



HAL
open science

An early origin of iron–sulfur cluster biosynthesis machineries before Earth oxygenation

Pierre Simon Garcia, Francesca D'angelo, Sandrine Ollagnier de Choudens, Macha Dussouchaud, Emmanuelle Bouveret, Simonetta Gribaldo, Frédéric Barras

► To cite this version:

Pierre Simon Garcia, Francesca D'angelo, Sandrine Ollagnier de Choudens, Macha Dussouchaud, Emmanuelle Bouveret, et al.. An early origin of iron–sulfur cluster biosynthesis machineries before Earth oxygenation. *Nature Ecology & Evolution*, 2022, 6, pp.1564-1572. 10.1038/s41559-022-01857-1. hal-03780728

HAL Id: hal-03780728

<https://hal.science/hal-03780728>

Submitted on 12 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

19 **Abstract**

20 Iron-sulfur (Fe-S) clusters are ubiquitous co-factors essential for life. It is largely
21 thought that the emergence of oxygenic photosynthesis and progressive oxygenation
22 of the atmosphere led to the origin of multiprotein machineries (ISC, NIF, and SUF)
23 assisting Fe-S cluster synthesis in the presence of oxidative stress and shortage of
24 bioavailable iron. However, previous analyses have left unclear the origin and
25 evolution of these systems. Here, we combine exhaustive homology searches with
26 genomic context analysis and phylogeny to precisely identify Fe-S cluster biogenesis
27 systems in over 10,000 archaeal and bacterial genomes. We highlight the existence
28 of two additional and clearly distinct “minimal” Fe-S cluster assembly machineries,
29 MIS and SMS, which we infer in the Last Universal Common Ancestor (LUCA), and
30 we experimentally validate SMS as a bona fide Fe-S cluster biogenesis system.
31 These ancestral systems were kept in Archaea whereas they went through stepwise
32 complexification in Bacteria to incorporate additional functions for higher Fe-S cluster
33 synthesis efficiency leading to SUF, ISC, and NIF. Horizontal gene transfers and
34 losses then shaped the current distribution of these systems, driving ecological
35 adaptations such as the emergence of aerobic lifestyles in archaea. Our results show
36 that dedicated machineries were in place early in evolution to assist Fe-S cluster
37 biogenesis, and that their origin is not directly linked to Earth oxygenation.

38 **Main**

39 Iron-Sulfur (Fe-S) clusters are ubiquitous and essential protein cofactors that allow
40 redox chemistry and Lewis acid-type catalysis for hundreds of structurally and
41 functionally diverse proteins in a wide array of processes associated with
42 bioenergetics, metabolism, DNA synthesis, and genetic regulation¹. Studies in model
43 systems, both prokaryotes and eukaryotes, have identified three Fe-S cluster
44 biogenesis machineries^{2,3}: ISC, considered as the housekeeping system, SUF,
45 acting as a backup in stress conditions, and NIF, specialized in the maturation of
46 nitrogenase in N₂ fixing bacteria^{2,3}. Eukaryotic Fe-S cluster assembly machineries
47 (ISC in mitochondria and SUF in photosynthetic lineages) are of endosymbiotic origin
48 from *Alphaproteobacteria* and *Cyanobacteria*, respectively⁴. Although made of
49 different components, these three systems facilitate Fe-S cluster biogenesis following
50 the same basic principles (Figure 1A): a cysteine desulfurase takes sulfur from L-
51 cysteine and transfers it as a persulfide onto a scaffold protein, which also receives
52 Fe²⁺ and electrons to build an Fe-S cluster that is then transferred to cellular apo-
53 protein targets via carrier proteins^{2,5}.

54 A largely held view posits that in early life, anaerobiosis, combined with high
55 availability of ferrous iron and sulfur, allowed spontaneous assembly of Fe-S
56 clusters^{6,7}. Subsequent oxygenation of the atmosphere following the emergence of
57 photosynthesis would have then led to increasingly unfavourable conditions for Fe-S
58 based biology as ferrous iron level decreased, reactive oxygen species formed, and
59 oxidized sulfur became more difficult to assimilate. This would have favoured the
60 emergence of protein machineries (ISC, SUF and NIF) enabling Fe-S cluster
61 biogenesis, distribution, and protection^{8,9}.

62 To date only two studies have tackled the issue of the origin and evolution of Fe-S
63 biogenesis machineries but reached contradictory conclusions. An early survey
64 suggested the presence of SUF in the Last Universal Common Ancestor (LUCA)
65 based on the presence of homologues of SufBC in both Archaea and Bacteria¹⁰. In
66 contrast, another study proposed that SUF originated in the Archaea, followed by
67 massive horizontal transfers to Bacteria based on a SufBC phylogeny¹¹. These
68 contrasting results derive probably by the limited taxonomic sampling used and the
69 lack of extensive phylogenomic analysis.

70 The uttermost importance of Fe-S-based biology for all living organisms prompted us
71 to carry with an exhaustive large-scale analysis including a larger fraction of
72 prokaryotic diversity, and notably the Archaea, for which very little is known about Fe-
73 S cluster synthesis.

74

75 **Expanded diversity of Fe-S cluster biogenesis reveals two additional systems**

76 To infer the evolutionary history of Fe-S cluster systems, it is first necessary to
77 precisely identify them, a task hampered by the fact that most components are part of
78 very large multiprotein families. Therefore, we combined exhaustive homology
79 searches with analysis of genomic context and phylogeny to delineate the protein
80 families corresponding to each component of the ISC, NIF, and SUF systems in
81 10,865 prokaryotic genomes covering all current bacterial and archaeal diversity
82 (Figure 1B, Tables S1, S2, and S3, Supplementary data). Our results show that the
83 cysteine desulfurase and scaffold are core components shared among SUF, ISC and
84 NIF machineries (Figure 1B, light blue and orange, respectively). Cysteine
85 desulfurases SufS, NifS, and IscS form distinct clades suggesting that they were
86 each recruited from the same large family and evolved independently. A similar
87 evolution can be inferred for the U-type scaffold proteins (SufU, NifU, and IscU). In
88 contrast, the SUF-type scaffold proteins belong to two separate families (SufBD and
89 SufC) (Figure 1B). These core components integrated additional proteins recruited
90 from various families to form the SUF, NIF and ISC systems (Figure 1B).

91 Interestingly, the combination of phylogenies and genomic context analyses revealed
92 the existence of two additional "minimal" Fe-S cluster biogenesis systems, which we
93 called MIS (Minimal Iron Sulfur) and SMS (SUF-like Minimal System) (Figure 1C).
94 They are each composed of two components, which form phylogenetically distinct
95 families and whose genes are adjacent in the genomes (Figure 1B and C). The MIS
96 system is made of a pair of IscS/IscU-like homologs (a cysteine desulfurase and a
97 scaffold, respectively), which we call MisS/MisU (Figure 1B and C). As a matter of
98 fact, MIS systems from the archaeon *Methanosarcina acetivorans* and the bacterium
99 *Helicobacter pylori* have been shown to catalyse Fe-S cluster biogenesis *in vitro*, but
100 were misidentified as ISC and NIF, respectively^{12,13}.

101 The second minimal system (SMS) is composed of two homologs of the SUF scaffold
102 proteins SufBC, which we called SmsCB (because of the order of the two genes,
103 Figure 1B and C). This system is totally uncharacterized and does not include any

104 component predicted to have a cysteine desulfurase activity, hence leaving open
105 what might be the source of sulfur. We therefore tested SmsCB from two archaea, *M.*
106 *acetivorans* and *Methanocaldococcus jannaschii*, for their capacity to promote Fe-S
107 cluster biogenesis. Using the bacterial two-hybrid system, we found that *M.*
108 *jannaschii* SmsB and SmsC form homo and heterodimers (Figure 2A, Extended Data
109 figure 1AB), suggesting the existence of a SmsCB complex. We then cloned *M.*
110 *acetivorans* and *M. jannaschii smsCB* on an expression plasmid into *E. coli* and
111 purified the corresponding complexes under anaerobic conditions (Figure 2B). In
112 both cases, a complex of SmsCB proteins was evidenced as SmsB co-eluted with
113 His-tagged SmsC. UV-visible spectra of both SmsCB complexes were recorded and
114 an absorption band at 420 nm, characteristic of an Fe-S cluster, was observed for in
115 each (Figure 2CD). Quantitative measurements showed 4.05 ± 0.02 Fe and $3.8 \pm$
116 0.02 S per *M. acetivorans* SmsCB dimer and 0.15 ± 0.02 Fe and 0.13 ± 0.03 S per
117 *M. jannaschii* SmsCB dimer. These results demonstrated that as-purified SmsCB
118 harbors Fe-S cluster(s). Next, Fe-S bound SmsCB complex from *M. acetivorans* was
119 co-incubated with non-active apo-aconitase. Analysis over time showed that
120 aconitase activity resumed, indicating that *M. acetivorans* SmsCB complex
121 reactivated aconitase by transferring an Fe-S cluster and therefore acted as a
122 scaffold (Figure 2E). In parallel, the Fe-S bound SmsCB complex from *M. jannaschii*
123 was reconstituted *in vitro* after incubation with iron and sulfur salts under reducing
124 conditions (Extended Data figure 1C). The resulting *M. jannaschii* holo-SmsCB
125 complex was also found to reactivate apo-aconitase (Figure 2F). Altogether, these
126 analyses show that SmsCB proteins bind Fe-S clusters and can promote reactivation
127 of a cellular apo-target by transferring their cluster.

128 To conclude, our phylogenomic and functional analyses revealed that the number
129 and the diversity of Fe-S cluster machineries are much larger than currently thought.

130

131 **Fe-S cluster biogenesis systems are unevenly distributed in Bacteria and** 132 **Archaea**

133 We then analysed how the five systems, ISC, SUF, NIF, MIS, and SMS are
134 distributed in prokaryotes. Except for some members of the *Lactobacillales*
135 (*Aerococcaceae*, *Lactobacillaceae*, *Leuconostocaceae*) and a few intracellular
136 bacteria (ex. *Mollicutes*), Fe-S biogenesis systems were found in all analysed
137 genomes (Table S1, 2 and S3, and Supplementary Discussion). They are notably

138 present in the bacterial Candidate Phyla Radiation (CPR) and the archaeal DPANN
139 clade (Figure 3), despite their highly streamlined genomes^{14,15}. In general, we
140 observe the co-occurrence of components of each system, indicating overall co-
141 evolution. This is the case of SufBCDSU in the SUF system, IscSU, IscX, HscAB,
142 Fdx, CyaY in the ISC system, and NifUS, IscAnif in the NIF system (Extended Data
143 figure 2 and 3). A few exceptions likely correspond to components endowed with
144 alternative functions other than Fe-S cluster assembly. This is the example of SufT,
145 which exhibits a wider taxonomic distribution than the other SUF components
146 (Extended Data figure 2 and 3, and Supplementary Discussion).

147 Mapping of the five systems onto the reference phylogenies of Bacteria and Archaea
148 shows a striking pattern. ISC and NIF are in fact limited to a few bacterial clades,
149 whereas SUF and SMS are much more widespread and generally mutually exclusive
150 in Bacteria and Archaea, respectively. Finally, MIS is present in both Bacteria and
151 Archaea (Figure 3, Extended Data figure 1 and 2). Outside very few exceptions,
152 mutual exclusion of the different systems refers to both core and accessory proteins
153 (Extended Data figures 2, 3, and Table S3).

154 Our results show that SUF is the most widely distributed Fe-S cluster system in
155 Bacteria, as opposed to the much more studied ISC (43.5% of all genomes analysed
156 having SUF versus 7.3% having ISC). We also observe an overall distinct
157 composition of SUF between *Terrabacteria* (mostly SufBCDSU), and the *Gracilicutes*
158 (either SufBCDSU or SufBCDEA) (Extended Data figure 2). Moreover, most bacteria
159 possess only one system, indicating that the presence of both SUF and ISC in *E. coli*
160 and other *Enterobacteria* is rather an exception than the rule among
161 *Gammaproteobacteria* (Extended Data figure 4). In contrast, in Archaea the MIS and
162 SMS systems largely co-occur, although they are never side by side in the genomes
163 (Figure 3).

164

165 **SMS and MIS are ancestral systems that date back to the Last Universal** 166 **Common Ancestor**

167 Our results establish the ubiquity of Fe-S cluster biogenesis across prokaryotes and
168 pose the question of what evolutionary processes shaped their current distribution.
169 We therefore analysed the phylogeny of each of the five systems (Methods). To
170 increase the phylogenetic signal, we concatenated the homologous components
171 shared by the SUF and SMS systems (SufBCD, and SmsBC, respectively). The

172 resulting tree shows a clear and well-supported separation of Bacteria (SufBCD) and
173 Archaea (SmsCB) (98% Bootstrap Value (BV)) consistent with their distinctive
174 genomic contexts (Figure 4A), and the monophyly of major clades indicating vertical
175 inheritance (Extended Data figure 5). Also, the phylogenies of SufS, SufU and SufE
176 are overall congruent with the SufBCD tree (Extended Data figures 6-8). We thus
177 infer that SMS is ancestral to Archaea and SUF is ancestral to Bacteria. Because
178 both SUF and SMS systems share only the scaffold components SufBD/SmsB and
179 SufC/SmsC, we conclude that at least a minimal system -similar to present-day SMS-
180 was already present in the Last Universal Common Ancestor (LUCA). The origin of
181 bacterial SufS remains unclear as two hypotheses can be formulated based on
182 phylogeny (Extended Data figure 6, Supplementary data). The presence of a small
183 sister clade of archaeal cysteine desulfurases in the phylogeny of the superfamily
184 (Supplementary data) may suggest that SufS was recruited to work in the bacterial
185 SUF system from an ancestral cysteine desulfurase already present in the LUCA.
186 Alternatively, the whole clade including the few archaeal cysteine desulfurases and
187 the bacterial SufS may have diverged from MisS. Concerning the other components,
188 SufU may have derived from MisU, and SufE would have appeared *de novo* in
189 Bacteria.

190 The phylogeny of the U-type scaffold protein superfamily being poorly resolved
191 (Supplementary data), we could not concatenate MisS and MisU for a combined
192 phylogenetic analysis of the MIS system. Yet the phylogeny of MisS shows the
193 monophyly of many phyla and the separation of Archaea and Bacteria (Figure 4B,
194 Extended Data figure 9). Together with the widespread distribution of MIS, these
195 results may indicate its presence in the LUCA. Still, one should remain cautious with
196 this interpretation as the short branch separating Bacteria and Archaea in the MisS
197 tree differ with those shown by other proteins inferred in the LUCA.

198 Interestingly, the NIF system, which can be clearly identified by the conserved
199 placement of *nifS* and *nifU* within the *nif* operon (Extended Data figure 9) appears to
200 have originated recently from MIS ancestors, as NifS and NifU branch within the
201 MisS and MisU clades, respectively (Figure 2A, 4B, and Extended Data figure 9).
202 Interestingly, no NIF Fe-S biogenesis system is found in the nitrogen fixing archaea
203 (i.e. methanogens such as *Methanosarcinales* or *Methanococcales*). The presence of
204 SMS in these archaeal lineages leaves open the possibility that SMS might
205 contribute to nitrogenase maturation in these species.

206 Finally, the phylogeny inferred from a concatenation of ISC components (IscU, IscS,
207 IscX, HscA, HscB, Fdx and CyaY), combined with their limited taxonomic distribution,
208 confirms a much more recent origin of this system within the *Gracilicutes*, followed by
209 vertical inheritance and several losses (Extended Data figure 10 and Sup
210 Discussion). IscS and IscU branch close to MisS and MisU, respectively (Figure 1B).
211 Together with the much older origin inferred for MIS, this suggests that the cysteine
212 desulfurase and scaffolds of ISC were recruited from MIS components. Given the
213 limited resolution of the cysteine desulfurase and U-type scaffold phylogenies
214 (Supplementary data), it remains unclear if IscSU and NifSU emerged from gene
215 duplication or horizontal gene transfers.

216 Altogether, our results show that the emergence of the machineries to assemble Fe-
217 S clusters is very ancient and occurred in the LUCA, therefore well before the GOE.

218

219 **Horizontal gene transfers drove major adaptations in Archaea and Bacteria**

220 Given the clear early divergence of the two systems, the presence of SUF in a few
221 archaea and SMS in a few bacteria (Figure 3) can be attributed to interdomain
222 horizontal gene transfers (HGTs) with replacement of the original version (Figure 4A
223 and Extended Data figure 5). Replacement of the original SMS by complete SUF
224 systems from different bacteria match closely the emergence of archaeal lineages
225 that adapted to aerobic lifestyles, such as marine group II (*Thalassoarchaea*) and III
226 (*Pontarchaea*), extreme halophiles (*Haloarchaea*), marine and soil ammonia
227 oxidizers (AOA, *Thaumarchaea*), the thermoacidophilic *Thermoplasmatales*, and the
228 aerobic members of *Crenarchaeota* and *Asgard* (Figure 3, Extended Data figures 5-
229 8). These results strongly suggest that bacteria were the first to conquer fully
230 oxygenated niches and that HGTs from them -likely favoured by the tight clustering of
231 components in the genomes - helped the archaea to diversify into a larger range of
232 environments. Conversely, HGTs introduced SMS systems from archaea into a few
233 bacterial lineages, possibly providing higher fit to strict anaerobic conditions
234 (Extended Data figure 5).

235 Strikingly, our results indicate that *Cyanobacteria* also likely replaced their original
236 SUF system (SufBCDSU) by a more complex SufBCDSEA acquired through HGT
237 from *Proteobacteria*, likely *Nitrosomonadales* (Figure 4A). A more detailed
238 phylogeny, including a larger taxonomic sampling, points to an early event, just after
239 the divergence of *Gloeobacter* (Supplementary figure 1). By integrating SufE, a sulfur

240 acceptor that allows to accumulate sulfides, thereby preventing free sulfide to harm
241 cell components, and SufA (ATC-II), a carrier that optimises Fe-S delivery under
242 oxidative stress, this transfer was likely an asset to fight oxidative stress. Therefore, it
243 may have allowed *Cyanobacteria* to cope with increasing levels of oxygen produced
244 by their own activity, *de facto* helping the emergence of oxygenic photosynthesis.
245 These observations challenge the alleged evolutionary connection between the origin
246 of Fe-S cluster biogenesis and the emergence of oxygenic photosynthesis, but also
247 fit with the idea that increasingly sophisticated systems arose to protect from the
248 harmful effect of oxygen. Interestingly, *Cyanobacteria* appear to have also acquired
249 their NIF system from *Proteobacteria* providing increased protection to the highly
250 oxygen-sensitive nitrogenase in diazotrophs of this phylum (Figure 4B and Extended
251 Data figure 9).

252

253 **A global scenario for the evolution of Fe-S cluster biogenesis**

254 Our results allow us to draw a comprehensive scenario for the origin and evolution of
255 Fe-S cluster biogenesis machineries (Figure 5). SMS and possibly MIS were already
256 present in the LUCA. MIS would have included a cysteine desulfurase (MisS) and a
257 U-type scaffold (MisU), while SMS would have included a SufBD superfamily scaffold
258 (SmsB) and an AAA+ ATPase (SmsC). These systems remained unchanged in the
259 Archaea, while they became progressively more complex in Bacteria.

260 The emergence of SUF may have occurred through the following steps prior to the
261 divergence of all Bacteria: (i) duplication of MisS or speciation from another member
262 of this large superfamily to give rise to a cysteine desulfurase (SufS); (ii) duplication
263 of SmsB to give rise to SufB and SufD, and speciation of SmsC to give rise to SufC;
264 (iii) duplication of MisU to give rise to SufU (and limiting its function to sulfur acceptor
265 only). Further complexification occurred in the *Gracilicutes*, with the recruitment of
266 SufA from the family of A-type carrier proteins, and the possible *de novo* origin of
267 SufE as an alternative sulfur acceptor (consistent with an almost mutually exclusive
268 pattern with SufU).

269 Subsequent independent multiple horizontal gene transfers occurred both intra and
270 inter domains. More specifically, archaeal SMS systems were transferred to a few
271 anaerobic bacteria, while bacterial SUF systems were transferred to a few aerobic
272 archaea. These transfers largely involved replacement of the native system.

273 ISC emerged within the *Gracilicutes* by the following steps: (i) probable duplication of
274 MisU and MisS to give rise to IscU and IscS; (ii) recruitment of DnaK and DnaJ
275 chaperones to give rise to HscA and HscB; (iii) recruitment of a ferredoxin (Fdx); (iii)
276 *de novo* emergence of a frataxin (CyaY) and IscX; (iv) addition of an A-type carrier
277 (IscA), possibly from several duplications or transfers of SufA. The ISC system was
278 mostly vertically inherited and lost many times independently in the *Gracilicutes*.
279 NIF emerged in the *Proteobacteria* probably by integrating components of the MIS
280 system (MisS and MisU) to give rise to NifS and NifU, and an A-type carrier (IscAnif).
281 These components were embedded into the large NIF operon involved in
282 nitrogenase synthesis and maturation. Like ISC, the NIF system was vertically
283 inherited but massively lost many times independently in *Proteobacteria*.
284 Finally, the NIF and SUF systems of *Cyanobacteria* were acquired early in this
285 phylum by transfer from *Proteobacteria*.

286

287 **Discussion**

288 Our analyses suggest that, despite thriving in stable and favourable environmental
289 conditions (i.e., in absence of oxygen and abundance of iron and sulfur), the LUCA
290 already harboured machineries to carry out Fe-S cluster biogenesis. These results
291 challenge previous hypotheses that stated an emergence of Fe-S cluster biogenesis
292 machineries in response to Earth oxygenation^{8,6,9}. In addition, they infer their
293 presence already in the LUCA and not an emergence in the Archaea as recently
294 proposed¹¹. Moreover, our large taxonomic sampling and integrated phylogenomic
295 analyses provide a new framework for the origin and subsequent evolution of Fe-S
296 cluster biogenesis systems across the two prokaryotic domains. Given the bacterial
297 origin of Fe-S cluster biogenesis machineries in eukaryotic chloroplasts and
298 mitochondria⁴, our results relate *de facto* to the whole diversity of life.

299 We infer that LUCA possessed at least one cysteine desulfurase MisS and two
300 scaffold systems, the single component MisU and the two-components SmsCB. The
301 fact that the ancestral SMS system was composed of a simple SmsCB scaffold
302 raises the question of what might have been the source of sulfur. One possibility is
303 that environmental exogenous sulfide was used as sulfur source instead of L-
304 cysteine, as proposed in the archaeon *Methanococcus maripaludis*¹⁶. However, the
305 SMS system is found in several present-day species thriving under anaerobic, non-
306 sulfide-rich environments. An alternative hypothesis is that MIS would have provided

307 sulfur to SMS through its cysteine desulfurase MisS. Since contemporary sulfur
308 acceptors such as SufE or CsdE share structural features with U-type scaffolds,
309 MisU could therefore have acted as a sulfur acceptor and SmsCB as the scaffold,
310 forming an ancestral common system in the LUCA. Such type of cooperation
311 between sulfur donor and scaffold systems that are not contiguous in the genomes
312 has been indeed observed between *csdAE* and *sufBCD* in *E. coli*¹⁷. This functional
313 hypothesis may be consistent with the observation that acquisition of SUF systems
314 by horizontal transfer coincided with the loss of both original MIS and SMS systems.
315 Interestingly, such combination of MIS and SMS was largely retained in the Archaea,
316 possibly linked to their more limited range of lifestyles (i.e., strictly anaerobic) with
317 respect to Bacteria. Further experimental work on archaeal SMS and MIS will
318 therefore provide key information to test these functional hypotheses, which in turn
319 will allow to understand the mechanisms of Fe-S cluster biosynthesis in the LUCA.
320 During the diversification of Bacteria, these ancestral systems underwent progressive
321 complexification through the integration of new components, leading to contemporary
322 SUF, ISC and NIF machineries. Concerning SUF, the emergence of additional
323 components may have served to increase efficiency for iron and sulfur capture, cope
324 with reactive oxygen species (ROS), and possibly mitigate sulfur derivative toxicity in
325 the presence of increasing levels of oxygen. However, in contrast with previous
326 hypotheses¹¹, we inferred that these complexifications occurred way before the GOE,
327 pointing to an early need to cope with traces of oxygen, also in agreement with the
328 ancestral inference of several oxygen-related enzymes in Bacteria^{18,19,20}. Horizontal
329 transfers of these specialized SUF systems then would have contributed to the
330 adaptation to aerobic niches in archaea. In this view, we hypothesize that the
331 acquisition of a SUF from *Proteobacteria* early in the evolution of *Cyanobacteria*
332 brought increased oxygen tolerance in this phylum, thereby favouring the very
333 emergence of oxygenic photosynthesis.

334 More recently in the evolution of Bacteria, tinkering through the recruitment of cellular
335 components from other processes, such as DnaKJ chaperones and ferredoxins,
336 likely enhanced the efficiency of biogenesis and distribution of Fe-S clusters
337 potentially under increasingly oxygenated conditions, giving rise to ISC. Finally,
338 cooperation and physical association with other components led to the highly
339 specialised NIF system, dedicated to the maturation of nitrogenase, a complex
340 enzyme containing multiple metal cofactors and which is extremely sensitive to O₂.

341 It may be asked if environmental adaptations are the only drivers of the variability of
342 Fe-S cluster biogenesis systems, their evolution and distribution. For example, while
343 the presence of SUF versus SMS is clearly linked to higher oxygen tolerance in both
344 bacteria and archaea, the distribution of SUF vs ISC in *Proteobacteria* does not seem
345 to be directly linked to a given ecological niche. It may therefore also be important to
346 assess the impact of enrichment and diversification in the client Fe-S cluster
347 proteome. An integrated view including both the evolution of the Fe-S cluster
348 machineries and the set of cellular client proteins is necessary to obtain a full
349 understanding of the co-evolution dynamics between both partners. Large-scale
350 bioinformatics surveys, combined with ecological considerations and functional tests
351 in a wider range of experimental models, is a powerful approach to reconstruct the
352 evolutionary history of lineages and their fundamental cellular machineries.
353 In conclusion, our results pave the way for future studies aimed at fully revealing Fe-
354 S biology as a key process for ancient and modern life.

355 **Material and Methods**

356 *Dataset assembly*

357 We assembled a database of prokaryotic proteomes gathered from NCBI (Table S1).
358 2,143 archaeal genomes and 182,122 bacterial genomes were available (November
359 2019). The genomes missing .faa files were annotated using Prodigal v2.6.3 (default
360 parameters)²¹. The redundancy between strains was eliminated by keeping one
361 genome per taxonomic ID, leading to 1,268 genomes for Archaea and 61,512
362 genomes for Bacteria. For Bacteria, we performed two additional steps of selection to
363 reduce the number of genomes. In the first step, we made a whole-genome
364 comparison using FastANI v1.0 (kmer=16, frag_length=3,000)²² and clustered the
365 genomes using a threshold of 95% identity. At this step, we selected one genome per
366 cluster by considering the number of proteins annotated as reviewed in Uniprot, the
367 representative status of genome (“representative genome”, “reference genome”, “na”),
368 the availability of annotation files (.faa, .gff) in NCBI, and the completeness status
369 (“Complete genome”, “Chromosome”, “Scaffold”, “Contig”). This allowed to reduce to
370 18,407 bacterial genomes. In the second step, genomes were classified into phyla
371 (genomes with no phylum annotation on NCBI taxonomy were placed as a separate
372 group). Then, we further subsampled the genomes in each phylum by performing
373 clustering based on the taxonomic marker RpoB and CDHIT v4.8.1 (default
374 parameters, percentage identity = 95%) and selecting one strain per cluster²³. This
375 led to a database of 9,596 representative bacterial proteomes. The proteome of
376 *Azotobacter vinelandii* was added manually because it is an important model for the
377 NIF system.

378 A search for homologues was performed, starting with a set of 26 prokaryotic
379 components of all Fe-S cluster biogenesis machineries (SUF, ISC, NIF). First, a
380 similarity search was done using BLASTP v2.8.1+²⁴ in an iterative way, using an e-
381 value threshold of $1e^{-4}$. The intermediate seeds for iterative BLASTP were selected
382 among the sequences with an e-value lower than the threshold. Sequences were
383 aligned using MAFFT v7.419 (auto option)²⁵. The alignment was then used to build
384 an HMM profile using HMMER v3.2.1²⁶ and to perform an HMM search on the
385 proteome database and hits with an e-value lower than 0.01 were kept. Sequences
386 selected by BLASTP and HMM were aligned using MAFFT v7.419 (auto option).
387 Alignments were manually curated to eliminate highly divergent sequences, and were

388 used to infer preliminary trees using FASTTREE v2.1.10 (LG+G4)²⁷, with and without
389 trimming using BMGE v1.12 (BLOSUM30)²⁸. Orthologue subfamilies were delineated
390 manually based on preliminary trees, genomic context, taxonomic distribution,
391 domain composition and length of sequences. More precisely, for each family, tree
392 has been inferred using FASTTREE v2.1.10 (LG+G4)²⁷, with and without trimming
393 using BMGE v1.12 (BLOSUM30)²⁸ and the different data have been mapped on it
394 (taxonomy, presence and vicinity of each other families involved in Fe-S cluster
395 biogenesis, domain composition and length of sequences). For huge multigenic
396 families, subtrees have been inferred to increase the phylogenetic signal in order to
397 better delineate subfamilies. For some cases, trees with a more restricted taxonomic
398 sampling have been inferred to increase the phylogenetic signal. The delineation of
399 subfamilies has been made considering all the different data (Supplementary data,
400 trees available in <https://itol.embl.de/shared/Pietrochki>).

401 To define each system in the taxonomic distribution, it should at least include a
402 scaffold protein and another component whose genes are neighbours in the
403 genomes: at least SufC and SufB or SufD for SUF; SmsB and SmsC for SMS; at
404 least IscU and HscA, HscB, Fdx, ATC-II (IscA), IscS or IscX for ISC; at least NifU and
405 NifS or IscA_{nif} in the NIF operon for NIF; MisU and MisS for MIS.

406

407 *Phylogenetic inference*

408 A representative subsample of strains was selected for tree inference. For archaeal
409 and bacterial reference phylogenies and for SUF/SMS and MIS phylogenies, up to
410 four genomes *per* major group were selected for Archaea and up to five genomes *per*
411 phylum for Bacteria (150 Archaea, 401 Bacteria, Table S2). For the detailed SUF
412 phylogeny of *Cyanobacteria*, we kept this sampling and enriched the *Cyanobacteria*
413 with one strain *per* species (598 genomes). For ISC, we enriched some phyla: one
414 strain *per* genus was selected for *Proteobacteria* and a larger selection for other
415 clades (947 genomes). For NIF, we selected one strain *per* genus (2,173 genomes).
416 For the *Gammaproteobacteria* reference tree, we selected two strains *per* family (137
417 genomes). After taxonomic sampling, sequences of each protein family were aligned
418 using MAFFT (-linsi option) and trimmed using BMGE (entropy threshold =0.95,
419 minimum length = 1, matrix = BLOSUM30).

420 Given the difficulty to test congruence on single gene trees with very large taxonomic
421 sampling²⁹, we chose the components to concatenate based on the following criteria:

422 (i) presenting a conserved synteny (ii) a co-occurrence; and (iii) an overall similar
423 topology of the single-gene tree upon visual inspection. For SUF/SMS, a
424 concatenation of three trimmed alignments was performed (SmsB + SufB, SmsB+
425 SufD and SmsC + SufC), keeping one species where a minimum of two markers
426 were present. This provided a supermatrix gathering 328 sequences and 919 amino
427 acid positions. For ISC, IscX, IscS, IscU, CyaY, Fdx, HscA and HscB, the trimmed
428 alignments were concatenated (keeping taxa with five markers minimum) providing a
429 supermatrix including 178 sequences and 1530 aa positions. For NIF, a
430 concatenation of NifS and NifU was assembled, using MisS and MisU as external
431 group (taxa with two markers minimum, 169 sequences and 656 positions). For each
432 supermatrix, we kept only one system per genome, favouring those closest to its
433 taxonomic group in individual gene trees (Table S3).

434 The bacterial species tree was inferred from a concatenation of RpoB, RpoC and If2
435 trimmed alignments (MAFFT+BMGE) leading to a supermatrix of 401 sequences and
436 2,574 aa positions. The archaeal species tree was inferred from a concatenation of
437 30 ribosomal proteins (rplA, rplB, rplC, rplD, rplE, rplF, rplK, rplM, rplN, rplO, rplR,
438 rplV, rplW, rplX, rpmC, rpsB, rpsC, rpsD, rpsE, rpsG, rpsH, rpsI, rpsJ, rpsK, rpsL,
439 rpsM, rpsN, rpsO, rpsQ, rpsS), RpoB and If2 trimmed alignments (MAFFT+BMGE)
440 leading to a supermatrix of 150 sequences and 5,978 aa positions, in a similar way
441 as previously described³⁰. The bacterial and archaeal trees were rooted as
442 previously described^{31,32}. The reference tree of *Gammaproteobacteria* was inferred
443 from a concatenation of RpoB, RpoC and If2 trimmed alignments (MAFFT+BMGE)
444 leading to a supermatrix of 137 sequences and 3,447 aa positions.

445 All trees were inferred using IQ-TREE v1.6.10³³ with the best-suited model selected
446 by the BIC criterium (Table S5). We used the PMSF (posterior mean site frequency)
447 option with 60 mixture categories³⁴, starting from phylogenies inferred with
448 homogeneous models. The robustness of the branches was assessed using non-
449 parametric bootstrap on 100 replicates of the original data. The genomic context was
450 visualized using GeneSpy v1.2³⁵ and the figures were generated using iTOL³⁶.

451

452 *Bacterial strains and plasmid construction*

453 The *E. coli* strains, plasmids, and oligonucleotides used in this study are listed in
454 Table S6. Bacterial strains were routinely grown in aeration at 37°C in Luria-Bertani

455 broth (LB) and solid media contained 1.5% agar. When required, ampicillin or
456 kanamycin were added at the concentration of 100 µg/ml and 50 µg/ml, respectively.
457 The nucleotide sequences were codon-optimized for expression in *E. coli* and genes
458 have been ordered from Twist Bioscience (TWB) (Table S7). For purification
459 experiments, the sequences of *smsCB* genes of *M. acetivorans* and *M. jannaschii*
460 were cloned in the pET6His-Tev vector³⁷. These constructions enabled the
461 production of both N-terminal 6his-tagged SmsC and untagged SmsB proteins. For
462 two-hybrid experiments, the sequences of SmsC and SmsB genes of *M. jannaschii*
463 were cloned in the pT18 and pT25 plasmids³⁸. Finally, all the plasmids were
464 sequenced using oligonucleotides T7 and T7term for pET vectors, or UT18 and UT25
465 oligonucleotides for two-hybrid vectors.

466

467 *Expression and purification of SMS (SmsCB) from Methanococcus jannaschii and* 468 *Methanosarcina acetivorans*

469 Each SmsCB was expressed under aerobic conditions in Luria Broth medium. Each
470 SmsCB was expressed under aerobic conditions in Luria Broth medium. Expression
471 was carried out in MG1655(DE3) for 3 hours at 37°C under 0.5 mM IPTG induction.
472 Harvested cells were resuspended anaerobically inside a glove box (Jacomex, O₂ <
473 2 ppm) in lysis buffer (100 mM Tris pH 8, 150 mM NaCl containing a protease
474 inhibitor cocktail (cOmplete Ultra tablets), sonicated (12 min, 10/30 sec ON/OFF,
475 power 50%) and resulting lysates were clarified by ultracentrifugation (40,000 g, 20
476 min, 4°C). Supernatants were loaded anaerobically onto a 5 mL Ni-NTA column.
477 Column was washed using washing buffer (100 mM Tris pH 8, 150 mM NaCl, 10 mM
478 imidazole) and proteins were eluted with washing buffer containing 400 mM
479 imidazole. Fractions containing proteins (SmsC and SmsB) were desalted to remove
480 imidazole, using NAP-25 columns equilibrated with desalting buffer (100 mM Tris-HCl
481 pH 8, 150 mM NaCl buffer), and were analyzed by UV-visible absorption
482 spectroscopy and for their iron and sulfur content^{39,40}.

483

484 *In vitro Fe-S cluster transfer to Aconitase*

485 To assess aconitase activation, apo-AcnB from *E. coli* (0.5 nmol) was pretreated with
486 DTT (desalting using a Microbiospin column (Biorad)) and then incubated
487 anaerobically in a glove box at 18 °C in aconitase buffer (50 mM Tris-HCl pH: 7.6)
488 with a 1.5-fold molar excess of the native SmsCB proteins from *M. acetivorans* (4

489 iron and 3.8 sulfur /SmsC2B2) or reconstituted SufCB from *M. jannaschii* (3.7 iron
490 and 3.5 sulfur /SmsC2B2) in order to provide 4 iron and 4 sulfur atoms per AcnB.
491 After 5, 15, and 30 min incubation, aconitase activity was assessed as described in
492 ⁴¹. Briefly, AcnB-SmsCB mixtures were added to 0.6 mM MnCl₂, 25 mM citrate, 0.5 U
493 isocitric dehydrogenase, 0.25 mM NADP⁺, 50 mM Tris-HCl, pH 7.6, in a 100 μL final
494 volume and NADPH formation was monitored at 340 nm by UV-vis. absorption
495 spectroscopy. The 100 % activity corresponds to the activity of the chemically
496 reconstituted AcnB prepared by incubating apo-AcnB with 5 molar excess of ferrous
497 iron and sulfur for 30 min in the presence of 5 mM DTT (2 μmol min⁻¹ mg⁻¹).

498

499 *UV-visible absorption spectroscopy*

500 UV-visible spectra were recorded with a Uvikon XL (Bio-Tek instrument)
501 spectrophotometer connected to the glove box with optical fibers.

502

503 *Bacterial two-hybrid*

504 We used the adenylate cyclase-based two-hybrid technique⁴² to test pairwise
505 interactions between SmsC and SmsB proteins of *M. jannaschii*. After co-
506 transformation of the BTH101 strain with the two plasmids expressing the fusions, LB
507 selection plates containing ampicillin and kanamycin were incubated at 30°C for 2
508 days. Then, overnight cultures in LB medium supplemented with ampicillin,
509 kanamycin and 0.5 mM IPTG were spotted on MacConkey plates containing 1%
510 maltose.

511

512 **Data Availability**

513 Supplementary data are available at
514 [https://figshare.com/articles/dataset/Garcia_et_al_2022_Supplementary_data/20209](https://figshare.com/articles/dataset/Garcia_et_al_2022_Supplementary_data/20209877)
515 877

516

517 **Acknowledgements**

518 We thank members of the SAME and of the EBMC units for discussion. This project
519 was supported by Institut Pasteur and CNRS and by grants from the ANR-10-LABX-
520 62-IBEID, the LabEx ARCANE (ANR-11-LABX-0003-01), and the CBH- EUR-GS
521 (ANR-17-EURE-0003).

522

523 **Author Contributions Statement**

524 P.S.G, F.D., E.B., S.O.d.C., S.G., F.B. designed the experiment. P.S.G, F.D.,
525 S.O.d.C., M.D. performed the experiments. P.S.G, F.D., S.O.d.C., E.B., S.G., F.B.
526 analysed the data. P.S.G, S.G., F.B. wrote the manuscript. P.S.G, S.G., F.B.
527 revised the manuscript.

528

529 **Competing Interests Statement**

530 The authors declare no competing interests.

531

532 **Figure Legends**

533 **Figure 1. A:** Main known steps of Fe-S cluster biogenesis. **B:** Schematic
534 representation of families of homologues involved in the steps of Fe-S cluster
535 synthesis depicted in A. The name of the superfamily/conserved domain is indicated
536 on the top. Each shape corresponds to a separate group based on phylogenies and
537 genomic context. Colours correspond to the function of proteins as in 1A, and grey
538 correspond to proteins/genes with additional minor roles (IscX, CyaY, SufT, SufU).
539 The fillings of the different clades indicate their belonging to a specific system (SUF:
540 unordered squares; ISC: slanted hatchings; NIF: ordered squares; SMS: quartering;
541 MIS: dots; SUF and ISC: fully coloured; No specific system: empty). The empty light
542 grey shapes represent sequences that could not be attributed to a specific family
543 because they lack conserved synteny or belong to paralogue families. The MIS and
544 SMS components are in bold. **C:** Genomic clusters of three known systems for Fe-S
545 cluster synthesis in Bacteria (ISC: *Escherichia coli*, NIF: *Azotobacter vinelandii* and
546 SUF: *Escherichia coli*), and the two newly identified MIS and SMS systems. Colours
547 correspond to the different biochemical roles as in A.

548

549 **Figure 2.** SMS is a *bona fide* Fe-S cluster biogenesis system. **A.** Bacterial two-hybrid
550 on *M. jannaschii* SmsCB proteins. Cultures of BTH101 strain transformed by pairs of
551 the 2-hybrid pT18 and T25 fusions with SmsC and SmsB genes from *M. jannaschii*
552 were dropped on MacConkey 1% maltose plate and incubated 2 days at 30°C. **B.**
553 SDS-PAGE of anaerobically purified *M. acetivorans* (left) and *M. jannaschii* (right). **C.**
554 UV-visible spectroscopy analysis of anaerobically purified *M. acetivorans* SmsCB
555 (1.6 mg/ml) in 100 mM Tris-HCL, 150 mM NaCl pH8. Iron and sulfur content: $4.05 \pm$
556 0.02 Fe and 3.8 ± 0.02 S / SmsC₂B₂. **D.** UV-visible spectroscopy analysis of

557 anaerobically purified *M. jannaschii* SmsCB (10 mg/ml) in 100 mM Tris-HCL, 150 mM
558 NaCl pH8. Iron and sulfur content: 0.15 ± 0.02 Fe and 0.13 ± 0.03 S/ SmsC₂B₂. **E, F.**
559 Iron-sulfur cluster transfer from SmsCB to aconitase B (AcnB), *M. acetivorans* and *M.*
560 *jannaschii* respectively. N = 3 independent experiments. Data are presented as mean
561 values +/- Standard Deviation. Apo AcnB (0.5 nmol) was incubated anaerobically
562 with as-purified SmsCB from *M. acetivorans* or reconstituted SmsCB from *M.*
563 *jannaschii*. Aconitase activity was measured at different time points as indicated
564 under the graph. Apo-AcnB (0.5 nmol) was incubated anaerobically with 1.5
565 equivalent of SmsCB in 50 mM Tris-HCl pH 7.6. The reaction mixture comprised 0.6
566 mM MnCl₂, 25 mM citrate, 0.5 U isocitrate dehydrogenase and 0.25 mM NADP⁺ was
567 added to the protein mixture in a final volume of 100 μ l. Aconitase activity was
568 measured by monitoring the absorption at 340 nm. As a positive control,
569 reconstituted AcnB was incubated 30 min. Reconstituted AcnB (Rec-AcnB) was
570 assayed. Rec-AcnB was obtained by incubating Apo-AcnB for 30 min. with a 5-molar
571 excess of iron and sulfide in the presence of 5 mM DTT.

572

573 **Figure 3.** Taxonomic distribution of the five different Fe-S cluster biogenesis systems
574 on the reference phylogenies (cladograms) of Bacteria (left panel) and Archaea (right
575 panel). Bootstrap values > 80% are indicated by dots on branches. Colour of leaves
576 corresponds to major high-ranking clades in Bacteria (*Terrabacteria*: marine blue,
577 *Gracilicutes*: light blue, CPR: dark blue, *Fusobacteria*: grey blue) and in Archaea
578 (Cluster I: orange and II: red). The presence of a system in a genome was inferred if
579 at least the corresponding scaffold and another component were present. Taxonomic
580 distribution: for each clade/phylum, the percentage of genomes possessing each
581 system is indicated by dots of different sizes in front of each leaf. The Venn diagrams
582 below each tree show the abundance and overlap of the different systems in
583 genomes across the different clades/phyla. A more detailed mapping of all the
584 components of each system in Bacteria and Archaea is presented in Extended Data
585 figures 2 and 3, respectively.

586

587 **Figure 4.** Phylogenetic tree of SUF/SMS and MIS/NIF systems. **A:** Phylogeny
588 (Cladogram) of SUF/SMS based on concatenation of SufB/SmsB, SufC/SmsC, and
589 SufD/SmsB. **B:** Phylogeny (Cladogram) of the MIS/NIF systems based on MisS/NifS.
590 Colours of branches correspond to Archaea (red) and Bacteria (blue). The central red

591 dot corresponds to the LUCA. The genomic context of the genes coding for SUF and
592 SMS components is indicated in front of the leaves, centred around SufB and SmsB
593 (brown and light brown, respectively). ATC-II (SufA): pink, SufC: yellow, SufD: dark
594 green, SufS: light blue, SufE: turquoise SufT: dark blue, SufU: violet, SmsC: light
595 green). The genomic context of genes coding for MIS/NIF components is centred
596 around MisS/NifS (red and green, respectively). MisU: blue, NifU: dark blue. The two
597 framed excerpts show the placement of the cyanobacterial Fe-S synthesis systems
598 SUF and NIF. For detailed trees, see Extended Data figures 5 and 9.

599

600 **Figure 5.** Inferred scenario for the origin and evolution of Fe-S synthesis
601 machineries, with the reconstruction of ancestral systems and major evolutionary
602 events. See text for details.

603

604 **Bibliography**

605

- 606 1. Fontecave M. Iron-sulfur clusters: Ever-expanding roles. *Nat Chem Biol*.
607 Published online 2006. doi:10.1038/nchembio0406-171
- 608 2. Roche B, Aussel L, Ezraty B, Mandin P, Py B, Barras F. Reprint of: Iron/sulfur
609 proteins biogenesis in prokaryotes: Formation, regulation and diversity.
610 *Biochim Biophys Acta - Bioenerg*. Published online 2013.
611 doi:10.1016/j.bbabi.2013.05.001
- 612 3. Braymer JJ, Freibert SA, Rakwalska-Bange M, Lill R. Mechanistic concepts of
613 iron-sulfur protein biogenesis in Biology. *Biochim Biophys Acta - Mol Cell Res*.
614 Published online 2021. doi:10.1016/j.bbamcr.2020.118863
- 615 4. Tsaousis AD. On the Origin of Iron/Sulfur Cluster Biosynthesis in Eukaryotes.
616 *Front Microbiol*. Published online 2019. doi:10.3389/fmicb.2019.02478
- 617 5. Py B, Barras F. Building Feg-S proteins: Bacterial strategies. *Nat Rev*
618 *Microbiol*. Published online 2010. doi:10.1038/nrmicro2356
- 619 6. Gao F. Iron–Sulfur Cluster Biogenesis and Iron Homeostasis in Cyanobacteria.
620 *Front Microbiol*. Published online 2020. doi:10.3389/fmicb.2020.00165
- 621 7. Wächtershäuser G. Groundworks for an evolutionary biochemistry: The iron-
622 sulphur world. *Prog Biophys Mol Biol*. Published online 1992.
623 doi:10.1016/0079-6107(92)90022-X
- 624 8. Imlay JA. Iron-sulphur clusters and the problem with oxygen. *Mol Microbiol*.

- 625 Published online 2006. doi:10.1111/j.1365-2958.2006.05028.x
- 626 9. Andreini C, Rosato A, Banci L. The relationship between environmental
627 dioxygen and iron-sulfur proteins explored at the genome level. *PLoS One*.
628 Published online 2017. doi:10.1371/journal.pone.0171279
- 629 10. Tokumoto U, Kitamura S, Fukuyama K, Takahashi Y. Interchangeability and
630 distinct properties of bacterial Fe-S cluster assembly systems: Functional
631 replacement of the *isc* and *suf* operons in *Escherichia coli* with the *nifSU*-like
632 operon from *Helicobacter pylori*. *J Biochem*. Published online 2004.
633 doi:10.1093/jb/mvh104
- 634 11. Boyd ES, Thomas KM, Dai Y, Boyd JM, Outten FW. Interplay between Oxygen
635 and Fe-S Cluster Biogenesis: Insights from the *Suf* Pathway. *Biochemistry*.
636 Published online 2014. doi:10.1021/bi500488r
- 637 12. Deere TM, Prakash D, Lessner FH, Duin EC, Lessner DJ. *Methanosarcina*
638 *acetivorans* contains a functional ISC system for iron-sulfur cluster biogenesis.
639 *BMC Microbiol*. Published online 2020. doi:10.1186/s12866-020-02014-z
- 640 13. Olson JW, Agar JN, Johnson MK, Maier RJ. Characterization of the *NifU* and
641 *NifS* Fe-S cluster formation proteins essential for viability in *Helicobacter pylori*.
642 *Biochemistry*. Published online 2000. doi:10.1021/bi001744s
- 643 14. Hug LA, Baker BJ, Anantharaman K, et al. A new view of the tree of life. *Nat*
644 *Microbiol*. Published online 2016. doi:10.1038/nmicrobiol.2016.48
- 645 15. Dombrowski N, Lee JH, Williams TA, Offre P, Spang A. Genomic diversity,
646 lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol Lett*.
647 Published online 2019. doi:10.1093/femsle/fnz008
- 648 16. Liu Y, Sieprawska-Lupa M, Whitman WB, White RH. Cysteine is not the sulfur
649 source for iron-sulfur cluster and methionine biosynthesis in the methanogenic
650 archaeon *Methanococcus maripaludis*. *J Biol Chem*. Published online 2010.
651 doi:10.1074/jbc.M110.152447
- 652 17. Trotter V, Vinella D, Loiseau L, De Choudens SO, Fontecave M, Barras F. The
653 *CsdA* cysteine desulphurase promotes Fe/S biogenesis by recruiting *Suf*
654 components and participates to a new sulphur transfer pathway by recruiting
655 *CsdL* (ex-*YgdL*), a ubiquitin-modifying-like protein. *Mol Microbiol*. Published
656 online 2009. doi:10.1111/j.1365-2958.2009.06954.x
- 657 18. Gribaldo S, Talla E, Brochier-Armanet C. Evolution of the haem copper
658 oxidases superfamily: a rooting tale. *Trends Biochem Sci*. Published online

- 659 2009. doi:10.1016/j.tibs.2009.04.002
- 660 19. Brochier-Armanet C, Talla E, Gribaldo S. The multiple evolutionary histories of
661 dioxygen reductases: Implications for the origin and evolution of aerobic
662 respiration. *Mol Biol Evol*. Published online 2009. doi:10.1093/molbev/msn246
- 663 20. Jabłońska J, Tawfik DS. The evolution of oxygen-utilizing enzymes suggests
664 early biosphere oxygenation. *Nat Ecol Evol*. Published online 2021.
665 doi:10.1038/s41559-020-01386-9
- 666 21. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal:
667 prokaryotic gene recognition and translation initiation site identification. *BMC*
668 *Bioinformatics*. 2010;11(1):119. doi:10.1186/1471-2105-11-119
- 669 22. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High
670 throughput ANI analysis of 90K prokaryotic genomes reveals clear species
671 boundaries. *Nat Commun*. Published online 2018. doi:10.1038/s41467-018-
672 07641-9
- 673 23. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-
674 generation sequencing data. *Bioinformatics*. Published online 2012.
675 doi:10.1093/bioinformatics/bts565
- 676 24. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a
677 new generation of protein database search programs. *Nucleic Acids Res*.
678 1997;25(17):3389-3402. Accessed August 3, 2018.
679 <http://www.ncbi.nlm.nih.gov/pubmed/9254694>
- 680 25. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version
681 7: Improvements in Performance and Usability. *Mol Biol Evol*. 2013;30(4):772-
682 780. doi:10.1093/molbev/mst010
- 683 26. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol*.
684 2011;7(10):e1002195. doi:10.1371/journal.pcbi.1002195
- 685 27. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-
686 Likelihood Trees for Large Alignments. Poon AFY, ed. *PLoS One*.
687 2010;5(3):e9490. doi:10.1371/journal.pone.0009490
- 688 28. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a
689 new software for selection of phylogenetic informative regions from multiple
690 sequence alignments. *BMC Evol Biol*. 2010;10(1):210. doi:10.1186/1471-2148-
691 10-210
- 692 29. Leigh JW, Schliep K, Lopez P, Baptiste E. Let them fall where they may:

- 693 Congruence analysis in massive phylogenetically messy data sets. *Mol Biol*
694 *Evol.* Published online 2011. doi:10.1093/molbev/msr110
- 695 30. Adam PS, Borrel G, Gribaldo S. Evolutionary history of carbon monoxide
696 dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes.
697 *Proc Natl Acad Sci U S A.* Published online 2018.
698 doi:10.1073/pnas.1716667115
- 699 31. Coleman GA, Davín AA, Mahendrarajah TA, et al. A rooted phylogeny resolves
700 early bacterial evolution. *Science (80-)*. Published online 2021.
701 doi:10.1126/science.abe0511
- 702 32. Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is
703 linked to a new root for the Archaea. *Proc Natl Acad Sci U S A.* Published
704 online 2015. doi:10.1073/pnas.1420858112
- 705 33. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and
706 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.
707 *Mol Biol Evol.* 2015;32(1):268-274. doi:10.1093/molbev/msu300
- 708 34. Wang HC, Minh BQ, Susko E, Roger AJ. Modeling Site Heterogeneity with
709 Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic
710 Estimation. *Syst Biol.* Published online 2018. doi:10.1093/sysbio/syx068
- 711 35. Garcia PS, Jauffrit F, Grangeasse C, Brochier-Armanet C. GeneSpy, a user-
712 friendly and flexible genomic context visualizer. *Bioinformatics.* Published
713 online 2019. doi:10.1093/bioinformatics/bty459
- 714 36. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new
715 developments. *Nucleic Acids Res.* Published online 2019.
716 doi:10.1093/nar/gkz239
- 717 37. Wahl A, My L, Dumoulin R, Sturgis JN, Bouveret E. Antagonistic regulation of
718 dgkA and plsB genes of phospholipid synthesis by multiple stress responses in
719 *Escherichia coli*. *Mol Microbiol.* Published online 2011. doi:10.1111/j.1365-
720 2958.2011.07641.x
- 721 38. Gully D, Bouveret E. A protein network for phospholipid synthesis uncovered
722 by a variant of the tandem affinity purification method in *Escherichia coli*.
723 *Proteomics.* Published online 2006. doi:10.1002/pmic.200500115
- 724 39. fish WW. Rapid Colorimetric Micromethod for the Quantitation of Complexed
725 Iron in Biological Samples. *Methods Enzymol.* Published online 1988.
726 doi:10.1016/0076-6879(88)58067-9

- 727 40. Beinert H. Semi-micro methods for analysis of labile sulfide and of labile sulfide
728 plus sulfane sulfur in unusually stable iron-sulfur proteins. *Anal Biochem*.
729 Published online 1983. doi:10.1016/0003-2697(83)90186-0
- 730 41. Gardner PR, Fridovich I. Inactivation-reactivation of aconitase in *Escherichia*
731 *coli*. A sensitive measure of superoxide radical. *J Biol Chem*. Published online
732 1992. doi:10.1016/s0021-9258(19)50343-x
- 733 42. Karimova G, Pidoux J, Ullmann A, Ladant D. A bacterial two-hybrid system
734 based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S*
735 *A*. Published online 1998. doi:10.1073/pnas.95.10.5752
- 736