Evolutionary plasticity and functional repurposing of the essential metabolic enzyme MoeA

Daniela Megrian[1,2#*], Mariano Martinez[1,3], Pedro M Alzari[1], & Anne Marie Wehenkel[1,3#]

1. Institut Pasteur, CNRS UMR 3528, Université Paris Cité, Structural Microbiology Unit, F-75015 Paris, France.
2. Institut Pasteur de Montevideo, Bioinformatics Unit, 11200 Montevideo, Uruguay.
3. Institut Pasteur, Université Paris Cité, Bacterial Cell Cycle Mechanisms Unit, F-75015 Paris, France.

Correspondence: dmegrian@pasteur.edu.uy and anne-marie.wehenkel@pasteur.fr

Abstract

MoeA, or gephyrin in higher eukaryotes, is crucial for molybdenum cofactor biosynthesis required in redox reactions. Gephyrin is a moonlighting protein also involved in postsynaptic receptor clustering, a feature thought to be a recent evolutionary trait. We showed previously that a repurposed copy of MoeA (Glp) is involved in bacterial cell division. To investigate how MoeA acquired multifunctionality, we used phylogenetic inference and protein structure analyses to understand the diversity and evolutionary history of MoeA. Glp-expressing Bacteria have at least two copies of the gene, and our analysis suggests that Glp has lost its enzymatic role. In Archaea we identified an ancestral duplication where one of the paralogs might bind tungsten instead of molybdenum. In eukaryotes, the acquisition of the moonlighting activity of gephyrin comprised three major events: first, MoeA was obtained from Bacteria by early eukaryotes, second, MogA was fused to the N-terminus of MoeA in the ancestor of opisthokonts, and finally, it acquired the function of anchoring GlyR receptors in neurons. Our results support the functional versatility and adaptive nature of the MoeA scaffold, which has been repurposed independently both in eukaryotes and bacteria to carry out analogous functions in network organization at the cell membrane.

Introduction

Most biological processes in all domains of life require redox reactions that rely on enzymes with associated metal ions or bound metal cofactors (Mayr et al., 2021). Available evidence suggests that the last universal common ancestor (LUCA) used these cofactors, including the pterin-based molybdenum cofactor (Moco) (Weiss et al., 2016). Molybdenum is an essential micronutrient for most living organisms, as its versatile redox chemistry allows molybdoenzymes to catalyze important reactions in biochemical cycles of carbon, nitrogen, and sulfur (Peng et al., 2018). Moco biosynthesis comprises three steps catalyzed by enzymes that are well-conserved in all domains of life (Mendel & Leimkühler, 2015). Many pathogenic bacteria such as *Mycobacterium tuberculosis*, *Salmonella enterica*, *Campylobacter jejuni* or *Haemophilus influenzae* use molybdoenzymes (sulfite dehydrogenases, S/N-oxides reductases, nitrate reductases, formate dehydrogenases) to facilitate adaptation of the pathogen to its environment by supporting energy generation, or by converting compounds generated in the host during inflammation (Zhong et al., 2020). In humans, deficiency in Moco biosynthesis causes a rare disease responsible for the loss of the enzymatic activity of all molybdoenzymes, leading to severe neurological damage and premature death (Schwarz, 2005).

Interestingly, secondary functions have been identified for proteins involved in Moco biosynthesis in animals (Tyagarajan & Fritschy, 2014) and more recently in *Corynebacteriales* (Martinez et al., 2023). Gephyrin is a well-documented moonlighting protein that was first discovered for its role in post-synaptic signaling (Kirsch et al., 1991) and only later identified as a fusion protein composed of MoeA and MogA, the two enzymes responsible for the last step in Moco biosynthesis (Stallmeyer et al., 1999). This gene fusion allowed gephyrin to acquire moonlighting functions as a scaffolding protein that binds to both the cytoskeleton and the glycine and GABA type A receptors, and plays a major structural role in synaptic signaling in the central nervous system (Tyagarajan & Fritschy, 2014) . The plant homolog of gephyrin, Cnx1, has also been described to interact with the cytoskeleton (Schwarz et al., 2000). These moonlighting properties of gephyrin homologs are an evolutionary trait thought to have been acquired in eukaryotes (Mayr et al., 2021). More recently, the discovery of a gephyrin-like protein (Glp) in bacteria (Martinez et al., 2023) raised the question of whether

the MoeA protein has the plasticity to adopt secondary functions and whether this plasticity is an ancient trait. *Corynebacteriales* contain two or more homologs of MoeA, one of them, Glp, plays an important role in cell division by directly binding the tubulin-like cytoskeletal protein FtsZ and an associated membrane protein GlpR (Martinez et al., 2023). Reminiscent of the eukaryotic gephyrin secondary function, Glp is thought to be similarly involved in network organization at the inner membrane of the corynebacterial septum (Martinez et al., 2023). Glp is phylogenetically distinct from *Escherichia coli* MoeA and seems to be the result of a duplication within the phylum Actinobacteria, though it is currently unknown if Glp has retained its enzymatic function (Martinez et al., 2023).

The seeming functional plasticity of MoeA raises several evolutionary questions. First, it is not clear how and when the protein fusion occurred during the evolutionary history of Eukaryotes, as this event has been studied independently only in a few plant, fungi and animal model organisms (Mayr et al., 2021). While this fusion is essential for the binding to - and network organization of - neurotransmitter receptors (Choii & Ko, 2015), it is not clear how widespread the fusion is in the tree of eukaryotes and what the effect of the fusion was in eukaryotes that do not have a nervous system. Moreover, it has not been reported how and when eukaryotes acquired MoeA. Besides eukaryotes, MoeA is ubiquitously present in Bacteria and Archaea (Zhang et al., 2011), but the functionality and the evolutionary history of this protein in the context of the tree of life has not been addressed in detail and the recent discovery of a gephyrin-like protein in bacteria (Martinez et al., 2023) suggests that functional repurposing of the MoeA scaffold is not a unique feature of higher eukaryotes. To understand the evolutionary history of MoeA, we present here the phylogenetic analysis of MoeA in all domains of life and show that both Archaea and Actinobacteria have independently undergone gene duplications and stably maintained two distinct clades over time. We show that the unique MoeA/gephyrin copy of Eukaryotes has a bacterial origin, and that the MogA-MoeA fusion occurred at least twice during the evolution of the Eukaryotes. Finally, we address the question on whether the MoeA homologs found in organisms that have more than one copy could be moonlighting proteins or repurposed enzymes that have lost their original catalytic function. Our analysis shows that the bacterial Glp homologs have a greatly altered active site in line with a possible loss of function, whereas in Archaea both copies seem to have a conserved active site, but possibly different substrate affinity. Our combined

phylogenetic and structural analyses emphasize the functional differences between MoeA homologs, and leads us to propose a scenario for the diversity and evolution of this protein.

Results

Overall distribution of MoeA and domain architecture in eukaryotes

The complex chemical transformations required for the biosynthesis of Moco are strictly conserved throughout all domains of life. To be catalytically active, molybdenum is scaffolded with a molybdopterin containing pterin (MPT) to form Moco (Mendel & Leimkühler, 2015). The biosynthesis of Moco comprises three major chemical rearrangements: (i) the circularization of GTP into cPMP, (ii) the transfer of sulfur to cPMP to generate MPT, and (iii) the insertion of molybdate into MPT to form Moco (Figure 1a) (Mendel & Leimkühler, 2015). These reactions are all catalyzed by highly conserved enzymes that occur individually (prokaryotes) or as multi-enzyme fusion proteins (eukaryotes) (Mendel & Leimkühler, 2015) (Figure 1b and Figure 1c). In prokaryotes, each step is catalyzed by the dual action of two individual proteins (Figure 1a) (Mendel & Leimkühler, 2015). In higher eukaryotes these pairs of proteins are fused to form 3 multi-domain proteins: MOCS1, MOCS2 and gephyrin (Figure 1b) (Mendel & Leimkühler, 2015). Plants represent an intermediate case, where only proteins MogA and MoeA are fused into the multi-domain gephyrin-homolog Cnx1 (Mendel & Leimkühler, 2015). Eukaryotic gephyrin contains two globular domains, G and E, respectively homologous to prokaryotic proteins MogA and MoeA, connected through a disordered linker region called C-domain that interacts with microtubules (Figure 1c) (Choii & Ko, 2015). Plant Cnx1 has a shorter C-domain, which may affect its quaternary structure, and has also been described to interact with the cytoskeleton, but in this case through its G-domain (Kaufholdt et al., 2016).

To infer the origin of MoeA in Eukaryotes we reconstructed a phylogeny including sequences obtained from the three domains of life (Figure 1d and Supplementary Figure 1). This phylogeny robustly places the Eukaryotes as a monophyletic group within Bacteria, suggesting that MoeA was acquired from Bacteria by the last eukaryotic common ancestor (LECA) or very early during the evolution of the Eukaryotes. The most evident difference between bacterial MoeA and animal gephyrin is the presence of the fused MogA (G-domain) to the N-ter of

gephyrin (Choii & Ko, 2015). This MogA-MoeA fusion is thought to have endowed gephyrin with its networking properties, as it allows for G-domain trimerization coupled to E-domain dimerization (Sola et al., 2004). To understand how and when the transition between MoeA and gephyrin happened during evolution, we investigated the domain distribution of MoeA proteins in Eukaryotes. Eukaryotic MoeA can be clustered roughly in two big groups: a first group belonging to algae, plants, and microbial eukaryotes (the Sar supergroup), and a second group belonging to protist clades Amoebozoa and Discoba, and the opisthokonts, which include fungi and animals (Figure 1e and Supplementary Figure 2) (Burki et al., 2020). Members of the first group have either the canonical MoeA architecture or an extra MogA domain, but in contrast to gephyrin, this domain is fused to the C-terminus of the protein. The extra MogA domain is present in most phyla of the Sar supergroup and Embryophyta (plants), but it is absent from algae and Sar phyla Pelagophyceae and Bacillariophyta. It is unclear whether the MoeA-MogA fusion in this group is ancestral and was lost later in some species, or alternatively whether the fusion happened at least three times during the evolution of these lineages.

A significant change happened in the ancestor of Amoebozoa, Discoba and the opisthokonts, as most members contain a MogA domain fused at the N-terminus of the MoeA domain (Figure 1e and Supplementary Figure 2). This fusion was an essential step for the transition between the canonical MoeA responsible for Moco biosynthesis, and the moonlighting gephyrin that is also responsible for postsynaptic clustering of neurotransmitter receptors in animals, as the N-terminal MogA domain is present in all studied animals. It is interesting to note that in fungi, MoeA was either kept in this fused form or was completely lost from the genome, as in the model organism *Saccharomyces cerevisiae* (Supplementary Figures 2 and 3). We looked for MoeA homologs in representative members of all fungal orders and found that MoeA is missing in all 19 analyzed genomes of Microsporidia and Cryptomycota, suggesting that it was lost in the ancestor of these groups (Supplementary Figure 3). We further identified several, possibly independent, losses scattered in the reference tree of Fungi, including members of Chytridiomycota, Blastocladiomycota, Zoomycota, Basidiomycota and Ascomycota (Supplementary Figure 3). It is not clear why and how several fungi could circumvent independently the need for the Moco cofactor.

Taken together these results show that Eukaryotic MoeA has a bacterial origin, and that the transition to moonlighting gephyrin involved changes in its domain architecture, in particular the fusion of the MogA domain to the N-terminus of MoeA.

Most Archaea contain two copies of MoeA from an ancestral MoeA duplication

Moco biosynthesis is widely conserved in Archaea and MoeA has been reported to be duplicated in some organisms like *Pyrococcus furiosus* (Bevers et al., 2008). Most Archaea are believed to prefer tungsten over molybdenum for the metal cofactor biosynthesis (Hagen, 2011), but it is not clear whether the MoeA duplication is related to this variation. To understand the evolutionary history of MoeA in the prokaryotic context, we carried out a phylogenetic analysis of all Archaea and Bacteria. We identified MoeA in all archaeal phyla except for most members of Methanomassiliicoccales, Aciduliprofundales and Poseidoniales -all of which belong to the candidate phyla Thermoplasmatota- and the members of DPANN superphyla (Figure 2a and Supplementary Table 1). The members of the DPANN are also known as nanoarchaea because of their greatly reduced cellular and genomic size. They live as epibionts and may have lost Moco biosynthesis enzymes because they can obtain the cofactor from the host (Castelle & Banfield, 2018). Interestingly, most Archaea have two copies of MoeA, and some members of Methanomicrobiales have up to six copies, obtained by independent and recent duplications. The phylogeny of MoeA shows two well-supported subtrees (MoeA1 and MoeA2) that contain most archaeal phyla and present a topology that roughly matches that of the species tree of Archaea (Figure 2b and Supplementary Figure 4). This indicates that MoeA was duplicated before the divergence of Archaea -or early during its evolution- and the two paralogs were maintained in most phyla, suggesting an important functional role for both copies. In most cases, the two paralogs are contiguous in the genome, suggesting their participation in related functions (Figure 2c).

We found two small bacterial clades branching within the clade formed by archaeal MoeA1 and MoeA2, both of which are phylogenetically distinct to the canonical bacterial MoeA clade that contains *E. coli* and most other bacterial species (Figure 2b). These smaller clades contain the same species, which are anaerobic or facultative aerobic, meso- or thermophilic, and were sampled from a wastewater treatment plant (*Brevefilum fermentans*) (McIlroy et al., 2017), a hot spring sulfur-turf (*Caldilinea aerophila*) (Sekiguchi et al., 2003), swine intestinal tract

(*Cloacibacillus porcorum*) (Looft et al., 2013), a methanogenic reactor treating protein-rich wastewater (*Coprothermobacter platensis*) (Etchebehere et al., n.d.), a methanogenic sludge (*Thermanaerovibrio acidaminovorans*) (Guangsheng et al., 1992) and hot aquatic environments (*Thermodesulfobacterium commune*) (Zeikus et al., 1983). The sister groups to both small bacterial clades correspond to the Methanomicrobiales (Figure 2b), an order of anaerobic archaea that produce methane and inhabit aquatic sediments, anaerobic sewage digestors and the gastrointestinal tract of animals (López-García & Moreira, 2006). This suggests that the Methanomicrobiales likely coexist or coexisted in the same environment with these bacteria. While it cannot be excluded that the two small bacterial clades correspond to an ancestral duplication, the fact that both homologs are continuous in the genome (Figure 2c), that they are phylogenetically distinct to the canonical bacterial MoeA, and that the species containing these homologs inhabit the same niches as Methanomicrobiales, suggests that these bacterial species could have obtained both MoeA copies from Methanomicrobiales in a single horizontal gene transfer (HGT) event. The bacterial acquisition of these archaeal proteins might have given them the ability to incorporate tungsten, instead of, or in addition to, molybdenum in the metal cofactor, which is rare in Bacteria (Peng et al., 2018). The placement of the Methanomicrobiales clades in the tree of Archaea is intriguing, as we would have expected it to branch together with Methanosarcinales (Adam et al., 2017). The support of the deepest branches in the phylogeny does not allow us to determine if this corresponds to a different evolutionary history of the MoeA copies of the Methanomicrobiales, or to an artifact, likely caused by long branch attraction (Figure 2b).

One of the archaeal paralogs is fused to a PBP domain and potentially binds tungsten

To understand the differences between the two MoeA paralogs in Archaea and to look for possible alternative functions of MoeA, we compared the sequences of the two archaeal paralogs. Most sequences in the MoeA1 clade are longer than the canonical *E. coli* MoeA (Figure 3a and Supplementary Figure 4). The analysis of these sequences shows the fusion of MoeA with a periplasmic-binding protein-like (PBP-like) domain in the C-terminal region (Figure 3a and Supplementary Figure 4). PBPs are nonenzymatic receptors used by prokaryotes to sense small molecules in the periplasm and transport them into the cytoplasm via ABC transporters (Borrok et al., 2009). Interestingly, the molybdate binding lipoprotein

ModA involved in molybdenum uptake is a PBP (Hagen, 2011). Molybdate uptake systems have been mostly studied in bacteria, where they consist of a three-protein machine encoded by the modABC cassette, in which ModA is a molybdate binding lipoprotein, ModB an integral membrane protein and ModC an ATP-binding cassette ABC-type transporter (Figure 1a) (Hagen, 2011).

In bacteria, ModA mediates the entrance of molybdenum into the cell, where it is incorporated into MPT-AMP by MoeA to form the Moco (Figure 1a) (Leimkühler, 2020). The archaeal PBP fusion to MoeA is likely cytoplasmic, as the canonical MoeA protein is found intracellularly, and there are no predicted signal peptides, that would suggest export into the periplasm, or predicted transmembrane domains between MoeA and the PBP domain that would suggest a communication through the membrane. The high-confidence AlphaFold atomic model of the MoeA-PBP dimer shows that the PBP-like domain sits on top of domain IV of MoeA (Figure 3b). This relative positioning allows for the formation of a continuous groove between the predicted ligand-binding site of the PBP and the MoeA active site (Figure 3b), suggesting that the PBP-like domain could facilitate the capture and channeling of the solute (molybdenum, tungsten, or other small molecules) into the active site of the dimer. To better characterize this PBP domain, we compared and investigated the presence or absence of other PBP proteins involved in the uptake of molybdenum and tungsten in Archaea: ModA (molybdenum and tungsten), WtpA (molybdenum and tungsten), and TupA (tungsten specific) (Hagen, 2011). We did not identify any of these proteins in Aciduliprofundales, Poseidoniales, Methanomassiliicoccales and the DPANN superphylum, in agreement with the absence of MoeA (Figure 3c and Supplementary Table 2). All ModA, WtpA, TupA and the PBP domain of MoeA1 are widely distributed in Archaea, and it is not clear if they have the same function but are regulated differently, or if they evolved specialized functions.

Interestingly, the PBP domain of the archaeal MoeA1 family is also found in one of the bacterial MoeA homologs that branches within Archaea (Figure 3a). While ModA is present in all bacterial phyla that contain a MoeA-like protein, the presence of the MoeA-PBP fusion protein as well as the tungsten transporters WtpA and TupA are restricted to a few phyla (Figure 3d and Supplementary Table 2). Surprisingly, MoeA-PBP and tungsten transporter TupA cooccur when present, and at least in the Chloroflexi *Brevefilum fermentans*, their coding sequences are located six genes apart. This putative functional link between MoeA-

PBP and TupA, suggests that MoeA-PBP might insert tungsten instead of molybdenum into MPT and might thus be involved in the biosynthesis of the tungsten cofactor instead of Moco. Considering the putative role of MoeA-PBP in the insertion of tungsten into MPT and the fact that the bacterial MoeA-PBP is phylogenetically related to the archaeal MoeA1 (Figure 3a), this would suggest that the archaeal MoeA1 is specialized in the biosynthesis of the tungsten cofactor in Archaea, while MoeA2 takes part in the biosynthesis of the molybdenum cofactor. The fact that the two copies of MoeA in Archaea are ancestral and widespread in the domain, might suggest an important role for both cofactors in the metabolism of Archaea.

Moonlighting, repurposing and specialization of MoeA

The functional canonical MoeA assembly is a homodimer that contains two catalytic sites with molybdopterin transferase enzymatic activity (Figure 4a). In animals, MoeA is a moonlighting protein, as the unique copy of the gene in the genomes carries both the canonical molybdopterin transferase enzymatic activity and has acquired additional functions in protein network organization at the post-synapse (Choii & Ko, 2015). Interestingly, we identified at least two copies of MoeA in Archaea and in some Bacteria (Figure 2a, Supplementary Figure 4 and Supporting Data), which may reflect specific physiological needs. For instance, Actinobacteria have systematically maintained two or more MoeA copies throughout their evolutionary history (Martinez et al., 2023). MoeA duplications could thus indicate functional redundancy, or alternatively, different paralogs could have evolved different functions. Indeed, protein redundancy conserved over very large evolutionary distances in prokaryotes seems unlikely as the evolution of genomes appears to be dominated by reduction, and duplicated genes become either specialized or are lost (Wolf & Koonin, 2013). Although the four domains of the MoeA monomeric structure are highly similar between all MoeA proteins, a conformational change between the domains leads to an important difference in the quaternary organization in the bacterial Glp homologs (Martinez et al., 2023). This conformational change between the domains translates into an opening of the dimer interface of Glp, that in turn generates the binding site for FtsZ and possibly GlpR (Martinez et al., 2023) (Figure 4a).

To explore the functional divergence or the possible dual role of all MoeA homologs in all the domains of life we computed AlphaFold high confidence structural models for each of the

main MoeA groups (archaeal MoeA1 and MoeA2, bacterial MoeA and Glp, and eukaryotic MoeA and gephyrin), and mapped the sequence conservation onto representative models (Figure 4a and Supplementary Figure 5). The predicted structures of MoeA revealed that each group has the same overall monomeric and dimeric structures (Figure 4a). In all homologs from Eukaryotes and Archaea, we observed a clear sequence conservation of the two protein regions that define the catalytic groove in the MoeA dimer, suggesting that they have a functional active site (Figure 4a). On the contrary, the putative active site of bacterial Glp homologs showed a low degree of conservation (Figure 4a and Supplementary Figure 5), strongly suggesting the loss of the Moco biosynthesis capability, in line with a scenario of evolutionary repurposing rather than moonlighting. Interestingly, both archaeal MoeA paralogs show a clear sequence conservation in the catalytic groove, suggesting the presence of a functional active site. However, the quaternary organization of the archaeal MoeA1 shows an opening at the dimer interface, which might reflect a specialization of the protein.

To understand whether quaternary structural rearrangements in the MoeA dimer could reflect a functional conservation or divergence, we compared the distances defined by the amino acids in the active site of representative structures of the different MoeA groups. For quantification purposes, we chose conserved representative amino acids from the active site, computed their relative distances, and performed a Principal Component Analysis (PCA) to identify if the distances between the residues in these regions can discriminate the different MoeA groups (Figure 4b and Supplementary Table 3). The PCA analysis based on the distances calculated for the active site separates Glp in the PC1 axis and archaeal MoeA1 in the PC2 axis from the other MoeA groups that cluster together (Figure 4b). The fact that Glp representatives are spread in the PC1 axis suggest a loss in the conservation of the active site structure, which is congruent with the lack of sequence conservation, supporting the hypothesis of the loss the Moco biosynthesis capability. On the other hand, the separation of archaeal MoeA1 suggests that the distances in the active site of MoeA1 are conserved but are different to the distances in the other groups (Figure 4b). This could be the consequence of the putative specialization of MoeA1 to bind tungsten instead of molybdenum, and/or the consequence of the physical constraints determined by the fusion of the PBP-like domain that sits on top of the active site.

Finally, the gephyrin homodimer contains the two catalytic sites with molybdopterin transferase enzymatic activity, as well as two binding sites for the GlyR and GABAA membrane neuroreceptors. To understand when during evolution MoeA acquired the potential to bind to these membrane receptors we analysed the protein sequence conservation on the key binding residues (Figure 5). The neuroreceptor binding site in gephyrin is very well conserved, however, this region is also well conserved in MoeA belonging to Sar, algae, and plants (Figure 5), organisms that lack a nervous system. This conservation is absent in the bacterial MoeA, indicating that it appeared in the LECA or very early during the evolution of the Eukaryotes. This result suggests that MoeA of Sar, algae, and plants might be able to bind other similar molecules in the same position, potentially an ancestral receptor, granting the non-animal eukaryotic MoeA another moonlighting function.

Discussion

Our results support the hypothesis that MoeA was present in the last universal common ancestor (LUCA), which suggests that LUCA had pterin-based cofactors, as it has been proposed before (Weiss et al., 2016). Our results indicate that during the evolution of life, MoeA was not only transmitted vertically to most species in all domains of life, but it also underwent duplications, horizontal gene transfers and fusions, which led to the repurposing, acquisition of moonlighting function and probably specializations of the protein (Figure 6). The most ancestral event we predict was a duplication of MoeA, which is reflected by the presence of two MoeA copies in most archaeal genomes, which form two separate clades in the phylogeny of MoeA (Figure 2 and 6). These clades follow roughly the species archaeal tree (Adam et al., 2017), suggesting that the two copies were inherited vertically in most archaea. The presence of two MoeA copies was reported before in some archaeal species (Bevers et al., 2008).

The history of MoeA in Bacteria seems less straightforward. Besides the largest clade, we identified two smaller clades of bacterial MoeA within methanogenic archaea (Figure 2 and Figure 6). Each of these clades can be evolutionarily associated to either archaeal MoeA1 or MoeA2 based on the length, domain architecture and tridimensional structure. However, it is not clear how bacteria obtained them. The topology of the phylogeny is compatible with two scenarios: either the duplication of MoeA happened before LUCA and the last bacterial

common ancestor (LBCA) had two ancestral MoeA copies that were later lost in most bacteria, or the LBCA had a single MoeA copy and some lineages acquired the archaeal MoeAs by HGT. The bacteria identified in these clades coexist with archaea in the same thermophilic and methanogenic niches (Etchebehere et al., n.d.; Guangsheng et al., 1992; Looft et al., 2013; López-García & Moreira, 2006; McIlroy et al., 2017; Sekiguchi et al., 2003; Zeikus et al., 1983), which supports both options. In the first scenario, both MoeA copies were obtained vertically by archaea and bacteria from LUCA. Bacteria that colonized other niches lost these genes, while a new MoeA could have been acquired from Archaea, and was later spread in the bacterial domain. The topology of the largest bacterial MoeA clade is not compatible with the species bacterial tree (Megrian et al., 2022), which might suggest that MoeA was spread in Bacteria by HGT. However, the resolution of a single gene tree covering the two prokaryotic domains has important limitations, especially at the nodes that connect phyla (Megrian et al., 2022) and can lead to misinterpretation of the events. In the second scenario, the LBCA had a single MoeA copy, and some bacterial species obtained, in a single event, the two archaeal MoeA genes by horizontal gene transfer from a methanogenic Archaea. It is important to highlight that both archaeal MoeA genes, when present, are contiguous in the archaeal and bacterial genomes, which is compatible with a single HGT event.

Independently on how these two MoeA genes were obtained, it seems that both have an important and non-redundant role, at least in Archaea, as both were kept during billions of years of evolution. The fact that MoeA has been reported to utilize tungsten as well as molybdenum, and the existence of homologous enzymes that can use the tungsten cofactor instead of Moco, leads us to put forward the hypothesis that one archaeal MoeA produces Moco, while the other produces a tungsten cofactor. The geological record suggests that tungsten was an essential element for the earliest life forms (Maia et al., 2017). The ocean in the early Earth was anoxic and sulfidic, and under these conditions tungsten forms soluble salt while molybdenum is insoluble (Maia et al., 2017). Around 2.5 billions years ago, the conditions of the ocean changed with the appearance of photosynthesizing bacteria. This produced a rise of dioxygen in the environment, which oxidized molybdenum-containing sulfide minerals and led to the accumulation of molybdenum in the oceans. This event probably forced the metabolisms of cellular organisms to adapt to the changing conditions of the ocean, and to start using tungsten instead of molybdenum, by duplicating and maybe specializing the machinery involved in the biosynthesis of the pterin-based cofactor and the

enzymes that use this cofactor (Maia et al., 2017). Nowadays, tungsten is mainly used by thermophilic anaerobic archaea (Maia et al., 2017), whose anoxic environments have higher tungsten than molybdenum bioavailability. However, these organisms, as well as most archaea and some bacteria have both MoeA1 and MoeA2, and the functional differences between them are still not clear.

It had been reported that in bacteria, molybdate is mainly taken up by the ModABC system, however, ModA can bind both molybdenum and tungsten (Hagen, 2011). Also, a homologous molybdenum and tungsten transporter, WtpABC, and a third tungsten-specific transporter, TupABC, have been identified, but their distribution was reported to be much more restricted (Hagen, 2011). In this work we confirmed that ModA is widespread in almost all bacteria, and that the presence of WtpA and TupA is scattered in the phylogeny of bacteria. Interestingly, we identified a ModA homolog fused to MoeA in some bacteria. This MoeA, which is the homologous to archaeal MoeA1, co-occur with transporter TupA, suggesting a functional and evolutive link. As TupA is a tungsten-specific transporter in Archaea, this result suggests that MoeA1 might be specific to tungsten and was either obtained or kept in bacteria that inhabit anoxic environments where molybdenum is less available than tungsten. In this regard, Mota et al. (Mota et al., 2011) studied the effects of molybdenum and tungstate on the expression levels of *moeA1* and *moeA2* of the bacterium *Desulfovibrio alaskensis*. The supplementation with tungsten did not affect the expression of *moeA2*, but decreased the expression of *moeA1*, while the supplementation with molybdenum did not affect the expression of both *moeA* genes. Using a different rationale, Malotky et al. (Malotky, 2002) expressed MoeA1 and MoeA2 of the archaeon *Pyrococcus furiosus* in an *E. coli* MoeA mutant strain, and observed that MoeA2 partially complements the mutant, suggesting that archaeal MoeA2 has a similar function to bacterial MoeA. This result agrees with the topology of our MoeA phylogeny, that places the largest clade of bacterial MoeA closer to archaeal MoeA2 (Figure 2 and Figure 6), and supports the hypothesis that MoeA2 uses molybdenum.

We showed that MoeA was obtained by early eukaryotes from Bacteria (Figure 1 and Figure 6), and that during the diversification of the Eukaryotes MoeA fused to MogA in two separate events, probably once in the C-terminus, and once in the N-terminus. Both types of fusion proteins have been reported to form networks and interact with the cytoskeleton, as it is the case of plant Cnx1 and animal gephyrin (Choii & Ko, 2015; Schwarz et al., 2000). We recently reported a similar case in the Actinobacteria, where an independent duplication of MoeA

within this phylum led to the specialization of Glp, one of the MoeA paralogs (Martinez et al., 2023). This protein binds to the bacterial tubulin homolog FtsZ, and acts as a protein scaffold to control cell division and morphogenesis. Differently to gephyrin, in this work we predict that Glp does not have a moonlighting function, as the catalytic activity seems to have been lost during the specialization. Our results support the functional versatility and adaptive nature of the MoeA protein, which has been repurposed independently in both eukaryotes and bacteria to carry out analogous functions in scaffolding and control at the inner membrane in dynamic systems, such as mammalian synaptic signaling and bacterial cell division. The potential of MoeA/gephyrin to create networks and to bind other proteins in eukaryotes that do not contain a nervous system, such as plants or fungi, reflected on their sequence conservation, opens up the question whether other cellular processes could be mediated by this versatile protein.

Overall, we propose an evolutionary scenario where MoeA was present in LUCA and is nowadays widespread in most species in all domains of life (Figure 6). During its evolutionary history, MoeA was subjected to independent duplications (and possibly HGTs), that led to its specialization, repurposing and acquisition of a moonlighting function. Besides its metabolic role, MoeA seems to have acquired networking capabilities in an independent manner, probably favoring the acquisition of novel and diverging functions, as it is the case for actinobacterial Glp and animal gephyrin. It remains an open question whether other MoeA homologs have other specialized or moonlighting functions, and whether this versatility exists in other proteins that maintained the folding while changed or acquired new functions.

Methods

Database assembly

To carry out a large-scale MoeA investigation in all domains of life, we assembled databases with genomes representing all bacterial, archaeal and eukaryotic diversity. For Bacteria, we assembled a database containing 81 genomes (five taxa per phylum), based on the taxonomic sampling in Martinez et al., 2023, and adding five actinobacterial taxa. For Archaea, we assembled a database containing 122 genomes representing all major phyla, based on the taxonomic sampling in (Pende et al., 2021), but excluding the genomes that are not annotated in the NCBI Genome database (Sayers et al., 2022). For Eukaryotes, we selected five taxa per

phylum (if available), from all eukaryotic annotated genomes in the NCBI Genome database (Sayers et al., 2022). We assembled a database containing the 129 genomes corresponding to these phyla, representing all diversity present at the NCBI as of October 2021. For Fungi, we assembled a database containing 171 genomes, including one representative of each fungal order with at least one annotated genome at the NCBI Genome database (Sayers et al., 2022) as of October 2021.

Homology searches and mapping

To study the taxonomic distribution of MoeA in all domains of life, we performed sensitive HMM homology searches against the Bacteria, Archaea, Eukaryotes and Fungi databases. First, we built HMM profiles based on the bacterial MoeA aligment provided in (Martinez et al., 2023), using the HMMBUILD tool from the HMMER package (Johnson et al., 2010). Then, we used these profiles to search for MoeA homologs in the four databases, using the HMMSEARCH tool from the HMMER package (Johnson et al., 2010) with by default parameters. To remove false positives, we used the Conserved Domain Database (CDD) online tool (Marchler-Bauer & Bryant, 2004) to identify hits that contain all three Pfam domains MoeA_N (pfam03453), MoCF_biosynth (pfam00994) and MoeA_C (pfam03454). We mapped the number of MoeA copies per archaeal genome on a schematic Archaea tree, obtained from (Pende et al., 2021) using iTOL (Letunic & Bork, 2019). To map the MoeA presence/absence in Fungi, first, we reconstructed a species Fungi phylogeny based on the DNA-directed RNA polymerase II subunit RPB2 protein. To identify RPB2 homologs in all fungal genomes, we used the JACKHMMER tool from the HMMER package (Johnson et al., 2010) and the Saccharomyces cerevisiae NCBI RefSeq sequence (NP_014794.3) as the query, against the Fungi database. We selected the best hit per genome, and we aligned the sequences with MAFFT (Katoh & Standley, 2013) using the L-INS-I algorithm, and we trimmed the alignment using trimAl (Capella-Gutiérrez et al., 2009), keeping the columns that contain less than 20% of gaps. We used this alignment to reconstruct a guide tree with IQ-TREE (Nguyen et al., 2015) using the Model Finder Plus (MFP) option. Then, we used this guide tree to reconstruct a maximum-likelihood tree with IQ-TREE (Nguyen et al., 2015) using the PMSF model, with ultrafast bootstrap supports calculated from 10.000 replicates, with a minimum correlation coefficient of 0.999. We mapped the MoeA presence/absence on this tree using iTOL (Letunic & Bork,

2019). Complete absences of MoeA in any of the genomes in the four databases were manually verified.

To identify ModA, WtpA and TupA homologs in all domains of life, we used the JACKHMMER tool from the HMMER package (Johnson et al., 2010), and the Escherichia coli ModA NCBI RefSeq sequence (NP_415284.1), Pyrococcus furiosus WtpA NCBI GenBank sequence (AAL80204.1), and Campylobacter jejuni TupA NCBI RefSeq sequence (YP_002344912.1) as the queries, against the Bacteria and Archaea databases. We aligned the three groups of hits separately with MAFFT (Katoh & Standley, 2013) using the L-INS-I algorithm, and we visually selected homolog sequences of each protein. We realigned these sequences, removed the columns with more than 20% of gaps, and built HMM profiles for each protein using the HMMBUILD tool from the HMMER package (Johnson et al., 2010). Then, we used these profiles to search for ModA, WtpA and TupA homologs in the Bacteria, Archaea and Eukaryotes databases using the HMMSEARCH tool from the HMMER package (Johnson et al., 2010) with by default parameters, and we selected the hits with an e-value above $1e\text{-}^{6}$.

Phylogenetic analyses

We reconstructed three MoeA phylogenies including the homologs identified in: (i) Bacteria, Archaea and Eukaryotes, (ii) Eukaryotes, and (iii) Bacteria and Archaea. To reconstruct these phylogenies, we aligned the protein sequences with MAFFT (Katoh & Standley, 2013) using the L-INS-I algorithm. We used these alignments to reconstruct a guide tree with IQ-TREE (Nguyen et al., 2015) using the Model Finder Plus (MFP) option. Then, we used these guide trees to reconstruct a maximum-likelihood trees with IQ-TREE (Nguyen et al., 2015) using the PMSF model, with ultrafast bootstrap supports calculated from 10.000 replicates, with a minimum correlation coefficient of 0.999.

We used the results of the CDD (Sayers et al., 2022) described in the previous section to identify extra domains, like the PBP, in some MoeA homologs. We mapped the domain organization of MoeA into the (ii) Eukaryotes, and (iii) Bacteria and Archaea phylogenies using iTOL (Letunic & Bork, 2019) and custom scripts.

Protein structure prediction and distance calculation

We predicted the structure of the dimeric form of ten representative MoeA homologs identified in all domains of life using AlphaFold (Jumper et al., 2021). To compare MoeA

519 structures and based on the alignment of all MoeA homologs in all domains of life, we
520 removed the N-terminal and C-terminal ends of each protein that do not align with *E. coli*
521 MoeA (see Supporting Data). All positions reported on MoeA structures refer to the equivalent
522 positions on *E. coli* MoeA based on the alignment, unless stated otherwise.
523 We classified MoeA structures into eight groups based on the phylogenetic analyses in the
524 previous section: archaeal MoeA1, archaeal MoeA2, bacterial MoeA, bacterial Glp, Sar, a
525 MoeA, Algae and plants MoeA, fungal MoeA, and animal Gephyrin. To map the sequence
526 conservation on a representative structure of each group we used the MoeA alignments
527 obtained in the previous section and software ChimeraX (Pettersen et al., 2021). The method
528 for calculating the sequence conservation is the entropy-based measure from software AL2CO
529 (Pei & Grishin, 2001). For the list of representative structures see Supplementary Table 4.
530 To evaluate the conservation of the distances between residues in the active site of MoeA, we
531 manually selected the residues on the active site surface (for details, see Supplementary Table
532 5). Then, we computed all distances between the residues in the active site for each predicted
533 protein structure, using the Python Bio.PDB package (Cock et al., 2009). Finally, we performed
534 a PCA analysis to compare the distances between the residues of interest in the different
535 MoeA groups.
536

537 **Data availability**

538 All data used to produce our results are provided as supporting data and can be found in
539 https://doi.org/10.17632/phw4knbn8m.1.
540

541 **Acknowledgements**

542 This work was supported in part by grants from the Agence Nationale de la Recherche (ANR,
543 France), contracts ANR-18-CE11-0017 (P.M.A.) and ANR-21-CE11-0003 (A.M.W.), and by
544 institutional grants from the Institut Pasteur, the CNRS, and Université Paris Cité. This work
545 used the computational and storage services (maestro cluster) provided by the IT department
546 at Institut Pasteur, Paris. Molecular graphics were done with ChimeraX, developed at UCSF
547 with support from NIH (R01-GM129325) and NIAID.

548 **Author contributions**

549 DM and AMW designed the research. DM carried out all bioinformatic analysis. DM, MM,
550 AMW and PMA carried out the structural analysis. DM and AMW wrote the paper. All authors
551 edited the paper.
552
553 **Competing interests**
554 The authors declare no competing financial interests.

Figure legends

Figure 1: Proteins involved in the biosynthesis of Moco. (a) Schematic representation of the steps involved in the entrance of molybdenum (MoO_4^{2-}) to the cell, and in the biosynthesis of Moco in *E. coli*. Protein names are indicated in colored boxes. **(b)** Comparison of proteins involved in Moco biosynthesis in a representative species of Bacteria (*E. coli*), plants (*A. thaliana*) and animals (*H. sapiens*). **(c)** Detail of the organization of MoeA and MogA domains in a representative species of Bacteria (*E. coli*), plants (*A. thaliana*) and animals (*H. sapiens*). Each line corresponds to an individual protein. Numbers indicate the length of the protein. Domain MoeA is indicated in dark green, and domain MogA in light green. **(d)** Maximum-likelihood phylogeny of MoeA/Gephyrin in Eukaryotes, Bacteria and Archaea. Monophyletic groups were collapsed into a single branch. Black dots indicate UFB > 90, gray dots indicate 80 < UFB <= 90 and branches without dots indicated UFB <= 80. The scale bar represents the average number of substitutions per site. For the detailed tree, see Supplementary Figure 1. **(e)** Domain organization of MoeA/Gephyrin in representative species of Eukaryotes. Domains indicated in gray correspond to domains different to MoeA or MogA. Higher taxonomic ranks are indicated in black boxes. The scale bar represents the average number of substitutions per site. For the detailed tree, see Supplementary Figure 2.

Figure 2: MoeA distribution in Archaea and Bacteria. (a) Phyletic pattern of the presence of MoeA in Archaea. Higher taxonomic ranks are indicated on the right. **(b)** Maximum-likelihood phylogeny of MoeA in Archaea and Bacteria. Monophyletic groups were collapsed into a single branch. Labels of branches that correspond to a collapsed group have bigger fonts than branches that correspond to single sequences. Bacteria phyla are indicated in green, and Archaea phyla are indicated in pink. Black dots indicate UFB > 90, gray dots indicate 80 < UFB <= 90 and branches without dots indicated UFB <= 80. The scale bar represents the average number of substitutions per site. For the detailed tree, see Supplementary Figure 4. **(c)** Genomic context of a representative archaeal MoeA and a bacterial MoeA that branch within Archaea.

584 **Figure 3: PBP-like domain fusion to archaeal MoeA1. (a)** Schematic representation of the
585 domain organization of MoeA in Archaea and Bacteria mapped on a schematic tree based on
586 the phylogeny in Figure 2b. For the detailed information, see Supplementary Figure 4. **(b)**
587 Alphafold protein structure of the MoeA1 dimer of archaeon *Archaeoglobus fulgidus*.
588 Monomer A is indicated in darker shades than monomer B. MoeA domains are indicated in
589 shades of orange, and the PBP domain is indicated in shades of blue. **(c)** Phyletic pattern of
590 the presence of molybdenum/tungsten related PBP proteins in Archaea. For the detailed
591 information, see Supplementary Table 2. **(d)** Phyletic pattern of the presence of
592 molybdenum/tungsten related PBP proteins in Bacteria. For the detailed information, see
593 Supplementary Table 2.

594

595 **Figure 4: Conservation analysis of the MoeA active site. (a)** On the left, the high-confidence
596 AlphaFold atomic model of the MoeA dimer of representatives of all domains of life. The two
597 symmetric active sites are indicated on the *E. coli* structure. Below, the Glp dimer of *C.*
598 *glutamicum* (PDB id 8bvf), indicating the FtsZ binding sites. On the right, the sequence
599 conservation of the MoeA active site mapped on a representative structure of each group. **(b)**
600 Plot of the first two components from the PCA analysis of the distances between residues
601 involved in the active site of MoeA in the different taxonomic groups. Each colored dot
602 represents a MoeA protein structure. For the detailed list of the distances, see Supplementary
603 Table 3.

604

605 **Figure 5: GlyR binding site conservation in Eukaryotes.**

606 On top, the protein structure of *Rattus norvegicus* gephyrin bound to a GlyR peptide (PDB id
607 4pd1) Below, sequence conservation of the MoeA membrane receptor binding site (dotted
608 circle) mapped on a representative structure of each group indicated on the left. On the right,
609 multiple sequence alignment of three fragments that form the binding site of MoeA/gephyrin
610 that are conserved in most Eukaryotes, but not conserved in Bacteria. Key residues are
611 indicated in red. All positions reported refer to the equivalent positions on E. coli MoeA.

612

Figure 6: **Scenario for the evolutionary history of MoeA in all domains of life.** All evolutionary events inferred in this work were mapped on a schematic phylogenetic tree based on the tree in Supplementary Figure 1.

Supplementary Table 1

Genbank protein ids of MoeA homologs identified in Archaea, Bacteria and Eukaryotes.

Supplementary Table 2

Genbank protein ids of ModA, MoeA-PBP, WtpA and TupA homologs identified in Archaea and Bacteria.

Supplementary Table 3

Pairwise distances between the residues of interest in the MoeA structures used for the PCA analyses.

Supplementary Table 4

List of Genbank protein ids of the selected structures used for mapping the sequence conservation.

Supplementary Table 5

List of residues positions of the active site included in the distance calculation analysis, relative to *E. coli* Genbank protein sequence AIZ54672.1.

637

638 **References**

639

640 Adam, P. S., Borrel, G., Brochier-Armanet, C., & Gribaldo, S. (2017). The growing tree of
641 Archaea: New perspectives on their diversity, evolution and ecology. In *ISME Journal* (Vol.
642 11, Issue 11, pp. 2407–2425). Nature Publishing Group.
643 https://doi.org/10.1038/ismej.2017.122
644 Bevers, L. E., Hagedoorn, P. L., Santamaria-Araujo, J. A., Magalon, A., Hagen, W. R., & Schwarz,
645 G. (2008). Function of MoaB proteins in the biosynthesis of the molybdenum and
646 tungsten cofactors. *Biochemistry*, *47*(3), 949–956. https://doi.org/10.1021/bi7020487
647 Borrok, M. J., Zhu, Y., Forest, K. T., & Kiessling, L. L. (2009). Structure-based design of a
648 periplasmic binding protein antagonist that prevents domain closure. *ACS Chemical*
649 *Biology*, *4*(6), 447–456. https://doi.org/10.1021/cb900021q
650 Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. G. B. (2020). The New Tree of Eukaryotes.
651 In *Trends in Ecology and Evolution* (Vol. 35, Issue 1, pp. 43–55). Elsevier Ltd.
652 https://doi.org/10.1016/j.tree.2019.08.008
653 Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated
654 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–
655 1973. https://doi.org/10.1093/bioinformatics/btp348
656 Castelle, C. J., & Banfield, J. F. (2018). Major New Microbial Groups Expand Diversity and Alter
657 our Understanding of the Tree of Life. In *Cell* (Vol. 172, Issue 6, pp. 1181–1197). Cell Press.
658 https://doi.org/10.1016/j.cell.2018.02.016
659 Choii, G., & Ko, J. (2015). Gephyrin: a central GABAergic synapse organizer. In *Experimental*
660 *and Molecular Medicine* (Vol. 47, Issue 4). Springer Nature.
661 https://doi.org/10.1038/EMM.2015.5
662 Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I.,
663 Hamelryck, T., Kauff, F., Wilczynski, B., & De Hoon, M. J. L. (2009). Biopython: Freely
664 available Python tools for computational molecular biology and bioinformatics.
665 *Bioinformatics*, *25*(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163
666 Etchebehere, C., Zorzopulos, ~ J, Soubes', M., & Muxi', L. (n.d.). Coprothermobacter platensis
667 S p m now, a new anaerobic proteolytic thermophilic bacterium isolated from an
668 anaerobic mesophilic sludge. In *Journal of Systematic Bacteriology* (Vol. 48).
669 Guangsheng, C., Plugge, C. M., Roelofsen, W., Houwen, F. P., & Stams, A. J. M. (1992).
670 Selenomonas acidaminovorans sp. nov., a versatile thermophilic proton-reducing
671 anaerobe able to grow by decarboxylation of succinate to propionate. In *Arch Microblol*
672 (Vol. 157).
673 Hagen, W. R. (2011). Cellular uptake of molybdenum and tungsten. In *Coordination Chemistry*
674 *Reviews* (Vol. 255, Issues 9–10, pp. 1117–1128).
675 https://doi.org/10.1016/j.ccr.2011.02.009
676 Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and
677 iterative HMM search procedure. *BMC Bioinformatics*, *11*.
678 https://doi.org/10.1186/1471-2105-11-431
679 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K.,
680 Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J.,
681 Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly

accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kaufholdt, D., Baillie, C. K., Bikker, R., Burkart, V., Dudek, C. A., Pein, L. von, Rothkegel, M., Mendel, R. R., & Hänsch, R. (2016). The molybdenum cofactor biosynthesis complex interacts with actin filaments via molybdenum insertase Cnx1 as anchor protein in Arabidopsis thaliana. *Plant Science*, *244*, 8–18. https://doi.org/10.1016/j.plantsci.2015.12.011

Kirsch, J., Langosch, D., Littauer, U. Z., Schmitt, B., & Betz, H. (1991). The 93-kDa Glycine Receptor-associated Protein Binds to Tubulin*. In *Inc OF BIOLOGICAL CHEMISTRY* (Vol. 266, Issue 33).

Leimkühler, S. (2020). The biosynthesis of the molybdenum cofactors in Escherichia coli. In *Environmental Microbiology* (Vol. 22, Issue 6, pp. 2007–2026). Blackwell Publishing Ltd. https://doi.org/10.1111/1462-2920.15003

Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), 256–259. https://doi.org/10.1093/nar/gkz239

Looft, T., Levine, U. Y., & Stanton, T. B. (2013). Cloacibacillus porcorum sp. nov., a mucin-degrading bacterium from the swine intestinal tract and emended description of the genus Cloacibacillus. *International Journal of Systematic and Evolutionary Microbiology*, *63*(PART6), 1960–1966. https://doi.org/10.1099/ijs.0.044719-0

López-García, P., & Moreira, D. (2006). Selective forces for the origin of the eukaryotic nucleus. *BioEssays*, *28*(5), 525–533. https://doi.org/10.1002/bies.20413

Maia, L. B., Moura, I., & Moura, J. J. G. (2017). Molybdenum and Tungsten-Containing Enzymes: An Overview. In Russ Hille (Ed.), *Molybdenum and Tungsten Enzymes: Biochemistry*. Royal Society of Chemistry. www.rsc.org

Malotky, E. L. (2002). *Functional Characterization of MoeA and MoeB Tungsten Cofactor Synthesis Proteins from the Hyperthermophilic Archaeon Pyrococcus furiosus*. Graduate Faculty of North Carolina State University.

Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: Protein domain annotations on the fly. *Nucleic Acids Research*, *32*(WEB SERVER ISS.). https://doi.org/10.1093/nar/gkh454

Martinez, M., Petit, J., Leyva, A., Sogues, A., Megrian, D., Rodriguez, A., Gaday, Q., Ben Assaya, M., Portela, M. M., Haouz, A., Ducret, A., Grangeasse, C., Alzari, P. M., Durán, R., & Wehenkel, A. M. (2023). Eukaryotic-like gephyrin and cognate membrane receptor coordinate corynebacterial cell division and polar elongation. *Nature Microbiology*, *8*(10), 1896–1910. https://doi.org/10.1038/s41564-023-01473-0

Mayr, S. J., Mendel, R. R., & Schwarz, G. (2021). Molybdenum cofactor biology, evolution and deficiency. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1868*(1). https://doi.org/10.1016/j.bbamcr.2020.118883

McIlroy, S. J., Kirkegaard, R. H., Dueholm, M. S., Fernando, E., Karst, S. M., Albertsen, M., & Nielsen, P. H. (2017). Culture-independent analyses reveal novel anaerolineaceae as abundant primary fermenters in anaerobic digesters treating waste activated sludge. *Frontiers in Microbiology*, *8*(JUN). https://doi.org/10.3389/fmicb.2017.01134

Megrian, D., Taib, N., Jaffe, A. L., Banfield, J. F., & Gribaldo, S. (2022). Ancient origin and constrained evolution of the division and cell wall (dcw) gene cluster across the Tree of Bacteria. *Nature Microbiology*.

Mendel, R. R., & Leimkühler, S. (2015). The biosynthesis of the molybdenum cofactors. *Journal of Biological Inorganic Chemistry*, *20*(2), 337–347. https://doi.org/10.1007/s00775-014-1173-y

Mota, C. S., Valette, O., González, P. J., Brondino, C. D., Moura, J. J. G., Moura, I., Dolla, A., & Rivas, M. G. (2011). Effects of molybdate and tungstate on expression levels and biochemical characteristics of formate dehydrogenases produced by Desulfovibrio alaskensis NCIMB 13491. *Journal of Bacteriology*, *193*(12), 2917–2923. https://doi.org/10.1128/JB.01531-10

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Pei, J., & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. In *BIOINFORMATICS* (Vol. 17, Issue 8).

Pende, N., Sogues, A., Megrian, D., Sartori-Rupp, A., England, P., Palabikyan, H., Rittmann, S. K. M. R., Graña, M., Wehenkel, A. M., Alzari, P. M., & Gribaldo, S. (2021). SepF is the FtsZ anchor in archaea, with features of an ancestral cell division system. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23099-8

Peng, T., Xu, Y., & Zhang, Y. (2018). Comparative genomics of molybdenum utilization in prokaryotes and eukaryotes. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-5068-0

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses. In *JSS Journal of Statistical Software* (Vol. 17). http://www.jstatsoft.org/

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

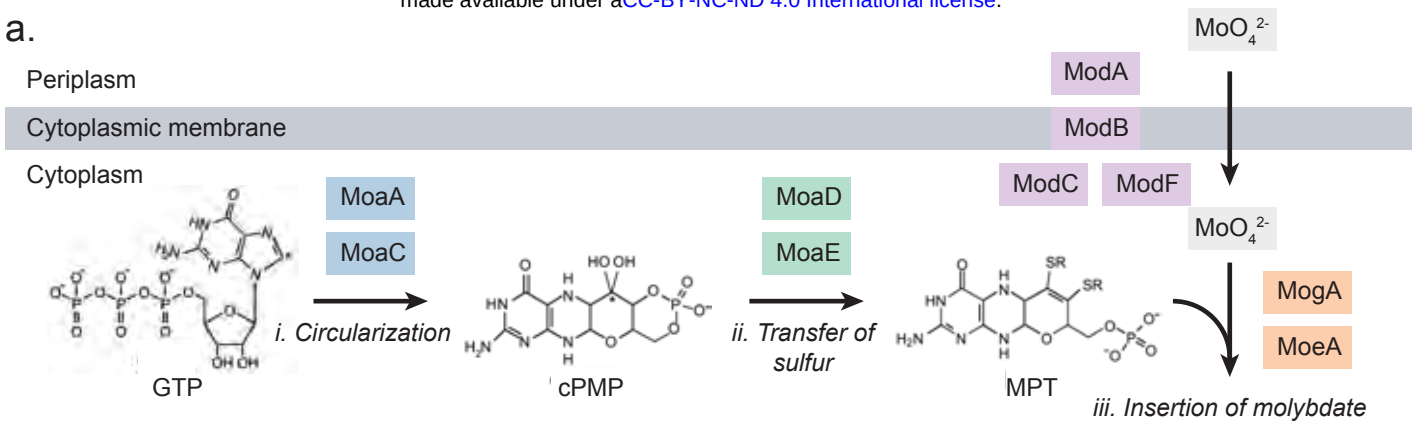
Schwarz, G. (2005). Molybdenum cofactor biosynthesis and deficiency. In *Cellular and Molecular Life Sciences* (Vol. 62, Issue 23, pp. 2792–2810). https://doi.org/10.1007/s00018-005-5269-y

Schwarz, G., Schulze, J., Bittner, F., Eilers, T., Kuper, J., Bollmann, G., Nerlich, A., Brinkmann, H., & Mendel, R. R. (2000). The Molybdenum Cofactor Biosynthetic Protein Cnx1 Complements Molybdate-Repairable Mutants, Transfers Molybdenum to the Metal Binding Pterin, and Is Associated with the Cytoskeleton. In *The Plant Cell* (Vol. 12). https://academic.oup.com/plcell/article/12/12/2455/6009386

Sekiguchi, Y., Yamada, T., Hanada, S., Ohashi, A., Harada, H., & Kamagata, Y. (2003). Anaerolinea thermophila gen. nov., sp. nov. and Caldilinea aerophila gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain bacteria at the subphylum level. *International Journal of Systematic and Evolutionary Microbiology*, *53*(6), 1843–1851. https://doi.org/10.1099/ijs.0.02699-0

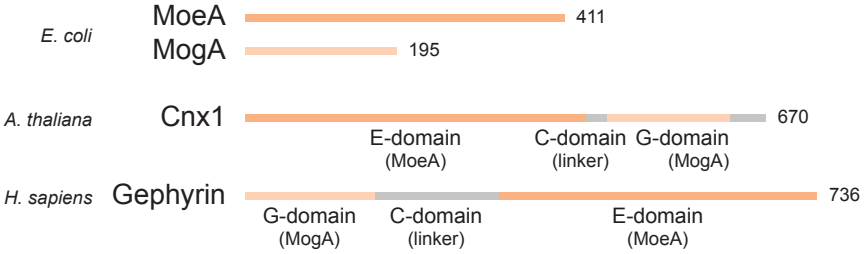
773 Sola, M., Bavro, V. N., Timmins, J., Franz, T., Ricard-Blum, S., Schoehn, G., Ruigrok, R. W. H.,
774 Paarmann, I., Saiyed, T., O'Sullivan, G. A., Schmitt, B., Betz, H., & Weissenhorn, W. (2004).
775 Structural basis of dynamic glycine receptor clustering by gephyrin. *EMBO Journal*,
776 *23*(13), 2510–2519. https://doi.org/10.1038/sj.emboj.7600256
777 Stallmeyer, B., Schwarz, G., Schulze, J., Nerlich, A., Reiss, J., Kirsch, J., & Mendel, R. R. (1999).
778 *The neurotransmitter receptor-anchoring protein gephyrin reconstitutes molybdenum*
779 *cofactor biosynthesis in bacteria, plants, and mammalian cells* (Vol. 96). www.pnas.org.
780 Tyagarajan, S. K., & Fritschy, J. M. (2014). Gephyrin: A master regulator of neuronal function?
781 In *Nature Reviews Neuroscience* (Vol. 15, Issue 3, pp. 141–156).
782 https://doi.org/10.1038/nrn3670
783 Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., & Martin,
784 W. F. (2016). The physiology and habitat of the last universal common ancestor. *Nature*
785 *Microbiology*, *1*(9). https://doi.org/10.1038/nmicrobiol.2016.116
786 Wolf, Y. I., & Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution.
787 *BioEssays*, *35*(9), 829–837. https://doi.org/10.1002/bies.201300037
788 Zeikus, J. G., Dawson, M. A., Thompson, T. E., Ingvorsen, K., & Hatchikian, E. C. (1983).
789 Microbial Ecology of Volcanic Sulphidogenesis: Isolation and Characterization of
790 Thevmodesulfobacteviurn commune gen. nov. and sp. nov. In *Journal of' General*
791 *Microbiology* (Vol. 129).
792 Zhang, Y., Rump, S., & Gladyshev, V. N. (2011). Comparative genomics and evolution of
793 molybdenum utilization. In *Coordination Chemistry Reviews* (Vol. 255, Issues 9–10, pp.
794 1206–1217). https://doi.org/10.1016/j.ccr.2011.02.016
795 Zhong, Q., Kobe, B., & Kappler, U. (2020). Molybdenum Enzymes and How They Support
796 Virulence in Pathogenic Bacteria. In *Frontiers in Microbiology* (Vol. 11). Frontiers Media
797 S.A. https://doi.org/10.3389/fmicb.2020.615860

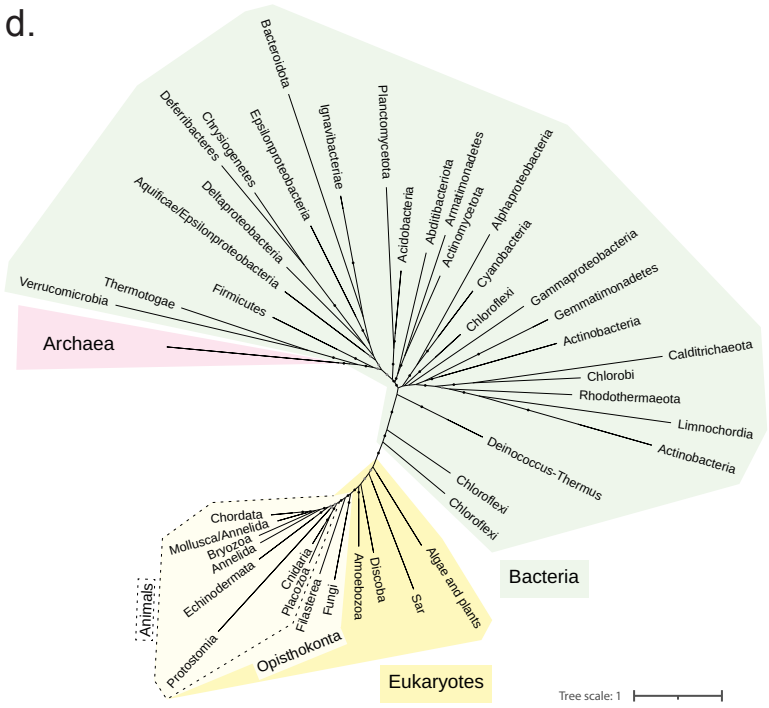
798



b.

E. coli	*A. thaliana*	*H. sapiens*
MoaA MoaC	Cnx2 Cnx3	MOCS1
MoaD MoaE	Cnx6 Cnx7	MOCS2
MogA MoeA	Cnx1	Gephyrin





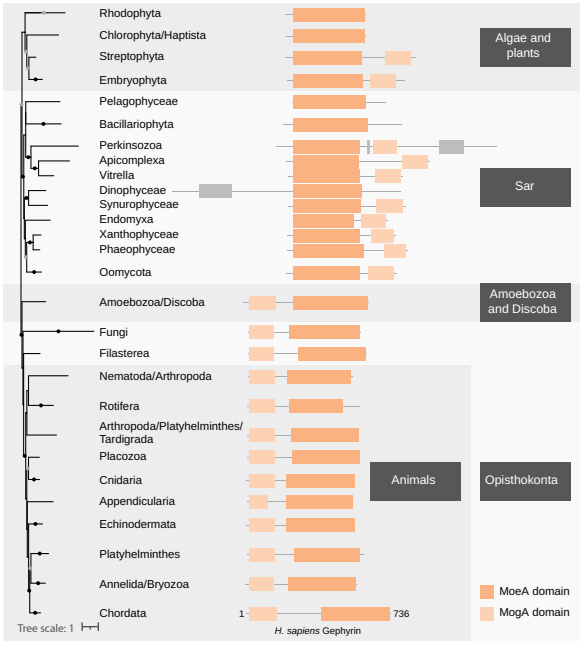


Figure 1: Proteins involved in the biosynthesis of Moco. (a) Schematic representation of the steps involved in the entrance of molybdenum (MoO_4^{2-}) to the cell, and in the biosynthesis of Moco in *E. coli*. Protein names are indicated in colored boxes. **(b)** Comparison of proteins involved in Moco biosynthesis in a representative species of Bacteria (*E. coli*), plants (A. thaliana) and animals (H. sapiens). **(c)** Detail of the organization of MoeA and MogA domains in a representative species of Bacteria (*E. coli*), plants (*A. thaliana*) and animals (*H. sapiens*). Each line corresponds to an individual protein. Numbers indicate the length of the protein. Domain MoeA is indicated in dark green, and domain MogA in light green. **(d)** Maximum-likelihood phylogeny of MoeA/Gephyrin in Eukaryotes, Bacteria and Archaea. Monophyletic groups were collapsed into a single branch. Black dots indicate UFB > 90, gray dots indicate 80 < UFB <= 90 and branches without dots indicated UFB <= 80. The scale bar represents the average number of substitutions per site. For the detailed tree, see Supplementary Figure 1. **(e)** Domain organization of MoeA/Gephyrin in representative species of Eukaryotes. Domains indicated in gray correspond to domains different to MoeA or MogA. Higher taxonomic ranks are indicated in black boxes. The scale bar represents the average number of substitutions per site. For the detailed tree, see Supplementary Figure 2. 1

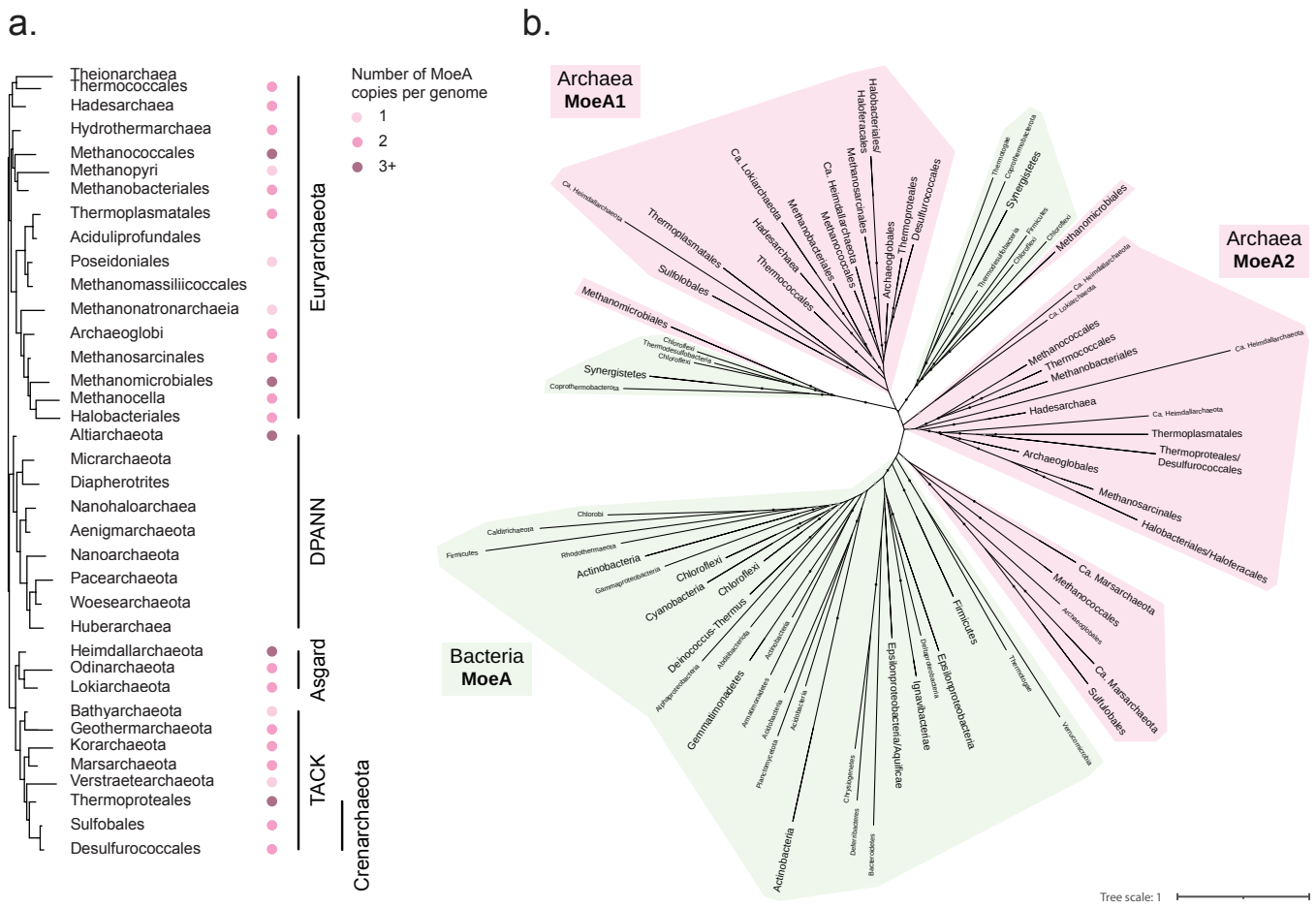


Figure 2: MoeA distribution in Archaea and Bacteria. **(a)** Phyletic pattern of the presence of MoeA in Archaea. Higher taxonomic ranks are indicated on the right. **(b)** Maximum-likelihood phylogeny of MoeA in Archaea and Bacteria. Monophyletic groups were collapsed into a single branch. Labels of branches that correspond to a collapsed group have bigger fonts than branches that correspond to single sequences. Bacteria phyla are indicated in green, and Archaea phyla are indicated in pink. Black dots indicate UFB > 90, gray dots indicate 80 < UFB <= 90 and branches without dots indicated UFB <= 80. The scale bar represents the average number of substitutions per site. For the detailed tree, see Supplementary Figure 4. **(c)** Genomic context of a representative archaeal MoeA and a bacterial MoeA that branch within Archaea.

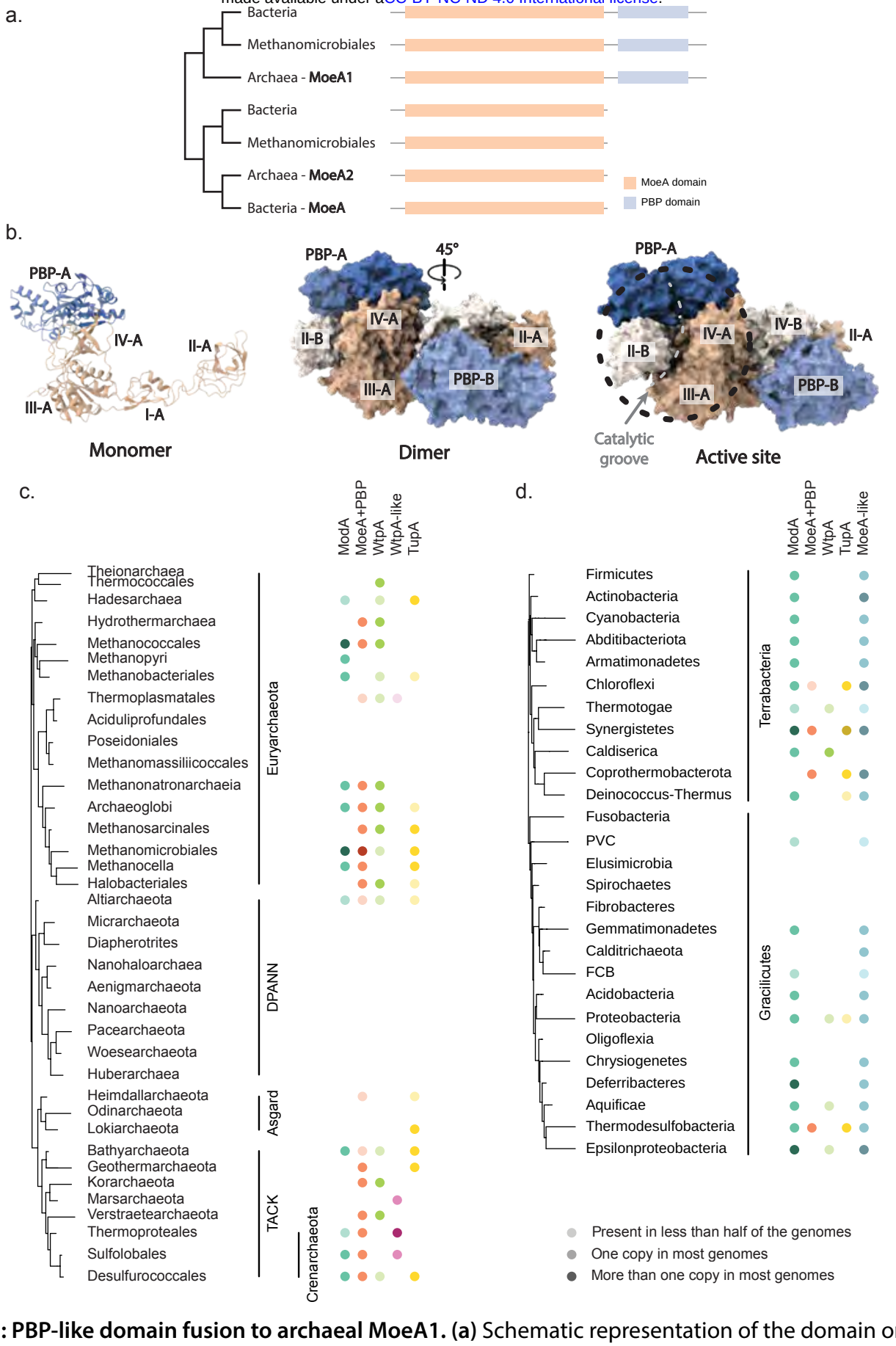


Figure 3: PBP-like domain fusion to archaeal MoeA1. **(a)** Schematic representation of the domain organization of MoeA in Archaea and Bacteria mapped on a schematic tree based on the phylogeny in Figure 2b. For the detailed information, see Supplementary Figure 4. **(b)** Alphafold protein structure of the MoeA1 dimer of archaeon *Archaeoglobus fulgidus*. Monomer A is indicated in darker shades than monomer B. MoeA domains are indicated in shades of orange, and the PBP domain is indicated in shades of blue. **(c)** Phyletic pattern of the presence of molybdenum/tungsten related PBP proteins in Archaea. For the detailed information, see Supplementary Table 2. **(d)** Phyletic pattern of the presence of molybdenum/tungsten related PBP proteins in Bacteria. For the detailed information, see Supplementary Table 2.

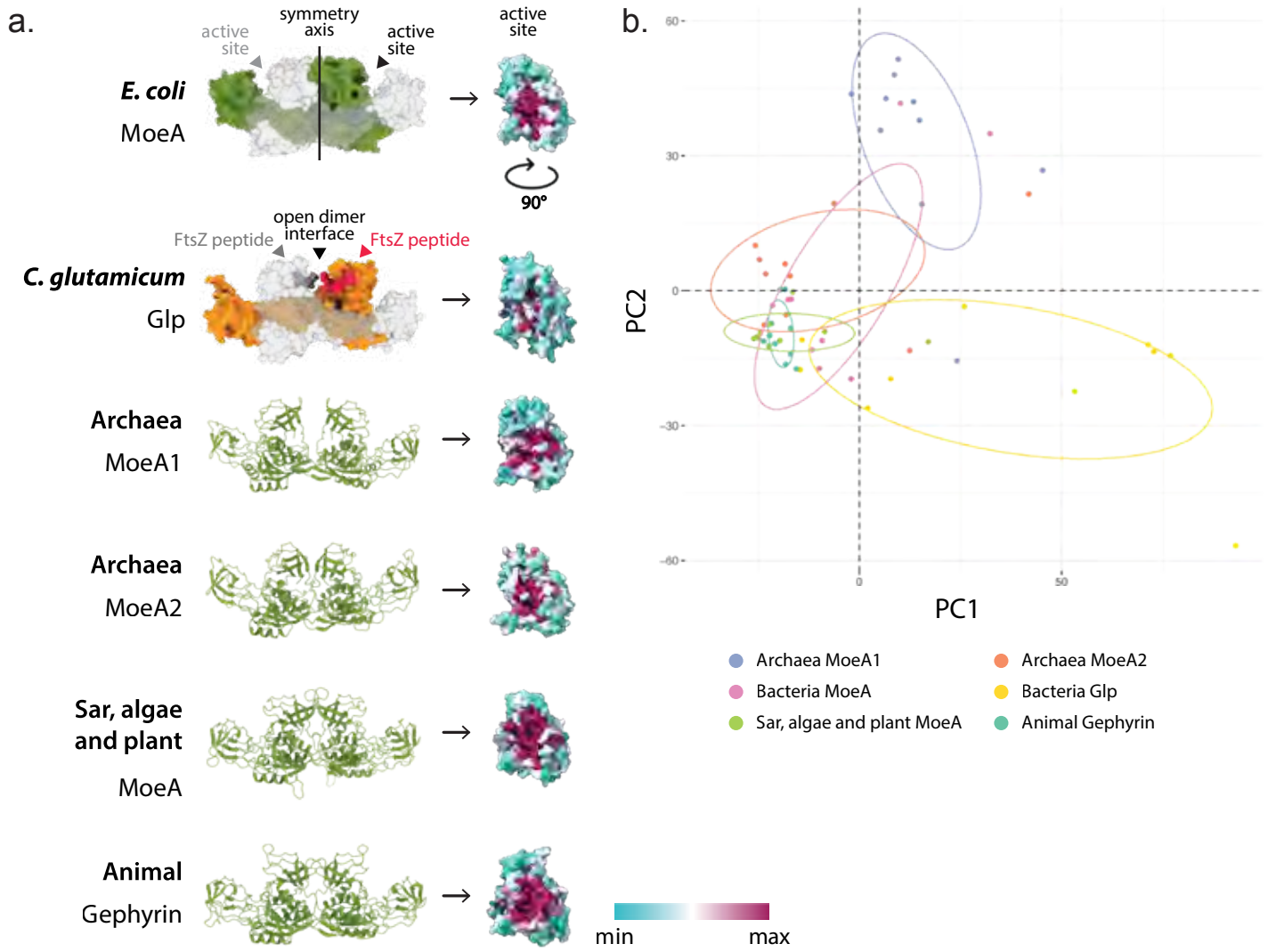


Figure 4: Conservation analysis of the MoeA active site. (a) On the left, the high-confidence AlphaFold atomic model of the MoeA dimer of representatives of all domains of life. The two symmetric active sites are indicated on the *E. coli* structure. Below, the Glp dimer of *C. glutamicum* (PDB id 8bvf), indicating the FtsZ binding sites. On the right, the sequence conservation of the MoeA active site mapped on a representative structure of each group. **(b)** Plot of the first two components from the PCA analysis of the distances between residues involved in the active site of MoeA in the different taxonomic groups. Each colored dot represents a MoeA protein structure. For the detailed list of the distances, see Supplementary Table 3.

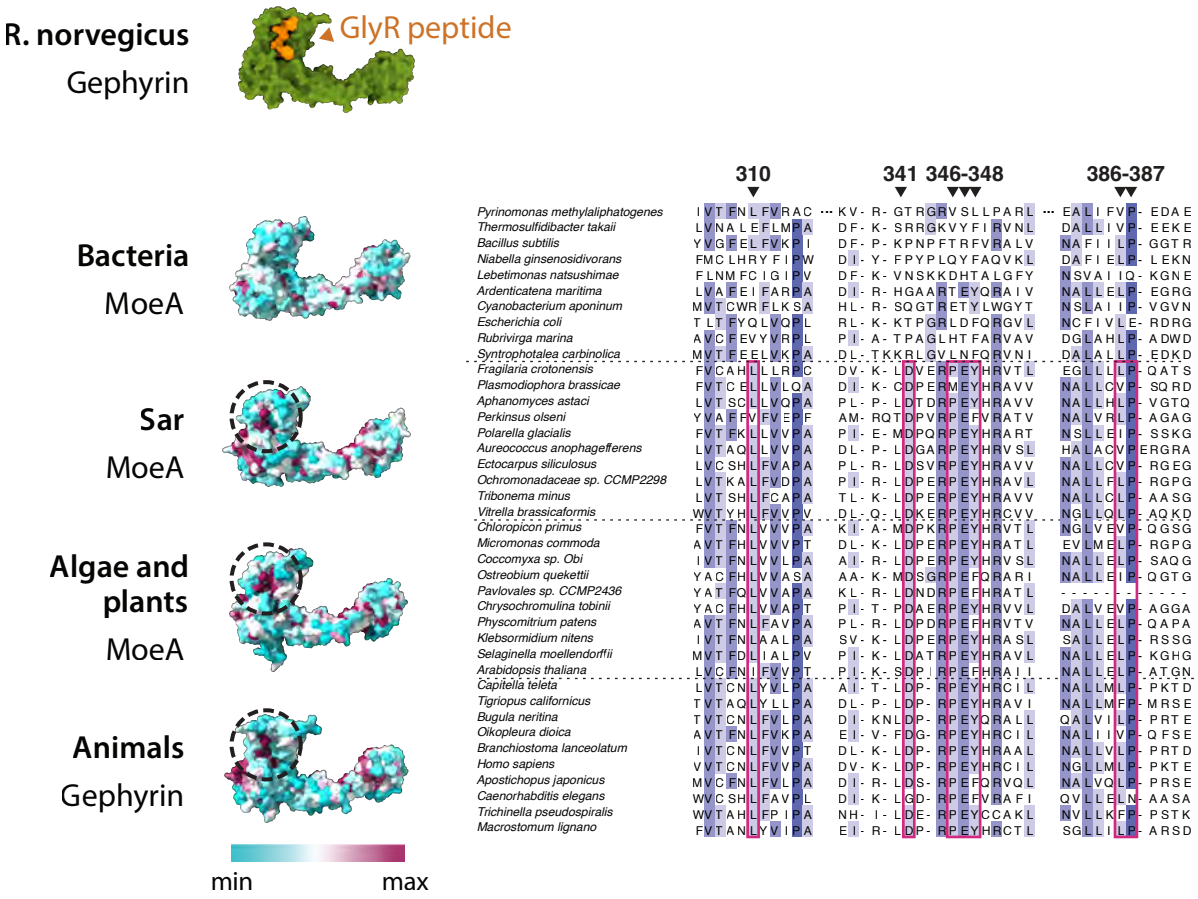


Figure 5: GlyR binding site conservation in Eukaryotes. On top, the protein structure of *Rattus norvegicus* gephyrin bound to a GlyR peptide (PDB id 4pd1) Below, sequence conservation of the MoeA membrane receptor binding site (dotted circle) mapped on a representative structure of each group indicated on the left. On the right, multiple sequence alignment of three fragments that form the binding site of MoeA/gephyrin that are conserved in most Eukaryotes, but not conserved in Bacteria. Key residues are indicated in red. All positions reported refer to the equivalent positions on *E. coli* MoeA.

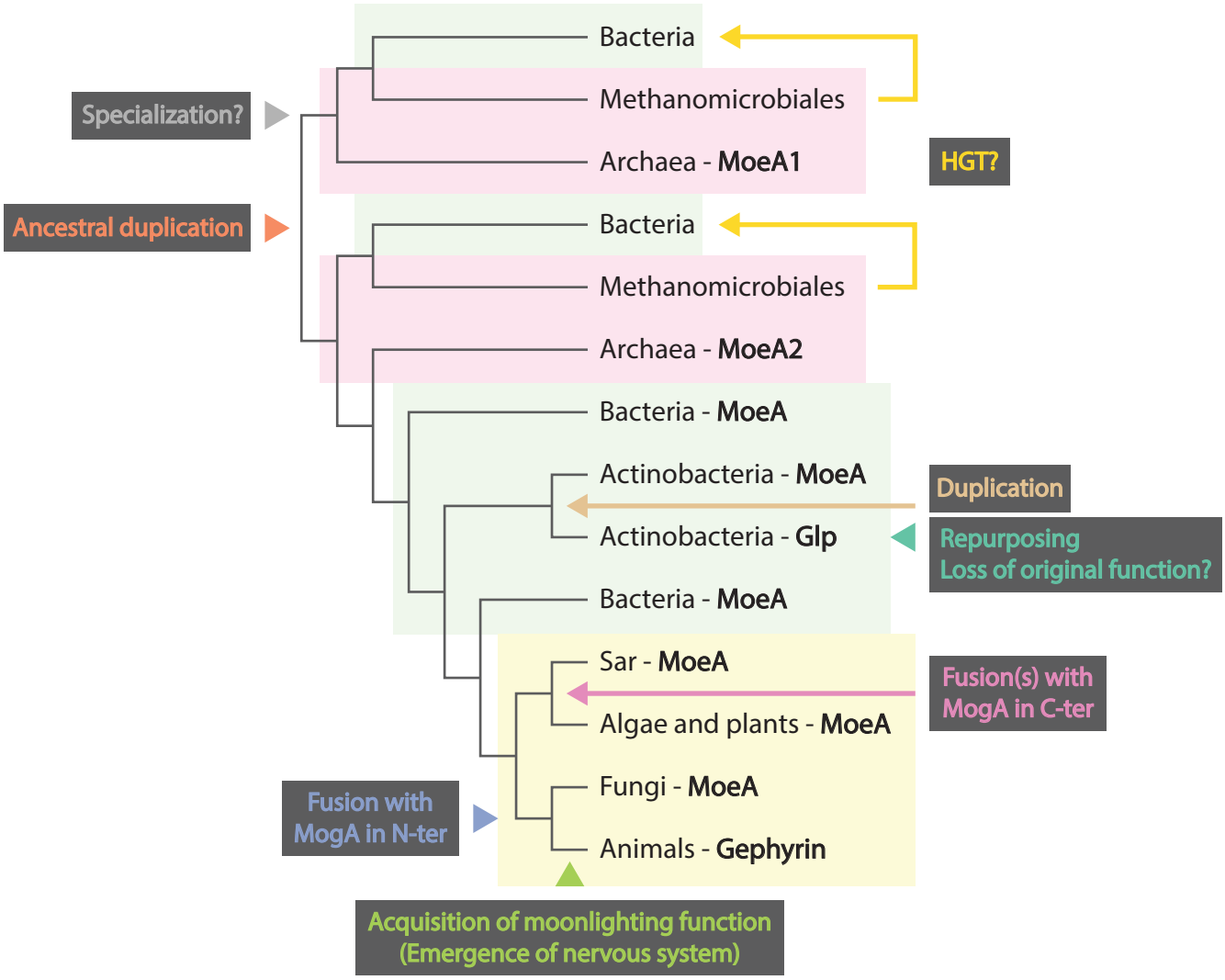


Figure 6: Scenario for the evolutionary history of MoeA in all domains of life. All evolutionary events inferred in this work were mapped on a schematic phylogenetic tree based on the tree in Supplementary Figure 1.