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-03780728v1

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.

An early origin of Iron-Sulfur cluster biosynthesis machineries before Earth oxygenation

Pierre Simon Garcia[1,2], Francesca D'Angelo*[1], Sandrine Ollagnier-de Choudens*[3], Macha Dussouchaud[1], Emmanuelle Bouveret[1], Simonetta Gribaldo[2 #], Frédéric Barras[1#]

[1] Institut Pasteur, Université Paris Cité, CNRS UMR6047, Department of Microbiology, Unit Stress Adaptation and Metabolism in enterobacteria, Paris, France

[2] Institut Pasteur, Université Paris Cité, CNRS UMR6047, Department of Microbiology, Unit Evolutionary Biology of the Microbial Cell, Paris, France

[3] Université Grenoble Alpes, CNRS, CEA, Laboratoire de Chimie et Biologie des Métaux, F-38000 Grenoble, France

for correspondence:

simonetta.gribaldo@pasteur.fr

frederic.barras@pasteur.fr

* These authors contributed equally to this work

Abstract

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

Main

Iron-Sulfur (Fe-S) clusters are ubiquitous and essential protein cofactors that allow redox chemistry and Lewis acid-type catalysis for hundreds of structurally and functionally diverse proteins in a wide array of processes associated with bioenergetics, metabolism, DNA synthesis, and genetic regulation[1]. Studies in model systems, both prokaryotes and eukaryotes, have identified three Fe-S cluster biogenesis machineries[2,3]: ISC, considered as the housekeeping system, SUF, acting as a backup in stress conditions, and NIF, specialized in the maturation of nitrogenase in N_2 fixing bacteria[2,3]. Eukaryotic Fe-S cluster assembly machineries (ISC in mitochondria and SUF in photosynthetic lineages) are of endosymbiotic origin from *Alphaproteobacteria* and *Cyanobacteria*, respectively[4]. Although made of different components, these three systems facilitate Fe-S cluster biogenesis following the same basic principles (Figure 1A): a cysteine desulfurase takes sulfur from L-cysteine and transfers it as a persulfide onto a scaffold protein, which also receives Fe^{2+} and electrons to build an Fe-S cluster that is then transferred to cellular apo-protein targets via carrier proteins[2,5].

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

To date only two studies have tackled the issue of the origin and evolution of Fe-S biogenesis machineries but reached contradictory conclusions. An early survey suggested the presence of SUF in the Last Universal Common Ancestor (LUCA) based on the presence of homologues of SufBC in both Archaea and Bacteria[10]. In contrast, another study proposed that SUF originated in the Archaea, followed by massive horizontal transfers to Bacteria based on a SufBC phylogeny[11]. These contrasting results derive probably by the limited taxonomic sampling used and the lack of extensive phylogenomic analysis.

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

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

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

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

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

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

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

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

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

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

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

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

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

Our results establish the ubiquity of Fe-S cluster biogenesis across prokaryotes and pose the question of what evolutionary processes shaped their current distribution.

We therefore analysed the phylogeny of each of the five systems (Methods). To increase the phylogenetic signal, we concatenated the homologous components shared by the SUF and SMS systems (SufBCD, and SmsBC, respectively). The

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

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

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

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

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

Horizontal gene transfers drove major adaptations in Archaea and Bacteria

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

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

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

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

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

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

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

ISC emerged within the *Gracilicutes* by the following steps: (i) probable duplication of MisU and MisS to give rise to IscU and IscS; (ii) recruitment of DnaK and DnaJ chaperones to give rise to HscA and HscB; (iii) recruitment of a ferredoxin (Fdx); (iii) *de novo* emergence of a frataxin (CyaY) and IscX; (iv) addition of an A-type carrier (IscA), possibly from several duplications or transfers of SufA. The ISC system was mostly vertically inherited and lost many times independently in the *Gracilicutes*.

NIF emerged in the *Proteobacteria* probably by integrating components of the MIS system (MisS and MisU) to give rise to NifS and NIfU, and an A-type carrier (IscAnif). These components were embedded into the large NIF operon involved in nitrogenase synthesis and maturation. Like ISC, the NIF system was vertically inherited but massively lost many times independently in *Proteobacteria*.

Finally, the NIF and SUF systems of *Cyanobacteria* were acquired early in this phylum by transfer from *Proteobacteria*.

Discussion

Our analyses suggest that, despite thriving in stable and favourable environmental conditions (i.e., in absence of oxygen and abundance of iron and sulfur), the LUCA already harboured machineries to carry out Fe-S cluster biogenesis. These results challenge previous hypotheses that stated an emergence of Fe-S cluster biogenesis machineries in response to Earth oxygenation[8,6,9]. In addition, they infer their presence already in the LUCA and not an emergence in the Archaea as recently proposed[11]. Moreover, our large taxonomic sampling and integrated phylogenomic analyses provide a new framework for the origin and subsequent evolution of Fe-S cluster biogenesis systems across the two prokaryotic domains. Given the bacterial origin of Fe-S cluster biogenesis machineries in eukaryotic chloroplasts and mitochondria[4], our results relate *de facto* to the whole diversity of life.

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

sulfur to SMS through its cysteine desulfurase MisS. Since contemporary sulfur acceptors such as SufE or CsdE share structural features with U-type scaffolds, MisU could therefore have acted as a sulfur acceptor and SmsCB as the scaffold, forming an ancestral common system in the LUCA. Such type of cooperation between sulfur donor and scaffold systems that are not contiguous in the genomes has been indeed observed between *csdAE* and *sufBCD* in *E. coli*[17]. This functional hypothesis may be consistent with the observation that acquisition of SUF systems by horizontal transfer coincided with the loss of both original MIS and SMS systems. Interestingly, such combination of MIS and SMS was largely retained in the Archaea, possibly linked to their more limited range of lifestyles (i.e., strictly anaerobic) with respect to Bacteria. Further experimental work on archaeal SMS and MIS will therefore provide key information to test these functional hypotheses, which in turn will allow to understand the mechanisms of Fe-S cluster biosynthesis in the LUCA.

During the diversification of Bacteria, these ancestral systems underwent progressive complexification through the integration of new components, leading to contemporary SUF, ISC and NIF machineries. Concerning SUF, the emergence of additional components may have served to increase efficiency for iron and sulfur capture, cope with reactive oxygen species (ROS), and possibly mitigate sulfur derivative toxicity in the presence of increasing levels of oxygen. However, in contrast with previous hypotheses[11], we inferred that these complexifications occurred way before the GOE, pointing to an early need to cope with traces of oxygen, also in agreement with the ancestral inference of several oxygen-related enzymes in Bacteria[18,19,20]. Horizontal transfers of these specialized SUF systems then would have contributed to the adaptation to aerobic niches in archaea. In this view, we hypothesize that the acquisition of a SUF from *Proteobacteria* early in the evolution of *Cyanobacteria* brought increased oxygen tolerance in this phylum, thereby favouring the very emergence of oxygenic photosynthesis.

More recently in the evolution of Bacteria, tinkering through the recruitment of cellular components from other processes, such as DnaKJ chaperones and ferredoxins, likely enhanced the efficiency of biogenesis and distribution of Fe-S clusters potentially under increasingly oxygenated conditions, giving rise to ISC. Finally, cooperation and physical association with other components led to the highly specialised NIF system, dedicated to the maturation of nitrogenase, a complex enzyme containing multiple metal cofactors and which is extremely sensitive to O_2.

It may be asked if environmental adaptations are the only drivers of the variability of Fe-S cluster biogenesis systems, their evolution and distribution. For example, while the presence of SUF versus SMS is clearly linked to higher oxygen tolerance in both bacteria and archaea, the distribution of SUF vs ISC in *Proteobacteria* does not seem to be directly linked to a given ecological niche. It may therefore also be important to assess the impact of enrichment and diversification in the client Fe-S cluster proteome. An integrated view including both the evolution of the Fe-S cluster machineries and the set of cellular client proteins is necessary to obtain a full understanding of the co-evolution dynamics between both partners. Large-scale bioinformatics surveys, combined with ecological considerations and functional tests in a wider range of experimental models, is a powerful approach to reconstruct the evolutionary history of lineages and their fundamental cellular machineries.

In conclusion, our results pave the way for future studies aimed at fully revealing Fe-S biology as a key process for ancient and modern life.

Material and Methods

Dataset assembly

We assembled a database of prokaryotic proteomes gathered from NCBI (Table S1). 2,143 archaeal genomes and 182,122 bacterial genomes were available (November 2019). The genomes missing .faa files were annotated using Prodigal v2.6.3 (default parameters)[21]. The redundancy between strains was eliminated by keeping one genome per taxonomic ID, leading to 1,268 genomes for Archaea and 61,512 genomes for Bacteria. For Bacteria, we performed two additional steps of selection to reduce the number of genomes. In the first step, we made a whole-genome comparison using FastANI v1.0 (kmer=16, frag_length=3,000)[22] and clustered the genomes using a threshold of 95% identity. At this step, we selected one genome per cluster by considering the number of proteins annotated as reviewed in Uniprot, the representative status of genome ("representative genome, "reference genome", "na"), the availability of annotation files (.faa, .gff) in NCBI, and the completeness status ("Complete genome", Chromosome", "Scaffold", "Contig"). This allowed to reduce to 18,407 bacterial genomes. In the second step, genomes were classified into phyla (genomes with no phylum annotation on NCBI taxonomy were placed as a separate group). Then, we further subsampled the genomes in each phylum by performing clustering based on the taxonomic marker RpoB and CDHIT v4.8.1 (default parameters, percentage identity = 95%) and selecting one strain per cluster[23]. This led to a database of 9,596 representative bacterial proteomes. The proteome of *Azotobacter vinelandii* was added manually because it is an important model for the NIF system.

A search for homologues was performed, starting with a set of 26 prokaryotic components of all Fe-S cluster biogenesis machineries (SUF, ISC, NIF). First, a similarity search was done using BLASTP v2.8.1+[24] in an iterative way, using an *e*-value threshold of $1e^{-4}$. The intermediate seeds for iterative BLASTP were selected among the sequences with an e-value lower than the threshold. Sequences were aligned using MAFFT v7.419 (auto option)[25]. The alignment was then used to build an HMM profile using HMMER v3.2.1[26] and to perform an HMM search on the proteome database and hits with an e-value lower than 0.01 were kept. Sequences selected by BLASTP and HMM were aligned using MAFFT v7.419 (auto option). Alignments were manually curated to eliminate highly divergent sequences, and were

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

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

Phylogenetic inference

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

Given the difficulty to test congruence on single gene trees with very large taxonomic sampling[29], we chose the components to concatenate based on the following criteria:

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

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

All trees were inferred using IQ-TREE v1.6.10[33] with the best-suited model selected by the BIC criterium (Table S5). We used the PMSF (posterior mean site frequency) option with 60 mixture categories[34], starting from phylogenies inferred with homogeneous models. The robustness of the branches was assessed using non-parametric bootstrap on 100 replicates of the original data. The genomic context was visualized using GeneSpy v1.2[35] and the figures were generated using iTOL [36].

Bacterial strains and plasmid construction

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

broth (LB) and solid media contained 1.5% agar. When required, ampicillin or kanamycin were added at the concentration of 100 µg/ml and 50 µg/ml, respectively. The nucleotide sequences were codon-optimized for expression in *E. coli* and genes have been ordered from Twist Bioscience (TWB) (Table S7). For purification experiments, the sequences of *smsCB* genes of *M. acetivorans* and *M. jannaschii* were cloned in the pET6His-Tev vector[37]. These constructions enabled the production of both N-terminal 6his-tagged SmsC and untagged SmsB proteins. For two-hybrid experiments, the sequences of SmsC and SmsB genes of *M. jannaschii* were cloned in the pT18 and pT25 plasmids[38]. Finally, all the plasmids were sequenced using oligonucleotides T7 and T7term for pET vectors, or UT18 and UT25 oligonucleotides for two-hybrid vectors.

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

Each SmsCB was expressed under aerobic conditions in Luria Broth medium. Each SmsCB was expressed under aerobic conditions in Luria Broth medium. Expression was carried out in MG1655(DE3) for 3 hours at 37°C under 0.5 mM IPTG induction. Harvested cells were resuspended anaerobically inside a glove box (Jacomex, O_2 < 2 ppm) in lysis buffer (100 mM Tris pH 8, 150 mM NaCl containing a protease inhibitor cocktail (cOmplete Ultra tablets), sonicated (12 min, 10/30 sec ON/OFF, power 50%) and resulting lysates were clarified by ultracentrifugation (40,000 *g*, 20 min, 4°C). Supernatants were loaded anaerobically onto a 5 mL Ni-NTA column. Column was washed using washing buffer (100 mM Tris pH 8, 150 mM NaCl, 10 mM imidazole) and proteins were eluted with washing buffer containing 400 mM imidazole. Fractions containing proteins (SmsC and SmsB) were desalted to remove imidazole, using NAP-25 columns equilibrated with desalting buffer (100 mM Tris-HCl pH 8, 150 mM NaCl buffer), and were analyzed by UV-visible absorption spectroscopy and for their iron and sulfur content[39,40].

In vitro Fe-S cluster transfer to Aconitase

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

iron and 3.8 sulfur /SmsC2B2) or reconstituted SufCB from *M. jannaschii* (3.7 iron and 3.5 sulfur /SmsC2B2) in order to provide 4 iron and 4 sulfur atoms per AcnB. After 5, 15, and 30 min incubation, aconitase activity was assessed as described in [41]. Briefly, AcnB-SmsCB mixtures were added to 0.6 mM $MnCl_2$, 25 mM citrate, 0.5 U isocitric dehydrogenase, 0.25 mM $NADP^+$, 50 mM Tris-HCl, pH 7.6, in a 100 µL final volume and NADPH formation was monitored at 340 nm by UV-vis. absorption spectroscopy. The 100 % activity corresponds to the activity of the chemically reconstituted AcnB prepared by incubating apo-AcnB with 5 molar excess of ferrous iron and sulfur for 30 min in the presence of 5 mM DTT (2 µmol min^{-1} mg^{-1}).

UV-visible absorption spectroscopy

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

Bacterial two-hybrid

We used the adenylate cyclase-based two-hybrid technique[42] to test pairwise interactions between SmsC and SmsB proteins of *M. jannaschii*. After co-transformation of the BTH101 strain with the two plasmids expressing the fusions, LB selection plates containing ampicillin and kanamycin were incubated at 30°C for 2 days. Then, overnight cultures in LB medium supplemented with ampicillin, kanamycin and 0.5 mM IPTG were spotted on MacConkey plates containing 1% maltose.

Data Availability

Supplementary data are available at https://figshare.com/articles/dataset/Garcia_et_al_2022_Supplementary_data/20209877

Acknowledgements

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

Author Contributions Statement

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

Competing Interests Statement

The authors declare no competing interests.

Figure Legends

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

Figure 2. SMS is a *bona fide* Fe-S cluster biogenesis system**. A.** Bacterial two-hybrid on *M. jannaschii* SmsCB proteins. Cultures of BTH101 strain transformed by pairs of the 2-hybrid pT18 and T25 fusions with SmsC and SmsB genes from *M. jannaschii* were dropped on MacConkey 1% maltose plate and incubated 2 days at 30°C. **B.** SDS-PAGE of anaerobically purified *M. acetivorans* (left) and *M. jannaschii* (right). **C.** UV-visible spectroscopy analysis of anaerobically purified *M. acetivorans* SmsCB (1.6 mg/ml) in 100 mM Tris-HCL, 150 mM NaCl pH8. Iron and sulfur content: 4.05 ± 0.02 Fe and 3.8 ± 0.02 S / $SmsC_2B_2$. **D.** UV-visible spectroscopy analysis of

anaerobically purified *M. jannaschii* SmsCB (10 mg/ml) in 100 mM Tris-HCL, 150 mM NaCl pH8. Iron and sulfur content: 0.15 ± 0.02 Fe and 0.13 ± 0.03 S/ $SmsC_2B_2$. **E, F.** Iron-sulfur cluster transfer from SmsCB to aconitase B (AcnB), *M. acetivorans* and *M. jannaschii* respectively. N = 3 independent experiments. Data are presented as mean values +/- Standard Deviation. Apo AcnB (0.5 nmol) was incubated anaerobically with as-purified SmsCB from *M. acetivorans* or reconstituted SmsCB from *M. jannaschii*. Aconitase activity was measured at different time points as indicated under the graph. Apo-AcnB (0.5 nmol) was incubated anaerobically with 1.5 equivalent of SmsCB in 50 mM Tris-HCl pH 7.6. The reaction mixture comprised 0.6 mM $MnCl_2$, 25 mM citrate, 0.5 U isocitrate dehydrogenase and 0.25 mM NADP+ was added to the protein mixture in a final volume of 100 µl. Aconitase activity was measured by monitoring the absorption at 340 nm. As a positive control, reconstituted AcnB was incubated 30 min. Reconstituted AcnB (Rec-AcnB) was assayed. Rec-AcnB was obtained by incubating Apo-AcnB for 30 min. with a 5-molar excess of iron and sulfide in the presence of 5 mM DTT.

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

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

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

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

Bibliography

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

Published online 2006. doi:10.1111/j.1365-2958.2006.05028.x

9. Andreini C, Rosato A, Banci L. The relationship between environmental dioxygen and iron-sulfur proteins explored at the genome level. *PLoS One*. Published online 2017. doi:10.1371/journal.pone.0171279

10. Tokumoto U, Kitamura S, Fukuyama K, Takahashi Y. Interchangeability and distinct properties of bacterial Fe-S cluster assembly systems: Functional replacement of the isc and suf operons in Escherichia coli with the nifSU-like operon from Helicobacter pylori. *J Biochem*. Published online 2004. doi:10.1093/jb/mvh104

11. Boyd ES, Thomas KM, Dai Y, Boyd JM, Outten FW. Interplay between Oxygen and Fe-S Cluster Biogenesis: Insights from the Suf Pathway. *Biochemistry*. Published online 2014. doi:10.1021/bi500488r

12. Deere TM, Prakash D, Lessner FH, Duin EC, Lessner DJ. Methanosarcina acetivorans contains a functional ISC system for iron-sulfur cluster biogenesis. *BMC Microbiol*. Published online 2020. doi:10.1186/s12866-020-02014-z

13. Olson JW, Agar JN, Johnson MK, Maier RJ. Characterization of the NifU and NifS Fe-S cluster formation proteins essential for viability in Helicobacter pylori. *Biochemistry*. Published online 2000. doi:10.1021/bi001744s

14. Hug LA, Baker BJ, Anantharaman K, et al. A new view of the tree of life. *Nat Microbiol*. Published online 2016. doi:10.1038/nmicrobiol.2016.48

15. Dombrowski N, Lee JH, Williams TA, Offre P, Spang A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol Lett*. Published online 2019. doi:10.1093/femsle/fnz008

16. Liu Y, Sieprawska-Lupa M, Whitman WB, White RH. Cysteine is not the sulfur source for iron-sulfur cluster and methionine biosynthesis in the methanogenic archaeon Methanococcus maripaludis. *J Biol Chem*. Published online 2010. doi:10.1074/jbc.M110.152447

17. Trotter V, Vinella D, Loiseau L, De Choudens SO, Fontecave M, Barras F. The CsdA cysteine desulphurase promotes Fe/S biogenesis by recruiting Suf components and participates to a new sulphur transfer pathway by recruiting CsdL (ex-YgdL), a ubiquitin-modifying-like protein. *Mol Microbiol*. Published online 2009. doi:10.1111/j.1365-2958.2009.06954.x

18. Gribaldo S, Talla E, Brochier-Armanet C. Evolution of the haem copper oxidases superfamily: a rooting tale. *Trends Biochem Sci*. Published online

2009. doi:10.1016/j.tibs.2009.04.002

19. Brochier-Armanet C, Talla E, Gribaldo S. The multiple evolutionary histories of dioxygen reductases: Implications for the origin and evolution of aerobic respiration. *Mol Biol Evol*. Published online 2009. doi:10.1093/molbev/msn246

20. Jabłońska J, Tawfik DS. The evolution of oxygen-utilizing enzymes suggests early biosphere oxygenation. *Nat Ecol Evol*. Published online 2021. doi:10.1038/s41559-020-01386-9

21. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):119. doi:10.1186/1471-2105-11-119

22. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. Published online 2018. doi:10.1038/s41467-018-07641-9

23. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*. Published online 2012. doi:10.1093/bioinformatics/bts565

24. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389-3402. Accessed August 3, 2018. http://www.ncbi.nlm.nih.gov/pubmed/9254694

25. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*. 2013;30(4):772-780. doi:10.1093/molbev/mst010

26. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol*. 2011;7(10):e1002195. doi:10.1371/journal.pcbi.1002195

27. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. Poon AFY, ed. *PLoS One*. 2010;5(3):e9490. doi:10.1371/journal.pone.0009490

28. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*. 2010;10(1):210. doi:10.1186/1471-2148-10-210

29. Leigh JW, Schliep K, Lopez P, Bapteste E. Let them fall where they may:

Congruence analysis in massive phylogenetically messy data sets. *Mol Biol Evol*. Published online 2011. doi:10.1093/molbev/msr110

30. Adam PS, Borrel G, Gribaldo S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc Natl Acad Sci U S A*. Published online 2018. doi:10.1073/pnas.1716667115

31. Coleman GA, Davín AA, Mahendrarajah TA, et al. A rooted phylogeny resolves early bacterial evolution. *Science (80-)*. Published online 2021. doi:10.1126/science.abe0511

32. Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci U S A*. Published online 2015. doi:10.1073/pnas.1420858112

33. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*. 2015;32(1):268-274. doi:10.1093/molbev/msu300

34. Wang HC, Minh BQ, Susko E, Roger AJ. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst Biol*. Published online 2018. doi:10.1093/sysbio/syx068

35. Garcia PS, Jauffrit F, Grangeasse C, Brochier-Armanet C. GeneSpy, a user-friendly and flexible genomic context visualizer. *Bioinformatics*. Published online 2019. doi:10.1093/bioinformatics/bty459

36. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res*. Published online 2019. doi:10.1093/nar/gkz239

37. Wahl A, My L, Dumoulin R, Sturgis JN, Bouveret E. Antagonistic regulation of dgkA and plsB genes of phospholipid synthesis by multiple stress responses in Escherichia coli. *Mol Microbiol*. Published online 2011. doi:10.1111/j.1365-2958.2011.07641.x

38. Gully D, Bouveret E. A protein network for phospholipid synthesis uncovered by a variant of the tandem affinity purification method in Escherichia coli. *Proteomics*. Published online 2006. doi:10.1002/pmic.200500115

39. fish WW. Rapid Colorimetric Micromethod for the Quantitation of Complexed Iron in Biological Samples. *Methods Enzymol*. Published online 1988. doi:10.1016/0076-6879(88)58067-9

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