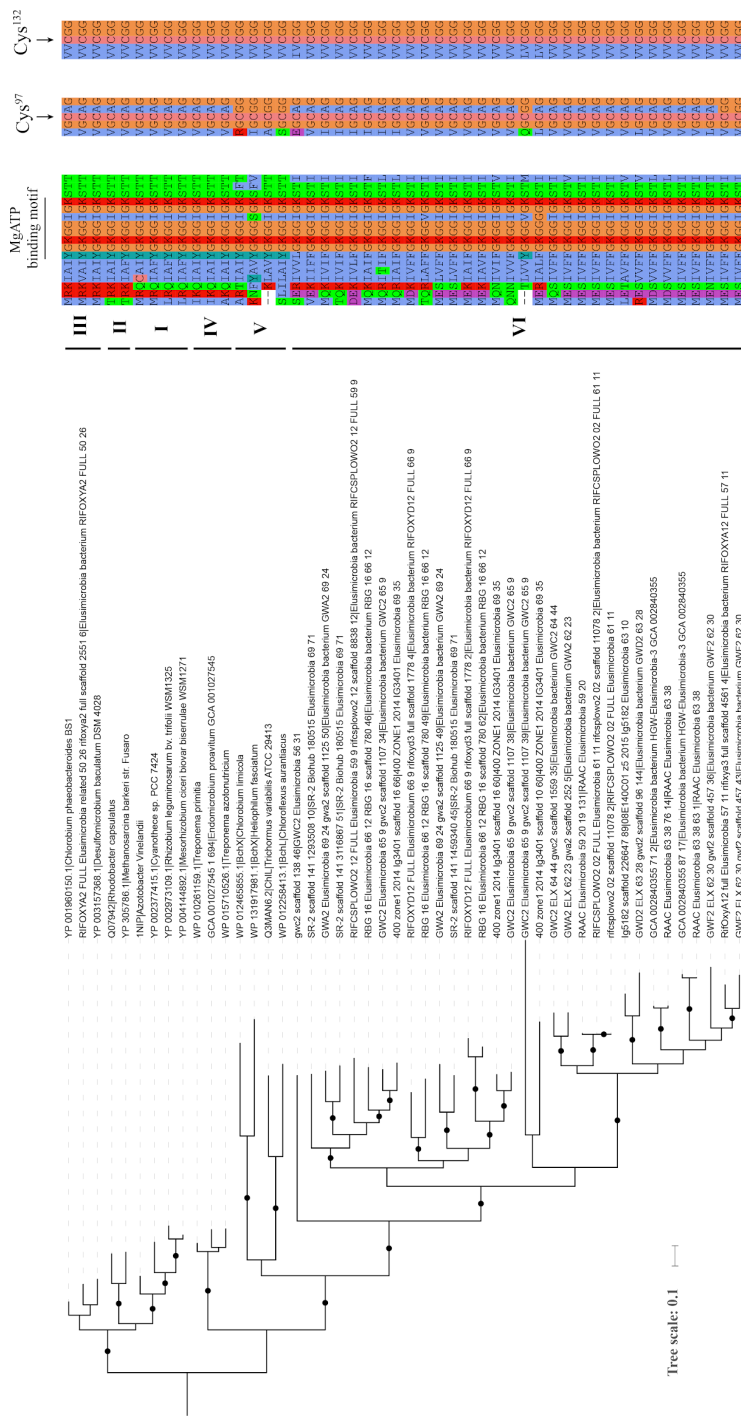


Figure S4. Phylogenetic tree of the *nifH* genes of *Elusimicrobia* and the nitrogenases of other bacterial species, and the conserved amino acids around the two crucial cysteine residues that coordinate the [Fe<sub>4</sub>S<sub>4</sub>] ligands in the Fe protein and for the binding of MgATP (GXGXXG). The cysteine residues, based on *A. vinelandii* numbering, are indicated with vertical arrows. Scale bar indicates substitutions per site.

## A. NifD

ZP\_0009772.2|Acetobacter\_vinelandii  
ZP\_00099859.2|Desulfobacterium\_hahnense\_DCB-2  
NP\_348897.1|Clostridium\_acetobutylicum\_ATCC\_824  
WP\_0157188.1|Elusimicrobia\_related\_50\_26\_rifoxya2\_full\_scaffold\_2551\_3  
ZP\_0007737.1|Elusimicrobia\_vinelandii  
ZP\_00092170.2|Acetobacter\_vinelandii  
ZP\_00011016.1|Rhodospseudomonas\_palisstis  
GCA\_001027445.1|898|Elusimicrobium\_proximum  
WP\_0157185.18|Troponeina\_azotolutricum  
WP\_0102581.85|Troponeina\_primilla  
WP\_0158171.45|Clostridium\_pasteurianum  
WP\_012486057.1|Rhodospseudomonas\_palisstis  
GCA\_001027445.1|898|Elusimicrobium\_proximum  
NP\_348897.1|Acetobacter\_vinelandii  
SR-2\_scaffold\_141\_1459340\_48  
SR-2\_scaffold\_141\_1459340\_16  
SR-2\_scaffold\_141\_1459340\_51  
GWA2\_Elusimicrobia\_69\_24\_gwa2\_scaffold\_1125\_50  
gwa2\_scaffold\_12344\_2  
400\_zone1\_2014\_ig3401\_scaffold\_107\_34  
RIFCSR\_OW02\_12\_FULL\_Elusimicrobia\_59\_9\_rifcsplow02\_12\_scaffold\_8838\_12  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_59\_9\_rifcsplow02\_12\_scaffold\_8838\_11  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_69\_24\_gwa2\_scaffold\_252\_6  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_57\_11\_rifoxya3\_full\_scaffold\_457\_37  
SR-2\_scaffold\_141\_1459340\_47  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_62\_6\_rifoxya2\_full\_scaffold\_84804\_1  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_59\_9\_rifcsplow02\_12\_scaffold\_8838\_3  
GWA2\_Elusimicrobia\_65\_9\_gwa2\_scaffold\_1107\_37  
GWA2\_Elusimicrobia\_65\_9\_gwa2\_scaffold\_1107\_37  
400\_zone1\_2014\_ig3401\_scaffold\_9\_20  
400\_zone1\_2014\_ig3401\_scaffold\_9\_3  
RBC\_16\_Elusimicrobia\_66\_12\_RBG\_16\_scaffold\_780\_161  
GWF2\_ELX\_62\_30\_gwf2\_scaffold\_457\_43  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_57\_11\_rifoxya2\_full\_scaffold\_4561\_4  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_61\_11\_rifcsplow02\_02\_scaffold\_11078\_2  
RAAC\_Elusimicrobia\_59\_20\_16\_131  
GWA2\_ELX\_62\_30\_gwa2\_scaffold\_1550\_35  
GWA2\_ELX\_62\_30\_gwa2\_scaffold\_1550\_35  
GWA2\_ELX\_62\_30\_gwa2\_scaffold\_1550\_35  
GWA2\_ELX\_62\_30\_gwa2\_scaffold\_1550\_35  
GWA2\_ELX\_62\_30\_gwa2\_scaffold\_1550\_35  
400\_zone1\_2014\_ig3401\_scaffold\_10\_59  
GWA2\_Elusimicrobia\_65\_9\_gwa2\_scaffold\_1107\_41  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_related\_50\_26\_rifoxya2\_full\_scaffold\_2551\_3  
GWA2\_Elusimicrobia\_69\_24\_gwa2\_scaffold\_1125\_51  
SR-2\_scaffold\_141\_1459340\_46  
GWA2\_Elusimicrobia\_65\_9\_gwa2\_scaffold\_1107\_59  
400\_zone1\_2014\_ig3401\_scaffold\_10\_58  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_66\_9\_rifoxya3\_full\_scaffold\_1778\_5

## B. NifK

WP\_01591391.0|Mesorhizobium\_japonicum  
ZP\_00099859.2|Desulfobacterium\_hahnense\_DCB-2  
NP\_488483.1|Nostoc\_sp\_PCC\_7120  
ZP\_0009773.1|Elusimicrobia\_vinelandii  
ZP\_00092170.2|Acetobacter\_vinelandii  
NP\_348898.1|Clostridium\_acetobutylicum\_ATCC\_824  
PI1347\_1|Clostridium\_pasteurianum  
ZP\_00089167.1|Acetobacter\_vinelandii  
NP\_276877.1|Methanothermobacter\_thermautotrophicus\_str\_Delta\_H  
NP\_276877.1|Methanothermobacter\_thermautotrophicus\_str\_Delta\_H  
ZP\_00092170.2|Acetobacter\_vinelandii  
FSY9022|Troponeina\_azotolutricum  
GWA2\_Elusimicrobia\_65\_9\_gwa2\_scaffold\_1107\_36  
gwa2\_scaffold\_12344\_6  
GWA2\_Elusimicrobia\_65\_9\_gwa2\_scaffold\_1107\_36  
400\_zone1\_2014\_ig3401\_scaffold\_9\_9  
400\_zone1\_2014\_ig3401\_scaffold\_9\_9  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_59\_9\_rifcsplow02\_12\_scaffold\_8838\_11  
GWA2\_Elusimicrobia\_65\_9\_gwa2\_scaffold\_1107\_33  
SR-2\_scaffold\_141\_1459340\_52  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_59\_9\_rifcsplow02\_12\_scaffold\_8838\_2  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_69\_24\_gwa2\_scaffold\_252\_6  
GWF2\_ELX\_62\_30\_gwf2\_scaffold\_457\_37  
SR-2\_scaffold\_141\_1459340\_16  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_59\_9\_rifcsplow02\_12\_scaffold\_8838\_2  
RAAC\_Elusimicrobia\_59\_20\_16\_132  
GWA2\_ELX\_62\_30\_gwa2\_scaffold\_252\_6  
GWA2\_ELX\_62\_30\_gwa2\_scaffold\_252\_6  
GCA\_002840355\_87\_18  
RAAC\_Elusimicrobia\_63\_38\_23\_47  
GWA2\_Elusimicrobia\_63\_38\_23\_47  
GWF2\_ELX\_62\_30\_gwf2\_scaffold\_457\_44  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_57\_11\_rifoxya3\_full\_scaffold\_4561\_3  
SR-2\_scaffold\_141\_1459340\_55  
RIFCSLOW02\_12\_FULL\_Elusimicrobia\_57\_11\_rifoxya3\_full\_scaffold\_1107\_40  
SR-2\_scaffold\_141\_1459340\_46  
400\_zone1\_2014\_ig3401\_scaffold\_6\_110  
GWA2\_Elusimicrobia\_69\_24\_gwa2\_scaffold\_1125\_46  
GWA2\_Elusimicrobia\_69\_24\_gwa2\_scaffold\_1125\_47

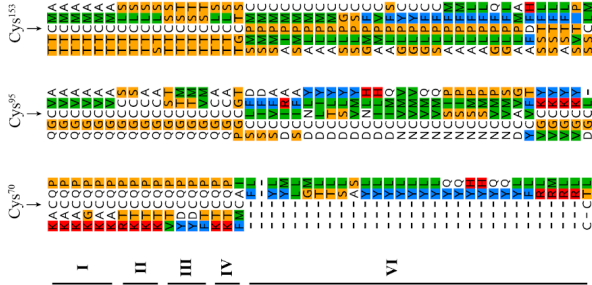
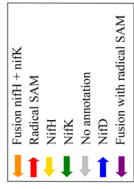
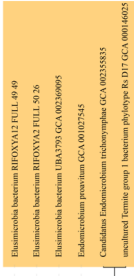
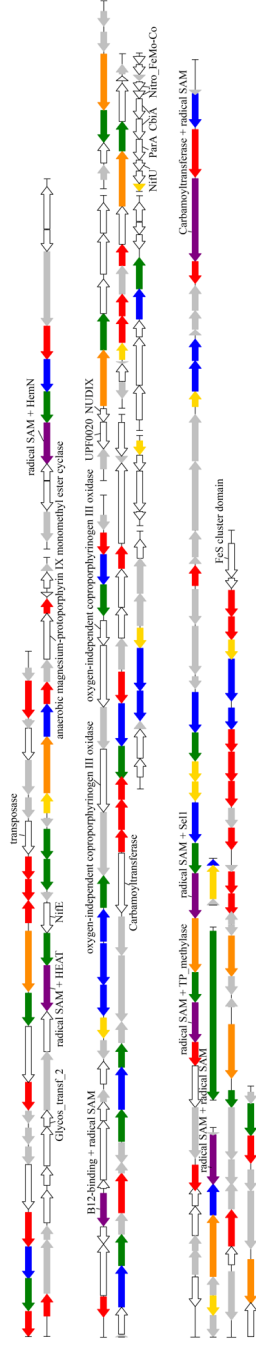
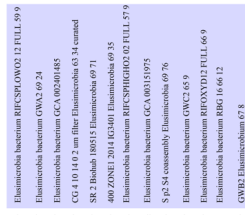


Figure S5. Conservation of the three crucial cysteine residues that coordinate the P-cluster in NifD (alpha) and NifK (beta) in *Elusimicrobia* and homologs in groups I, II, III and VI. The cysteine residues, based on *A. vinelandii* numbering, are indicated with vertical arrows.

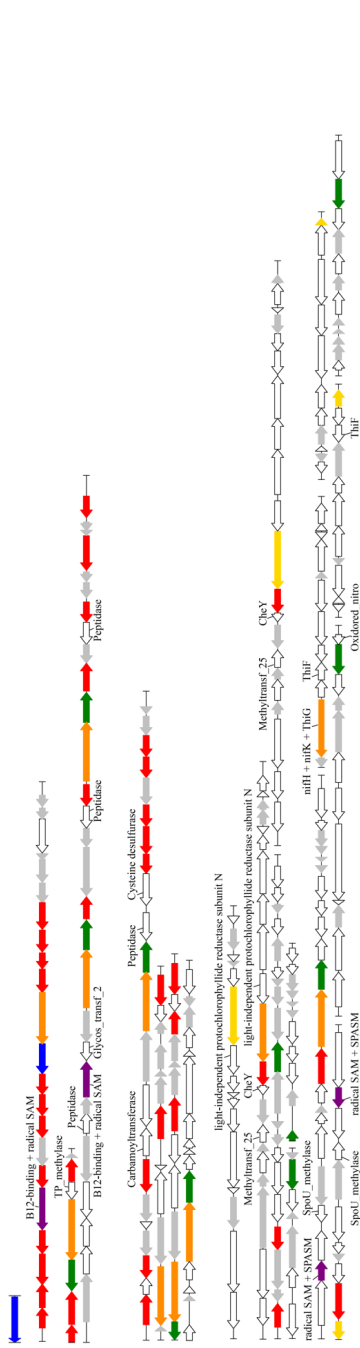
# Lineage I



# Lineage IV



# Lineage V



Tree scale: 0.1



Figure S6. Genomic neighborhood of nitrogenase in genomes from lineages I, IV and V. The 20 genes before and after the *nifH* genes are displayed when possible. Genes are colored according to their annotations. The maximum likelihood tree was constructed based on a concatenated alignment of 16 ribosomal proteins under an LG+gamma model of evolution (Figure 1). Scale bars indicate substitutions per site.

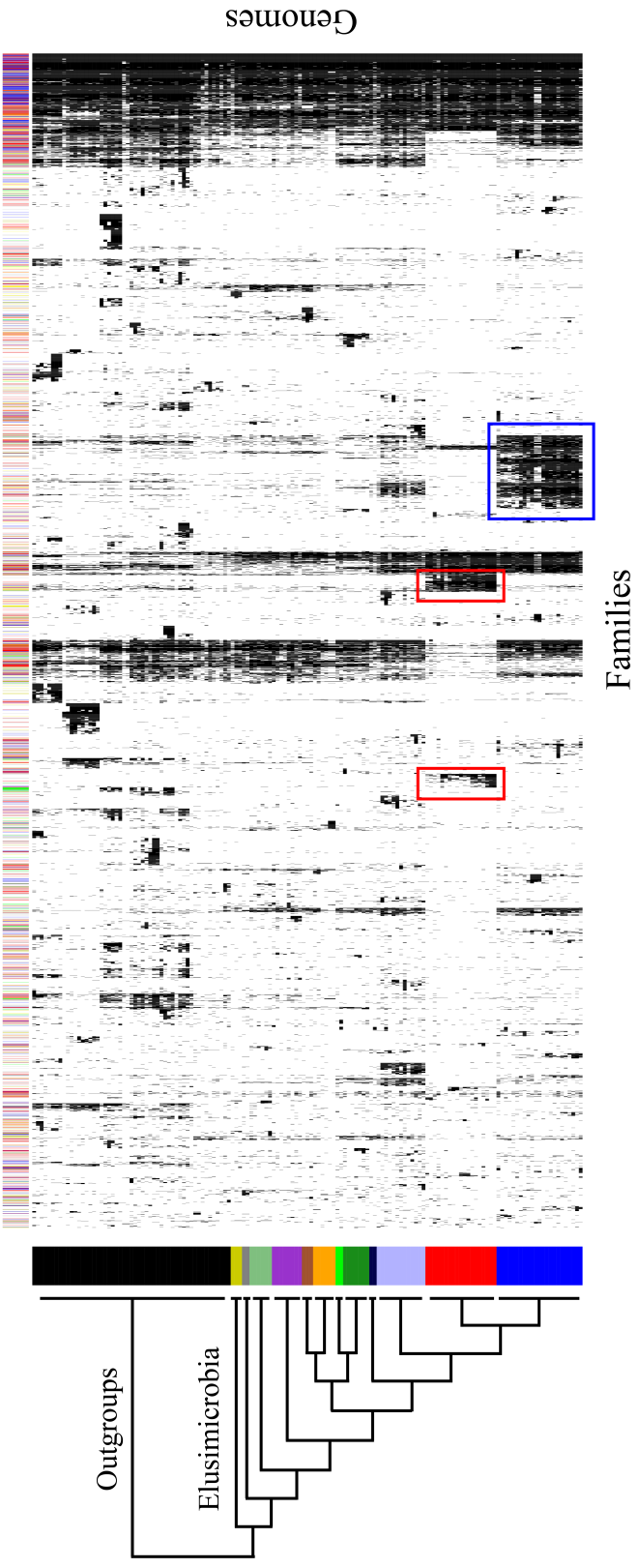


Figure S7. The distribution of 6,608 protein families across the 110 genomes. The distribution of 6,608 protein families (columns) in 110 genomes (rows). Genomes were sorted according to the phylogeny in Figure 1. Families are clustered based on the presence (black) / absence (white) profiles (using Jaccard distance and complete-linkage method). The colored top bar corresponds to the functional category of families (Metabolism: red, Genetic Information Processing: blue, Cellular Processes: green, Environmental Information Processing: yellow, Organismal systems: orange, Unclassified: grey, Unknown: white). Specific clusters of interest are highlighted with colored boxes, discussed in the main text.

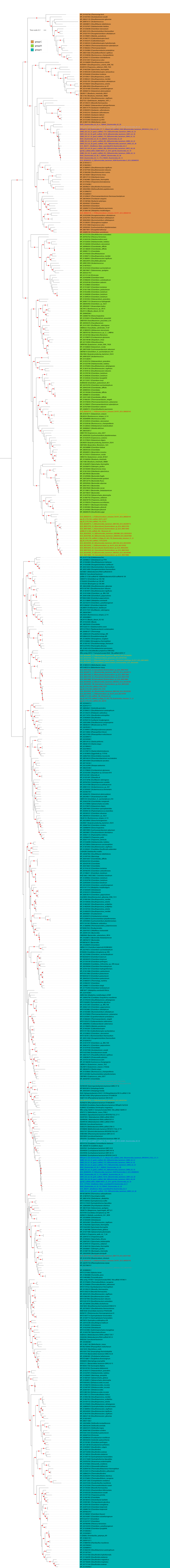


Figure S8. Phylogenetic tree of the [FeFe] hydrogenase (Maximum likelihood tree under an LG+I+G4 model of evolution). *Elusimicrobia* have group A, B and C [FeFe] hydrogenases. *Elusimicrobia* sequences are indicated with colored font. Scale bar indicates substitutions per site. The bootstraps are indicated by red circles when  $\geq 85$ . The tree was made using sequences from (Greening et al. 2016; Matheus Carnevali et al. 2019). The sequences and tree are available with full bootstrap values in Fasta and Newick formats in Supplementary Data.

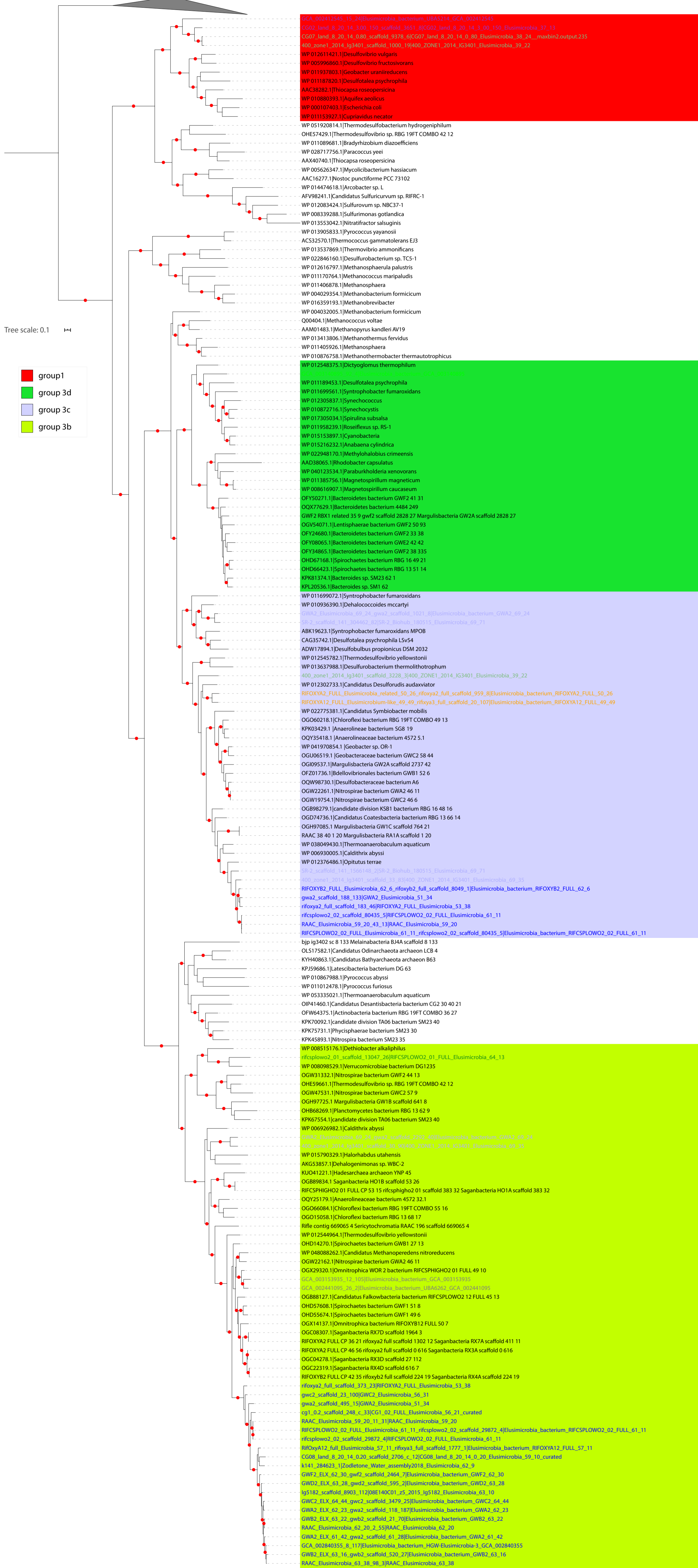
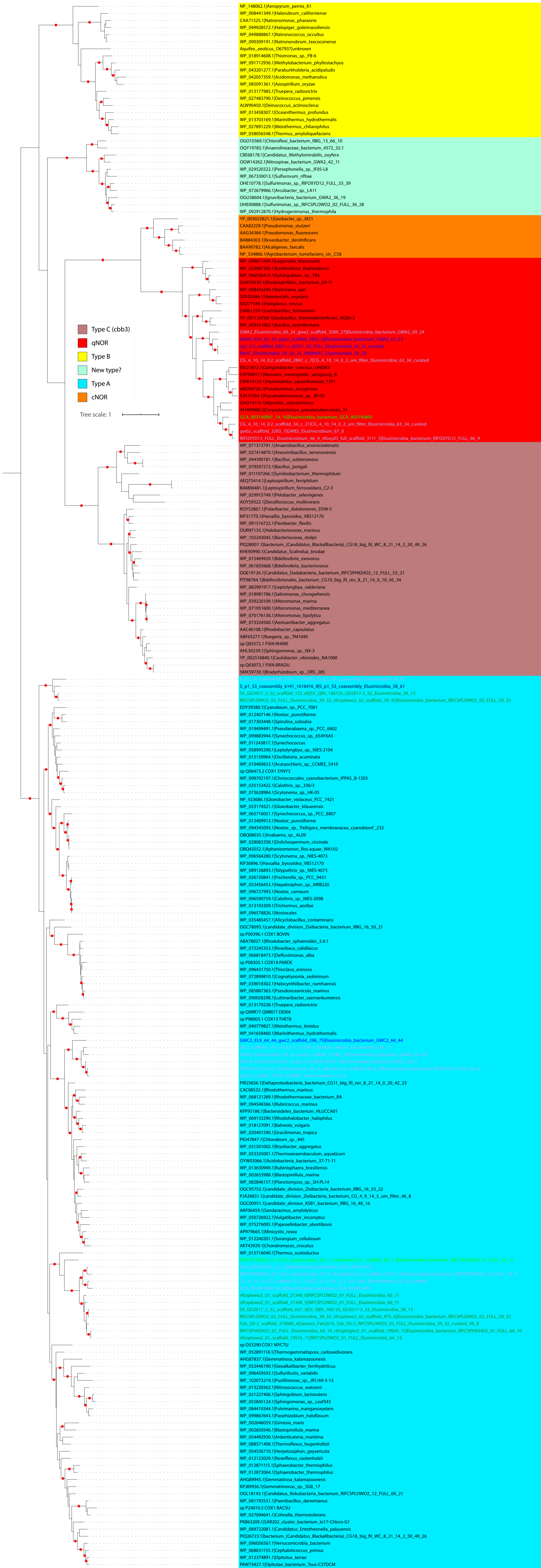


Figure S9. Phylogenetic tree of the groups 1, 2 and 3 of [NiFe] hydrogenase (Maximum likelihood tree under a LG+G4 model of evolution). *Elusimicrobia* have group 1, group 3b, 3c and 3d [NiFe] hydrogenases. *Elusimicrobia* sequences are indicated with colored font. Scale bar indicates substitutions per site. The bootstraps are indicated by red circles when  $\geq 85$ . The tree was made using sequences from (Matheus Carnevali et al. 2019). The sequences and tree are available with full bootstrap values in Fasta and Newick format in Supplementary Data.





Figure S10. Phylogenetic tree of the group 4 [NiFe] hydrogenase (Maximum likelihood tree under a LG+I+G4 model of evolution). *Elusimicrobia* have Mbh-Mrp, Mbh-Mrp-Oxidoreductase, Ech and Hydrogenase-related complex [NiFe] hydrogenases. They also have NADH dehydrogenase. *Elusimicrobia* sequences are indicated with colored font. Scale bar indicates substitutions per site. The bootstraps are indicated by red circles when  $\geq 85$ . The tree was made using sequences from (Matheus Carnevali et al. 2019). The sequences and tree are available with full bootstrap values in Fasta and Newick format in Supplementary Data.



- Type C (cbb3)
- qNOR
- Type B
- New type?
- Type A
- cNOR

Tree scale: 1

NP\_148062.1|Aeropyrum\_pemix\_K1  
WP\_008441349.1|Halorubrum\_californiense  
CAA71525.1|Natronomonas\_pharaonis  
WP\_049928572.1|Halopiger\_goleimassiliensis  
WP\_049888867.1|Natronococcus\_occultus  
WP\_090309191.1|Natronorubrum\_texcoconense  
Aquilifex\_aeolicus\_O67937|unknown  
WP\_018914608.1|Thiomonas\_sp\_FB-6  
WP\_091712936.1|Methylobacterium\_phyllostachyos  
WP\_043201277.1|Paraburkholderia\_acidipaludis  
WP\_042057359.1|Acidomonas\_methanolica  
WP\_085091361.1|Azospirillum\_oryzae  
WP\_013177985.1|Truepera\_radiovictrix  
WP\_027483790.1|Deinococcus\_pimensis  
ALW90450.1|Deinococcus\_actinoscleris  
WP\_013458307.1|Oceanithermus\_profundus  
WP\_013703169.1|Marinithermus\_hydrothermalis  
WP\_027891229.1|Meiothermus\_chliarophilus  
WP\_038056548.1|Thermus\_amyloliquefaciens  
OGO10369.1|Chloroflexi\_bacterium\_RBG\_13\_66\_10  
OQY19783.1|Anaerolineaceae\_bacterium\_4572\_32.1  
CBE68178.1|Candidatus\_Methylomirabilis\_oxifera  
OGW14262.1|Nitrospirae\_bacterium\_GWA2\_42\_11  
WP\_029520322.1|Persephonella\_sp\_IF05-L8  
WP\_067330013.1|Sulfurovum\_riftiae  
OHE10778.1|Sulfurimonas\_sp\_RIFOXD12\_FULL\_33\_39  
WP\_072679966.1|Arcobacter\_sp\_LA11  
OGU38604.1|Ignavibacteria\_bacterium\_GWA2\_36\_19  
OHE00888.1|Sulfurimonas\_sp\_RIFCSLOWO2\_02\_FULL\_36\_28  
WP\_092912870.1|Hydrogenimonas\_thermophila  
YP\_003022821.1|Geobacter\_sp\_M21  
CAA82229.1|Pseudomonas\_stutzeri  
AAG34384.1|Pseudomonas\_fluorescens  
BAB84303.1|Roseobacter\_denitrificans  
BAA90782.1|Alcaligenes\_faecalis  
NP\_534866.1|Agrobacterium\_tumefaciens\_str\_CS8  
WP\_006871449.1|Legionella\_drancourtii  
WP\_025987383.1|Burkholderia\_thailandensis  
WP\_066556414.1|Sphingobium\_sp\_TKS  
OJW29330.1|Rhodospirillales\_bacterium\_69-11  
WP\_008456294.1|Natriema\_gari  
SDF02486.1|Halorientalis\_regularis  
SDZ77349.1|Haloplanus\_vescus  
SJM61259.1|Lactobacillus\_fermentum  
YP\_001124768.1|Geobacillus\_thermodenitrificans\_NG80-2  
WP\_003331062.1|Bacillus\_azotoformans  
GWA2\_Elusimicrobia\_69\_24\_gwa2\_scaffold\_3589\_27|Elusimicrobia\_bacterium\_GWA2\_69\_24  
GWA2\_ELX\_61\_21\_gwa2\_scaffold\_3625\_6|Elusimicrobia\_bacterium\_GWA2\_61\_21  
gq1\_02\_scaffold\_3001\_11|CG1\_02\_FULL\_Elusimicrobia\_09\_21\_curated  
BAC\_Elusimicrobia\_09\_21\_2|BAC\_Elusimicrobia\_09\_21  
CG\_4\_10\_14\_0\_2\_scaffold\_2897\_c\_7|CG\_4\_10\_14\_0\_2\_um\_filter\_Elusimicrobia\_63\_34\_curated  
ERJ21813.1|Campylobacter\_conciscus\_UNSW3  
CRY99017.1|Neisseria\_meningitidis\_serogroup\_B  
CBW14123.1|Haemophilus\_painfuluentiae\_T3T1  
ABD94720.1|Pseudomonas\_aeruginosa  
SJM17569.1|Pseudoalteromonas\_sp\_JB197  
GAQ14115.1|Myroides\_odoratimimus  
AFH89980.2|Corynebacterium\_pseudotuberculosis\_31  
GCA\_003140805\_14\_10|Elusimicrobia\_bacterium\_GCA\_003140805  
CG\_4\_10\_14\_0\_2\_scaffold\_34\_c\_21|CG\_4\_10\_14\_0\_2\_um\_filter\_Elusimicrobia\_63\_34\_curated  
gwb2\_scaffold\_3283\_7|GWB2\_Elusimicrobium\_67\_8  
RIFOXD12\_FULL\_Elusimicrobium\_66\_9\_rifoxd3\_full\_scaffold\_3111\_3|Elusimicrobia\_bacterium\_RIFOXD12\_FULL\_66\_9  
WP\_071313791.1|Anaerobacillus\_arseniciselenatis  
WP\_027414870.1|Aneurinibacillus\_terranovensis  
WP\_044390181.1|Bacillus\_subterraneus  
WP\_079507272.1|Bacillus\_jeotgali  
WP\_011197266.1|Symbiobacterium\_thermophilum  
AEQ75414.1|Leptospirillum\_ferriphilum  
BAM06481.1|Leptospirillum\_ferrooxidans\_C2-3  
WP\_029915749.1|Pelobacter\_selenigenes  
AOY59322.1|Desulfococcus\_multivorans  
KOY52867.1|Polaribacter\_dokdonensis\_DSU-5  
KIF31770.1|Hassallia\_byssoidea\_VB512170  
WP\_091516722.1|Flexibacter\_flexilis  
OUR97135.1|Halobacteriovorax\_marinus  
WP\_102243045.1|Bacteriovorax\_stolpii  
PIQ28007.1|bacterium\_[Candidatus\_Blackallbacteria]\_CG18\_big\_fil\_WC\_8\_21\_14\_2\_50\_49\_26  
KHE09090.1|Candidatus\_Scalindua\_brodiae  
WP\_015469929.1|Bdellovibrio\_exovorus  
WP\_061835668.1|Bdellovibrio\_bacteriovorus  
OGE19126.1|Candidatus\_Dadabacteria\_bacterium\_RIFCSPHIGHO2\_12\_FULL\_53\_21  
PIT98784.1|Bdellovibrionales\_bacterium\_CG10\_big\_fil\_rev\_8\_21\_14\_0\_10\_45\_34  
WP\_082901917.1|Leptolyngbya\_valderiana  
WP\_018981786.1|Salinimonas\_chungwhensis  
WP\_039220109.1|Alteromonas\_marina  
WP\_071951600.1|Alteromonas\_mediterranea  
WP\_070176136.1|Alteromonas\_lipolytica  
WP\_073324560.1|Aestuariaibacter\_aggregatus  
AAC46108.1|Rhodobacter\_capsulatus  
ABF65277.1|Ruegeria\_sp\_TM1040  
sp\_Q05572.1|FIXN RHIME  
AHL30239.1|Sphingomonas\_sp\_NX-3  
YP\_002516840.1|Caulobacter\_vibrioides\_NA1000  
sp\_Q03073.1|FIXN BRADU  
SMX59739.1|Bradyrhizobium\_sp\_OR5\_285  
GCA\_003140805\_14\_10|Elusimicrobia\_bacterium\_GCA\_003140805  
S\_p1\_S3\_coassembly\_k141\_1616416\_8|S\_p1\_S3\_coassembly\_Elusimicrobia\_58\_61  
SS\_GD2017\_2\_S2\_scaffold\_122\_60|SS\_QB3\_180125\_GD2017-2\_S2\_Elusimicrobia\_38\_13  
RIFCSLOWO2\_02\_FULL\_Elusimicrobia\_39\_32\_rifcsplow2\_02\_scaffold\_39\_42|Elusimicrobia\_bacterium\_RIFCSLOWO2\_02\_FULL\_39\_32  
EDY39380.1|Cyanobium\_sp\_PCC\_7001  
WP\_012407146.1|Nostoc\_punctiforme  
WP\_017303448.1|Spirulina\_subsalsa  
WP\_019499491.1|Pseudanabaena\_sp\_PCC\_6802  
WP\_099883944.1|Synechococcus\_sp\_65AY6A5  
WP\_011243817.1|Synechococcus  
WP\_058995290.1|Leptolyngbya\_sp\_NIES-2104  
WP\_015150964.1|Oscillatoria\_acuminata  
WP\_010469823.1|Acaryochloris\_sp\_CCME5\_5410  
sp\_Q06473.2|COX1 SYNY3  
WP\_099702197.1|Chroococcales\_cyanobacterium\_IPPA5\_B-1203  
WP\_035153422.1|Calothrix\_sp\_336/3  
WP\_073628984.1|Scytonema\_sp\_HK-05  
NP\_923686.1|Gloeobacter\_violaceus\_PCC\_7421  
WP\_023174521.1|Gloeobacter\_kilauensis  
WP\_065716051.1|Synechococcus\_sp\_PCC\_8807  
WP\_012409913.1|Nostoc\_punctiforme  
WP\_094345093.1|Nostoc\_sp\_'Peltigera\_membranacea\_cyanobiont'\_232  
OBQ08035.1|Anabaena\_sp\_AL09  
WP\_028083358.1|Dolichospermum\_circinale  
OBQ43552.1|Aphanizomenon\_flos-aquae\_WA102  
WP\_096564280.1|Scytonema\_sp\_NIES-4073  
KIF36896.1|Hassallia\_byssoidea\_VB512170  
WP\_089126893.1|Tolypothrix\_sp\_NIES-4075  
WP\_026720841.1|Fischerella\_sp\_PCC\_9431  
WP\_053456453.1|Hapalosiphon\_sp\_MRB220  
WP\_096727993.1|Nostoc\_carneum  
WP\_096590759.1|Calothrix\_sp\_NIES-2098  
WP\_013193309.1|Trichormus\_azollae  
WP\_096578826.1|Nostocales  
WP\_035485457.1|Alicyclobacillus\_contaminans  
OGC78093.1|candidate\_division\_Zixibacteria\_bacterium\_RBG\_16\_50\_21  
sp\_P00396.1|COX1 BOVIN  
ABA78027.1|Rhodobacter\_sphaeroides\_24.1  
WP\_072245353.1|Roseibaca\_caldilacus  
WP\_066818473.1|Defluviimonas\_alba  
sp\_P08305.1|COX1A PARDE  
WP\_096431750.1|Thioclava\_arenosa  
WP\_072899810.1|Cognatyyoonia\_sediminum  
WP\_039018302.1|Halocynthiibacter\_namhaensis  
WP\_085887363.1|Pseudoceaneicola\_marinus  
WP\_090028298.1|Lutimariibacter\_saemankumensis  
WP\_013179238.1|Truepera\_radiovictrix  
sp\_Q9RR77\_Q9RR77 DEIRA  
sp\_P98005.1|COX13 THET8  
WP\_040779827.1|Meiothermus\_timidus  
WP\_041658460.1|Marinithermus\_hydrothermalis  
GWC2\_ELX\_64\_44\_gwc2\_scaffold\_296\_75|Elusimicrobia\_bacterium\_GWC2\_64\_44  
SR2\_scaffold\_141\_1015313\_157|SR2\_Biotub\_180515\_Elusimicrobia\_69\_71  
GWA2\_Elusimicrobia\_09\_24\_gwa2\_scaffold\_16796\_13|Elusimicrobia\_bacterium\_GWA2\_69\_24  
GWC2\_Elusimicrobia\_05\_9\_gwc2\_scaffold\_0933\_7|Elusimicrobia\_bacterium\_GWC2\_65\_9  
RIFOXD12\_FULL\_Elusimicrobium\_66\_9\_rifoxd3\_full\_scaffold\_8618\_3|Elusimicrobia\_bacterium\_RIFOXD12\_FULL\_66\_9  
gwb2\_scaffold\_37434\_2|GWB2\_Elusimicrobium\_67\_8  
PIR25656.1|Deltaproteobacteria\_bacterium\_CG11\_big\_fil\_rev\_8\_21\_14\_0\_20\_42\_23  
CAC08532.1|Rhodothermus\_marinus  
WP\_068121289.1|Rhodothermaceae\_bacterium\_RA  
WP\_094548366.1|Rubricoccus\_marinus  
KPP95186.1|Bacteroidetes\_bacterium\_HLUCCA01  
WP\_069133290.1|Rhodohalobacter\_halophilus  
WP\_018127091.1|Balneola\_vulgaris  
WP\_020401590.1|Gracilimonas\_tropica  
PIQ47847.1|Chlorobium\_sp\_445  
WP\_031501002.1|Bryobacter\_aggregatus  
WP\_053335001.1|Thermoanaerobaculum\_aquaticum  
OYW03066.1|Acidobacteria\_bacterium\_37-71-11  
WP\_013630949.1|Rubinisphaera\_brasiliensis  
WP\_002653988.1|Blastopirellula\_marina  
WP\_082846157.1|Planctomyces\_sp\_SH-PL14  
OGC95752.1|candidate\_division\_Zixibacteria\_bacterium\_RBG\_16\_50\_22  
PJA26831.1|candidate\_division\_Zixibacteria\_bacterium\_CG\_4\_9\_14\_3\_um\_filter\_46\_8  
OGC00951.1|candidate\_division\_KSB1\_bacterium\_RBG\_16\_48\_16  
AKF06459.1|Sandaracinus\_amyolyticus  
WP\_050726922.1|Vulgatibacter\_incomptus  
WP\_075276095.1|Pajaroellobacter\_abortibovis  
APR79665.1|Minicystis\_rosea  
WP\_012240201.1|Sorangium\_cellulosum  
AKT43929.1|Chondromyces\_crocalus  
WP\_015716040.1|Thermus\_scotoductus  
RIFCSLOWO2\_01\_FULL\_Elusimicrobia\_59\_12\_rifcsplow2\_01\_scaffold\_8371\_8|Elusimicrobia\_bacterium\_RIFCSLOWO2\_01\_FULL\_59\_12  
GCA\_003151975\_12\_12|Elusimicrobia\_bacterium\_GCA\_003151975  
RIFCSPHIGHO2\_02\_FULL\_Elusimicrobia\_57\_9\_rifcsphigh2\_02\_scaffold\_37722\_12|Elusimicrobia\_bacterium\_RIFCSPHIGHO2\_02\_FULL\_57\_9  
CG\_4\_10\_14\_0\_2\_scaffold\_76\_c\_53|CG\_4\_10\_14\_0\_2\_um\_filter\_Elusimicrobia\_63\_34\_curated  
GCA\_002401485\_86\_8|Elusimicrobia\_bacterium\_GCA\_002401485  
rifcsplow2\_01\_scaffold\_21348\_6|RIFCSLOWO2\_01\_FULL\_Elusimicrobia\_60\_11  
rifcsplow2\_01\_scaffold\_21348\_5|RIFCSLOWO2\_01\_FULL\_Elusimicrobia\_60\_11  
SS\_GD2017\_2\_S2\_scaffold\_637\_4|SS\_QB3\_180125\_GD2017-2\_S2\_Elusimicrobia\_38\_13  
RIFCSLOWO2\_02\_FULL\_Elusimicrobia\_39\_32\_rifcsplow2\_02\_scaffold\_973\_6|Elusimicrobia\_bacterium\_RIFCSLOWO2\_02\_FULL\_39\_32  
S36\_SO-2\_scaffold\_379086\_4|Genascl\_Feb2018\_S36\_SO-2\_RIFCSLOWO2\_02\_FULL\_Elusimicrobia\_39\_32\_curated\_49\_8  
RIFCSPHIGHO2\_01\_FULL\_Elusimicrobia\_64\_10\_rifcsphigh2\_01\_scaffold\_19950\_15|Elusimicrobia\_bacterium\_RIFCSPHIGHO2\_01\_FULL\_64\_10  
rifcsplow2\_01\_scaffold\_19516\_11|RIFCSLOWO2\_01\_FULL\_Elusimicrobia\_64\_13  
sp\_O53290\_COX1 MYCTU  
WP\_052891118.1|Thermogemmatispora\_carboxidivorans  
AHG87837.1|Gemmatirosa\_kalamazoonesis  
WP\_052446190.1|Geoalkalibacter\_ferrihydriticus  
WP\_096459593.1|Sulfurifustis\_variabilis  
WP\_102073219.1|Pusillimonas\_sp\_JR1/69-3-13  
WP\_013220362.1|Nitrosococcus\_watsonii  
WP\_021227406.1|Sphingobium\_lactosutens  
WP\_055845124.1|Sphingomonas\_sp\_Leaf343  
WP\_084410344.1|Fulvimarina\_manganosydans  
WP\_099867643.1|Pararhizobium\_haloflavum  
WP\_002646059.1|Gimesia\_maris  
WP\_002650546.1|Blastopirellula\_marina  
WP\_054492930.1|Ardenticatena\_maritima  
WP\_088571408.1|Thermoflexus\_hugenholtzii  
WP\_054536710.1|Herpetosiphon\_geysericola  
WP\_012122029.1|Roseiflexus\_castenholzii  
WP\_012871115.1|Sphaerobacter\_thermophilus  
WP\_012872064.1|Sphaerobacter\_thermophilus  
AHG89945.1|Gemmatirosa\_kalamazoonesis  
KPJ89936.1|Gemmatimonas\_sp\_SG8\_17  
OGL18143.1|Candidatus\_Rokubacteria\_bacterium\_RIFCSLOWO2\_12\_FULL\_69\_21  
WP\_081793531.1|Paenibacillus\_darwinianus  
sp\_P24010.3|COX1 BACSU  
WP\_027094641.1|Cohnella\_thermotolerans  
PKB63209.1|SAR202\_cluster\_bacterium\_lo17-Chloro-G1  
WP\_089722081.1|Candidatus\_Entotheonella\_palauensis  
PIQ26763.1|bacterium\_[Candidatus\_Blackallbacteria]\_CG18\_big\_fil\_WC\_8\_21\_14\_2\_50\_49\_26  
WP\_096056567.1|Verrucomicrobia\_bacterium  
WP\_068631155.1|Cephalotococcus\_primus  
PAW73427.1|Opitutae\_bacterium\_Tous-CSTDCM

Figure S11. Phylogenetic tree of the catalytic subunit I of heme-copper oxygen reductases (Maximum likelihood tree under a LG+F+I+G4 model of evolution). *Elusimicrobia* have an aa<sub>3</sub>-type heme-copper oxygen reductase (Type A) and nitric oxide reductase that uses quinol as the electron donor (qNOR). *Elusimicrobia* sequences are indicated with colored font. The bootstraps are indicated by red circles when  $\geq 85$ . The tree was made using sequences from (Matheus Carnevali et al. 2019). Scale bar indicates substitutions per site. The sequences and tree are available with full bootstrap values in Fasta and Newick format in Supplementary Data.



Figure S12. Phylogenetic analysis of the dimethyl sulfoxide (DMSO) reductase superfamily (Maximum likelihood tree under a LG+F+G4 model of evolution). *Elusimicrobia* have an alternative complex III, a tetrathionate reductase (TtrA), cytoplasmic and periplasmic nitrite oxidoreductase (NXR), formate dehydrogenase and nitrate reductase (NapA and NasA). *Elusimicrobia* sequences are indicated with colored font. Scale bar indicates substitutions per site. The bootstraps are indicated by red circles when  $\geq 85$ . The tree was made using sequences from (Castelle et al. 2013; Matheus Carnevali et al. 2019). The sequences and tree are available with full bootstrap values in Fasta and Newick format in Supplementary Data.

## SUPPLEMENTARY DATASET

**Supplementary Dataset 1.** List of the 94 genomes used in this study. For each genome (column A), the NCBI genome accession, completeness and the contamination based on single copy genes are displayed in columns B, C and D respectively. Columns from E to J informs about the concatenated ribosomal proteins. Environments are provided in columns K and L, pubmed PMID in column M. The GTDB classification is shown in column N, classification based on 16S rRNA in column O. Final classification is shown in column P.

**Supplementary Dataset 2.** List of the 17 genomes from Myxococcales and Gracilibacteria genomes with the group VI nifH. Genome names and taxonomy are provided in columns A and C. When available, NCBI assembly accession is provided in column B.

**Supplementary Dataset 3.** Metabolic capacities across the 94 Elusimicrobia genomes. Genome names and environments are provided in columns A and B. Each metabolic feature is considered either complete (indicated by 1 and a red cell), partial (indicated by P and an orange cell) or absent (indicated by 0 and a white cell). Completeness of the pathways is based on literature and annotations (see the row 2 for partial information and the materials and methods for full details).