An archaeal origin of the Wood-Ljungdahl H₄MPT branch and the emergence of bacterial methylotrophy

Authors: Panagiotis S. Adam^{1,2,†}, Guillaume Borrel¹, Simonetta Gribaldo^{1*}

Affiliations:

¹Department of Microbiology, Unit Evolutionary Biology of the Microbial Cell, Institut Pasteur, Paris, France

²Université Paris Diderot, Sorbonne Paris Cité, Paris, France

[†]Present address: Group for Aquatic Microbial Ecology, Biofilm Center, Faculty of Chemistry, University of Duisburg-Essen, Essen, Germany

*Correspondence to: simonetta.gribaldo@pasteur.fr

The tetrahydromethanopterin (H₄MPT) methyl branch of the Wood-Ljungdahl (WL) pathway is shared by archaeal and bacterial metabolisms that greatly contribute to the global carbon budget and greenhouse gas fluxes: methanogenesis and methylotrophy, including methanotrophy¹⁻³. It has been proposed that the H₄MPT branch dates back to the Last Universal Common Ancestor (LUCA)⁴⁻⁶. Interestingly, it has been recently identified in a number of recently sequenced and mostly uncultured non-methanogenic and non-methylotrophic archaeal and bacterial lineages, where its function remains unclear^{5,7}. Here, we have examined the distribution and phylogeny of the enzymes involved in the H₄MPT branch and the biosynthesis of its cofactors in over 6400 archaeal and bacterial genomes. We find that a full WL-H₄MPT pathway is widespread in Archaea and likely ancestral to this domain, whereas this is not the case for Bacteria. Moreover, the inclusion

of recently sequenced lineages leads to an important shortening of the branch separating Archaea and Bacteria with respect to previous phylogenies of the H₄MPT branch. Finally, the genes for the pathway are co-localized in many of the recently sequenced archaeal lineages, similar to bacteria. Together, these results weaken the LUCA hypothesis and rather favor an origin of the H₄MPT branch in Archaea and its subsequent transfer to Bacteria. We propose a scenario for its potential initial role in the first bacterial recipients, and its evolution up to the emergence of aerobic methylotrophy. Finally, we discuss how an ancient horizontal transfer not only triggered the emergence of key metabolic processes but also important transitions in Earth's history.

The WL pathway consists of the reversible reduction of CO₂ into the carbonyl and methyl moieties of acetyl-CoA through two separate (carbonyl and methyl) branches⁸ (Figure 1). There exist two versions of the methyl branch, one using tetrahydrofolate (H₄F), and the other methanofuran (MF) and tetrahydromethanopterin (H₄MPT) as cofactor carriers of reduced carbon. Archaea use the H₄MPT methyl branch either in the reductive direction for carbon fixation and/or CO₂-reducing methanogenesis, or in reverse for anaerobic acetyl-CoA or methane oxidation⁷. Other uses have been proposed, such as participation in the oxidation of short-chain alkanes^{3,9}. Five enzymatic steps are involved: in the reductive direction, (1) CO₂ is reduced to MF-bound formyl (Fwd enzyme complex), (2) formyl is transferred to H₄MPT (Ftr), cyclized into (3) methenyl (Mch), and then progressively reduced to (4) methylene (Mtd/Hmd) and (5) methyl (Mer) (Figure 1). Then, the H₄MPT-bound methyl is either transferred to coenzyme M and reduced to methane (methanogenesis), or combined with CO and CoA to form acetyl-CoA in the carbonyl branch of the WL pathway (CO₂ fixation). In methylotrophic Proteobacteria, the H₄MPT branch is entered at methylene-H₄MPT through the formaldehyde-activating enzyme

(Fae) and operates in the oxidative direction (via bMtd and Mch) before being diverted into the H₄F methyl branch of the WL pathway (Fhc), the Serine cycle, and the Calvin cycle (Figure 1). The H₄MPT branch appears as a module in aerobic methylotrophy or as standalone thus playing roles in carbon assimilation, formaldehyde detoxification, and energy conservation¹⁰.

The initial discovery of the H₄MPT branch in methylotrophic Proteobacteria 20 years ago gave rise to two alternative hypotheses on its evolutionary origin: either it was acquired via horizontal transfer from methanogenic archaea or it was inherited from the Last Universal Common Ancestor (LUCA)^{11,12}. The LUCA scenario gained traction when the H₄MPT branch was also found in Planctomycetes, where it is supposed to perform formaldehyde oxidation⁴. Two main arguments were put forward: (i) the distinct separation between Bacteria and Archaea in phylogenies of some H₄MPT components (Figure 2a) and cofactor biosynthesis genes, and (ii) the lack of gene co-localization in archaeal methanogens as opposed to Bacteria, making a concurrent transfer inherently difficult^{4,13,14}. However, considering the large number of independent losses of the pathway in Bacteria implied by the LUCA hypothesis, the transfer scenario has remained a valid alternative¹³.

The presence of an H₄MPT branch has since been highlighted in members of several bacterial lineages. These include some Gemmatimonadetes, proposed to be aerobic methylotrophs¹⁵, and the candidate division NC10, including *Candidatus* Methylomirabilis oxyfera which is anaerobic and performs methane oxidation coupled to denitrification¹⁶. Nonetheless, many bacteria with an H₄MPT branch are non-methylotrophic and their use of the pathway is unknown^{2,5}. Even among Proteobacteria there exist some non-methylotrophs (e.g. members of the Burkholderiales) which probably use the branch for formaldehyde oxidation¹⁰. The H₄MPT branch has also been recently found in a number of uncultured, non-methanogenic archaea, including Persephonearchaea

(MSBL1), Hadesarchaea, Theionarchaea, Altiarchaeales, Thorarchaeota, Bathyarchaeota, and Lokiarchaeota⁹. This wider distribution has been proposed to support the LUCA scenario^{2,5}. Moreover, a recent survey of genes potentially present in the LUCA found some components of the H₄MPT methyl branch, although not the complete pathway⁶. However, a comprehensive phylogenetic analysis including all the recently discovered lineages has been lacking, and the evolution of several H₄MPT and MF biosynthetic genes has never been studied.

By screening a local databank of 643 archaeal and 5,796 bacterial genomes, we could infer the presence of a complete or almost complete H₄MPT methyl branch in the majority of analyzed archaeal lineages, while its distribution in Bacteria appears to be much more restricted (Supplementary Data 1). We found bacterial homologs of the Fwd complex, Ftr, and Mch, but no homologs of Mtd. Concerning Mer, it is a member of the vast luciferase-like family, and we could not identify it from our homology searches. However, it has never been reported in Bacteria. Although most single gene phylogenies could not be completely resolved because of poor sequence conservation, a few intradomain transfers, and the presence of paralogous copies (Supplementary Figures 1-7), three subunits of the Fwd complex (FwdABC) had a congruent history and could be concatenated to increase their phylogenetic signal (Methods). The FwdABC phylogeny is overall resolved (Figure 2b and Supplementary Figure 8). Interestingly, the addition of the recently sequenced lineages leads to a shortening of the branch separating Archaea and Bacteria to one tenth and one fifth with respect to that obtained using lineages that were available in 1999¹² and 2004^{4,13} respectively (Figure 2a). Such a short branch is unlike those usually observed for genes inferred in the LUCA (see for example¹⁷). Similar results were obtained for Mch (1999: 0.8941, 2004: 0.3426, this study: 0.0793 substitutions per site, respectively). Strikingly, the genes for the H₄MPT branch are often co-localized in the newly added archaeal

lineages (Figure 2c), consistently with recent observations². This is unlike methanogens but similar to Bacteria, including the non-methylotrophs (Figure 2c). These results (evolutionary proximity of archaeal and bacterial homologs and gene co-localization in archaea) weaken the two main arguments previously put forward in favor of the LUCA scenario, and rather support a horizontal transfer.

Regarding the direction of the transfer, a number of arguments favor one from Archaea to Bacteria. The archaeal part of the FwdABC phylogeny is overall consistent with the known relationships within this domain¹⁸, in particular the presence of clades corresponding to the Euryarchaeota and TACK-Asgard and the monophyly of major lineages within (Figure 2b). Despite being less resolved and containing a few horizontal gene transfers, the phylogenies of the other components of the pathway (Ftr, Mch, Mtd, Mer) also show overall similar clades, suggestive of a similar evolutionary history (Supplementary Figures 1, 2, 4 and 6, and their legends for additional discussion). Coupled with the wide taxonomic distribution of the H₄MPT methyl branch in Archaea (Supplementary Data 1), this strongly suggests its presence in the Last Archaeal Common Ancestor (LACA). In addition, almost all archaeal lineages that possess the H₄MPT branch also have the carbonyl branch (Supplementary Data 1), granting them a complete WL pathway. In contrast, the two branches never coexist in Bacteria, with the exception of some Firmicutes and one δ -Proteobacteria which are homoacetogens/acetate oxidizers^{19,20} (Supplementary Data 1). Because a functional carbonyl branch was likely present in the LUCA, and therefore in the ancestor of all Bacteria²¹, this lack of co-occurrence could be explained by the fact that the H₄MPT branch in Bacteria was indeed acquired later from Archaea.

Under any of the current three alternative rootings of the archaeal phylogeny (i.e. in between Euryarchaeota and the TACK, at the DPANN, or within Euryarchaeota; see¹⁸ and references

therein), the bacterial homologs stem from within the archaeal radiation, and in particular close to the Altiarchaeales, Hadesarchaea, Persephonearchaea (MSBL1), and Theionarchaea (Figure 2b). The hypothesis of a transfer from Archaea to Bacteria is also supported by the overall similar topologies of the Ftr and Mch trees (Supplementary Figures 1 and 2) to the FwdABC tree.

An argument against the transfer scenario could be the fact that until recently, Archaea were known to use enzymes non-homologous to the bacterial ones for the methylene-methenyl oxidoreduction (Mtd and Hmd). Surprisingly though, several members of the recently sequenced archaeal lineages analyzed here possess a bMtd homolog together with or instead of Mtd (Supplementary Data 1). Also, these bMtd homologs are commonly clustered in both archaeal and bacterial genomes with the other H₄MPT branch genes (Figure 2c). Although the bMtd phylogeny is poorly resolved (Supplementary Figure 3), these results suggest that bMtd was also likely part of the original transfer from Archaea to Bacteria.

Under the hypothesis that the H₄MPT branch was transferred from Archaea to Bacteria, it is expected that this event would have involved also enzymes necessary for the biosynthesis of the MF and H₄MPT cofactors. It should be mentioned that many steps of folate and H₄MPT biosynthesis are catalyzed by different enzymes across Bacteria and Archaea. Some of the genes involved in H₄MPT cofactors biosynthesis in Archaea (e.g. MptD, MptE) are present in a few bacteria but generally not in the ones that possess the H₄MPT branch²², and thus they were omitted from our analysis.

Along with the discovery of the H₄MPT branch in Planctomycetes, it was observed that in Bacteria these genes are clustered with some genes related to methanol oxidation and to the biosynthesis of pterin cofactors^{4,13,14,23}. Two genes encode dihydromethanopterin reductase

(DmrX/DmrB/AfpA) and β -ribofuranosylaminobenzene 5'-phosphate synthase (MptG). Ten other genes had unknown or only predicted functions and were referred to as numbered open reading frames (ORFs)^{4,13,14,23}. The phylogenies of some of these ORFs and MptG were found to be similar to those of the H₄MPT branch components^{4,13,14}, suggesting a common evolutionary history. Mutagenesis experiments and bioinformatic annotation in *Methylobacterium extorquens* suggested a role in H₄MPT biosynthesis for most of these ORFs^{14,23}. Since then, the genes for MF and to some extent H₄MPT cofactor biosynthesis have been characterized in the archaeon *Methanocaldococcus jannaschii* and other methanogens, and their correspondence to the *M. extorquens* ORFs has been assigned (see Supplementary Figures 9-10 and their legends for additional discussion and references).

Similar to the H₄MPT branch, we analyzed the distribution and phylogeny of these twelve genes (10 ORFs, DmrX/DmrB/AfpA, MptG) in our archaeal and bacterial genomes (Supplementary Data 1 and Supplementary Figures 11-20). Seven of the original ORFs correspond to the characterized enzymes in the MF (Orf1/MfnD, Orf9/MfnF, Orf21/MnfE, Orf22/MnfB) and putatively in the H₄MPT (Orf5/MptN, Orf20, OrfY) biosynthesis pathways (Figure 2c, Supplementary Figures 9-10 and their legends for additional discussion and references). Moreover, we could putatively assign two of the remaining ORFs (Orf7, 17) to yet uncharacterized reactions of H₄MPT biosynthesis, allowing us to propose an almost complete pathway (Figure 2c, Supplementary Figure 10 and its legend for additional discussion and references).

Despite the lack of resolution and the presence of horizontal transfers, the phylogenies of these enzymes (Supplementary Figures 11-20) are very similar to those of the H₄MPT branch. This suggests that the same horizontal transfer from Archaea might have introduced the H₄MPT

branch in bacteria along with the necessary cofactor biosynthesis enzymes, an event that would have involved at least 18 genes. This scenario is also supported by co-localization of cofactor biosynthesis and H₄MPT branch genes in both Bacteria and Archaea (Figure 2c).

Interestingly, most cofactor biosynthesis genes have a wide distribution in Archaea with respect to Bacteria (Supplementary Data 1). Their presence in various Crenarchaeota and -for some- in Thermococcales has been known^{4,13} but we also found homologs in Heimdallarchaeota and Aigarchaeota (Supplementary Data 1). These genes might represent functionally-repurposed remnants of an ancestral H₄MPT branch which was lost in these lineages.

It has been proposed that bacterial methylotrophy likely arose from the combination of different modules, with the H₄MPT branch predating it^{2,10}. The inclusion of current bacterial and archaeal diversity in our analysis allows us to draw a more comprehensive scenario for the assembly of methylotrophy around the H₄MPT branch (Figure 3). From our results, and as discussed above, it is very likely that a complete WL pathway was already present in the LACA (Figure 3, black dot). The H₄MPT branch (Fwd, Ftr, Mch, and possibly bMtd) plus the cofactor biosynthesis enzymes (but not the carbonyl branch) were then transferred to Bacteria (Figure 3, red dot), and then spread through a combination of vertical inheritance followed by multiple independent losses and further intra-domain transfers. Planctomycetes seem to have acquired at least part of the pathway (Fwd, Ftr) independently two or three times (Figure 2b and Supplementary Figure 8, and Supplementary Figure 1, respectively). A key event was the addition of Fae to the pathway before the divergence of Planctomycetes and the NC10 phylum (Figure 3, green dot). As Fae allows linkage of methyl compound oxidation into formaldehyde to the H₄MPT branch, this event would have led to the use of the pathway for formaldehyde oxidation and eventually for (denitrifying) methylotrophy in NC10. There exist multiple Fae homologs (Fae-like proteins) in

Bacteria (Supplementary Data 1), but the canonical Fae is likely the original component, as it co-localizes with the H₄MPT genes (Figure 2c), is widespread in methylotrophs (Supplementary Data 1), and is the only one whose mutant cannot be complemented by other homologs^{4,14}. The Fae homologs in bacteria might also have originated through subsequent transfers from Archaea, where the enzyme is far more widespread (Supplementary Data 1 and Supplementary Figure 7). At this point, methylene-H₄MPT was still fully oxidized to CO₂. The next milestone would have been the loss of the FwdD subunit (Figure 3, orange dot). In organisms without FwdD, like α , β , γ -Proteobacteria and most Planctomycetes, the FwdABC subunits and Ftr form a complex called Fhc^{13,24} (Figure 1). Fhc performs the formyl transfer to MF analogs and then lyses it to formate instead of oxidizing it fully to CO₂, since it probably lacks the formate dehydrogenase activity of the FwdBD subcomplex²⁵. Formate fluxes could be diverted to CO₂ (using Fdh) and to methylene-H₄F, to be used in the Serine Cycle or Calvin Cycle²⁶, thus creating the link between the H₄MPT branch and the methylotrophic assimilation pathways (Figure 1), leading to the emergence of contemporary aerobic methylotrophy (Figure 3).

Almost all archaeal lineages that possess the H₄MPT branch also have the carbonyl branch, granting them a complete WL pathway, while this is not the case in Bacteria (Supplementary Data 1). Thus, one major question that remains to be answered is what the potential function of the H₄MPT methyl branch in the pre-methylotrophic (without Fae) bacterial lineages might have been⁵. In the phylogenies of the H₄MPT branch enzymes, the bacterial lineages closest to the putative initial transfer are the Limitata, Synergistetes, and Firmicutes (Clostridia, Halanaerobiales) (Figure 2b and Supplementary Figures 2, 3, 8).

Limitata are a deep-branching uncultured clade related to Synergistetes (Supplementary Figure 21) and are currently represented by two mostly complete (GBS1: 100%, 42_11: 84.1%) metagenome-assembled genomes (MAGs) placed under NCBI's "Unclassified Bacteria".

We analyzed the metabolic potential of the Limitata MAGs to gain some clues about the role of the H₄MPT branch (Figure 4, Supplementary Data 2). In Limitata, the H₄MPT branch might potentially work in either the oxidative or the reductive direction. In the reductive direction, the branch could serve as a carbon-fixing electron sink during fermentative growth, as is the case in some Firmicutes²⁷. Recently it was also proposed that an uncultured member of the δ -Proteobacteria can autotrophically fix carbon through the H₄F branch leading into the Glycine Cleavage System (GCS)²⁸. However, given that in other bacteria (Proteobacteria, Planctomycetes, NC10) the H₄MPT branch operates oxidatively, a reductive direction seems unlikely in Limitata.

In the oxidative direction, methylene is fully oxidized to CO₂. In the absence of Fae, though formaldehyde can spontaneously condense with H₄MPT at low rates²⁹, the most plausible way to link methylene-H₄MPT with the rest of the metabolism (amino acid fermentation, pyruvate from glycolysis) is through serine and glycine (Figure 4). Normally, one methylene-H₄F is produced during the conversion of serine to glycine by the serine hydroxymethyltransferase, and then the GCS converts the glycine to another methylene-H₄F, CO₂, and NH₃. Nevertheless, according to our inference, in Limitata the serine hydroxymethyltransferase and the GCS should be able to produce methylene-H₄MPT instead of methylene-H₄F (Figure 4). For serine hydroxymethyltransferase, this indeed seems to occur in Archaea, such as *Methanosphaera stadmanae* and *Methanococcus thermolithotrophicus*³⁰. The full oxidation process is expected to produce five reducing equivalents per serine molecule, or three per glycine molecule.

Limitata possess Hox-like and Mbh-like hydrogenases that are possibly used for hydrogen production to regenerate cofactors and dispose of excess reductants resulting from fermentation³¹, including those that would be produced by the H₄MPT branch working in the oxidative direction³¹ (Figure 4). In lineages with the H₄MPT branch, such as *Halanaerobium* (Firmicutes) and *Anaerobaculum* (Synergistetes, the closest relatives of Limitata) (Supplementary Data 1), hydrogen production associated to fermentation has also been reported. In these bacteria, presence of the H₄MPT reactions in addition to normal fermentation might be related to their sizeable biohydrogen yield which is consumed when co-cultured with a methanogen^{32,33}. Hydrogen production ties in to the ecological roles of these bacteria: for example *Anaerobaculum* has been found in oil pipes where it seems to act as a syntrophic partner to a methanogen³⁴. Since Limitata genomes originate from environments similar to those of *Anaerobaculum* and *Halanaerobium* (oil fields, hot spring sediments), they probably fulfill a similar role. From these considerations, one possibility is that the original role of the H₄MPT methyl branch in non-methylotrophic Bacteria was to increase their hydrogen production, which in turn would promote CO₂ fixation by surrounding bacteria or archaea with a WL pathway and would lead to higher biomass and carbon turnover.

Methane and carbon dioxide, two end products of the WL pathway, represent centerpieces of the global carbon cycle, and have been of crucial importance to global climate change through geological time^{1,35}. Having determined the presence of a complete WL pathway in the LACA (²¹ and this study) lends credibility to previous work that places the origin of methanogenesis at 3.46 Ga or even earlier³⁶. In contrast, the origin of methylotrophy is harder to determine. In the past, the extremely ¹³C-depleted (-45 to -60‰) isotopic signatures found in the late Archean have been referred to as the “Age of Global Methanotrophy” and attributed over time to aerobic

methylotrophs, archaeal anaerobic methane oxidizers (ANME), and more recently to multiple uses of the WL pathway, including acetogenesis³⁷. However, methylotrophs do not produce carbon isotopic signatures in that range^{38,39}.

In recent years, the taxonomic range of (*bona fide*) bacterial methylotrophs has expanded through the discovery of methanol dehydrogenases and the H₄MPT branch in lineages outside the Proteobacteria, namely Gemmatimonadetes and Acidobacteria (^{2,15} and this study). This origin of aerobic methylotrophy seems to correspond to the loss of FwdD (Figure 4) and is exemplified by these lineages sometimes being on long branches in our phylogenies, possibly resulting from rapid adaptation (Figure 2b, Supplementary Figures 2, 11). This is especially interesting in the case of α,β,γ -Proteobacteria, as some divergence dating studies have placed their origin close to 2.5 Ga⁴⁰. It could then be speculated that the emergence of aerobic methylotrophy is linked to the Great Oxygenation Event (GOE) and increased oxygen concentrations in the atmosphere. Shortly after the GOE, global temperatures plummeted, causing the Paleoproterozoic glaciations. In some models, this climate shift was driven by the collapse of the methane greenhouse that contributed to sustaining the temperature of the Archean Earth⁴¹. Various reasons for this collapse have been proposed, such as a strong decrease of methanogen populations due to nickel depletion⁴², and/or an increase in anaerobic methane oxidation spurred by higher sulfate levels⁴³. More recently, the contribution of aerobic methanotrophy as a suppressor of methane fluxes slightly prior to the GOE has also been considered⁴⁴. The emergence of aerobic methylotrophy, which includes methanotrophy, near the GOE could have therefore contributed to the collapse of the methane greenhouse.

In conclusion, our study underlines how an ancient inter-domain transfer of the H₄MPT methyl branch and its cofactor biosynthesis not only triggered the emergence of key metabolic processes in the carbon cycle but also important transitions in the history of our planet.

Methods

Homology searches

A local databank of genome sequences from 643 Archaea and 5,796 Bacteria was assembled, representing all genomes (one per species) present at the National Center for Biotechnology Information (NCBI) as of June 2016, and included some retroactively added Bacteria and Archaea recently shown to possess genes of the H₄MPT pathway, some genomes manually downloaded from the Joint Genome Institute/Integrated Microbial Genomes (JGI/IMG)⁴⁵ and the ggkbase⁴⁶ (ggkbase.berkeley.edu/). For some genomes with only nucleotide sequences, ORFs were predicted by using Prodigal⁴⁷ with default parameters. A full list of taxa used to build this databank is provided in Supplementary Data 3.

Homology searches for the components of the H₄MPT branch and MF and H₄MPT biosynthesis genes (FwdA, B, C, D, Ftr, Mch, bMtd, Mtd, Hmd, Fae, Mer, DmrX, MptG, MfnB, D, E, F, OrfY, 5, 7, 17, 19, 20) were performed as follows: from the Kyoto Encyclopedia of Genes and Genomes (KEGG)⁴⁸ orthology page of each enzyme or known homologs from the literature, up to five sequences from both Bacteria and Archaea, preferably characterized, were aligned and used to create preliminary hidden Markov model (HMM) profiles. These were used to perform hidden Markov model-based searches with HMMer⁴⁹ against our local genome databanks of Bacteria and Archaea separately. The first 20–30 hits in descending e-value order from each domain were used to create a second round of domain-specific HMM profiles for each protein.

Usually a cutoff of e^{-4} was applied unless there were too many hits ($>>1000$) or a very clear demarcation of homologs. In this case, we set the cutoff manually based on the e -value slope and annotations. In some cases when putative homologs (by annotation) were present above the e^{-4} cutoff, a separate HMM profile based on a larger sampling of sequences was created to pick these divergent homologs. Either way, the resulting datasets were manually refined through preliminary phylogenetic analysis to remove unrelated sequences. Very close homologs from genomes belonging to members of the same species/genus were removed from the final datasets to reduce computational burden. Preliminary Fwd trees were used to distinguish the clades corresponding to characterized tungsten (Fwd), molybdenum (Fmd), and a separate divergent clade referred here as “tertiary” Fwd homologs (containing only BD subunits, see legend to Supplementary Figure 22). Since Fmd and tertiary Fwd exist only in Archaea, only the Fwd homologs were retained for the FwdABC concatenation (FwdD being too restricted in bacteria to be included). The clustering of the genes involved in the H_4 MPT branch in archaeal and bacterial genomes was checked in Genespy⁵⁰ and supplemented using MacSyFinder⁵¹ and HMM profiles of the related proteins.

Phylogenomic analysis

Single protein datasets were aligned with MUSCLE⁵² with default parameters and trimmed with BMGE 1.1⁵³ using the BLOSUM30 substitution matrix to select unambiguously aligned amino acid positions. To check for phylogenetic congruence among FwdA, B, C subunits and among the MfnB, E, F subunits, single gene trees were calculated in IQTree under the TEST option with 100 bootstrap replicates⁵⁴ and collapsed at nodes showing $<80\%$ bootstrap support. These trees were tested for congruence against a strictly bifurcating tree of a blind concatenation, by

performing an Internode Certainty (IC) test⁵⁵ in RaxML⁵⁶. Sequences responsible for incongruences at order level or above (in Archaea: Methanoflorentaceae were treated as part of Methanocellales) or phylum level (in Bacteria: α , β , γ -Proteobacteria were treated as a single phylum, and in Firmicutes the deep-branching order Halanerobiales was treated as separate from Clostridia) were removed and the procedure was repeated until no further incongruence was found. The resulting datasets were then concatenated. The concatenated Fwd dataset (FwdABC, 857 aa) contained no more than 6% missing taxa per protein, and the concatenated Mfn dataset (MfnBEF, 360 aa) contained no more than 3.4% missing taxa per protein. An Archaea-only dataset was also created including Fmd and tertiary Fwd for the four main (non-ferredoxin) subunits (FwdABCD). A separate round of congruence testing was performed, which resulted in a concatenated dataset (FwdABCD, 1192 aa) containing no more than 22% missing taxa per gene. The larger number of missing taxa is due to the tertiary Fwd only containing subunits B and D.

To determine the effect of the addition of novel lineages on the length of the branch separating Archaea and Bacteria, we reconstructed the FwdABC phylogeny using the lineages that were available in 1999¹² (Methanococcales, Methanobacteriales, Methanopyrales, Methanosarcinales, Archaeoglobales, α , β , γ -Proteobacteria), and 2004⁴ (same as 1999 plus Planctomycetes) but with all the genomes currently available for them. Phylogenies were calculated in IQTree as described above with 1000 ultrafast bootstrap replicates⁵⁷.

No concatenation was possible for the other genes (Ftr, Mch, Mtd, Hmd, bMtd, Fae, Mer, DmrX, MptG, MfnD, OrfY, 5, 7, 17, 19, 20). Although their topologies were overall similar, there were some differences due to taxonomic distribution, a few horizontal transfers or poor signal.

Phylogenies were inferred in PhyloBayes⁵⁸ under the CAT+GTR+ Γ 4 model for the Fwd

concatenations and LG+ Γ 4 for the single-gene datasets and the MfnBEF concatenation, because of the smaller number of positions, and in IQTree under the TEST option with 100 bootstrap replicates for all datasets. For Bayesian analyses, four independent Markov chain Monte Carlo chains were run until convergence (with the exception of Orf7 that never got to a standard deviation below 0.8) and checked by sampling every two cycles with a 25% burn-in. All phylogenies are technically unrooted but for display purposes we have placed in each case an ad hoc root according to one of the suggested roots of Archaea i.e. between Euryarchaeota and TACK-Asgard or within Euryarchaeota¹⁸.

For the reference phylogeny of Bacteria, a Bayesian phylogeny was built in PhyloBayes under the CAT+GTR+ Γ 4 model using a concatenation of RNA polymerase subunits B, B', and IF-2 (2,337-aa positions). The tree was rooted according to¹⁷.

All datasets relative to these analyses (original alignments, trimmed datasets, and trees in Newick format) are provided in Supplementary Data 4.

Limitata metabolic annotation

As with all genomes in our local database for which protein sequences were not available, the nucleotide sequences of bacterium GBS-1 (txid: 1662110) were downloaded and open reading frames were predicted with Prodigal under default settings⁴⁷. The resulting amino-acid sequences were assigned arbitrary accession numbers. Functional annotations for all open reading frames and metabolic processes were predicted using BlastKOALA (family_eukaryotes+genus_prokaryotes, and species_prokaryotes databases)⁵⁹, supplemented by BLASTp⁶⁰ searches against the non-redundant database at NCBI. Open reading frames were also predicted on the RAST server⁶¹ with default options, and for those the metabolic annotation was

done on both RAST and BlastKOALA. The metabolism of the second Limitata representative (bacterium 42_11 (txid: 1635281)) for which amino acid sequences were available, was annotated in the same way. All outputs are included in Supplementary Data 2. Genome completeness and contamination (respectively 100% and 0% for GBS-1, 84.1% and 0.15% for 42_11) were estimated in CheckM⁶². For comparison, *Halanaerobium hydrogeniformans* (txid: 656519), *Anaerobaculum hydrogeniformans* (txid: 592015), and *Anaerobaculum mobile* (txid: 891968) were run in BlastKOALA (family_eukaryotes+genus_prokaryotes) and the results are also included in Supplementary Data 2.

Figure Legends:

Figure 1. Schematic of the WL pathway reactions, and their relationships with other methylotrophy modules. In the H₄MPT methyl branch, Fwd denotes both the tungsten (Fwd) or molybdenum (Fmd) formylmethanofuran dehydrogenase complex (Fwd/FmdABCD). Ftr is the formylmethanofuran:tetrahydromethanopterin formyltransferase. Mch is the methenyltetrahydromethanopterin cyclohydrolase. Mtd/Hmd denote F₄₂₀-dependent methylenetetrahydromethanopterin dehydrogenase and H₂-forming methylenetetrahydromethanopterin dehydrogenase, respectively. bMtd refers to the bacterial methylene tetrahydromethanopterin dehydrogenase (MtdA/B/C, which denote homologous enzyme classes, not subunits) to avoid confusion with the non-homologous archaeal Mtd. Mer is the methylenetetrahydromethanopterin reductase. Fhc stands for the bacterial enzyme complex composed of FwdABC+Ftr.

The dashed arrow in the H₄F methyl branch indicates that spontaneous condensation of formaldehyde (HCHO) with H₄F can occur but is a very minor contribution to assimilatory fluxes²⁶. Blue arrows denote reactions occurring in Archaea. Green arrows denote reactions occurring in methylotrophs. The glutathione (GSH)-dependent oxidation and ribulose monophosphate (RuMP) cycle are alternative pathways of formaldehyde assimilation found in methylotrophs, sometimes coexisting with the H₄MPT branch. Note that throughout the text, unless specifically noted, MF and H₄MPT are used as umbrella terms for all of their analogs which are produced through the same fundamental biosynthetic pathways, including methanofurans a-e, methylofuran, dephosphotetrahydromethanopterin, and tetrahydrosarcinapterin. See also Supplementary Figures 9 and 10 and their legends for details on MF and H₄MPT biosynthetic pathways.

Figure 2. (a) Schematic comparison of the length of the branch (average number of substitutions per site) separating Archaea and Bacteria in the phylogeny based on a concatenation of the three Fwd subunits common to Bacteria and Archaea (FwdABC) according to the taxonomic sampling available in 1999, 2004, and this study; (b) Bayesian phylogeny based on FwdABC (337 sequences and 859 aa). The phylogeny was inferred in PhyloBayes⁵⁸ under the CAT+GTR+Γ4 model. Values at nodes are posterior probabilities. Black circles denote branches with posterior probability of at least 0.95. Scale bar represents the average number of substitutions per site. Some clades are collapsed, for clarity (the full tree is provided in Supplementary Figure S8). The tree is rooted according to one of the three possible roots for the Archaea, namely that between Euryarchaeota and TACK-Asgard, but bacteria emerge from within the archaeal radiation under all alternative rootings (see main text). The placement of

Persephonearchaea, Theionarchaea and Altiarchaeales with respect to the root has not been yet specifically tested¹⁸. For simplicity, the molybdenum-containing homolog (Fmd, or secondary) as well as additional divergent Fwd homologs (here referred to as “tertiary”) present in Archaea (Supplementary Data 1) were analyzed separately and are presented in Supplementary Figure 22. W1-2 and W3-4 refer to the primary and secondary tungsten-containing Fwd clusters of Methanomicrobiales based on range of distribution (Supplementary Data 1); (c) Co-localization of H₄MPT branch and cofactor biosynthesis genes. Gene order is mostly conserved (FwdDBA-Ftr-FwdC-Mch) between the uncultured archaeal and bacterial lineages close to the putative transfer event. The gene-to-biosynthetic reaction correspondence is noted with reaction numbers (H1 to H15 and M1 to M7/M7’) and given in the table on the right for brevity. These reactions are presented in detail in Supplementary Figure 9 & 10. “O” is abbreviation for “ORF”. Slashes indicate a contig edge. Figure was produced in GeneSpy⁵⁰ by comparing a region of 10 kbp centered around FwdA (or Mch when two sub-clusters existed in separate contigs).

Figure 3. Scenario for the origin and evolution of the H₄MPT methyl branch in Bacteria.

Colored circles denote the major events during the assembly of methylotrophy in Bacteria, as discussed in the main text. For simplicity the MF and H₄MPT cofactor biosynthesis pathways are not shown, but they follow the same evolutionary history as the H₄MPT branch (transferred at the red circle). Black circle: Initial configuration of the WL pathway in the LACA with a carbonyl and an H₄MPT branch. Red circle: the H₄MPT methyl branch is transferred from Archaea to Bacteria (FwdABCD, Ftr, Mch, and possibly bMtd). Fermentation products can enter the pathway through serine hydroxymethyltransferase and the Glycine Cleavage System. Green circle: Fae appears in Bacteria (possibly also transferred from Archaea). Methylated compounds

can now enter the pathway through formaldehyde. Orange circle: FwdD is lost. FwdABC become associated with Ftr (Fhc complex), forcing the reaction into the H₄F branch.

HCHO, formaldehyde; HCOOH, formate; CH₃-, methyl-; CH₂-, methylene-.

Figure 4. Metabolic prediction of Limitata with main biochemical reactions. The reconstruction is derived from bacterium GBS-1 (Supplementary Data 2). The metabolism of bacterium 42_11 is almost identical, but was not used here, as the genome is incomplete. Apart from the GCS (see text for discussion), the H₄MPT methyl branch can be also entered through other reactions (not shown), for instance at methylene-H₄MPT/H₄F through thymidylate synthase (ThyX) when converting dTMP to dUMP (in pyrimidine metabolism), through dehydropantoate hydroxymethyltransferase from 2—dehydropantoate (in pantothenate and CoA biosynthesis), or by diverting methionine into homocysteine and methyl-H₄MPT/H₄F (homocysteine methyltransferase) which is oxidized to methylene-H₄MPT/H₄F (methylene-H₄F reductase, not present in Limitata but in *Halanaerobium*). However, all these alternatives require intermediates of biosynthetic pathways, making it unlikely that they are consistently used as H₄MPT branch input beyond some contribution during nucleotide or amino acid catabolism.

References

1. Singh, B. K., Bardgett, R. D., Smith, P. & Reay, D. S. Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nat. Rev. Microbiol.* **8**, 779–790 (2010).
2. Chistoserdova, L. & Kalyuzhnaya, M. G. Current Trends in Methyлотrophy. *Trends Microbiol.* **26**, 703–714 (2018).

3. Borrel, G. *et al.* Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat. Microbiol.* **4**, 603–613 (2019).
4. Chistoserdova, L. *et al.* The enigmatic planctomycetes may hold a key to the origins of methanogenesis and methylotrophy. *Mol. Biol. Evol.* **21**, 1234–1241 (2004).
5. Chistoserdova, L. Wide distribution of genes for tetrahydromethanopterin/methanofuran-linked C1 transfer reactions argues for their presence in the common ancestor of bacteria and archaea. *Frontiers in Microbiology* **7**, 1425 (2016).
6. Weiss, M. C. *et al.* The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).
7. Borrel, G., Adam, P. S. & Gribaldo, S. Methanogenesis and the Wood-Ljungdahl Pathway: An Ancient, Versatile, and Fragile Association. *Genome Biol. Evol.* **8**, 1706–1711 (2016).
8. Fuchs, G. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* **65**, 631–658 (2011).
9. Laso-Pérez, R. *et al.* Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396–401 (2016).
10. Chistoserdova, L. Modularity of methylotrophy, revisited. *Environmental Microbiology* **13**, 2603–2622 (2011).
11. Chistoserdova, L., Vorholt, J. A., Thauer, R. K. & Lidstrom, M. E. C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic archaea. *Science* (80-.). **281**, 99–102 (1998).
12. Vorholt, J. A., Chistoserdova, L., Stoltyar, S. M., Thauer, R. K. & Lidstrom, M. E. Distribution of tetrahydromethanopterin-dependent enzymes in methylotrophic bacteria

- and phylogeny of methenyl tetrahydromethanopterin cyclohydrolases. *J. Bacteriol.* **181**, 5750–5757 (1999).
13. Bauer, M. *et al.* Archaea-like genes for C1-transfer enzymes in Planctomycetes: Phylogenetic implications of their unexpected presence in this phylum. *J. Mol. Evol.* **59**, 571–586 (2004).
 14. Kalyuzhnaya, M. G. *et al.* Analysis of gene islands involved in methanopterin-linked C1 transfer reactions reveals new functions and provides evolutionary insights. *J. Bacteriol.* **187**, 4607–4614 (2005).
 15. Butterfield, C. N. *et al.* Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ* **4**, e2687 (2016).
 16. Ettwig, K. F. *et al.* Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**, 543–548 (2010).
 17. 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.* **112**, 6670–6675 (2015).
 18. Adam, P. S., Borrel, G., Brochier-Armanet, C. & Gribaldo, S. The growing tree of Archaea: New perspectives on their diversity, evolution and ecology. *ISME J.* **11**, 2407–2425 (2017).
 19. Ramamoorthy, S. *et al.* Desulfosporosinus lacus sp. nov., a sulfate-reducing bacterium isolated from pristine freshwater lake sediments. *Int. J. Syst. Evol. Microbiol.* **56**, 2729–2736 (2006).
 20. Davidova, I. A. *et al.* Dethiosulfatarculus sandiegensis gen. nov., sp. nov., isolated from a methanogenic paraffin-degrading enrichment culture and emended description of the family Desulfarculaceae. *Int. J. Syst. Evol. Microbiol.* **66**, 1242–1248 (2016).

21. Adam, P. S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl. Acad. Sci.* **115**, E5837 (2018).
22. Sousa, F. L. & Martin, W. F. Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim. Biophys. Acta - Bioenerg.* **1837**, 964–981 (2014).
23. Chistoserdova, L., Rasche, M. E. & Lidstrom, M. E. Novel Dephosphotetrahydromethanopterin Biosynthesis Genes Discovered via Mutagenesis in *Methylobacterium extorquens* AM1. *J. Bacteriol.* **187**, 2508–2512 (2005).
24. Pomper, B. K., Saurel, O., Milon, A. & Vorholt, J. A. Generation of formate by the formyltransferase/hydrolase complex (Fhc) from *Methylobacterium extorquens* AM1. *FEBS Lett.* **523**, 133–137 (2002).
25. Wagner, T., Ermler, U. & Shima, S. The methanogenic CO₂ reducing-and-fixing enzyme is bifunctional and contains 46 [4Fe-4S] clusters. *Science (80-.).* **354**, 114–117 (2016).
26. Crowther, G. J., Kosaly, G. & Lidstrom, M. E. Formate as the main branch point for methylotrophic metabolism in *Methylobacterium extorquens* AM1. *J. Bacteriol.* **190**, 5057–5062 (2008).
27. Bar-Even, A., Noor, E. & Milo, R. A survey of carbon fixation pathways through a quantitative lens. *Journal of Experimental Botany* **63**, 2325–2342 (2012).
28. Figueroa, I. A. *et al.* Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO₂ fixation pathway. *Proc. Natl. Acad. Sci.* 201715549 (2017). doi:10.1073/pnas.1715549114
29. Vorholt, J. A., Marx, C. J., Lidstrom, M. E. & Thauer, R. K. Novel formaldehyde-

- activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol. *J. Bacteriol.* **182**, 6645–6650 (2000).
30. Lin, Z. & Sparling, R. Investigation of serine hydroxymethyltransferase in methanogens. *Can. J. Microbiol.* **44**, 652–656 (1998).
 31. Schwartz, E., Fritsch, J. & Friedrich, B. in *The Prokaryotes* 119–199 (2013).
doi:10.1007/978-3-642-30141-4_65
 32. Begemann, M. B., Mormile, M. R., Sitton, O. C., Wall, J. D. & Elias, D. A. A streamlined strategy for biohydrogen production with *Halanaerobium hydrogeniformans*, an alkaliphilic bacterium. *Front. Microbiol.* **3**, 93 (2012).
 33. Maune, M. W. & Tanner, R. S. Description of *Anaerobaculum hydrogeniformans* sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus *Anaerobaculum*. *Int. J. Syst. Evol. Microbiol.* **62**, 832–838 (2012).
 34. Liang, R., Grizzle, R. S., Duncan, K. E., McInerney, M. J. & Suflita, J. M. Roles of thermophilic thiosulfate-reducing bacteria and methanogenic archaea in the biocorrosion of oil pipelines. *Front. Microbiol.* **5**, 89 (2014).
 35. Kasting, J. F. Methane and climate during the Precambrian era. *Precambrian Res.* **137**, 119–129 (2005).
 36. Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519 (2006).
 37. Slotznick, S. P. & Fischer, W. W. Examining Archean methanotrophy. *Earth Planet. Sci. Lett.* **441**, 52–59 (2016).
 38. Summons, R. E., Jahnke, L. L. & Roksandic, Z. Carbon isotopic fractionation in lipids

- from methanotrophic bacteria: Relevance for interpretation of the geochemical record of biomarkers. *Geochim. Cosmochim. Acta* **58**, 2853–2863 (1994).
39. Jahnke, L. L., Summons, R. E., Hope, J. M. & Des Marais, D. J. Carbon isotopic fractionation in lipids from methanotrophic bacteria II: The effects of physiology and environmental parameters on the biosynthesis and isotopic signatures of biomarkers. *Geochim. Cosmochim. Acta* **63**, 79–93 (1999).
 40. Battistuzzi, F. U. & Hedges, S. B. A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* **26**, 335–343 (2009).
 41. Haqq-Misra, J. D., Domagal-Goldman, S. D., Kasting, P. J. & Kasting, J. F. A revised, hazy methane greenhouse for the Archean Earth. *Astrobiology* **8**, 1127–1137 (2008).
 42. Konhauser, K. O. *et al.* Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. *Nature* **458**, 750–753 (2009).
 43. Catling, D. C., Claire, M. W. & Zahnle, K. J. Anaerobic methanotrophy and the rise of atmospheric oxygen. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **365**, 1867–1888 (2007).
 44. Daines, S. J. & Lenton, T. M. The effect of widespread early aerobic marine ecosystems on methane cycling and the Great Oxidation. *Earth Planet. Sci. Lett.* **434**, 42–51 (2016).
 45. Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012).
 46. Wrighton, K. C. *et al.* Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* (80-.). **337**, 1661–1665 (2012).
 47. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

48. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
49. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).
50. Garcia, P. S., Jauffrit, F., Grangeasse, C. & Brochier-Armanet, C. GeneSpy, a user-friendly and flexible genomic context visualizer. *Bioinformatics* **35**, 329–331 (2019).
51. Abby, S. S., Néron, B., Ménager, H., Touchon, M. & Rocha, E. P. C. MacSyFinder: A program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* **9**, e110726 (2014).
52. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
53. 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.* **10**, 210 (2010).
54. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
55. Kobert, K., Salichos, L., Rokas, A. & Stamatakis, A. Computing the Internode Certainty and Related Measures from Partial Gene Trees. *Mol. Biol. Evol.* **33**, 1606–1617 (2016).
56. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
57. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2:

- improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2017).
58. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
 59. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* **428**, 726–731 (2016).
 60. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
 61. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, D206–D214 (2014).
 62. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

Author notes: S.G. and P.S.A. designed research; P.S.A. performed all phylogenomic analyses; P.S.A. and G.B. performed genome synteny and metabolic reconstruction analyses; P.S.A., G.B., and S.G. analyzed the data and wrote the paper.

Data availability: All raw data (HMM profiles, initial and trimmed alignments, full trees in Newick format) are available in Supplementary Data 4.

Acknowledgments: P.S.A. is supported by a PhD fellowship from Paris Diderot University and by funds from the PhD Programme “Frontières du Vivant (FdV)–Programme Bettencourt.” G.B.

acknowledges support from the Institut Pasteur through a Roux-Cantarini fellowship. S.G.
acknowledges funding from the French National Agency for Research Grant ArchEvol (ANR-16-CE02-0005-01).

Competing interests: The authors declare no competing interests.

Figure 1

Figure 2

Figure 3

Figure 4