Aquatic Elusimicrobia are metabolically diverse compared to gut microbiome Elusimicrobia and some have novel nitrogenase-like gene clusters

Raphaël Méheust[1,2], Cindy J. Castelle[1,2], Paula B. Matheus Carnevali[1,2], Ibrahim F. Farag[3], Christine He[1], Lin-Xing Chen[1,2], Yuki Amano[4], Laura A. Hug[5], and Jillian F. Banfield[1,2].

[1]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, USA
[2]Innovative Genomics Institute, Berkeley, CA, USA
[3]School of Marine Science and Policy, University of Delaware, Lewes DE 19968
[4]Nuclear Fuel Cycle Engineering Laboratories, Japan Atomic Energy Agency, Tokai-mura, Ibaraki, Japan
[5]Department of Biology, University of Waterloo, ON, Canada

Abstract

Currently described members of Elusimicrobia, a relatively recently defined phylum, are animal-associated and rely on fermentation. However, free-living Elusimicrobia have been detected in sediments, soils and groundwater, raising questions regarding their metabolic capacities and evolutionary relationship to animal-associated species. Here, we reconstructed 94 draft-quality, non-redundant genomes from diverse animal-associated and natural environments. Genomes group into 12 clades, 10 of which previously lacked reference genomes. Groundwater-associated Elusimicrobia are predicted to be capable of heterotrophic or autotrophic lifestyles, reliant on oxygen or nitrate/nitrite-dependent respiration of fatty acids, or a variety of organic compounds and Rnf-dependent acetogenesis with hydrogen and carbon dioxide as the substrates. Genomes from two clades of groundwater-associated Elusimicrobia often encode a new homologous group of nitrogenase-like proteins that co-occur with an extensive suite of radical SAM-based proteins. We identified similar genomic loci in genomes of bacteria from the Gracilibacteria and Myxococcus phyla and predict that the gene clusters reduce a tetrapyrrole, possibly to form a novel cofactor. The animal-associated Elusimicrobia clades nest phylogenetically within two groundwater-associated clades. Thus, we propose an evolutionary trajectory in which some Elusimicrobia adapted to animal-associated lifestyles from groundwater-associated species via genome reduction.

Introduction

Elusimicrobia is an enigmatic bacterial phylum. The first representatives were termite gut-associated (M. Ohkuma and Kudo 1996), although one 16S rRNA gene sequence was identified in a contaminated aquifer (Dojka et al. 1998). Initially referred to as Termite Group 1 (TG1) (Hugenholtz, Goebel, and Pace 1998), the phylum was renamed as Elusimicrobia after isolation of a strain from a beetle

35 larvae gut (Geissinger et al. 2009). Currently available sequences form a monophyletic group distinct from
36 Planctomycetes, Verrucomicrobia, Chlamydia, Omnitrophica, Desantisbacteria (Probst et al. 2017),
37 Kiritimatiellaeota (Spring et al. 2016) and Lentisphaerae (Cho et al. 2004). Together, these phyla form the
38 PVC superphylum.

39 The current understanding of Elusimicrobia mostly relies on a single taxonomic class for which the
40 name "Endomicrobia" has been proposed (Stingl et al. 2005). Endomicrobia comprise abundant members
41 of the hindgut of termites, cockroaches and other insects, as well as in rumen where they occur as
42 endosymbionts (Moriya Ohkuma et al. 2007), or ectosymbionts (Izawa et al. 2017) of various flagellated
43 protists, or as free-living bacteria (Mikaelyan et al. 2017). Of the three Endomicrobia genomes that have
44 been described, all belong to the genus *Endomicrobium*. One is for a free-living *E. proavitum* (H. Zheng,
45 Dietrich, and Brune 2017) and two are *E. trichonymphae* Rs-D17 endosymbionts, genomovar Ri2008
46 (Hongoh et al. 2008) and genomovar Ti2015 (Izawa et al. 2016). The fourth Elusimicrobia genome
47 available is for *Elusimicrobium minutum* strain Pei191^{T} (D. P. R. Herlemann et al. 2009), which was
48 isolated from a beetle larva gut (Geissinger et al. 2009). *E. minutum* forms a distinct monophyletic lineage
49 of gut-adapted organisms for which the name Elusimicrobiaceae was proposed (Geissinger et al. 2009).
50 Cultivation and genome-based studies revealed that *E. proavitum* and *E. minutum* strain Pei191^{T} are strictly
51 anaerobic ultramicrobacteria capable of glucose fermentation (Daniel P. R. Herlemann, Geissinger, and
52 Brune 2007; Geissinger et al. 2009; H. Zheng et al. 2016a).

53 Despite prior focus on gut-associated Elusimicrobia, phylogenetic analysis of environmental
54 sequences revealed numerous novel and distinct lineages from a wide range of habitat types, including soils,
55 sediments and groundwater (Daniel P. R. Herlemann, Geissinger, and Brune 2007). Here, we analyzed 94
56 draft-quality and non-redundant genomes from diverse environments and identified 12 lineages, including
57 10 that previously lacked genomic representatives. We predict numerous traits that constrain the roles of
58 Elusimicrobia in biogeochemical cycles, identify a new nitrogenase-like homolog, and infer the
59 evolutionary processes involved in adaptation to animal-associated lifestyles.

60

61 **Results and Discussion**

62

63 <u>Selection and phylogenetic analysis of a non-redundant set of genomes</u>

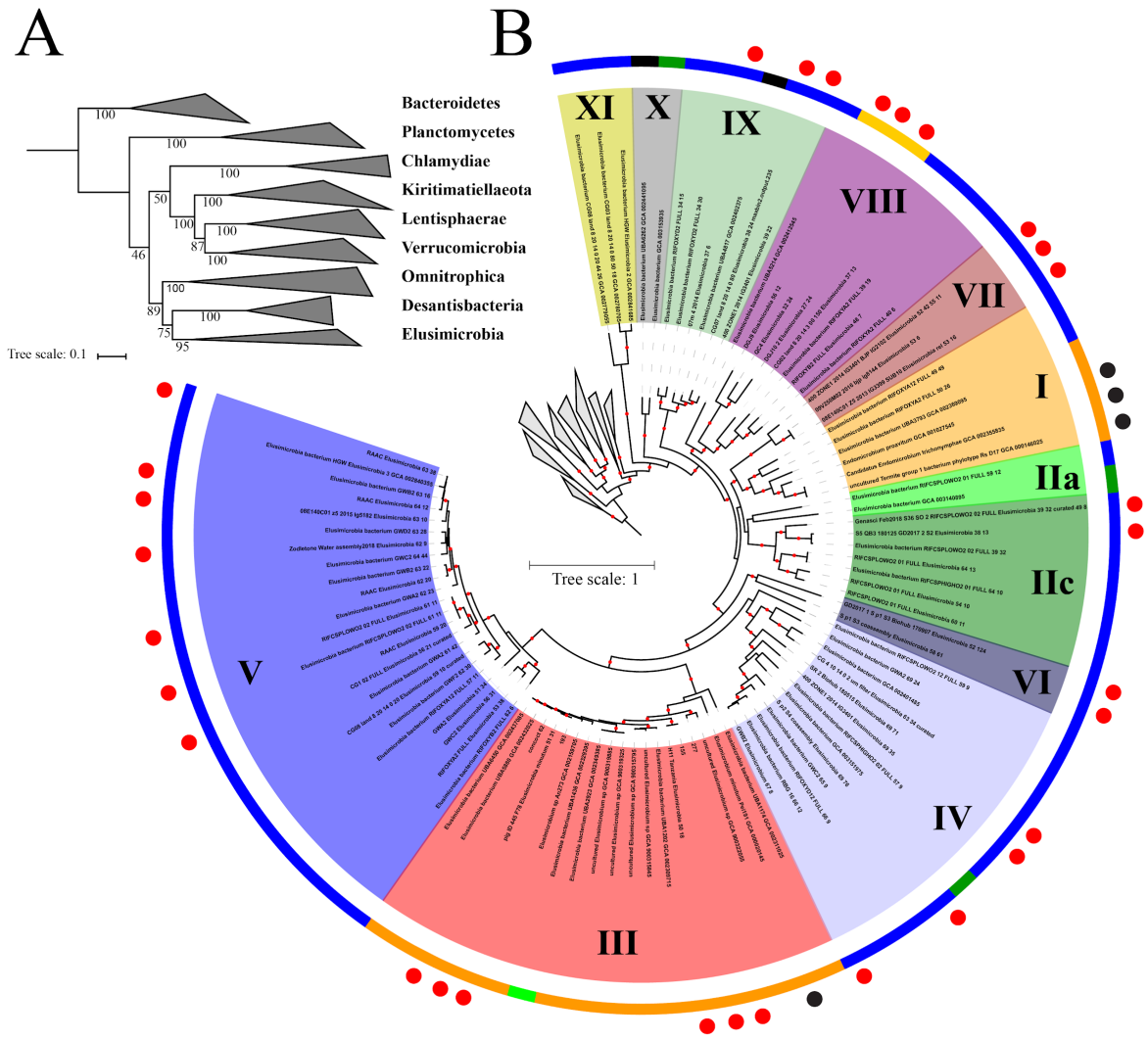
64

65 As of June 2018, 120 Elusimicrobia genomes were available at NCBI. We augmented these
66 genomes with 30 newly reconstructed genomes from metagenomic samples from Zodletone spring
67 (Oklahoma), groundwater or aquifer sediments (Serpens Ridge, California; Rifle, Colorado; Genasci dairy,
68 Modesto, California; Horonobe, Japan), hot spring sediments (Tibet, China), and the arsenic impacted

69 human gut (see Methods). Genomes with >70% completeness and <10% contamination were clustered at
70 ≥ 99% average nucleotide identity (ANI), and 113 genomes representative of the clusters
71 were selected based on completeness and contamination metrics. Among them, 19 genomes missed many
72 ribosomal proteins and thus were excluded, resulting in a final dataset of 94 non-redundant Elusimicrobia
73 genomes of good quality (median completeness of 92%, 69 genomes were >90% completeness, **Table S1**).
74 Phylogenetic analyses based on 16 ribosomal proteins (RPs) (see Methods) confirmed that all 94
75 genomes belong to the Elusimicrobia phylum, which forms a supported monophyletic group (**Figure 1A**,
76 bootstrap: 95) with Desantisbacteria and Omnitrophica as sibling phyla (**Figure 1A**, bootstrap: 89). The
77 bootstrap support (75) was insufficient to confirm that Desantisbacteria or Omnitrophica are the phyla most
78 closely related to Elusimicrobia.
79 Of the 94 Elusimicrobia genomes, 23 are from intestinal habitats and 71 come from other habitats,
80 mostly groundwater-associated (**Figure 1B**). Although some groundwater genomes are publicly available
81 (Anantharaman et al. 2016), none have been analyzed in detail. The Elusimicrobia genomes were further
82 classified into lineages, consistent with the previous classification based on 16S rRNA gene sequences
83 (Daniel P. R. Herlemann, Geissinger, and Brune 2007; Geissinger et al. 2009). The 16S rRNA genes were
84 binned for 25 of the 94 genomes, and belong to lineages I (Endomicrobia), IIc, III (Elusimicrobiaceae), IV,
85 V and VI. We were not able to classify four genomes based on their 16S rRNA genes due to a lack of
86 resolution in the basal branches (**Figure S1**). In order to improve the phylogenetic tree, we added two
87 additional 16S rRNA gene sequences from genomes within the 113 clusters that were not originally chosen
88 as representative genomes. These two sequences belong to lineages IIa and IIc. Thus, the 16S rRNA gene
89 sequences investigated here span 7 of the 8 previously defined lineages (**Figure S1**) (Daniel P. R.
90 Herlemann, Geissinger, and Brune 2007; Geissinger et al. 2009).
91 All of the gut-associated genomes clustered into the two known lineages, i.e., Endomicrobia
92 (lineage I) and Elusimicrobiaceae (lineage III). The Endomicrobia lineage formed a distinct clade (n=6,
93 bootstrap: 100) that includes the three earliest described Endomicrobium genomes and a new genome from
94 the rumen of sheep (Shi et al. 2014). Elusimicrobiaceae, which includes genomes related to *Elusimicrobium*
95 *minutum,* contains genomes from animal habitats, with the exception of one genome from palm oil mill
96 effluent. Of note, two newly reported Elusimicrobiaceae genomes in this group are from the gut
97 microbiomes of humans living in Bangladesh (Devoto et al. 2019) and in Tanzania (Rampelli et al. 2015).
98 This is consistent with the recent finding that Elusimicrobia genomes are present in the gut microbiomes of
99 non-westernized populations (Pasolli et al. 2019). As previously suggested (Geissinger et al. 2009), the
100 lineage III clade (n = 19) is nested within three clades comprising bacteria from groundwater environments
101 (lineages IV, V and VI, n = 38). Based on extrapolated genome sizes, Elusimicrobiaceae genomes were
102 significantly smaller than those of sibling lineages IV, V and IV (**Figure S2**). The overall placement and

short branch lengths within the Elusimicrobiaceae raise the possibility that the Elusimicrobiaceae adapted and diversified recently from ancestral groundwater genomes of lineages IV, V and VI.

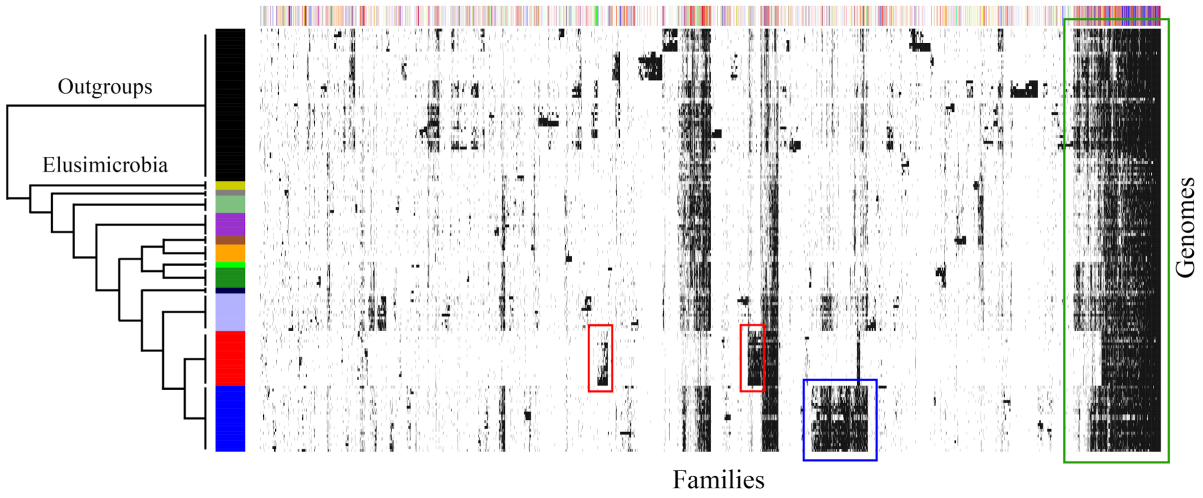
Eleven genomes clustered into lineages IIa (n=2, bootstrap: 41) and IIc (n=9, bootstrap: 100). The lack of bootstrap support does not allow us to confirm the relationship between the two lineages. None of the 94 genomes were classified as lineage IIb, due to either to a true absence of genomes sequences or the presence of lineage IIb genomes in the dataset that lack 16S rRNA gene sequences. Finally, the 22 remaining genomes defined 5 new clades (VII, VIII, IX, X and XI), of which four are basal to previously reported Elusimicrobia lineages. Sixteen of these 22 genomes are from groundwater, and the other 6 genomes found in sediment, oil sand and peat.



115 **Figure 1. Phylogenetic placement of newly reconstructed Elusimicrobia genomes. A.** Relationships
116 between the phyla of the PVC superphylum. **B.** Placement of the 94 genomes of Elusimicrobia. The
117 maximum likelihood tree was constructed based on a concatenated alignment of 16 ribosomal proteins
118 under an LG+gamma model of evolution. Nodes with bootstrap values ≥ 85 are
119 indicated by red circles. The outside color bar indicates the environment of origin (blue: groundwater,
120 orange: animal-associated, dark green: peat, black: oil sand, green: palm oil effluent, yellow: sediment from
121 hot springs in Tibet). Black circles represent the four genomes described in previous studies while red
122 circles represent the 30 new genomes added in this study. The remaining genomes comes from the NCBI
123 genome database. Scale bars indicate substitutions per site.

124 Protein clustering revealed lineage-specific genomic contents.

125 The availability of good quality genomes allowed us to compare their gene contents. We clustered
126 the protein sequences for all genomes to generate groups of homologous proteins (protein families) in a
127 two-step procedure (see Methods) (**Figure 2**). The objective was to compare the proteomes across the
128 genomes. Our approach is agnostic and unbiased by preconceptions about the importance of genes, and
129 allows us to cluster protein sequences with no homology within annotation databases. This resulted in 6,608
130 protein families that were present in at least three distinct genomes. We constructed an array of the genomes
131 versus the protein families, and sorted the families based on profiles of genome presence/absence.
132 Numerous protein families group together due to co-existence in multiple genomes (**Figure 2**).
133

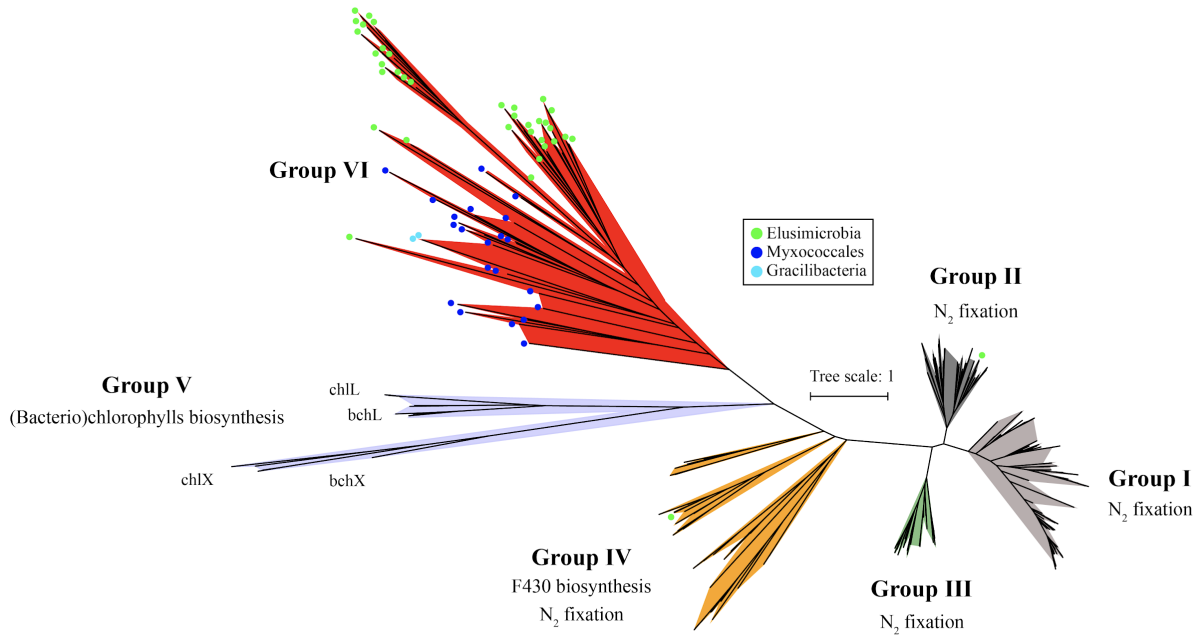


134
135 **Figure 2. The distribution of 6,608 protein families across the 110 genomes.** The distribution of 6,608
136 protein families (columns) in 110 genomes (rows). Genomes were sorted according to the phylogeny in
137 **Figure 1**. Families are clustered based on the presence (black) / absence (white) profiles (using Jaccard
138 distance and complete-linkage method). The colored top bar corresponds to the functional category of

139 families (Metabolism: red, Genetic Information Processing: blue, Cellular Processes: green, Environmental
140 Information Processing: yellow, Organismal systems: orange, Unclassified: grey, Unknown: white).
141 Specific clusters of interest are highlighted with colored boxes, discussed in the main text.
142
143 A block of 521 protein families is broadly conserved across the 94 Elusimicrobia genomes (far
144 right side of **Figure 2**). Unsurprisingly, the 521 families are involved in well-known functions, including
145 replication, transcription and translation, basic metabolism, and environmental interactions (green box in
146 **Figure 2**). Beside these widespread families, several blocks of families are fairly lineage specific. This is
147 particularly apparent for the gut-associated Elusimicrobiaceae lineage (lineage III) which lacks numerous
148 families abundant in other lineages but also contains 161 families abundant in Elusimicrobiaceae but rare
149 or absent in other lineages (red boxes in **Figure 2**). These families may be associated with adaptation to gut
150 environments (see below). Unlike lineage III, lineages IV and V encode extended sets of protein families
151 (blue boxes in **Figure 2**) consistent with their larger genome size (**Figure S2**). Other lineages also have
152 enriched group of families, although it is less apparent than for lineages III, IV and V (**Figure 2**). Overall,
153 the patterns of presence/absence of protein families are consistent with the lineages defined by the
154 ribosomal protein phylogeny (e.g., genomes from the same lineage tend to have a similar protein families
155 set) and may reflect different metabolic strategies.
156
157 <u>A novel nitrogenase-like gene may act as a cofactor</u>
158 A previous study reported the nitrogen fixing ability of *Endomicrobium proavitum* by an unusual
159 nitrogenase belonging to group IV (H. Zheng et al. 2016a) (**Figure 3**). Dinitrogenase reductase, encoded
160 by *nifH*, donates reducing equivalents to a complex encoded by *nifD* and *nifK*. The case reported in
161 *Endomicrobium proavitum* is the only evidence supporting the role of group IV *nifH* in N_2 fixation (H.
162 Zheng et al. 2016a). Canonically, N_2 fixing NifH proteins belong to groups I, II and III (Raymond et al.
163 2004) while group IV nitrogenase-like NifH proteins are implicated in the biosynthesis of cofactor F430,
164 the prosthetic group of methyl coenzyme M reductase, which catalyzes methane release in the final step of
165 methanogenesis (Moore et al. 2017; K. Zheng et al. 2016) NifH type, phylogenetically defined as group V,
166 is involved in chlorophyll biosynthesis (Burke, Hearst, and Sidow 1993). In this case, protochlorophyllide is
167 converted to chlorophyllide via the BchLNB complex in which BchL is the NifH homolog and BchN and BchB
168 are the NifD and NifK homologs, respectively.
169 We searched the Elusimicrobia genomes for *nifH*, *nifD* and *nifK* and identified homologues in 24
170 Elusimicrobia genomes from lineages I, IV and V. We also searched for accessory genes required for the
171 assembly of the M- and P- metallo clusters required by the nitrogenase complex. Two genomes from lineage
172 I, RIFOXYA2_FULL_50_26 and the previously reported *Endomicrobium proavitum*, contain both the

173 nitrogenase subunits and the accessory proteins. The NifH protein from RIFOXYA2_FULL_50_26 places
174 phylogenetically into NifH group III, and thus is plausibly associated with N_2 fixation (Raymond et al.
175 2004).

176 The other NifH proteins all belong to lineages IV and V of Elusimicrobia and place
177 phylogenetically outside of all previously described groups of NifH (**Figure 3**). Instead, the NifH sequences
178 from 22 Elusimicrobia genomes form a new group that is sibling to group V (**Figure 3**). Consistent with
179 the new placement, these 22 *nifH*-encoding genomes lack the accessory genes required for the assembly of
180 the M- and P- metallo clusters of nitrogenase. We designate these as group VI, and infer that these
181 nitrogenase-like proteins may have a distinct and perhaps previously undescribed biological role. The group
182 VI NifH proteins contain the GXGXXG consensus motif for the binding of MgATP and two Cys residues
183 (Cys^{97} and Cys^{132}) that bridge the two subunits through a $[Fe_4S_4]$ cluster (**Figure S3**). However, the
184 associated *nifD*- and *nifK-like* genes are highly divergent from true nitrogenase genes (**Figure 4**).
185 Importantly, they lack some conserved cysteine motifs that are involved in the attachment of the P-clusters
186 (Raymond et al. 2004) (**Figure S4**).



187

188 **Figure 3. Maximum likelihood phylogeny of the nitrogenase subunit NifH.** The tree was inferred using
189 an LG+I+G4 model of evolution. The tree was made using reference sequences from (Raymond et al. 2004).
190 The green circles highlight the sequences present in Elusimicrobia genomes. Scale bar indicates
191 substitutions per site. The sequences and tree are available with full bootstrap values in Fasta and Newick
192 format in the Supplementary Data.

193

194 Interestingly, nitrogenase-like complexes from group IV and V each modify a tetrapyrrole molecule
195 by reducing a carbon-carbon double bound (Hu and Ribbe 2015; Moore et al. 2017; Burke, Hearst, and
196 Sidow 1993). Biosynthesis of cofactor F430 involves the sirohydrochlorin precursor and biosynthesis of
197 chlorophyll involves a protoporphyrin precursor, both of which derive from uroporphyrinogen III (also a
198 building block for cobalamin). Thus, we speculate that the novel group VI Elusimicrobia NifH-like proteins
199 might also reduce a tetrapyrrole molecule. We searched the Elusimicrobia genomes for evidence of
200 uroporphyrinogen III synthesis by looking for genes encoding porphobilinogen synthase,
201 hydroxymethylbilane synthase and uroporphyrinogen synthase. These three enzymes are part of the three-
202 enzymatic-step core pathway from 5-aminolevulinate to uroporphyrinogen III (Dailey et al. 2017). We
203 found indications of the capacity for tetrapyrroles production in 16 genomes (**Table S3, Figure 4**). In a
204 subset of genomes with the novel group IV *nifH* genes (five genomes of lineage IV, **Table S3**), we identified
205 the capacity to produce uroporphyrinogen III (for example, see Elusimicrobia bacterium GWA2_69_24).
206 The absence of precursor biosynthesis pathways in other genomes of Elusimicrobia predicted to have the
207 capacity to make nitrogenase clusters does not rule out an analogous function, as many bacteria scavenge
208 such molecules (e.g., cobalamin; (Rosnow et al. 2018) or haem (Naoe et al. 2016)).

209 We examined the genomic neighborhoods for insights into the function of the novel group VI
210 nitrogenase-like proteins in Elusimicrobia from lineages IV and V. The lineage IV genomes encode several
211 copies of *nifK*, *nifD* and *nifH* whereas most genomes of lineage V genomes have only one copy of each
212 subunit (**Figure S5**). Strikingly, many adjacent genes encode radical S-adenosylmethionine (SAM)
213 proteins. Radical SAM proteins have many functions, including catalysis of methylation, isomerization,
214 sulfur insertion, ring formation, anaerobic oxidation, and protein radical formation, and also in the
215 biosynthesis of cofactors, DNA and antibiotics (Sofia et al. 2001; Mehta et al. 2015). The copy number of
216 radical SAM genes varies greatly across the genomes, from no radical SAM genes to 13 copies in close
217 proximity to the nitrogenase-like gene cluster in the genome of GWC2_Elusimicrobia_56_31. Several
218 radical SAM genes are fused with B12-binding domains (SR-2 Biohub_180515 Elusimicrobia_69_71,
219 GWC2_Elusimicrobia_56_31 and Elusimicrobia bacterium GWF2_62_30) and/or HemN_C domain
220 (Elusimicrobia bacterium GWA2 _69_24) (**Figure S5**). The B12-binding domain is involved in binding
221 cobalamin while the HemN_C domain has been suggested to bind the substrate of coproporphyrinogen III
222 oxidase (Layer et al. 2003). Both substrates have tetrapyrrole structures, which is consistent with the
223 phylogenetic position of the NifH near the groups IV and V NifH clades (**Figure 3**). Other evidence
224 suggesting that the substrate of the NifH reductase is a tetrapyrrole includes the presence of two *cbiA* genes
225 in the genome 400_ZONE1_2014_IG3401_Elusimicrobia_69_35. The *cbiA* gene encodes an amidase that
226 is involved in amidation within the cobalamin pathway. Two genomes (Elusimicrobia bacterium GWC2 65

227 9 and Elusimicrobia bacterium RIFOXYA12 FULL 57 11) also encode a corrin/porphyrin methylase near
228 the nitrogenase subunits.

229 Finally, we constructed a Hidden Markov Model using the new group VI Elusimicrobia NifH
230 sequences and searched for homologous sequences in genomes from other phyla. Interestingly, we found
231 homologs of the group VI NifH in two genomes of Gracilibacteria, a phylum within the bacterial Candidate
232 Phyla Radiation (Brown et al. 2015), and in fourteen myxococcales genomes from marine and freshwater
233 environments (**Figure 3, Table S2**). Together, these observations suggest that the enzymes likely do not
234 perform nitrogen fixation, but have an alternative function that may be related to the biosynthesis of a
235 tetrapyrrole cofactor similar to chlorophyll or F430 cofactors.

236

237 Overall metabolic potential in Elusimicrobia genomes

238

239 Groundwater-associated Elusimicrobia display substantial metabolic versatility, with the genomic
240 potential for both autotrophic and heterotrophic-based lifestyles (**Figure 4**). Most of the groundwater-
241 associated genomes may have the capacity for sugar fermentation to acetate, butyrate and/or ethanol, which
242 generates ATP via substrate phosphorylation (**Table S3**). We also identified a large variety of membrane-
243 bound, cytoplasmic [NiFe] and [FeFe] hydrogenases (**Table S3**), some of which may be electron
244 bifurcating. Other electron bifurcating complexes, such as the transhydrogenase (NfnAB) or the Bcd/EtfAB
245 complex are also present in these genomes (**Table S3**), suggesting the ability to minimize free energy waste
246 and to optimize energy conservation (Peters et al. 2016). This large repertoire of metabolic capacities differs
247 from their animal habitat-residing counterparts (lineages I and III), which rely only on fermentation
248 (Hongoh et al. 2008; D. P. R. Herlemann et al. 2009; H. Zheng, Dietrich, and Brune 2017; H. Zheng et al.
249 2016b) (**Figure 4**).

250

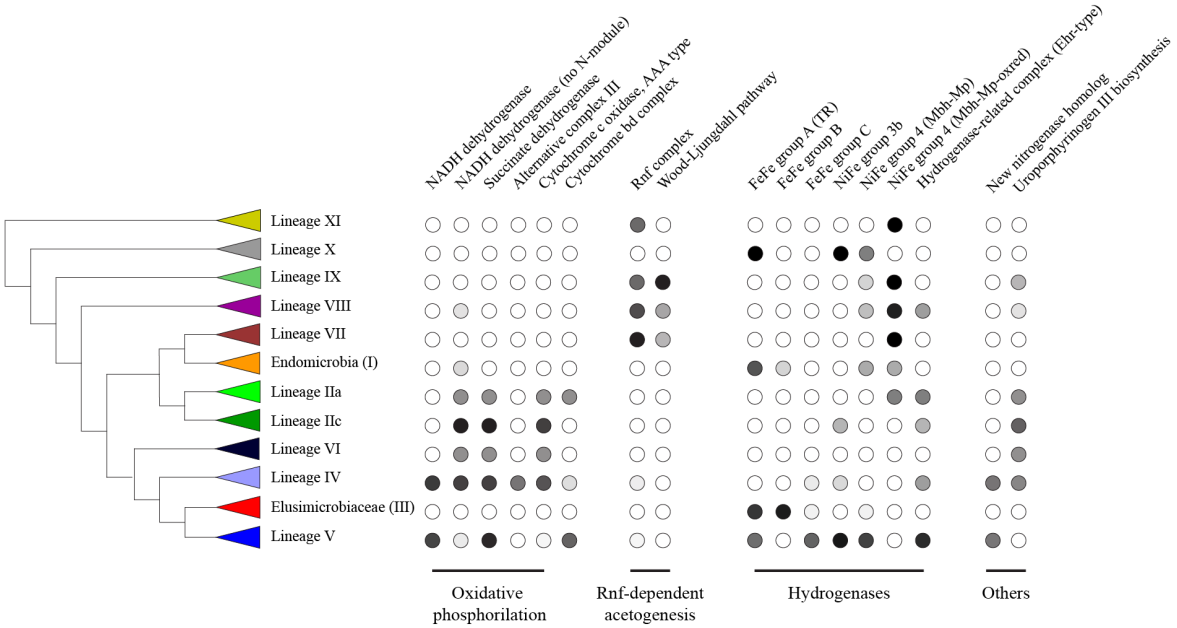


Figure 4. Schematic representation of the relationship of the Elusimicrobia lineages and the distribution of selected metabolic features across the different members of the phylum (see Table S3 for more details). Grey scale of the filled circles represents the prevalence of a function within a given lineage, with black = 100% of genomes and white = 0% of genomes.

Diverse respiratory strategies in groundwater-associated Elusimicrobia

Oxidative phosphorylation and the TCA cycle are commonly encoded in genomes from lineages IIa, IIc and IV, V and VI (**Figure 5A**). However, lineage V is different in terms of the completeness of the electron transport chain (ETC). Lineage V genomes encode a canonical NADH dehydrogenase (complex I with 14 subunits) and succinate dehydrogenase (complex II) that link electron transport to oxygen as a terminal electron acceptor via the high-affinity oxygen cytochrome *bd* oxidase (complex IV). Surprisingly, these genomes consistently lack a complex III (bc_1-complex, cytochrome c reductase).

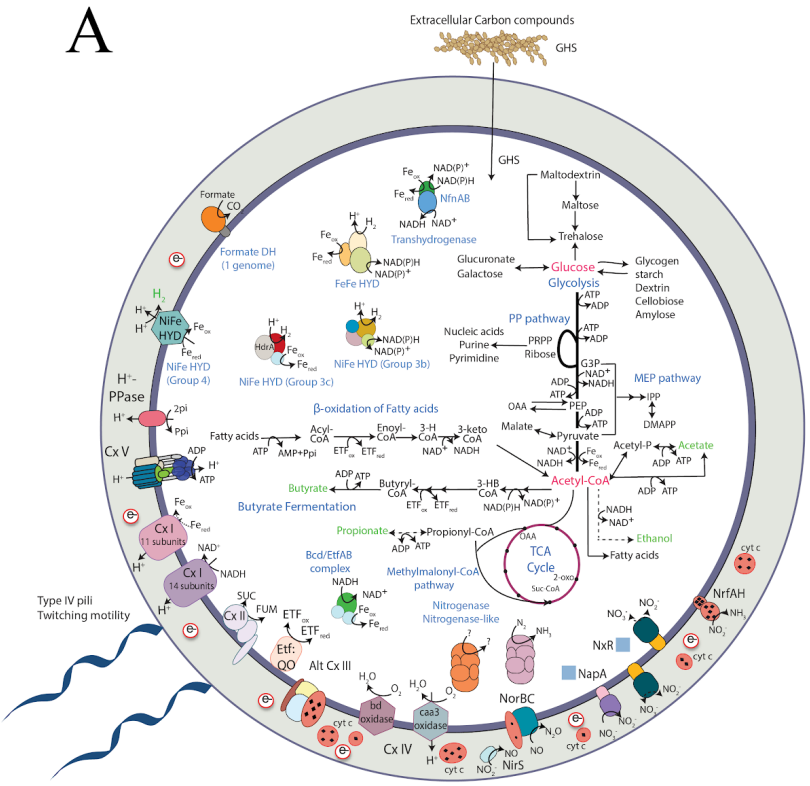
Group V Elusimicrobia typically have a set of different types of hydrogenases, but hydrogenases are sparsely distributed in other lineages. Some groundwater-associated lineage V genomes have trimeric group A [FeFe] hydrogenases directly downstream from a monomeric group C hydrogenase, related to those seen in Ignavibacterium album and Caldithrix abyssi (Greening et al. 2016) (Figure S6). In general, [FeFe] hydrogenases can either use or produce H_2 whereas the group C hydrogenases are co-transcribed with regulatory elements and are predicted to sense H_2 (Shaw, Hogsett, and Lynd 2009). In most group C hydrogenases this sensing occurs via a Per-ARNT-Sim (PAS) domain (Calusinska et al. 2010; Greening et

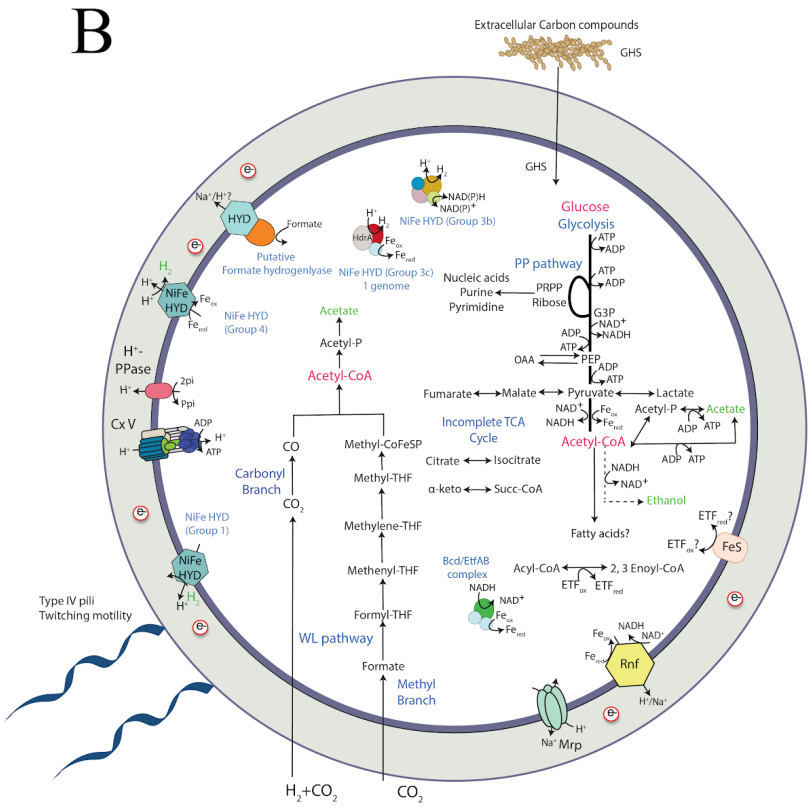
271 al. 2016). However, the Elusimicrobia hydrogenase seems to be fused to a histidine kinase HATPase
272 domain, probably fulfilling this sensory function. Additionally, the genomic region in the vicinity of these
273 hydrogenases includes several regulatory genes encoding proteins such as Rex, which is known to sense
274 NADH to NAD^+ ratios for transcriptional regulation (Y. Zheng et al. 2014).

275 Groundwater-associated genomes of lineage V also encode genes for different types of [NiFe]
276 hydrogenases. Seven genomes encode cytoplasmic group 3b (NADP-reducing) [NiFe] hydrogenases that
277 are likely bidirectional (also known as sulfhydrogenase). Four other genomes also have cytoplasmic group
278 3c (methyl viologen-reducing) [NiFe] hydrogenases (seen in one other genome in this clade), probably
279 involved in H_2 utilization (**Figure S7**). Additionally, most genomes in lineage V encode a membrane-
280 bound group 4 [NiFe] hydrogenases of the Mbh-Mrp type likely involved in H_2 production (**Figure S8**).
281 Membrane-bound hydrogenases are known to oxidize reduced ferredoxin, and the presence of antiporter-
282 like subunits suggests that they may be involved in ion translocation across the membrane and the
283 generation of a membrane potential (Schut et al. 2013). These genomes also encode hydrogenase-related
284 complexes (Ehr), the role of which is still unknown, although it has been suggested that they may interact
285 with oxidoreductases and quinones and may be involved in ion translocation across the membrane
286 (Marreiros et al. 2013; Carnevali et al. 2019).

287 Groundwater and peat-associated lineages IIa, IIc, IV, and VI have somewhat distinct respiratory
288 capacities compared to the groundwater-associated lineage V Elusimicrobia. All have complex I lacking
289 the diaphorase N-module (*nuoEFG* genes), which is hypothesized to use reduced ferredoxin instead of
290 NADH (Battchikova, Eisenhut, and Aro 2011). In addition to complex I without the N-module, lineage IV
291 also have complex I with an N-module. All genomes from these lineages have complex II (succinate
292 dehydrogenase) and complex IV (cytochrome c oxidase, type A; **Figure S9**). This lineage IV is also
293 different in that it encodes an alternative complex III and a few genomes contain a canonical complex III
294 (cytochrome c reductase). Unlike lineage V, which has signature hydrogenases, some of these genomes
295 have a set of hydrogenases and others do not (**Table S3**).

296 The genomes of some members of the groundwater and peat-associated lineages IIa, IIc, IV, V and
297 VI have nitrogen metabolic capacities that are rare in other lineages. Specifically, they encode a
298 nitrite/nitrate oxidoreductase (NXR, **Figure S10**) or nitrate reductase (**Figure S10**). We also identified other
299 enzymes involved in the nitrogen cycle, such as quinol-dependent nitric-oxide reductase (qNOR, **Figure**
300 **S9**), cytochrome-c nitrite reductase (NfrAH), copper and/or ferredoxin nitrite reductase(s) (NirK and NirS)
301 (**Table S3**). In summary, across the phylum, Elusimicrobia may use oxygen or nitrogen compounds as
302 terminal electron acceptors for energy conservation.


A
Extracellular Carbon compounds
GHS
Formate
CO
Formate DH
(1 genome)
NfnAB
Transhydrogenase
FeFe HYD
Maltodextrin
Maltose
Trehalose
Glucuronate
Galactose
Glucose
Glycolysis
Glycogen
starch
Dextrin
Cellobiose
Amylose
PP pathway
Nucleic acids
Purine
Pyrimidine
PRPP
Ribose
NiFe HYD
(Group 4)
NiFe HYD (Group 3c)
NiFe HYD (Group 3b)
MEP pathway
IPP
DMAPP
β-oxidation of Fatty acids
Fatty acids
Acyl-CoA
Enoyl-CoA
3-H CoA
3-keto CoA
Malate
Pyruvate
Acetyl-P
Acetate
Acetyl-CoA
Butyrate
Butyrate Fermentation
Butyryl-CoA
3-HB CoA
Propionate
Propionyl-CoA
Ethanol
Fatty acids
TCA
Cycle
Bcd/EtfAB
complex
Methylmalonyl-CoA
pathway
Nitrogenase
Nitrogenase-like
H+-
PPase
Cx V
Cx I
11 subunits
Cx I
14 subunits
Cx II
Etf:
QO
Alt Cx III
bd
oxidase
caa3
oxidase
Cx IV
NorBC
NirS
NrfAH
NxR
NapA
cyt c
Type IV pili
Twitching motility


B
Extracellular Carbon compounds
GHS
HYD
Formate
Putative
Formate hydrogenlyase
NiFe HYD (Group 3b)
NiFe HYD (Group 3c)
1 genome
NiFe HYD
(Group 4)
Glucose
Glycolysis
PP pathway
Nucleic acids
Purine
Pyrimidine
PRPP
Ribose
Acetate
Acetyl-P
Acetyl-CoA
H+-
PPase
Cx V
Fumarate
Malate
Pyruvate
Lactate
Incomplete TCA
Cycle
Citrate
Isocitrate
α-keto
Succ-CoA
Ethanol
Fatty acids?
FeS
CO
Carbonyl
Branch
Methyl-CoFeSP
Methyl-THF
Methylene-THF
Methenyl-THF
Formyl-THF
WL pathway
Formate
Methyl
Branch
NiFe HYD
(Group 1)
Bcd/EtfAB
complex
Acyl-CoA
2, 3 Enoyl-CoA
Rnf
Mrp
Type IV pili
Twitching motility

304 **Figure 5. Overview of the metabolic capabilities of non-gut-associated Elusimicrobia. A. Lineage IV**
305 **and V, (B). Lineage IX.** Abbreviations not defined in the text: GHS, glycoside hydrolases; TCA,
306 tricarboxylic acid cycle; 2-oxo, 2-oxoglutarate; Suc-CoA, succinyl-CoA; DH, dehydrogenase; HYD,
307 hydrogenase; PPi, pyrophosphate; Pi, inorganic phosphate; PRPP, Phosphoribosyl pyrophosphate; G3P,
308 glyceraldehyde-3-phosphate; PEP, phosphoenolpyruvate; OAA, oxaloacetate; MEP pathway, 2-*C*-methyl-
309 D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway; IPP, isopentenyl pyrophosphate;
310 DMAPP, dimethylallylpyrophosphate; FeS, FeS oxidoreductase; ETF, electron transfer flavoprotein;
311 ETF:QO, Electron-transferring flavoprotein dehydrogenase/ubiquinone oxidoreductase; BCD/ETFAB,
312 butyryl-CoA dehydrogenase/electron-transfer flavoprotein complex; 3HB-CoA, 3-hydroxybutyryl-CoA
313 dehydrogenase; 3-H-CoA, 3-hydroxyacyl-CoA; 3-keto-CoA, 3-ketoacyl-CoA; H+ -PPase, proton-
314 translocating pyrophosphatase; Cx V, ATP synthase; Cx I, NADH dehydrogenase; Cx II, succinate
315 dehydrogenase/fumarate reductase; Alt Cx III, alternative complex III; Cx IV, cytochrome c oxidase or
316 cytochrome bd oxidase; NorBC, nitric-oxide reductase; NirS, cytochrome cd1-type nitrite reductase; cyt c,
317 c-type cytochrome; NapA, periplasmic nitrate reductase; NxR, nitrite/nitrate oxidoreductase; NrfAH, nitrite
318 reductase; Mrp, multi-subunit Na^+/H^+ antiporter; WL pathway, Wood–Ljungdahl pathway; THF,
319 tetrahydrofolate; Rnf, Ferredoxin:NAD+-Oxidoreductase.
320
321 **RNF-dependent acetogenesis in groundwater-associated Elusimicrobia**
322 Other groundwater-associated Elusimicrobia in lineages VII, VIII, IX, X and XI lack the capacity
323 to reduce oxygen or nitrate. Instead, lineages VII, VIII and IX encode the Wood-Ljungdahl pathway (WLP)
324 for the reduction of carbon dioxide to acetyl-coenzyme A (CoA) and energy conservation (**Figure 5B**). The
325 coupling of the WLP with cytochromes and quinones is often used to generate a proton potential across the
326 membrane for the generation of ATP via an ATP synthase in acetogens (Schuchmann and Müller 2016).
327 However, cytochromes and quinones are missing in these lineages. Instead, these genomes encode the Rnf
328 complex, a sodium-motive ferredoxin:NAD oxidoreductase. The Rnf complex generates a sodium ion
329 potential across the cell membrane in *Acetobacterium woodii* (Poehlein et al. 2012). These observations
330 suggest that lineages VII, VIII and IX are capable of autotrophic growth with molecular hydrogen and
331 carbon dioxide as substrates.
332 Members of lineages VII, VIII, IX, XI, and a few genomes in lineage I have membrane-bound
333 group 4 [NiFe] hydrogenases of the Mbh-Mrp-oxidoreductase type (**Figure S8**). The function of this
334 oxidoreductase is unknown, though it has been seen before in other organisms such as the Thermococcales
335 (Schut et al. 2013) and Saganbacteria (Carnevali et al. 2019). Both RNF complexes and group 4 NiFe
336 membrane-bound hydrogenases could extrude ions across the membrane and contribute to the creation of
337 a membrane potential.

338

339 **Gut associated Elusimicrobia**

340 The genomes and general metabolic characteristics of gut-related Elusimicrobia (lineages I and III)
341 have been well described previously (D. P. R. Herlemann et al. 2009; H. Zheng, Dietrich, and Brune 2017;
342 Hongoh et al. 2008; Izawa et al. 2016). However, here we extend prior work by detailing the hydrogenases
343 of these bacteria to enable comparison to groundwater-associated Elusimicrobia (**Figure 4**). In brief, prior
344 work indicates that gut-associated Elusimicrobia rely on fermentation (Daniel P. R. Herlemann, Geissinger,
345 and Brune 2007; Geissinger et al. 2009; H. Zheng et al. 2016b; H. Zheng, Dietrich, and Brune 2017; Hongoh
346 et al. 2008). Most gut-related Elusimicrobia have [FeFe] hydrogenases in the 'ancestral' group B, which
347 are typically found in anaerobic bacteria (*e.g.,* Clostridia, Actinobacteria, and Bacteroidetes) that inhabit
348 gastrointestinal tracts and anoxic soils and sediments (Calusinska et al. 2010; Greening et al. 2016) (**Figure**
349 **S6**). The majority of [FeFe] hydrogenases are part of group A, and are involved in fermentative metabolism.
350 Group B [FeFe] hydrogenases remain largely uncharacterized. It has been suggested that they could couple
351 ferredoxin oxidation to H_2 evolution, because fermentative organisms rely on ferredoxin as a key redox
352 equivalent (Greening et al. 2016). In general, group B enzymes are monomeric. Many of the gut associated
353 Elusimicrobia genomes also harbor a trimeric group A [FeFe] hydrogenase. Group A [FeFe] hydrogenases
354 contain NADH-binding domains in the beta subunit and are known to be involved in electron bifurcation
355 or the reverse reaction, electron confurcation (**Figure S6**). In the case of electron bifurcation, [FeFe]
356 hydrogenases oxidize H_2 by coupling the exergonic reduction of NAD^+ with the endergonic reduction of
357 oxidized ferredoxin. In electron confurcation, these [FeFe] hydrogenases produce H_2 by using the exergonic
358 oxidation of ferredoxin with protons (H^+) to drive the endergonic oxidation of NADH with H^+
359 (Schuchmann, Chowdhury, and Müller 2018; Müller, Chowdhury, and Basen 2018; Buckel and Thauer
360 2018). Thus, we conclude that many gut-associated Elusimicrobia can potentially both produce and use H_2.

361

362 **Conclusions**

363

364 The Elusimicrobia phylum was previously represented by just four genomes from gut-associated
365 habitats. We expanded the genomic sampling of this phylum with new sequences, mainly from groundwater
366 environments. We predicted novel nifH-like proteins that co-occur with an extensive suite of radical SAM-
367 based proteins in groundwater-associated Elusimicrobia from lineages IV and V. Based on gene context
368 and protein phylogeny, we predict that the nifH-like proteins reduce a tetrapyrrole. Given the functions of
369 related clades of genes, the nifH-like protein may synthesize a novel cofactor, but further investigations are
370 needed to determine the product and its function.

371 Our phylogenetic results suggest that gut-associated Elusimicrobia of lineages I and III evolved
372 from two distinct groups of groundwater-associated ancestors, lineages IV, V and group VII, respectively.
373 Based on extant members of groundwater groups, the ancestors of gut-associated Elusimicrobia that likely
374 had larger repertoires of metabolic capacities. The gut-associated lineages likely underwent strong genome
375 reduction following the transition from nutrient-poor (groundwater) to nutrient-rich (gut) environments.
376 Genome reduction following formation of a long-term association with another organism is generally well
377 documented for many endosymbionts (for example of insects; (McCutcheon and Moran 2011) or in lineage
378 I; (H. Zheng, Dietrich, and Brune 2017)). Here we report the genome reduction and the habitat shifts of
379 free-living organisms. This is particularly interesting as Elusimicrobia of lineage III have been recently
380 found in human gut (Pasolli et al. 2019).
381

382 **Methods**

383

384 *Elusimicrobia genomes collection*

385

386 Six genomes were obtained from groundwater samples from Genasci Dairy Farm, located in Modesto, CA
387 and from Serpens Ridge, a private property located in Middletown, CA. Over 400 L of groundwater were
388 filtered from monitoring well 5 on Genasci Dairy Farm, located in Modesto, CA, over a period ranging
389 from March 2017 to June 2018. Over 400 L of groundwater were filtered from Serpens Ridge, a private
390 property located in Middletown, CA in November 2017. DNA was extracted from all filters using Qiagen
391 DNeasy PowerMax Soil kits and ~10 Gbp of 150 bp, paired end Illumina reads were obtained for each
392 filter. Scaffolds were binned on the basis of GC content, coverage, presence of ribosomal proteins,
393 presence/copies of single copy genes, tetranucleotide frequency, and patterns of coverage across samples.
394 Bins were obtained using manual binning on ggKbase (Raveh-Sadka et al. 2015), Maxbin2 (Wu, Simmons,
395 and Singer 2016), CONCOCT (Alneberg et al. 2014), Abawaca1, and Abawaca2
396 (https://github.com/CK7/abawaca), with DAS Tool (Sieber et al., n.d.) used to choose the best set of bins
397 from all programs. All bins were manually checked to remove incorrectly assigned scaffolds using
398 ggKbase.

399

400 Three genomes were obtained from sediment samples of Tibet hot springs in 2016. The samples were
401 collected as previously described (Song et al. 2013), followed by genomic DNA extraction using the
402 FastDNA SPIN kit (MP Biomedicals, Irvine, CA, United States) according to the manufacturer's
403 instructions. The metagenomic sequencing was performed by an Illumina HiSeq2500 platform with PE150
404 bp kits. Quality filter, *de novo* assembly, genome binning and modification were detailed in (Chen et al.
405 2019).

406

407 One genome was obtained from sediment samples from Zodletone spring in California. Sediment samples
408 were collected from of Zodletone spring, an anaerobic sulfide and sulfur-rich spring in Western Oklahoma.
409 Detailed site descriptions and metagenomic sequencing were previously reported in (Youssef et al. 2019).
410 Metagenomic datasets can be located at NCBI GenBank SRA under project PRJNA364990. Metagenomic
411 datasets were assembled using MegaHit (Li et al. 2015) applying default parameters. Scaffolds were binned
412 using ggKbase (Raveh-Sadka et al. 2015), Maxbin2 (Wu, Simmons, and Singer 2016), and CONCOCT
413 (Alneberg et al. 2014). Genomic bins generated from all binning tools were de-replicated, aggregated and
414 best ones were selected using DAStool (Sieber et al., n.d.).

415

416 Seven genomes were obtained from groundwater samples collected from the Mizunami and the Horonobe
417 Underground Research Laboratories (URL) in Japan. For sampling refer to (Ino et al. 2018) and (Hernsdorf
418 et al. 2017), and for DNA processing and sequencing methods refer to (Hernsdorf et al. 2017). In brief,
419 Over 400 L of groundwaters were collected from boreholes at 200 and 400 m deep of Mizunami URL and
420 15 to 43 L of groundwaters were collected from boreholes at 140 and 250 m deep of Horonobe URL, by
421 filtration through 0.22 m GVWP filters. Genomic DNA was extracted using an Extrap Soil DNA Kit Plus
422 version 2 (Nippon Steel and Sumikin EcoTech Corporation), libraries prepared using TruSeq Nano DNA
423 Sample Prep Kit (Illumina), and 150 bp paired-end reads with a 550 bp insert size were sequenced by
424 Hokkaido System Science Co. using Illumina HiSeq2500 (Hernsdorf et al. 2017). All sample data were
425 assembled de novo using IDBA-UD (Peng et al. 2012) with the following parameters: -mink 40, -maxk
426 100, -step 20 and -pre_correction. Assembled scaffolds >1000 bp were binned by a combination of
427 phylogenetic profiles, read coverage and nucleotide content (that is, GC proportion and tetranucleotide
428 signatures) within the binning interface of ggKbase (Raveh-Sadka et al. 2015).
429
430 Thirteen additional genomes were also assembled from raw data of previous studies following the methods
431 described in (Devoto et al. 2019) (**Table S1**).
432
433 *Genome completeness assessment and de-replication.*
434 Genome completeness and contamination were estimated based on the presence of single-copy
435 genes (SCGs) as described in (Anantharaman et al. 2016). Genome completeness was estimated using 51
436 SCGs, following (Anantharaman et al. 2016). Genomes with completeness > 70% and contamination <
437 10% (based on duplicated copies of the SCGs) were considered as draft-quality genomes. Genomes were
438 de-replicated using dRep (version v2.0.5 with ANI > 99%) (Olm et al. 2017). The most complete genome
439 per cluster was used in downstream analyses.
440
441 *16S rRNA gene phylogeny*
442 To construct a comprehensive 16S ribosomal RNA gene tree, all gene sequences assigned to the
443 Elusimicrobia in the SILVA database (Quast et al. 2013) were exported. Sequences longer than 750 bp
444 were clustered at 97% identity using uclust, and the longest representative gene for each cluster included in
445 the phylogenetic tree. All 16S rRNA genes associated with a binned Elusimicrobia genome were added to
446 the SILVA reference set, along with an outgroup of Omnitrophica, Planctomycetes, Verrucomicrobia, and
447 Chlamydiae rRNA gene sequences from organisms with published genomes (37 outgroup sequences total).
448 The final dataset comprised 700 sequences, which were aligned together using the SILVA SINA alignment
449 algorithm. Common gaps and positions with less than 3% information were stripped from the alignment,

450 for a final alignment of 1,593 columns. A maximum likelihood phylogeny was inferred using RAxML
451 (Stamatakis 2014) version 8.2.4 as implemented on the CIPRES high performance computing cluster
452 (Miller, Pfeiffer, and Schwartz 2010), under the GTR model of evolution and the MRE-based bootstopping
453 criterion. A total of 408 bootstrap replicates were conducted, from which 100 were randomly sampled to
454 determine support values.
455
456 *Concatenated 16 ribosomal proteins phylogeny*
457 A maximum-likelihood tree was calculated based on the concatenation of 16 ribosomal proteins
458 (L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17, and S19). Homologous protein
459 sequences were aligned using Muscle (Edgar 2004), and alignments refined to remove regions of ambiguity
460 by removing columns with less than 3% information and manually removing abberrent N and C-terminal
461 extensions. The protein alignments were concatenated, with a final alignment of 147 genomes and 2,388
462 positions. The tree was inferred using RAxML (Stamatakis 2014) (version 8.2.10) (as implemented on the
463 CIPRES web server (Miller, Pfeiffer, and Schwartz 2010)), under the LG plus gamma model of evolution,
464 and with the number of bootstraps automatically determined via the MRE-based bootstopping criterion. A
465 total of 108 bootstrap replicates were conducted, from which 100 were randomly sampled to determine
466 support values.
467
468 *Protein clustering*
469 _____Protein clustering into families was achieved using a two-step procedure as previously described
470 in (Méheust et al. 2018). A first protein clustering was done using the fast and sensitive protein sequence
471 searching software MMseqs2 (version 9f493f538d28b1412a2d124614e9d6ee27a55f45) (Steinegger and
472 Söding 2017). An all-vs-all search was performed using e-value: 0.001, sensitivity: 7.5 and cover: 0.5. A
473 sequence similarity network was built based on the pairwise similarities and the greedy set cover algorithm
474 from MMseqs2 was performed to define protein subclusters. The resulting subclusters were defined as
475 subfamilies. In order to test for distant homology, we grouped subfamilies into protein families using an
476 HMM-HMM comparison procedure as follows: the proteins of each subfamily with at least two protein
477 members were aligned using the result2msa parameter of mmseqs2, and, from the multiple sequence
478 alignments, HMM profiles were built using the HHpred suite (version 3.0.3) (Soding 2005). The
479 subfamilies were then compared to each other using hhblits (Remmert et al. 2011) from the HHpred suite
480 (with parameters -v 0 -p 50 -z 4 -Z 32000 -B 0 -b 0). For subfamilies with probability scores of ≥ 95% and
481 coverage ≥ 0.50, a similarity score (probability × coverage) was used as weights of the input network in
482 the final clustering using the Markov Clustering algorithm (Enright, Van Dongen, and Ouzounis 2002),
483 with 2.0 as the inflation parameter. These clusters were defined as the protein families.

484

485 *Functional annotation*

486 Protein sequences were functionally annotated based on the accession of their best Hmmsearch
487 match (version 3.1) (E-value cut-off 0.001) (Eddy 1998) against an HMM database constructed based on
488 ortholog groups defined by the KEGG (Kanehisa et al. 2016) (downloaded on June 10, 2015). Domains
489 were predicted using the same Hmmsearch procedure against the Pfam database (version 31.0) (Punta et al.
490 2012). The domain architecture of each protein sequence was predicted using the DAMA software (version
491 1.0) (default parameters) (Bernardes et al. 2016). SIGNALP (version 4.1) (parameters: -f short -t gram-)
492 (Petersen et al. 2011) was used to predict the putative cellular localization of the proteins. Prediction of
493 transmembrane helices in proteins was performed using TMHMM (version 2.0) (default parameters)
494 (Krogh et al. 2001). The transporters were predicted using BLASTP (version 2.6.0) (Altschul et al. 1997)
495 against the TCDB database (downloaded on February 2019) (keeping the best hit, e-value cut-off 1e-20)
496 (Saier et al. 2016).

497

498 *Phylogenetic analyses of individual protein sequences.*

499 Each individual tree was built as follows. Sequences were aligned using MAFFT (version 7.390)
500 (--auto option) (Katoh and Standley 2016). Alignment was further trimmed using Trimal (version 1.4.22)
501 (--gappyout option) (Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009). Tree reconstruction was
502 performed using IQ-TREE (version 1.6.6) (as implemented on the CIPRES web server (Miller, Pfeiffer,
503 and Schwartz 2010)) (Nguyen et al. 2015), using ModelFinder (Kalyaanamoorthy et al. 2017) to select the
504 best model of evolution, and with 1000 ultrafast bootstrap (Hoang et al. 2018).

Acknowledgements

Support was provided by grants from the Lawrence Berkeley National Laboratory's Genomes-to-Watershed Scientific Focus Area. The U.S. Department of Energy (DOE), Office of Science, and Office of Biological and Environmental Research funded the work under contract DE-AC02-05CH11231 and the DOE carbon cycling program DOE-SC10010566, the Innovative Genomics Institute at Berkeley and the Chan Zuckerberg Biohub. The Ministry of Economy, Trade and Industry of Japan funded a part of the work as "The project for validating assessment methodology in geological disposal system". Teruki Iwatsuki, Kazuki Hayashida, Toshihiro Kato, Mitsuru Kubota, Kazuya Miyakawa, and Akihito Mochizuki assisted with groundwater sampling at Mizunami and Horonobe Underground Research Laboratories, Japan Atomic Energy Agency (JAEA).C.H. acknowledges the Camille and Henry Dreyfus Foundation Postdoctoral Program in Environmental Chemistry for a Fellowship.

Author contributions

J.F.B., L.A.H and R.M. conceived the study. C.J.C., P.M.C., L.A.H. and R.M. analysed the genomic data. R.M. and L.A.H. performed the phylogenetic analyses. I.F. performed the Cazy analysis. L-X.C., Y.A., C.H., I.F. provided the new genomes. R.M., J.F.B, C.J.C and P.M.C. wrote the manuscript with input from all authors. All documents were edited and approved by all authors.

Competing interests

J.F.B. is a founder of Metagenomi. The other authors declare no competing interests.

Materials and Correspondence

Correspondence and material requests should be addressed to jbanfield@berkeley.edu.

Data availability

The genomes of the herein analysed Elusimicrobia have been made publicly available on NCBI and on the ggkbase database (https://ggkbase.berkeley.edu/non_redundant_elusimicrobia_2018/organisms). Detailed annotations of the metabolic repertoire are provided in Supplementary Table S3 accompanying this paper. Raw data files (phylogenetic trees and fasta sequences) are made available via figshare under the following link: https://figshare.com/articles/Elusimicrobia_analysis/8939678.

References

Alneberg, Johannes, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z. Ijaz, Leo Lahti, Nicholas J. Loman, Anders F. Andersson, and Christopher Quince. 2014. "Binning Metagenomic Contigs by Coverage and Composition." *Nature Methods* 11 (11): 1144–46.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25 (17): 3389–3402.

Anantharaman, Karthik, Christopher T. Brown, Laura A. Hug, Itai Sharon, Cindy J. Castelle, Alexander J. Probst, Brian C. Thomas, et al. 2016. "Thousands of Microbial Genomes Shed Light on Interconnected Biogeochemical Processes in an Aquifer System." *Nature Communications* 7 (1): 13219.

Battchikova, Natalia, Marion Eisenhut, and Eva-Mari Aro. 2011. "Cyanobacterial NDH-1 Complexes: Novel Insights and Remaining Puzzles." *Biochimica et Biophysica Acta* 1807 (8): 935–44.

Bernardes, J. S., F. R. J. Vieira, G. Zaverucha, and A. Carbone. 2016. "A Multi-Objective Optimization Approach Accurately Resolves Protein Domain Architectures." *Bioinformatics* 32 (3): 345–53.

Brown, Christopher T., Laura A. Hug, Brian C. Thomas, Itai Sharon, Cindy J. Castelle, Andrea Singh, Michael J. Wilkins, Kelly C. Wrighton, Kenneth H. Williams, and Jillian F. Banfield. 2015. "Unusual Biology across a Group Comprising More than 15% of Domain Bacteria." *Nature* 523 (7559): 208–11.

Buckel, Wolfgang, and Rudolf K. Thauer. 2018. "Flavin-Based Electron Bifurcation, Ferredoxin, Flavodoxin, and Anaerobic Respiration With Protons (Ech) or NAD (Rnf) as Electron Acceptors: A Historical Review." *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2018.00401.

Burke, D. H., J. E. Hearst, and A. Sidow. 1993. "Early Evolution of Photosynthesis: Clues from Nitrogenase and Chlorophyll Iron Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 90 (15): 7134–38.

Calusinska, Magdalena, Thomas Happe, Bernard Joris, and Annick Wilmotte. 2010. "The Surprising Diversity of Clostridial Hydrogenases: A Comparative Genomic Perspective." *Microbiology* 156 (Pt 6): 1575–88.

Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25 (15): 1972–73.

Carnevali, Paula B. Matheus, Frederik Schulz, Cindy J. Castelle, Rose S. Kantor, Patrick M. Shih, Itai Sharon, Joanne M. Santini, et al. 2019. "Hydrogen-Based Metabolism as an Ancestral Trait in Lineages Sibling to the Cyanobacteria." *Nature Communications* 10 (1): 463.

Castelle, Cindy J., Laura A. Hug, Kelly C. Wrighton, Brian C. Thomas, Kenneth H. Williams, Dongying Wu, Susannah G. Tringe, Steven W. Singer, Jonathan A. Eisen, and Jillian F. Banfield. 2013. "Extraordinary Phylogenetic Diversity and Metabolic Versatility in Aquifer Sediment." *Nature Communications*. https://doi.org/10.1038/ncomms3120.

Chen, Lin-Xing, Basem Al-Shayeb, Raphaël Méheust, Wen-Jun Li, Jennifer A. Doudna, and Jillian F. Banfield. 2019. "Candidate Phyla Radiation Roizmanbacteria From Hot Springs Have Novel and Unexpectedly Abundant CRISPR-Cas Systems." *Frontiers in Microbiology* 10 (May): 928.

Cho, Jang-Cheon, Kevin L. Vergin, Robert M. Morris, and Stephen J. Giovannoni. 2004. "Lentisphaera Araneosa Gen. Nov., Sp. Nov, a Transparent Exopolymer Producing Marine Bacterium, and the Description of a Novel Bacterial Phylum, Lentisphaerae." *Environmental Microbiology* 6 (6): 611–21.

Dailey, Harry A., Tamara A. Dailey, Svetlana Gerdes, Dieter Jahn, Martina Jahn, Mark R. O'Brian, and Martin J. Warren. 2017. "Prokaryotic Heme Biosynthesis: Multiple Pathways to a Common Essential Product." *Microbiology and Molecular Biology Reviews: MMBR* 81 (1).

https://doi.org/10.1128/MMBR.00048-16.
Devoto, Audra E., Joanne M. Santini, Matthew R. Olm, Karthik Anantharaman, Patrick Munk, Jenny Tung, Elizabeth A. Archie, et al. 2019. "Megaphages Infect Prevotella and Variants Are Widespread in Gut Microbiomes." *Nature Microbiology* 4 (4): 693–700.
Dojka, M. A., P. Hugenholtz, S. K. Haack, and N. R. Pace. 1998. "Microbial Diversity in a Hydrocarbon- and Chlorinated-Solvent-Contaminated Aquifer Undergoing Intrinsic Bioremediation." *Applied and Environmental Microbiology* 64 (10): 3869–77.
Eddy, S. R. 1998. "Profile Hidden Markov Models." *Bioinformatics* 14 (9): 755–63.
Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5): 1792–97.
Enright, A. J., S. Van Dongen, and C. A. Ouzounis. 2002. "An Efficient Algorithm for Large-Scale Detection of Protein Families." *Nucleic Acids Research* 30 (7): 1575–84.
Geissinger, Oliver, Daniel P. R. Herlemann, Erhard Mörschel, Uwe G. Maier, and Andreas Brune. 2009. "The Ultramicrobacterium 'Elusimicrobium Minutum' Gen. Nov., Sp. Nov., the First Cultivated Representative of the Termite Group 1 Phylum." *Applied and Environmental Microbiology* 75 (9): 2831–40.
Greening, Chris, Ambarish Biswas, Carlo R. Carere, Colin J. Jackson, Matthew C. Taylor, Matthew B. Stott, Gregory M. Cook, and Sergio E. Morales. 2016. "Genomic and Metagenomic Surveys of Hydrogenase Distribution Indicate H2 Is a Widely Utilised Energy Source for Microbial Growth and Survival." *The ISME Journal* 10 (3): 761–77.
Herlemann, Daniel P. R., Oliver Geissinger, and Andreas Brune. 2007. "The Termite Group I Phylum Is Highly Diverse and Widespread in the Environment." *Applied and Environmental Microbiology* 73 (20): 6682–85.
Herlemann, D. P. R., O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz, and A. Brune. 2009. "Genomic Analysis of 'Elusimicrobium Minutum,' the First Cultivated Representative of the Phylum 'Elusimicrobia' (Formerly Termite Group 1)." *Applied and Environmental Microbiology*. https://doi.org/10.1128/aem.02698-08.
Hernsdorf, Alex W., Yuki Amano, Kazuya Miyakawa, Kotaro Ise, Yohey Suzuki, Karthik Anantharaman, Alexander Probst, David Burstein, Brian C. Thomas, and Jillian F. Banfield. 2017. "Potential for Microbial H2 and Metal Transformations Associated with Novel Bacteria and Archaea in Deep Terrestrial Subsurface Sediments." *The ISME Journal* 11 (8): 1915–29.
Hoang, Diep Thi, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, and Le Sy Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35 (2): 518–22.
Hongoh, Yuichi, Vineet K. Sharma, Tulika Prakash, Satoko Noda, Todd D. Taylor, Toshiaki Kudo, Yoshiyuki Sakaki, Atsushi Toyoda, Masahira Hattori, and Moriya Ohkuma. 2008. "Complete Genome of the Uncultured Termite Group 1 Bacteria in a Single Host Protist Cell." *Proceedings of the National Academy of Sciences of the United States of America* 105 (14): 5555–60.
Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. "Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity." *Journal of Bacteriology* 180 (18): 4765–74.
Hu, Yilin, and Markus W. Ribbe. 2015. "Nitrogenase and Homologs." *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry* 20 (2): 435–45.
Ino, Kohei, Alex W. Hernsdorf, Uta Konno, Mariko Kouduka, Katsunori Yanagawa, Shingo Kato, Michinari Sunamura, et al. 2018. "Ecological and Genomic Profiling of Anaerobic Methane-Oxidizing Archaea in a Deep Granitic Environment." *The ISME Journal* 12 (1): 31–47.
Izawa, Kazuki, Hirokazu Kuwahara, Kumiko Kihara, Masahiro Yuki, Nathan Lo, Takehiko Itoh, Moriya Ohkuma, and Yuichi Hongoh. 2016. "Comparison of Intracellular 'Ca. Endomicrobium Trichonymphae' Genomovars Illuminates the Requirement and Decay of Defense Systems against Foreign DNA." *Genome Biology and Evolution* 8 (10): 3099–3107.
Izawa, Kazuki, Hirokazu Kuwahara, Kaito Sugaya, Nathan Lo, Moriya Ohkuma, and Yuichi Hongoh. 2017. "Discovery of Ectosymbiotic Endomicrobium Lineages Associated with Protists in the Gut of

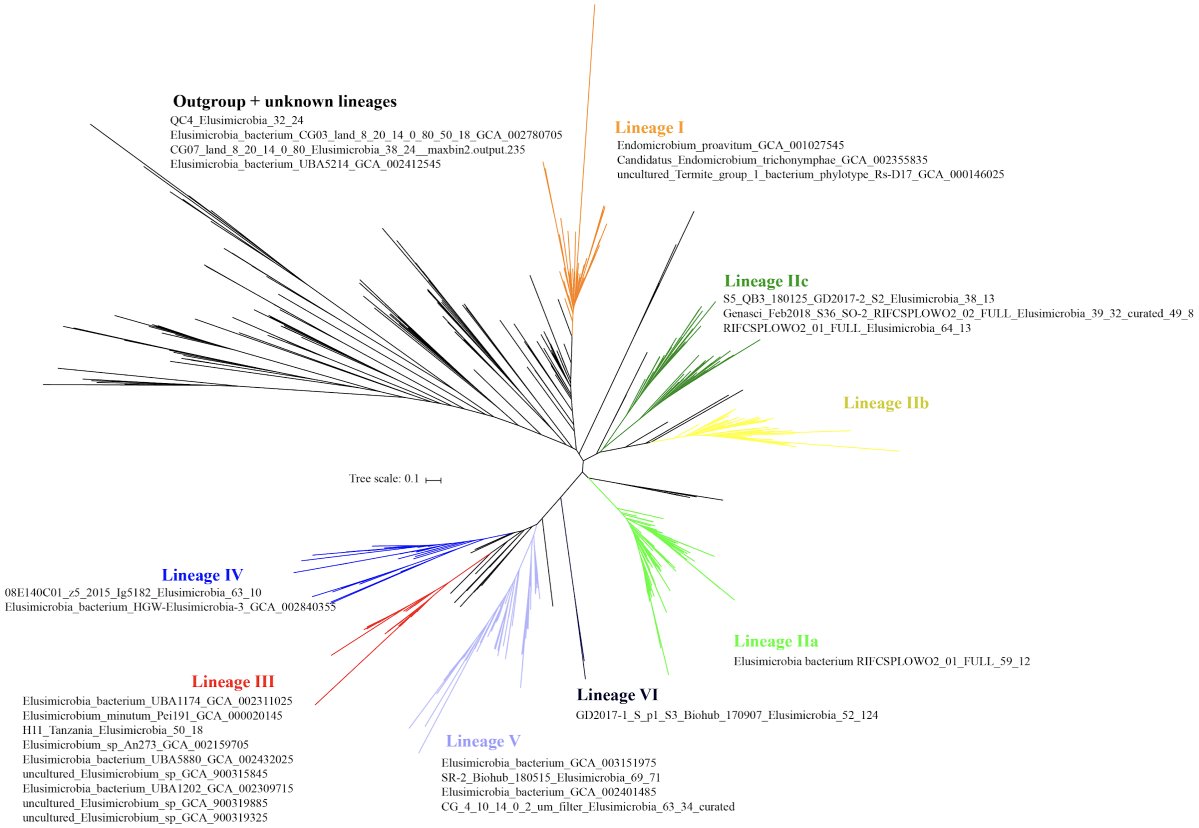
Stolotermitid Termites." *Environmental Microbiology Reports* 9 (4): 411–18.
Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K. F. Wong, Arndt von Haeseler, and Lars S. Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14 (6): 587–89.
Kanehisa, Minoru, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. 2016. "KEGG as a Reference Resource for Gene and Protein Annotation." *Nucleic Acids Research* 44 (D1): D457–62.
Katoh, Kazutaka, and Daron M. Standley. 2016. "A Simple Method to Control over-Alignment in the MAFFT Multiple Sequence Alignment Program." *Bioinformatics* 32 (13): 1933–42.
Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes." *Journal of Molecular Biology* 305 (3): 567–80.
Layer, Gunhild, Jürgen Moser, Dirk W. Heinz, Dieter Jahn, and Wolf-Dieter Schubert. 2003. "Crystal Structure of Coproporphyrinogen III Oxidase Reveals Cofactor Geometry of Radical SAM Enzymes." *The EMBO Journal* 22 (23): 6214–24.
Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. 2015. "MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph." *Bioinformatics* 31 (10): 1674–76.
Marreiros, Bruno C., Ana P. Batista, Afonso M. S. Duarte, and Manuela M. Pereira. 2013. "A Missing Link between Complex I and Group 4 Membrane-Bound [NiFe] Hydrogenases." *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1827 (2): 198–209.
McCutcheon, John P., and Nancy A. Moran. 2011. "Extreme Genome Reduction in Symbiotic Bacteria." *Nature Reviews. Microbiology* 10 (1): 13–26.
Méheust, Raphaël, David Burstein, Cindy J. Castelle, and Jillian F. Banfield. 2018. "Biological Capacities Clearly Define a Major Subdivision in Domain Bacteria." *bioRxiv*. https://doi.org/10.1101/335083.
Mehta, Angad P., Sameh H. Abdelwahed, Nilkamal Mahanta, Dmytro Fedoseyenko, Benjamin Philmus, Lisa E. Cooper, Yiquan Liu, Isita Jhulki, Steven E. Ealick, and Tadhg P. Begley. 2015. "Radical S-Adenosylmethionine (SAM) Enzymes in Cofactor Biosynthesis: A Treasure Trove of Complex Organic Radical Rearrangement Reactions." *The Journal of Biological Chemistry* 290 (7): 3980–86.
Mikaelyan, Aram, Claire L. Thompson, Katja Meuser, Hao Zheng, Pinki Rani, Rudy Plarre, and Andreas Brune. 2017. "High-Resolution Phylogenetic Analysis of Endomicrobia Reveals Multiple Acquisitions of Endosymbiotic Lineages by Termite Gut Flagellates." *Environmental Microbiology Reports*. https://doi.org/10.1111/1758-2229.12565.
Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. "Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees." In *2010 Gateway Computing Environments Workshop (GCE)*, 1–8.
Moore, Simon J., Sven T. Sowa, Christopher Schuchardt, Evelyne Deery, Andrew D. Lawrence, José Vazquez Ramos, Susan Billig, et al. 2017. "Elucidation of the Biosynthesis of the Methane Catalyst Coenzyme F430." *Nature* 543 (7643): 78–82.
Müller, Volker, Nilanjan Pal Chowdhury, and Mirko Basen. 2018. "Electron Bifurcation: A Long-Hidden Energy-Coupling Mechanism." *Annual Review of Microbiology* 72 (September): 331–53.
Naoe, Youichi, Nozomi Nakamura, Akihiro Doi, Mia Sawabe, Hiro Nakamura, Yoshitsugu Shiro, and Hiroshi Sugimoto. 2016. "Crystal Structure of Bacterial Haem Importer Complex in the Inward-Facing Conformation." *Nature Communications* 7 (November): 13411.
Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74.
Ohkuma, M., and T. Kudo. 1996. "Phylogenetic Diversity of the Intestinal Bacterial Community in the Termite Reticulitermes Speratus." *Applied and Environmental Microbiology* 62 (2): 461–68.
Ohkuma, Moriya, Tomoyuki Sato, Satoko Noda, Sadaharu Ui, Toshiaki Kudo, and Yuichi Hongoh. 2007. "The Candidate Phylum âTermite Group 1â of Bacteria: Phylogenetic Diversity, Distribution, and Endosymbiont Members of Various Gut Flagellated Protists." *FEMS Microbiology Ecology*.

https://doi.org/10.1111/j.1574-6941.2007.00311.x.
Olm, Matthew R., Christopher T. Brown, Brandon Brooks, and Jillian F. Banfield. 2017. "dRep: A Tool for Fast and Accurate Genomic Comparisons That Enables Improved Genome Recovery from Metagenomes through de-Replication." *The ISME Journal* 11 (12): 2864–68.
Pasolli, Edoardo, Francesco Asnicar, Serena Manara, Moreno Zolfo, Nicolai Karcher, Federica Armanini, Francesco Beghini, et al. 2019. "Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle." *Cell* 176 (3): 649–62.e20.
Peng, Yu, Henry C. M. Leung, S. M. Yiu, and Francis Y. L. Chin. 2012. "IDBA-UD: A de Novo Assembler for Single-Cell and Metagenomic Sequencing Data with Highly Uneven Depth." *Bioinformatics* 28 (11): 1420–28.
Petersen, Thomas Nordahl, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2011. "SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions." *Nature Methods* 8 (10): 785–86.
Peters, John W., Anne-Frances Miller, Anne K. Jones, Paul W. King, and Michael Ww Adams. 2016. "Electron Bifurcation." *Current Opinion in Chemical Biology* 31 (April): 146–52.
Poehlein, Anja, Silke Schmidt, Anne-Kristin Kaster, Meike Goenrich, John Vollmers, Andrea Thürmer, Johannes Bertsch, et al. 2012. "An Ancient Pathway Combining Carbon Dioxide Fixation with the Generation and Utilization of a Sodium Ion Gradient for ATP Synthesis." *PloS One* 7 (3): e33439.
Probst, Alexander J., Cindy J. Castelle, Andrea Singh, Christopher T. Brown, Karthik Anantharaman, Itai Sharon, Laura A. Hug, et al. 2017. "Genomic Resolution of a Cold Subsurface Aquifer Community Provides Metabolic Insights for Novel Microbes Adapted to High CO2 Concentrations." *Environmental Microbiology* 19 (2): 459–74.
Punta, Marco, Penny C. Coggill, Ruth Y. Eberhardt, Jaina Mistry, John Tate, Chris Boursnell, Ningze Pang, et al. 2012. "The Pfam Protein Families Database." *Nucleic Acids Research* 40 (Database issue): D290–301.
Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41 (Database issue): D590–96.
Rampelli, Simone, Stephanie L. Schnorr, Clarissa Consolandi, Silvia Turroni, Marco Severgnini, Clelia Peano, Patrizia Brigidi, Alyssa N. Crittenden, Amanda G. Henry, and Marco Candela. 2015. "Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota." *Current Biology: CB* 25 (13): 1682–93.
Raveh-Sadka, Tali, Brian C. Thomas, Andrea Singh, Brian Firek, Brandon Brooks, Cindy J. Castelle, Itai Sharon, et al. 2015. "Gut Bacteria Are Rarely Shared by Co-Hospitalized Premature Infants, regardless of Necrotizing Enterocolitis Development." *eLife* 4 (March). https://doi.org/10.7554/eLife.05477.
Raymond, Jason, Janet L. Siefert, Christopher R. Staples, and Robert E. Blankenship. 2004. "The Natural History of Nitrogen Fixation." *Molecular Biology and Evolution* 21 (3): 541–54.
Remmert, Michael, Andreas Biegert, Andreas Hauser, and Johannes Söding. 2011. "HHblits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment." *Nature Methods* 9 (2): 173–75.
Rosnow, Joshua J., Sungmin Hwang, Bryan J. Killinger, Young-Mo Kim, Ronald J. Moore, Stephen R. Lindemann, Julie A. Maupin-Furlow, and Aaron T. Wright. 2018. "A Cobalamin Activity-Based Probe Enables Microbial Cell Growth and Finds New Cobalamin-Protein Interactions across Domains." *Applied and Environmental Microbiology* 84 (18). https://doi.org/10.1128/AEM.00955-18.
Saier, Milton H., Vamsee S. Reddy, Brian V. Tsu, Muhammad Saad Ahmed, Chun Li, and Gabriel Moreno-Hagelsieb. 2016. "The Transporter Classification Database (TCDB): Recent Advances." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv1103.
Schuchmann, Kai, Nilanjan Pal Chowdhury, and Volker Müller. 2018. "Complex Multimeric [FeFe] Hydrogenases: Biochemistry, Physiology and New Opportunities for the Hydrogen Economy."

Frontiers in Microbiology. https://doi.org/10.3389/fmicb.2018.02911.
Schuchmann, Kai, and Volker Müller. 2016. "Energetics and Application of Heterotrophy in Acetogenic Bacteria." *Applied and Environmental Microbiology* 82 (14): 4056–69.
Schut, Gerrit J., Eric S. Boyd, John W. Peters, and Michael W. W. Adams. 2013. "The Modular Respiratory Complexes Involved in Hydrogen and Sulfur Metabolism by Heterotrophic Hyperthermophilic Archaea and Their Evolutionary Implications." *FEMS Microbiology Reviews* 37 (2): 182–203.
Shaw, A. Joe, David A. Hogsett, and Lee R. Lynd. 2009. "Identification of the [FeFe]-Hydrogenase Responsible for Hydrogen Generation in Thermoanaerobacterium Saccharolyticum and Demonstration of Increased Ethanol Yield via Hydrogenase Knockout." *Journal of Bacteriology* 191 (20): 6457–64.
Shi, Weibing, Christina D. Moon, Sinead C. Leahy, Dongwan Kang, Jeff Froula, Sandra Kittelmann, Christina Fan, et al. 2014. "Methane Yield Phenotypes Linked to Differential Gene Expression in the Sheep Rumen Microbiome." *Genome Research* 24 (9): 1517–25.
Sieber, Christian M. K., Alexander J. Probst, Allison Sharrar, Brian C. Thomas, Matthias Hess, Susannah G. Tringe, and Jillian F. Banfield. n.d. "Recovery of Genomes from Metagenomes via a Dereplication, Aggregation, and Scoring Strategy." https://doi.org/10.1101/107789.
Soding, J. 2005. "Protein Homology Detection by HMM-HMM Comparison." *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bti125.
Sofia, Heidi J., Guang Chen, Beth G. Hetzler, Jorge F. Reyes-Spindola, and Nancy E. Miller. 2001. "Radical SAM, a Novel Protein Superfamily Linking Unresolved Steps in Familiar Biosynthetic Pathways with Radical Mechanisms: Functional Characterization Using New Analysis and Information Visualization Methods." *Nucleic Acids Research* 29 (5): 1097–1106.
Song, Zhao-Qi, Feng-Ping Wang, Xiao-Yang Zhi, Jin-Quan Chen, En-Min Zhou, Feng Liang, Xiang Xiao, et al. 2013. "Bacterial and Archaeal Diversities in Yunnan and Tibetan Hot Springs, China." *Environmental Microbiology*. https://doi.org/10.1111/1462-2920.12025.
Spring, Stefan, Boyke Bunk, Cathrin Spröer, Peter Schumann, Manfred Rohde, Brian J. Tindall, and Hans-Peter Klenk. 2016. "Characterization of the First Cultured Representative of Verrucomicrobia Subdivision 5 Indicates the Proposal of a Novel Phylum." *The ISME Journal* 10 (12): 2801–16.
Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13.
Steinegger, Martin, and Johannes Söding. 2017. "MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets." *Nature Biotechnology* 35 (11): 1026–28.
Stingl, Ulrich, Renate Radek, Hong Yang, and Andreas Brune. 2005. "'Endomicrobia': Cytoplasmic Symbionts of Termite Gut Protozoa Form a Separate Phylum of Prokaryotes." *Applied and Environmental Microbiology* 71 (3): 1473–79.
Wu, Yu-Wei, Blake A. Simmons, and Steven W. Singer. 2016. "MaxBin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets." *Bioinformatics* 32 (4): 605–7.
Youssef, Noha H., Ibrahim F. Farag, C. Ryan Hahn, Hasitha Premathilake, Emily Fry, Matthew Hart, Krystal Huffaker, et al. 2019. "Candidatus Krumholzibacterium Zodletonense Gen. Nov., Sp Nov, the First Representative of the Candidate Phylum Krumholzibacteriota Phyl. Nov. Recovered from an Anoxic Sulfidic Spring Using Genome Resolved Metagenomics." *Systematic and Applied Microbiology*. https://doi.org/10.1016/j.syapm.2018.11.002.
Zheng, Hao, Carsten Dietrich, and Andreas Brune. 2017. "Genome Analysis of Endomicrobium Proavitum Suggests Loss and Gain of Relevant Functions during the Evolution of Intracellular Symbionts." *Applied and Environmental Microbiology* 83 (17). https://doi.org/10.1128/AEM.00656-17.
Zheng, Hao, Carsten Dietrich, Renate Radek, and Andreas Brune. 2016a. "Endomicrobium Proavitum, the First Isolate ofEndomicrobiaclass. Nov. (phylumElusimicrobia) - an Ultramicrobacterium with an Unusual Cell Cycle That Fixes Nitrogen with a Group IV Nitrogenase." *Environmental*

Microbiology. https://doi.org/10.1111/1462-2920.12960.
———. 2016b. "E Ndomicrobium Proavitum, the First Isolate of E Ndomicrobia Class. Nov.(phylum E Lusimicrobia)--an Ultramicrobacterium with an Unusual Cell Cycle That Fixes Nitrogen with a G Roup IV Nitrogenase." *Environmental Microbiology* 18 (1): 191–204.
Zheng, Kaiyuan, Phong D. Ngo, Victoria L. Owens, Xue-Peng Yang, and Steven O. Mansoorabadi. 2016. "The Biosynthetic Pathway of Coenzyme F430 in Methanogenic and Methanotrophic Archaea." *Science* 354 (6310): 339–42.
Zheng, Y., J. Kahnt, I. H. Kwon, R. I. Mackie, and R. K. Thauer. 2014. "Hydrogen Formation and Its Regulation in Ruminococcus Albus: Involvement of an Electron-Bifurcating [FeFe]-Hydrogenase, of a Non-Electron-Bifurcating [FeFe]-Hydrogenase, and of a Putative Hydrogen-Sensing [FeFe]-Hydrogenase." *Journal of Bacteriology*. https://doi.org/10.1128/jb.02070-14.

804 SUPPLEMENTARY FIGURES



805

806 Figure S1. Phylogenetic classification of 27 Elusimicrobia genomes based on their 16S rRNA gene
807 sequences. Reference sequences from the SILVA database anchor the tree. Scale bar indicates substitutions
808 per site.

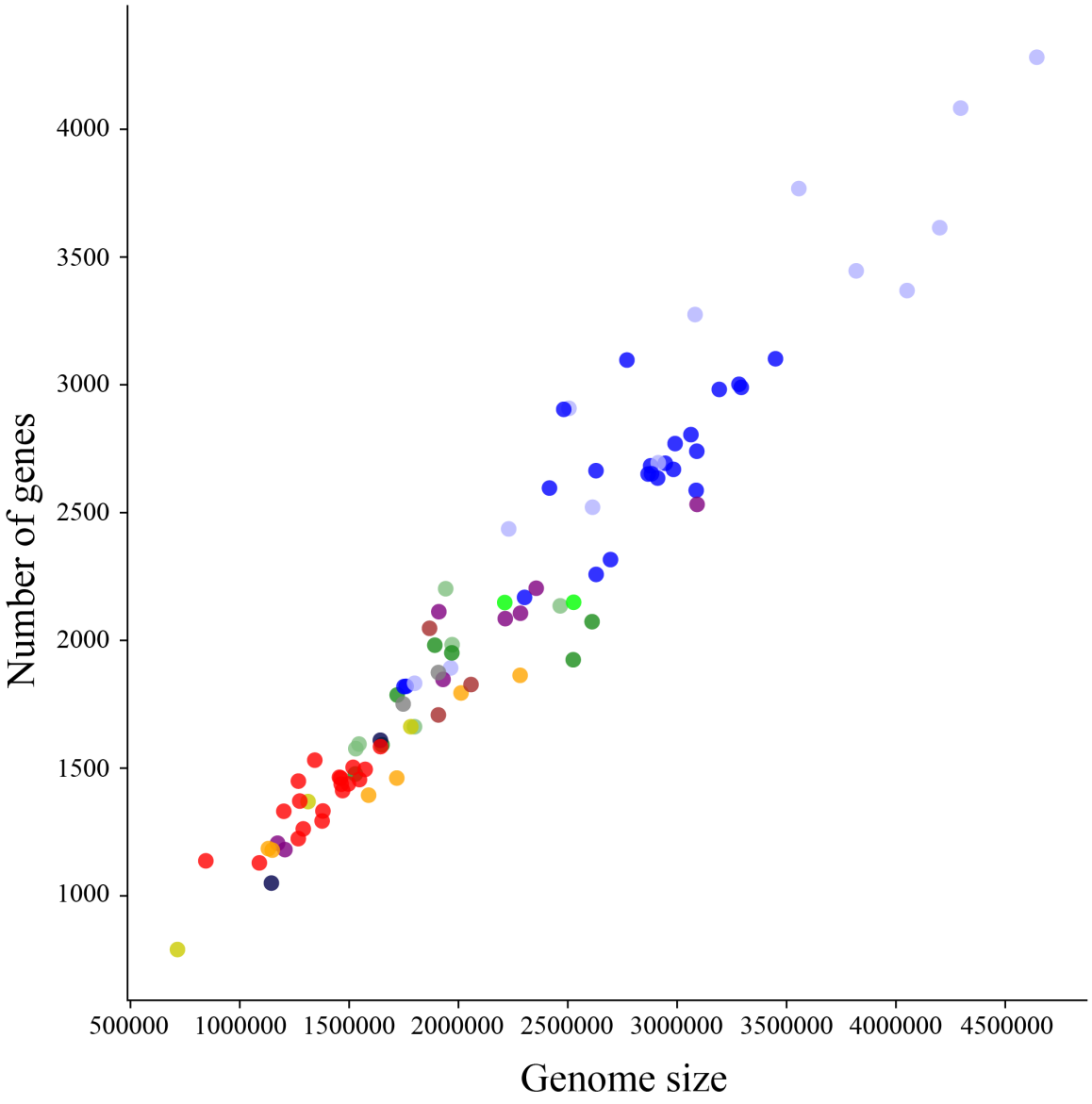


Figure S2. Number of genes and size (in base pairs) of the 94 Elusimicrobia genomes. Each genome is represented by a dot of the color of the lineage it has been classified to (see Figure 1 for color scheme).

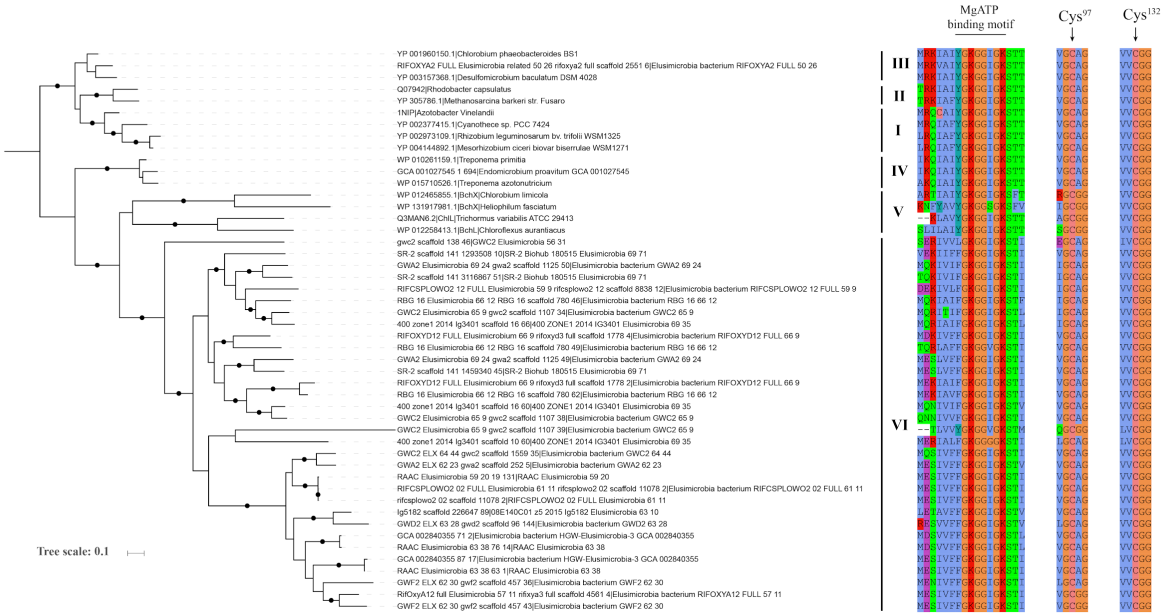


Figure S3. Phylogenetic tree of the nifH genes of Elusimicrobia and the nitrogenases of other bacterial species, and the conserved amino acids around the crucial cysteine residues that coordinate the [Fe_4S_4] ligands in the Fe protein and for the binding of MgATP. The cysteine residues, based on *A. vinelandii* numbering, are indicated with vertical arrows. Scale bar indicates substitutions per site.

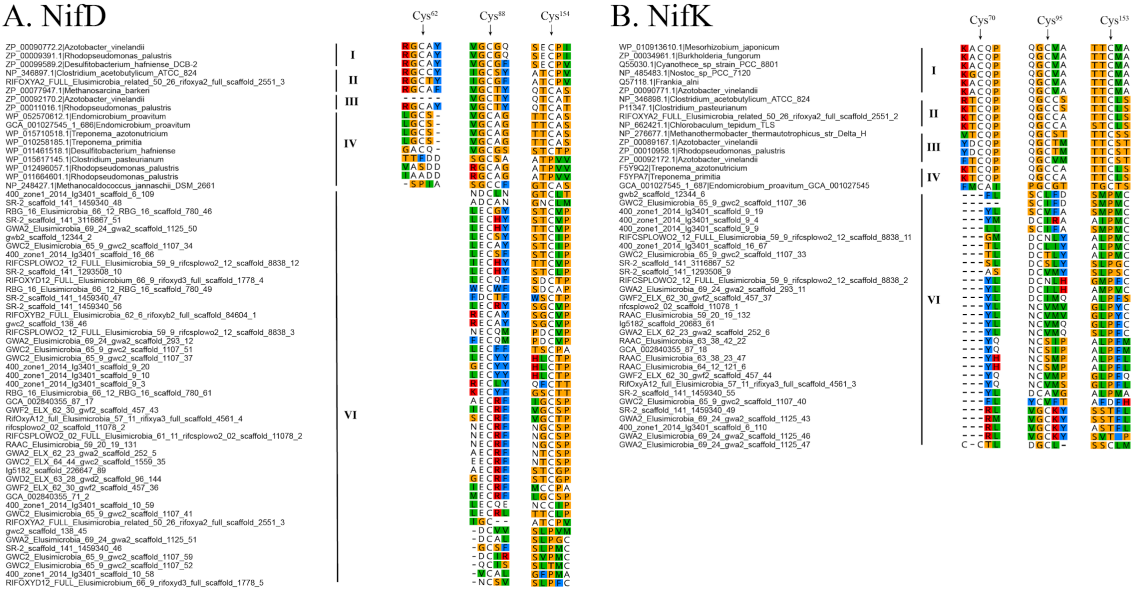


Figure S4. A. Conservation of crucial cysteine residues in NifD (alpha) and NifK (beta) in Elusimicrobia and homologs in groups I, II, III and VI. The P-cluster ligands, based on *A vinelandii* numbering, are indicated with vertical arrows.

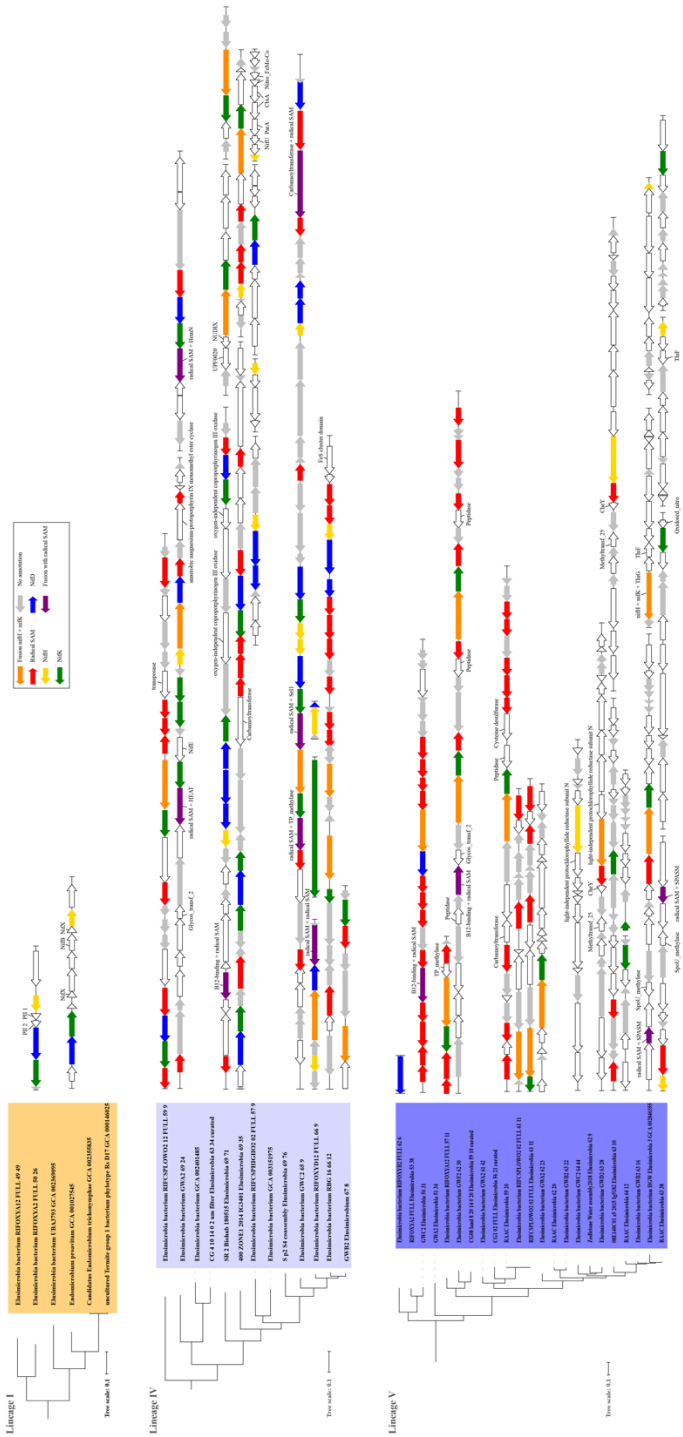


Figure S5. Genomic neighborhood of nitrogenase in genomes from lineages I, IV and V.

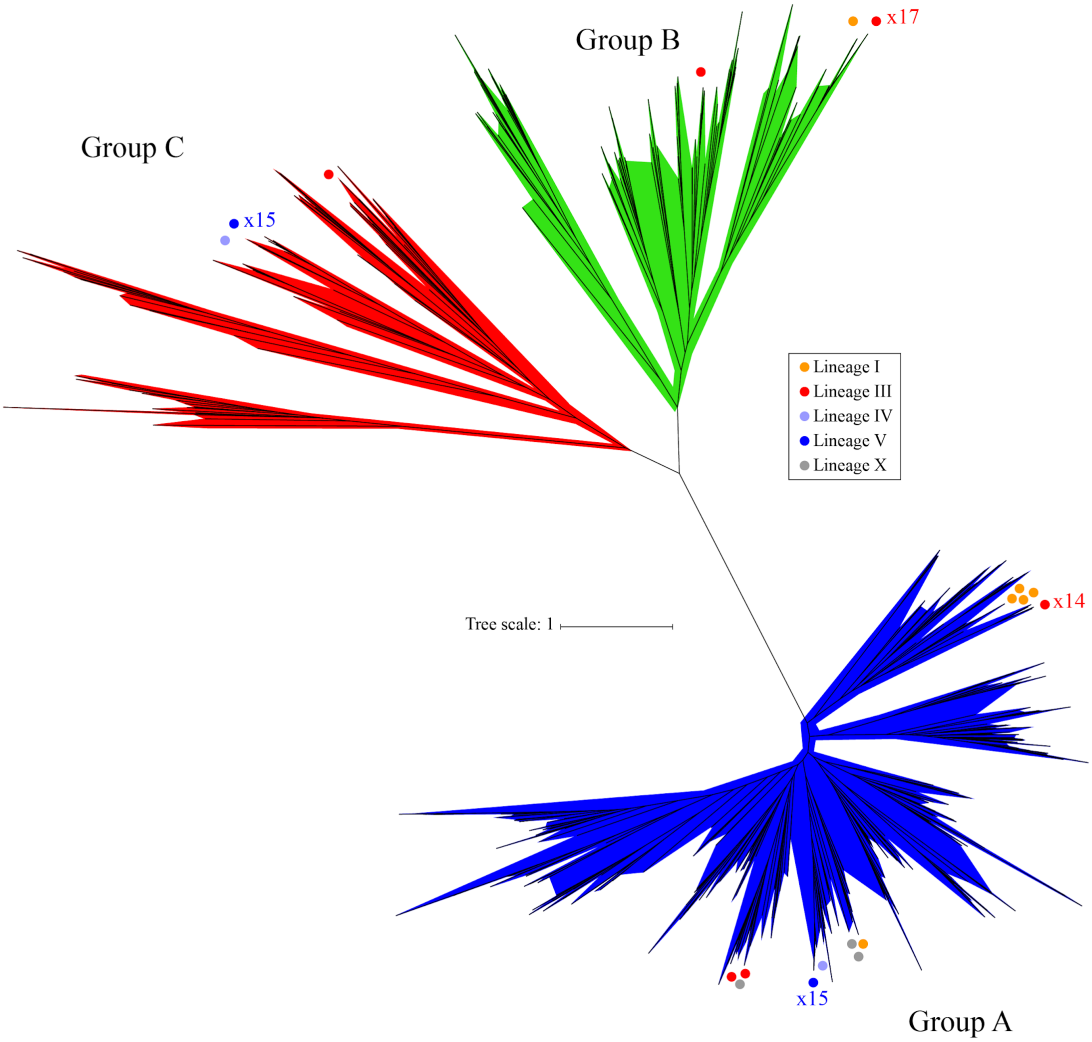


Figure S6. Phylogenetic tree of the [FeFe] hydrogenase (Maximum likelihood tree under an LG+I+G4 model of evolution). Elusimicrobia have group A, B and C [FeFe] hydrogenases. Elusimicrobia sequences are indicated with colored dots. Scale bar indicates substitutions per site. The tree was made using sequences from (Carnevali et al. 2019; Greening et al. 2016). The sequences and tree are available with full bootstrap values in Fasta and Newick formats in Supplementary Data.

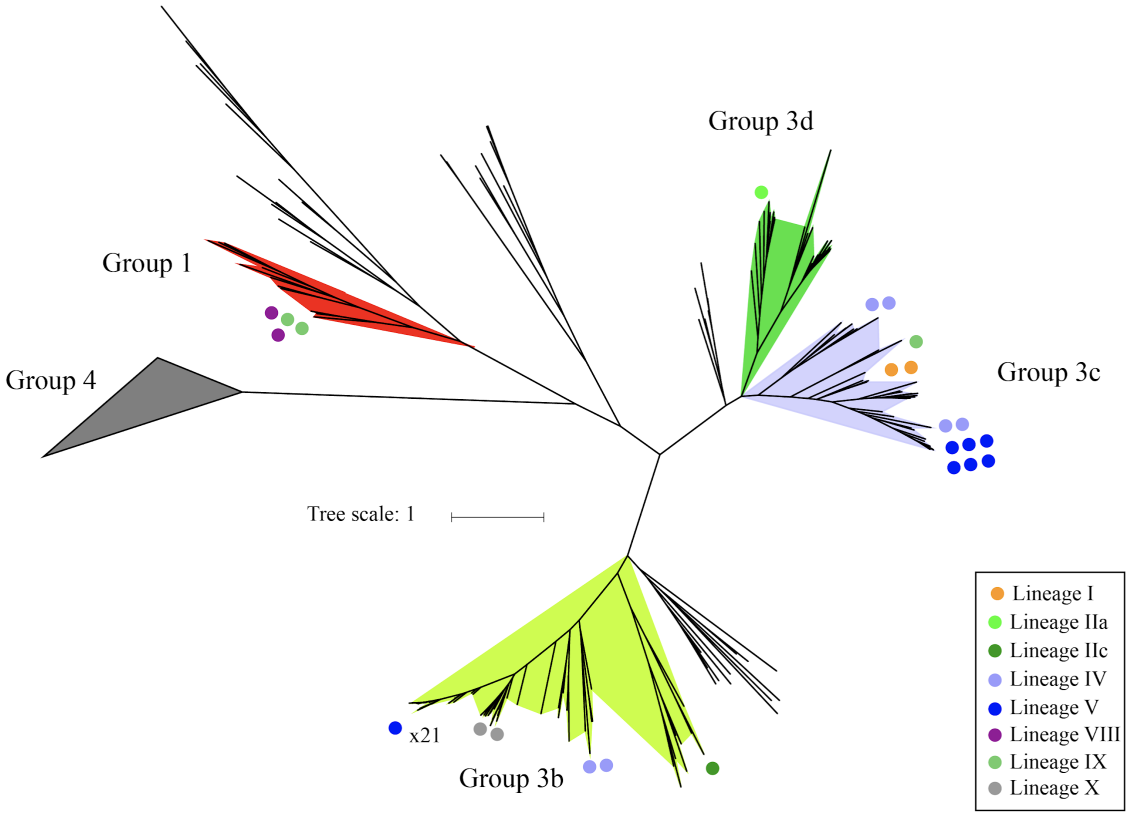
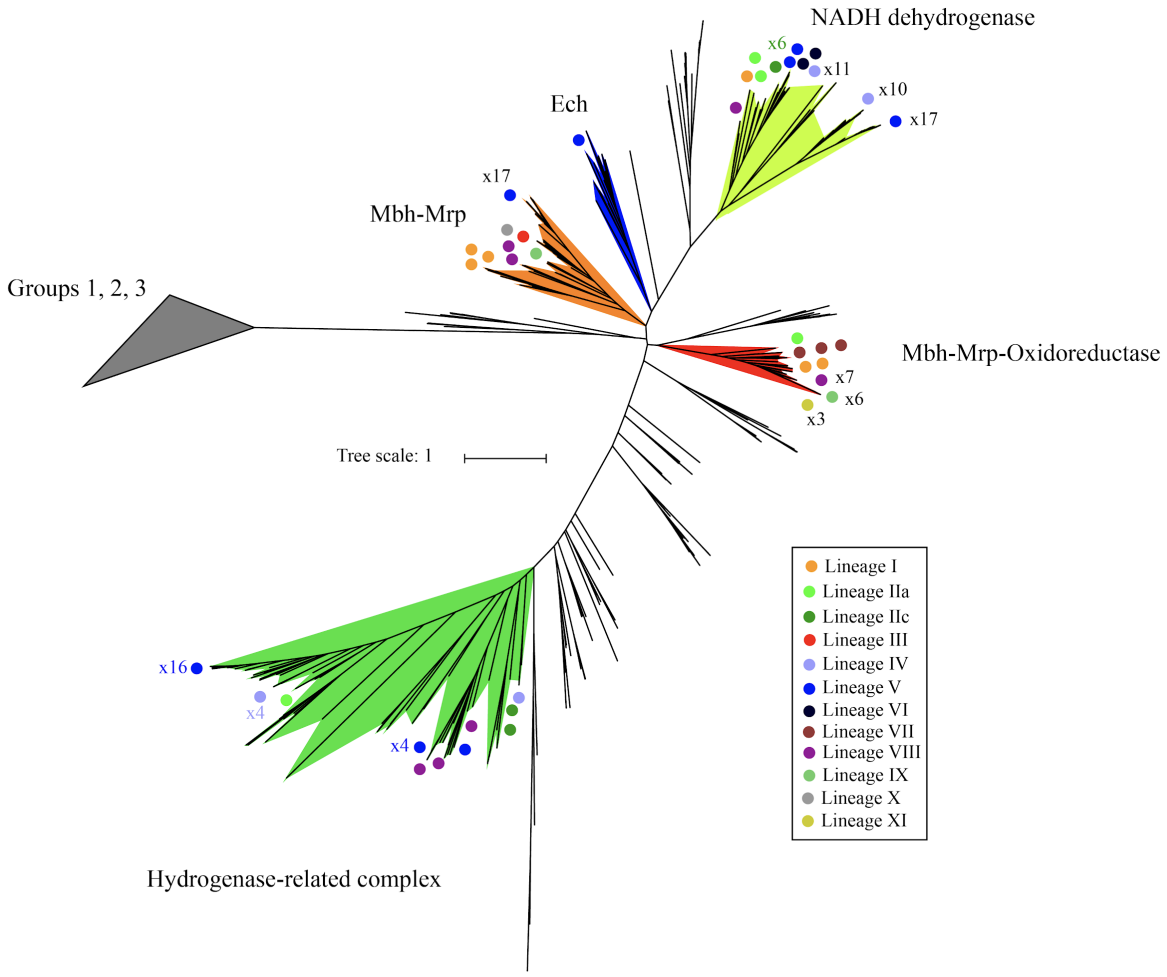


Figure S7. Phylogenetic tree of the groups 1, 2 and 3 of [NiFe] hydrogenase (Maximum likelihood tree under a LG+G4 model of evolution). Elusimicrobia have group 1, group 3b, 3c and 3d [NiFe] hydrogenases. Elusimicrobia sequences are indicated with colored dots. Scale bar indicates substitutions per site. The tree was made using sequences from (Carnevali et al. 2019). The sequences and tree are available with full bootstrap values in Fasta and Newick format in Supplementary Data.



858

859 Figure S8. Phylogenetic tree of the group 4 [NiFe] hydrogenase (Maximum likelihood tree under a
860 LG+I+G4 model of evolution). Elusimicrobia have Mbh-Mrp, Mbh-Mrp-Oxidoreductase, Ech and
861 Hydrogenase-related complex [NiFe] hydrogenases. They also have NADH dehydrogenase. Elusimicrobia
862 sequences are indicated with colored dots. Scale bar indicates substitutions per site. The tree was made
863 using sequences from (Carnevali et al. 2019). The sequences and tree are available with full bootstrap values
864 in Fasta and Newick format in Supplementary Data.

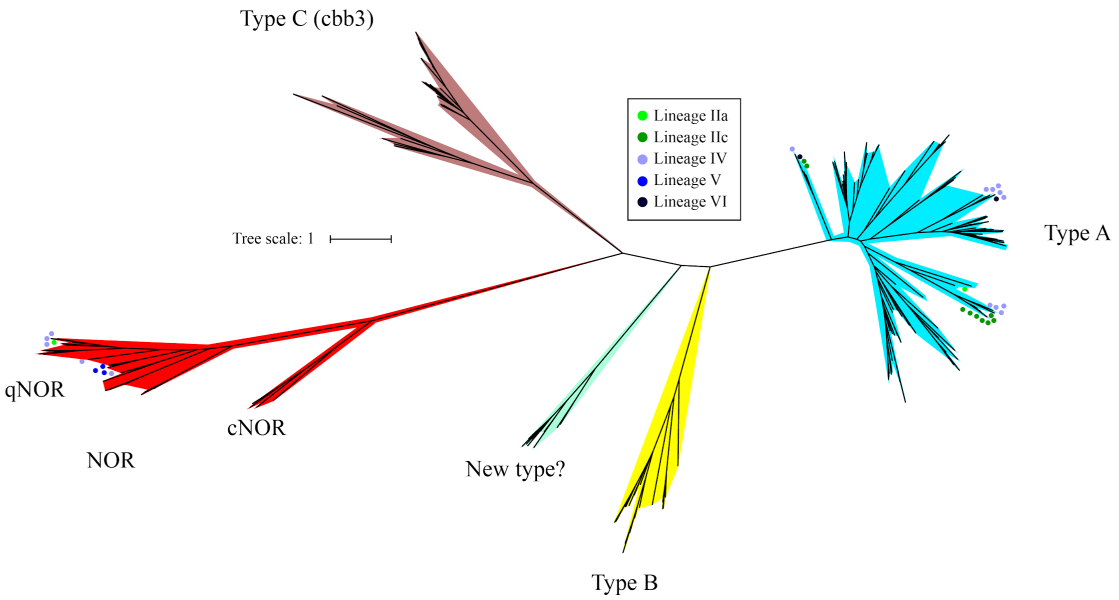


Figure S9. Phylogenetic tree of the catalytic subunit I of heme-copper oxygen reductases (Maximum likelihood tree under a LG+F+I+G4 model of evolution). Elusimicrobia have an aa_3-type heme-copper oxygen reductase (Type A) and nitric oxide reductase that uses quinol as the electron donor (qNOR). Elusimicrobia sequences are indicated with colored dots. The tree was made using sequences from (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.

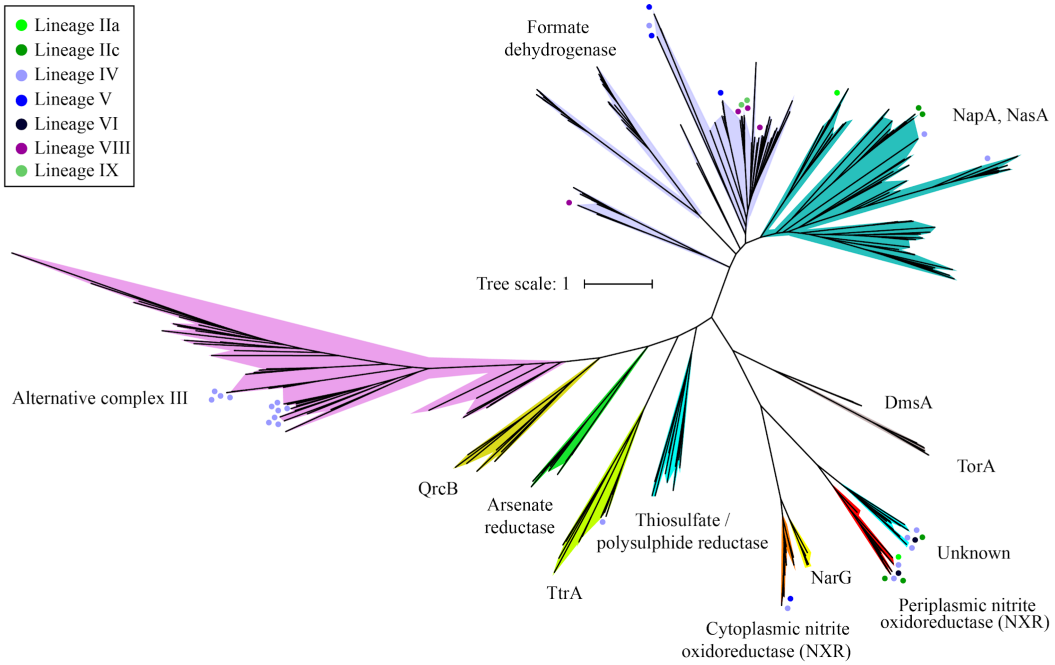


Figure S10. Phylogenetic analysis of the dimethyl sulfoxide (DMSO) reductase superfamily (Maximum likelihood tree under a LG+F+G4 model of evolution). Elusimicrobia have an alternative complex III, a tetrathionate reductase (TtrA), cytoplasmic and periplasmic nitrite oxidoreductase (NXR), formate dehydrogenase and nitrate reductase (NapA and NasA). Elusimicrobia sequences are indicated with colored dots. Scale bar indicates substitutions per site. The tree was made using sequences from (Castelle et al. 2013; Carnevali et al. 2019). The sequences and tree are available with full bootstrap values in Fasta and Newick format in Supplementary Data.

Table S1. List of the 94 genomes used in this study.

Table S2. List of Myxococcales and Gracilibacteria genomes with the group VI nifH.

Table S3. Metabolic capacities across the 94 Elusimicrobia genomes.