



HAL
open science

Bacterial origin of thymidylate and folate metabolism in Asgard Archaea

Jonathan Filée, Hubert F Becker, Lucille Mellottee, Zihui Li,
Jean-Christophe Lambry, Ursula Liebl, Hannu Myllykallio

► **To cite this version:**

Jonathan Filée, Hubert F Becker, Lucille Mellottee, Zihui Li, Jean-Christophe Lambry, et al.. Bacterial origin of thymidylate and folate metabolism in Asgard Archaea. 2022. hal-03855599

HAL Id: hal-03855599

<https://cnrs.hal.science/hal-03855599>

Preprint submitted on 16 Nov 2022

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

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

17

18

Abstract

19 **Little is known about the evolution and biosynthetic function of DNA precursor and**
20 **the folate metabolism in the Asgard group of archaea. As Asgard occupy a key**
21 **position in the archaeal and eukaryotic phylogenetic trees, we have exploited very**
22 **recently emerged genome and metagenome sequence information to investigate**
23 **these central metabolic pathways. Our genome-wide analyses revealed that the**
24 **recently cultured Asgard archaeon *Candidatus* Prometheoarchaeum syntrophicum**
25 **strain MK-D1 (*Psyn*) contains a complete folate-dependent network for the**
26 **biosynthesis of DNA/RNA precursors, amino acids and syntrophic amino acid**
27 **utilization. Altogether our experimental and computational data suggest that**
28 **phylogenetic incongruences of functional folate-dependent enzymes from Asgard**
29 **archaea reflect their persistent horizontal transmission from various bacterial**
30 **groups, which has rewired the key metabolic reactions in an important and recently**
31 **identified archaeal phylogenetic group. We also experimentally validated the**
32 **functionality of the lateral gene transfer of *Psyn* thymidylate synthase ThyX. This**
33 **enzyme uses bacterial-like folates efficiently and is inhibited by mycobacterial**
34 **ThyX inhibitors. Our data raise the possibility that the thymidylate metabolism,**
35 **required for *de novo* DNA synthesis, originated in bacteria and has been**
36 **independently transferred to archaea and eukaryotes. In conclusion, our study has**
37 **revealed that recent prevalent lateral gene transfer has markedly shaped the**
38 **evolution of Asgard archaea by allowing them to adapt to specific ecological niches.**

39

40 Introduction

41 The canonical thymidylate synthase ThyA (EC 2.1.1.45) was once considered the
42 only enzyme capable of catalyzing the *de novo* methylation of the essential DNA precursor
43 dTMP (deoxythymidine 5'-monophosphate or thymidylate) from dUMP (deoxyuridine 5'-
44 monophosphate). However, combined *in silico* and experimental approaches led to the
45 discovery of a new metabolic pathway that operates in the methylation of DNA precursors
46 in numerous microbial species ^{1, 2, 3}, relying on a novel thymidylate synthase, ThyX (EC
47 2.1.1.148). No sequence or structural homology exists between ThyA (found in $\approx 65\%$ of
48 microbial genomes) and ThyX flavoproteins ($\approx 35\%$)⁴. Differently from the homodimeric
49 ThyA proteins, the active site of ThyX flavoenzymes, which accommodates dUMP, NADPH
50 and the carbon donor methylene tetrahydrofolate (CH₂H₄folate, a vitamin B9 derivative),
51 is located at the interface of three subunits of the homotetrameric protein complex⁵.
52 Consequently, formation of the ThyX tetramer is necessary for catalytic activity.

53 In the unique reductive methylation reaction catalyzed by homodimeric ThyA,
54 CH₂H₄folate functions both as a source of carbon (C₁-carrier) and reducing equivalents⁶,
55 thus leading to the formation of dihydrofolate (H₂folate) as the oxidation product of
56 tetrahydrofolate (H₄folate) (Figure 1a, left panel). H₂folate is subsequently reduced to
57 tetrahydrofolate (H₄folate) by dihydrofolate reductase (DHFR) FoaA, as only reduced
58 folate derivatives are functional in intermediary metabolism. Consequently, ThyA and
59 FoaA form a functionally coupled adaptive unit embedded within the bacterial folate
60 metabolism network⁷. CH₂H₄folate is then reformed in a reversible reaction by serine
61 hydroxymethyltransferase (SHMT) GlyA using H₄folate and serine as substrates.

62 ThyX also uses CH₂H₄folate as a C₁ carrier, but acquires the reducing hydride from
63 NADPH and not from H₄folate. ThyX catalyzes dTMP formation using flavin-mediated

64 hydride transfer from pyridine nucleotides, which is required for reduction of the
65 methylene group, thus directly resulting in the formation of H₄folate (Fig. 1a, right panel).
66 The ThyX reaction mechanism appears less catalytically efficient than that used by ThyA
67 enzymes⁸, but maintains the folate in its reduced form (as H₄folate) at the end of the
68 catalytic cycle. Consequently, *thyX*-containing organisms do not have an absolute
69 requirement for FoaA in their thymidylate metabolism (Fig. 1a), explaining why the *foaA*
70 gene is frequently absent from *thyX*-containing organisms⁹. It remains unclear how two
71 independent solutions for the synthesis of the central building block of DNA evolved. In
72 addition, thymidylate synthase homologs are known to participate in biosynthesis of
73 nucleoside antibiotics and modified nucleotides in bacteriophages⁹. Phylogenetic
74 analyses have revealed the existence of frequent and multiple lateral gene transfers
75 during the evolution of ThyA and ThyX, sometimes associated with non-homologous
76 replacement^{1,10}. *ThyX* appears overrepresented in genome-reduced and/or slow-growing
77 prokaryotes^{8, 11}, and changes in ThyX function of human microbiota associate with
78 obesity-associated deficits in inhibitory control towards responses to stimuli (e.g. food)¹².

79 Despite the essential role of thymidylate synthase for DNA synthesis, data and
80 details on the experimental characterization of archaeal thymidylate synthases remain
81 scarce. Consequently, the functionality of the archaeal thymidylate synthases has only
82 been sporadically followed up to date. While the halophilic archaeon *Haloferax volcanii*
83 contains the gene for canonical thymidylate synthase ThyA, in the closely related
84 halophilic archaeon *Halobacterium salinarum* genetic evidence was provided for the
85 presence of functional *thyX*³. Biochemically dTMP biosynthesis has been demonstrated in
86 cell-free extracts of two different archaeal species^{13, 14}: *Methanosarcina thermophila* and
87 *Sulfolobus solfataricus* (now referred to as *Saccharolobus solfataricus*). Although the
88 corresponding enzymes were not purified or identified in this study, dTMP formation was

89 detected from dUMP using externally added isotope-labeled formaldehyde and either
90 chemically modified folates present in the cell extract (tetrahydrosarcinapterin, *M.*
91 *thermophila*) or added synthetic fragments of sulfopterin (*S. solfataricus*). Indeed,
92 biochemical studies have indicated the presence of at least six distinct chemically
93 modified, but thermodynamically similar, C₁ carriers that function in archaeal central
94 biosynthetic networks and/or energy-yielding reactions¹⁵. These chemical modifications
95 differ among various archaeal species and likely replace chemically distinct
96 tetrahydrofolate derivatives in many archaea. Recent phylogenetic studies have
97 suggested an archaeal origin for the energy-producing Wood-Ljungdahl pathway that is
98 dependent on tetrahydromethanopterin (H₄MPT), frequently found in methanogenic
99 archaea¹⁶. This archaeal pathway may have contributed to the early origin of
100 methanogenesis and the emergence of the use of *e.g.* reduced one-carbon compounds as
101 carbon source (methylotrophy) in bacteria.

102 An archaeal superphylum called Asgard was recently discovered that includes the
103 closest known archaeal relatives of eukaryotes^{17, 18, 19}. This striking discovery has
104 solidified the two-domains tree of life hypothesis in which Eukaryotes have emerged from
105 within the archaeal tree²⁰. Even if the true identity of the archaeal ancestors of Eukaryotes
106 is still being debated, Asgard archaea occupy a pivotal position in the archaeal/eukaryotic
107 phylogenetic trees^{17, 19}. Thus, studying the thymidylate and folate metabolism in the
108 Asgard group is of particular interest to better understand the origins and the evolution
109 of these essential pathways participating in DNA, RNA, and protein synthesis, as well as
110 in catabolic reactions. This is also underlined by the fact that pioneering genomic studies
111 on Lokiarchaeota suggested the absence of biosynthetic capacity for
112 tetrahydromethanopterin and tetrahydrofolate co-factors, but the presence of some
113 folate-dependent enzymes implicated in amino acid utilization¹⁸.

114 In this study, we have exploited the recent increase in Asgard metagenome and
115 genome sequence information to investigate thymidylate synthases and folate-dependent
116 metabolic networks in more than 140 complete or nearly complete Asgard genomes. Our
117 sequence similarity searches revealed a high level of similarity between the protein
118 sequences of bacterial and Asgard folate-dependent enzymes, including thymidylate
119 synthases. Our genomic and phylogenetic analyses indicated that Asgard archaea have
120 ‘hijacked’ at several independent occasions bacterial folate-dependent enzymes and
121 pathways to support their central metabolism. Detailed experimental analyses revealed
122 that the protein encoded by *thyX* from the recently cultured Asgard archaeon *Candidatus*
123 *Prometheoarchaeum syntrophicum* strain MK-D1¹⁸ (further referred to as *Psyn*) is fully
124 functional as thymidylate synthase in bacterial cells and efficiently interacts with
125 bacterial folates. In conclusion, our combined experimental and computational data
126 suggest that the patchy phylogenetic distribution and phylogenetic incongruences of
127 functional folate-dependent enzymes from Asgard archaea reflect their independent
128 horizontal transmission from various bacterial groups. These lateral (horizontal) gene
129 transfer (LGT) events have potentially rewired the key metabolic reactions in an
130 important archaeal phylogenetic group.

131 **Results**

132 **Identification of a ThyX orthologue in *Candidatus Prometheoarchaeum*** 133 ***syntrophicum* strain MK-D1 (*Psyn*)**

134 As an initial approach to investigate the Asgard archaeal thymidylate and C₁
135 metabolism, we searched for *thyX* and *thyA* homologs in sequence databases for archaeal
136 genomes or metagenome-assembled genomes (MAGs). These studies led to the
137 identification of both *thyX* and *thyA* sequences in the genomes of the understudied and

138 diverse group of archaea. Interestingly, our similarity searches identified a ThyX
139 orthologue in *Psyn*, which is up to 51.44% identical to bacterial ThyX sequences at the
140 protein sequence level (e.g. *Zixibacteria*, e-value $\approx 10^{-78}$ or *Calditrichaeota* bacterium, e-
141 value $\approx 10^{-76}$). The observed e-values are very low and are adjusted to the large sequence
142 database size. Therefore, the quality of the observed hits is very high. As *Psyn* is the only
143 known example of *Asgard* archaea that can be cultivated and its genome, lacking *thyA*, is
144 completely sequenced, we concentrated our efforts on this archaeal ThyX.

145 Transcriptome analyses of RNA extracted from enriched cultures of *Psyn* indicated
146 that *thyX* from this strain is expressed [Reads Per Kilobase of transcript, per Million
147 mapped reads (RPKM) value of 225.37 using the data set from the sequence read archive
148 (SRA) DRR199588]. This gene encodes a protein with a predicted molecular mass of 35,
149 294 Da. Its genomic environment (Fig. 1b and Supplementary Figure 1) comprises an
150 upstream gene coding for a Nif3-like protein with a length of 259 residues (29,026 Da)
151 and a downstream gene encoding a domain of unknown function DUF2095. Physical
152 association of *Psyn thyX* with the Nif3 family encoding gene is of interest, as this family of
153 proteins may correspond to GTP cyclohydrolase 1 type 2, which converts GTP to
154 dihydroneopterin triphosphate and may function in folic acid synthesis. The structure-
155 based sequence alignment of *Psyn ThyX* with *Thermotoga maritima* ThyX (PDB structure
156 5CHP) predicts the conservation of functionally important residues involved in folate-,
157 nucleotide- and flavin-binding (Fig. 1b). The marked degree of sequence similarity
158 allowed the construction of a high-quality structural model for *Psyn ThyX* indicating its
159 functional significance as thymidylate synthase and nucleotide and folate binding protein.
160 More specifically, we constructed a model of the ThyX homotetramer based upon PDB
161 structures 1O26, 3N0B and 6J61 as templates and using the protein structure modeling
162 program Modeller²¹ (Fig. 1c, left panel). Structural superposition of the model with PDB

163 structure 3GT9 using the Chimera software²² allowed the addition of FAD, dUMP and
164 folate molecules into the active site of the model (Fig. 1c, middle panel). Importantly, the
165 model suggests the highly plausible transfer of 1C units via the N5 atom of the FAD
166 cofactor to the accepting carbon of dUMP (Fig. 1c, right panel).

167 ***thyX* and *thyA* are present in Asgard archaea**

168 As our sequence similarity searches suggested a close link between *Psyn* and
169 bacterial ThyX sequences, we obtained more detailed insight into the thymidylate
170 synthase gene distribution and phylogeny in Asgard archaea. In particular, in addition to
171 the complete genome of cultivated *Psyn*, we analyzed more than 140 MAGs of uncultivated
172 Asgard archaea (Fig. 2). Our analyses revealed that *thyX* is present in only a small subset
173 of lineages with a patchy gene distribution. We found *thyX* genes to be present in 33 out
174 of a total of 141 Asgard (meta)genomes: 6/52 in the Loki lineage, 6/16 in Heimdall, 0/5
175 in Hel, 1/1 in Odin, 6/38 in Thor, 1/1 in Wukong, 1/2 in Borr, 1/12 in Hod, 3/3 in Gerd,
176 7/10 in Hermod, 0/1 in Kari and 2/2 in Baldr. In contrast, *thyA* genes are widely
177 distributed among Asgard archaea: 93 out of a total of 141 (36/52 for Loki, 13/16 for
178 Heimdall, 5/5 for Hel, 1/1 for Odin, 31/38 for Thor, 0/1 for Kari, 1/2 for Wukong, 1/2 for
179 Borr, 3/12 for Hod, 0/3 for Gerd, 1/10 for Hermod, 1/2 for Baldr). Approximately twenty
180 Asgard MAGs encode for both *thyX* and *thyA* genes, although *thyX* genes appear often
181 partial or fragmented in this latter case. In a few cases, both *thyX* and *thyA* appear to be
182 absent, likely reflecting the fact that some of the analysed MAGs correspond to partially
183 assembled genomes.

184 **Automated prediction of gene transfers between *Psyn* and bacteria**

185 Considering the high level of sequence similarity of *Psyn* ThyX and bacterial
186 orthologs, we estimated the possible extent of gene transfer between bacteria and *Psyn*

187 using HGTector. This automated method is well suited for discovering potential recent
188 gene transfers²³ by analyzing the hit distribution patterns from the similarity searches.
189 The method performs a systematic analysis to detect genes that do not appear to have a
190 typical vertical history of the analysed organisms but instead have a putative horizontal
191 origin. This approach efficiently detects atypical genes that are highly likely to be
192 horizontal gene transfers (HGT). As this method uses all sequence data available in the
193 GenBank, it allows exhaustive detection of potential recent gene transfers.

194 Using HGTector, we predicted 149 HGT events [[supplementary table 1](#)
195 (sheet “potential gene transfers”)], including *Psyn thyX*, when all the \approx 4000 *Psyn*
196 predicted protein-coding genes were analyzed (Fig. 3). In this plot, each dot represents
197 one gene, and likely horizontally transferred genes are indicated in yellow. As depicted in
198 Fig. 3, many *Psyn* folate-dependent genes (purple legends in Fig. 3a, more detailed figure
199 with annotations is available as [supplementary Fig. 3](#)) were also likely transferred from
200 bacteria. The BlastKOALA annotation²⁴ of the potentially transferred genes indicates their
201 functional distribution in amino acid, nucleotide, and carbohydrate metabolism (Fig. 3b
202 [and supplementary table 1](#)). This is of interest considering syntrophic amino acid
203 utilization of *Psyn*¹⁸.

204 ***thyX* was transferred from bacteria to diverse Asgard archaea**

205 To better understand the origin and evolution of the thymidylate synthase genes
206 in Asgard archaea, we have reconstructed their phylogenies. Note that Asgard sequences
207 are indicated in red, other archaea in blue, bacteria in black and Eukarya in green. The
208 *ThyX* phylogeny (Fig. 4 and [supplementary Fig. 4](#) for a detailed version of the tree with
209 the taxon names) shows that Asgard sequences are deeply polyphyletic; they are
210 scattered into several monophyletic groups. Four of them are related to different bacterial

211 groups and appear distantly related to other archaeal sequences. The *Psyn* ThyX clusters
212 along with some, but not all, Loki and Heimdall sequences, together with bacteria
213 belonging to the *Deinococcus/Thermus* group, as well as with several alpha-
214 proteobacteria. Additional *Heimdall* and *Borrarchaeota* sequences appear scattered in
215 this tree with varying affinities for diverse bacterial lineages. The three remaining Asgard
216 groups, including all of the thorarchaeal sequences, are related to other archaeal
217 sequences, but do not form a monophyletic cluster, indicating possible lateral gene
218 transfers among archaeal species. The only eukaryotic ThyX sequence from *Dictyostelium*
219 branches with alpha-proteobacteria, suggesting a transfer from bacteria to eukaryotes.
220 This phylogenetic analysis supports the existence of multiple and independent lateral
221 (horizontal) gene transfers of the *thyX* genes from highly diverse bacteria into Asgard
222 genomes.

223 **ThyA corresponds to the ancestral thymidylate synthase in *Asgard* archaea**

224 In contrast to ThyX, the ThyA phylogeny (Fig. 4 and supplementary Fig. 5 for a
225 detailed version of the tree with the taxon names) indicates that Asgard sequences form
226 a monophyletic group within various archaea. The ThyA sequences of the Asgard archaea
227 are distributed into four clusters that appear related to diverse methanogenic archaea.
228 This pattern is incompatible with the occurrence of multiple, recent, and independent
229 lateral gene transfers from bacteria as observed above with ThyX (Fig. 4). Finally, the
230 Asgard ThyA sequences appear distant from bacterial and eukaryotic sequences,
231 suggesting that the ancestors of the Asgard archaea contained ThyA. However, this tree
232 topology is also compatible with the more complex scenarios of *i)* the Asgard ancestor
233 inheriting its *thyA* gene from a bacterium, followed by subsequent gene transfers to

234 diverse methanogens, or *ii*) initial *thyA* transfers to methanogenic archaea and
235 subsequent ones to the Asgard ancestor.

236 ***Psyn thyX* is functional in *Escherichia coli***

237 Next, we investigated whether *Psyn thyX* functionally complements growth defects
238 of an *E. coli* strain specifically impaired in thymidylate synthase activity. Towards this goal
239 we designed and constructed a synthetic plasmid, pTwist-*Psyn-ThyX*, where the
240 transcription of *Psyn thyX* is under control of a synthetic T5 promoter, carrying the *lac*
241 operator (*lacO*) sequences. The plasmid also contains an ampicillin resistance marker and
242 a p15A replication origin, as well as a λ t0 terminator, located after the *Psyn thyX* gene
243 (Fig. 5a). As the T5 promoter is recognized by native *E. coli* RNA polymerase, this plasmid
244 expresses an N-terminal His-tagged *Psyn ThyX* protein in any *E. coli* strain.

245 The ability of *Psyn thyX* to permit thymidine-independent growth of the *E. coli*
246 thymidine-auxotroph strain FE013⁸ (Δ *thyA::aphA3*, derived from wild type MG1655) was
247 scored after three days at room temperature or 37°C in the presence of 1mM IPTG using
248 either minimal M9 or thymidine-deprived rich medium L⁺ (see the methods section). Fig.
249 5b shows the formation of individual colonies of *Psyn thyX*/FE013 in the absence of
250 thymidine in the presence of IPTG and appropriate antibiotics (plate on left, streaks 1 and
251 2). Under these thymidine limiting conditions, the strain carrying the control plasmid
252 lacking the insert did not form individual colonies (streak 3, Fig. 5b). Altogether the
253 results of this inter-kingdom complementation experiment indicate that this archaeal
254 *ThyX* protein is fully functional in a bacterial strain and must use *E. coli* folate derivatives
255 for *de novo* thymidylate synthesis.

256

257

258 **Enzymatic activity of *PsynThyX***

259 The recombinant histidine-tagged *PsynThyX* was produced in soluble form and
260 efficiently purified by one-step affinity chromatography (Figure 6a). Western blot
261 analysis using an anti-Histidine-Tag monoclonal antibody identified that revealed a single
262 band with an apparent molecular mass of ~37 kDa (Figure 6b). Thymidylate synthases
263 *ThyX* are dUMP-dependent NADPH oxidases^{25,26}, differently from thymidylate synthase
264 *ThyA*. The formation of dTMP catalyzed by *ThyX* enzymes involves two half-reactions,
265 oxidation and transfer of the methylene group from CH₂H₄Folate to dUMP, to form the
266 final product dTMP. Using a spectrophotometric biochemical assay, we found that *Psyn*
267 *ThyX* catalyzes the NADPH oxidation to NADP⁺ only in the presence of dUMP (Figure 6c).
268 The oxidation of NADPH was assayed with 0.4 μM *PsynThyX* using saturating amounts of
269 FAD co-factor (50 μM) and the substrates dUMP (20 μM) and NADPH (750 μM). The
270 specific activity of *PsynThyX* (0.030 UI.mg⁻¹) is somewhat lower than that of
271 *Mycobacterium tuberculosis* *ThyX* (0.044 UI.mg⁻¹) or *Paramecium bursaria chlorella virus-*
272 *1* (PBCV-1) *ThyX* (0.043 UI.mg⁻¹). However, this level of activity is sufficient for genetic
273 complementation (Fig. 5b). Under these experimental conditions, the hyperbolic
274 saturation curves show a nanomolar affinity of the nucleotide substrate for the enzyme
275 [$K_{m, dUMP} = 235 \pm 35$ nM (Figure 6d)].

276 **Inhibition of *PsynThyX* by folate analogs**

277 Our genetic complementation tests indicate that *Psyn ThyX* must use bacterial
278 folates for thymidylate synthesis. To provide additional support for this notion, we further
279 characterized the *PsynThyX* enzymatic activity by performing kinetics measurements in
280 the presence of molecules that bind to the folate binding pocket of bacterial *ThyX* proteins.
281 These inhibitory studies of *Psyn ThyX* were performed using H₄folate (reaction product),

282 CH₂H₄folate (substrate), and the tight-binding *Mycobacterium tuberculosis* ThyX inhibitor
283 2716^{27, 28}, which all inhibit the NADPH oxidase activity of bacterial ThyX. This analysis
284 revealed that the three bacterial folate analogs tested substantially inhibit *Psyn* ThyX
285 activity compared to an assay without the addition of any molecule (Fig. 6e). Note that the
286 molecule 2716, a potent inhibitor of *Mycobacterium tuberculosis* ThyX, was solubilized in
287 DMSO and results need to be compared to the control condition in presence of 1% DMSO.
288 The three folate analogs presented a percentage of inhibition of *Psyn*ThyX over 50% with
289 high reproducibility (Figure 6f). Altogether our genetic (Fig. 5) and biochemical (Fig. 6)
290 data indicate that bacterial folate-like molecules are efficiently utilized by archaeal *Psyn*
291 ThyX for *de novo* synthesis.

292 **Distribution of the folate-mediated one-carbon metabolism enzymes in *Asgard*** 293 **archaea**

294 As THF derivatives are important biological cofactors that play a wide role in the
295 biosynthesis of DNA, RNA, and proteins, our observations obtained using *Psyn*ThyX
296 prompted us to investigate in more detail *Psyn* folate-dependent biosynthetic networks
297 in genomes and metagenomes of Asgard archaea.

298 Our analyses indicated the sporadic presence of many C₁ generating/transferring
299 and folate-interconverting enzymes in Asgard archaea (Fig. 2). In addition to *thyX* and
300 *thyA* genes, we have analyzed the distribution of *folA*, 5-formyltetrahydrofolate cyclo-
301 ligase (MTHFS), *folD*, methylenetetrahydrofolate reductase (MTHFR), *metH*, formate-
302 tetrahydrofolate ligase (FTHFS), and *purH* genes in the *Psyn* genome and other Asgard
303 metagenomic assemblies (Fig. 2 and [supplementary table 1](#)). Many of these were
304 predicted as likely transfer events by the HGTector (Fig. 3 and [supplementary Fig. 3](#)).
305 ThyA appears the most universal folate-dependent *Asgard* enzyme, whereas the majority

306 of the other folate-related genes have a very sporadic distribution. Nevertheless, the *fold*
307 and *purH* genes are frequently found in *Asgard* archaea except for the *Heimdall* and
308 *Odinarchaeota*, where they are absent or partially deleted.

309 The phylogenies of FTHFS and PurH show that a substantial fraction of the Asgard
310 sequences are grouped with other Archaea (Fig. 7 and supplementary Figs 6-11 for a
311 detailed version of the trees with the taxon names). However, several Asgard sequences
312 are only distantly related to other archaeal sequences and located close to bacterial ones.
313 The Asgard Fola, MTHFS, MTHFR, and MetH sequences appear all related to bacterial
314 sequences, indicating additional multiple and independent events of lateral interkingdom
315 acquisition. By contrast, the Fold sequences ([supplementary figure 11](#)), which are widely
316 distributed in Asgard genomes, form a monophyletic group with few bacterial sequences.
317 This supports the idea that Fold may be the only ancestral folate-dependent enzyme in
318 *Asgard* archaea, without obvious evidence of LGT. Therefore, with the possible exception
319 of Fold, the phylogenetic trees of the folate-related enzymes analyzed above strongly
320 support multiple LGT events from Bacteria (Fig. 7), as Asgard sequences appear as
321 polyphyletic groups that are scattered into the bacterial subtrees in most of the obtained
322 phylogenies.

323 **Reconstruction of the complete folate-mediated one-carbon metabolism network** 324 **in *Psyn***

325 *Psyn* is unique among *Asgard* archaea, as this cultivated species contains, based
326 upon the MAGs analyzed, a remarkable number of predicted folate-dependent enzymes
327 (Fig. 2 and [supplementary table 1](#)). This prompted us to perform a detailed reconstruction
328 of enzymatic reactions participating in an interdependent reaction network required for
329 chemically activating and transferring one-carbon (C₁) units in *Psyn*. Our automated

330 genome-wide analyses, together with highly sensitive manual similarity searches,
331 revealed a high metabolic potential of *Psyn* to use folate derivatives for *de novo* synthesis of
332 not only thymidylate but also of inosine-5'-monophosphate (IMP), and remethylation of
333 homocysteine to methionine (Fig. 8a). Moreover, we determined that both carbon dioxide
334 (at the level of 10-formyl THF) and serine (at the level of CH₂H₄ THF) could provide a
335 feasible source for one-carbon units entering the folate-mediated metabolic network in
336 this species. The complete glycine cleavage system and its folate-binding component GcvT
337 also provide an alternative source of CH₂H₄folate from glycine. According to our metabolic
338 reconstruction studies, the folate network moreover participates in histidine catabolism
339 by converting histidine to glutamate and 5-formyl THF. Many of these predictions are well
340 supported by the strong association of genes encoding folate-dependent enzymes with
341 the physically and functionally linked genes of, among others, glycine cleavage or histidine
342 catabolism (Fig. 8b).

343 Discussion

344 We initiated this study to obtain experimental and evolutionary information on
345 thymidylate synthases in Asgard archaea. This led to the first experimental
346 characterization of the archaeal flavin-dependent thymidylate synthase ThyX² from *Psyn*.
347 Our genetic and biochemical experiments reveal that this enzyme functions as a robust
348 thymidylate synthase in bacterial cells (Fig. 5). Our biochemical results further indicate
349 that *Psyn* ThyX catalyzes dUMP dependent NADPH oxidation, which is significantly
350 inhibited by bacterial folate analogues (Fig. 6), as has been previously shown for bacterial
351 or viral ThyX proteins (for a recent review, see⁹). Considering that archaea are widely
352 thought to use C₁ carriers chemically distinct from their bacterial counterparts, our
353 biochemical observations are unexpected.

354 Our phylogenetic studies indicated for the first time that Asgard archaeal *Psyn*
355 *ThyX* originated in Bacteria (Fig. 4). An alternative explanation could be the possible
356 contamination of the *Psyn* genome assembly with foreign DNA. To exclude this possibility,
357 we analysed the genomic environment of the *Psyn thyX* gene (Fig. 1 and [supplementary](#)
358 [Figure 1](#)). Genes surrounding *Psyn thyX* are unambiguously of archaeal origin, as almost
359 all of the 18 genes upstream or downstream have their first BLAST hit matching with an
360 archaeon. Moreover, 14 of these genes have more than 75% of their top100 BLAST hits
361 belonging to archaea. This analysis indicates that the *Psyn thyX* gene is inserted into an
362 archaeal-like genomic context and is not the result of contamination or an assembly
363 artifact. Thus, this gene represents a *bona fide* thymidylate synthase gene acquired
364 laterally from bacteria. Taken together, our phyletic and phylogenetic analyses of
365 thymidylate synthase genes in Asgard archaea reveal a complex evolutionary history
366 involving many lateral gene transfers from bacteria and subsequent homologous and non-
367 homologous replacements. The large distribution of *thyA* in Asgard metagenomes and the
368 phylogenetic clustering of thymidylate synthase genes (Figs 4 and 5) along with the other
369 archaeal sequences suggest that *thyA* is the ancestral thymidylate synthase, likely in
370 Asgard and possibly in other archaea. This indicates that in *Psyn* the *thyX* gene likely
371 replaced the ancestral Asgard *thyA*; a process known as “non-homologous
372 displacement”²⁹. The non-homologous displacement of the *thyA* gene by *thyX* has already
373 been suggested for diverse bacterial and archaeal species^{2, 10}, but genetic (Fig. 5) and
374 biochemical (Fig. 6) validation of the functionality of the acquired gene via an
375 interkingdom LGT has been lacking.

376 The phylogeny of *Psyn ThyX*, together with its biochemical properties, prompted
377 us to investigate more globally the distribution and phylogeny of folate-dependent
378 enzymes in Asgard archaea. We first observed that *Psyn* carries many folate-dependent

379 enzymes, which raised the possibility that a complex interdependent reaction network
380 operates in this species to chemically activate and transfer one-carbon units. The detailed
381 reconstruction of the *Psyn* one-carbon metabolism (Fig. 8) is fully consistent with the
382 reported lifestyle of this species in anaerobic marine methane seep sediments. In
383 particular, a syntrophic amino acid utilization of *Psyn* using serine, glycine and histidine
384 is highly feasible, as these amino acids can readily provide one carbon units for the folate
385 metabolism at different oxidation levels. *Psyn* likely uses folate derivatives for the
386 synthesis of thymidylate, purine, and methionine. In many cases, a strong association of
387 the genes for the folate-dependent enzymes with additional functionally linked genes, for
388 instance, glycine cleavage or histidine catabolism, was also observed (Fig. 8b). We also
389 noticed that, in agreement with a previous study¹⁸, biosynthetic pathways for folic acid
390 and tetrahydromethanopterin appear incomplete in *Psyn*. However, the growth medium
391 of *Psyn* is supplemented with folic acid-containing vitamin supplement¹⁸, and many
392 bacteria are known to transport folates^{30, 31, 32}. These observations raised the possibility
393 of *Psyn* cross-feeding with the folates produced by its symbiotic partners, as has been
394 previously observed for bacterial communities^{31, 33}.

395 Strikingly, most, if not all, folate-dependent *Psyn* enzymes appear to have a
396 bacterial origin (Figs. 4 and 7). The transfer of genes from Bacteria to Archaea is well
397 documented³⁴. However, our study has revealed a remarkable multigene transfer that
398 globally influences the central metabolism of this enigmatic species. Our phylogenetic
399 analyses did not suggest a well-defined single donor for these transferred genes that are
400 scattered around the *Psyn* genome. Therefore, these transfers likely have occurred on
401 multiple and sequential occasions. Our observations agree well with the previous
402 suggestion that LGT of bacterial metabolic networks may provide an adaptive benefit to
403 *Psyn* in response to changing environments³⁵. Our observations are of general interest as

404 our HGTECTOR analysis (Fig. 3) indicated that at least 4% of *Psyn* genes correspond to
405 transfer events including many genes implicated *e.g.* in amino acid and nucleotide
406 metabolism. Our analyses may underestimate the true extent of LGT for this species, as
407 the initial discovery of the first Asgard genome -a lokiarchaeal species- reported a
408 substantial fraction of genes with bacterial affinities (29% of all genes)³⁶. Widespread
409 interdomain LGTs have also been proposed in uncultured planktonic thaumarchaeota and
410 euryarchaeota³⁷. Association of *Psyn* with co-cultured bacteria might have favoured the
411 gene exchanges, possibly influencing microbial population structure under natural
412 conditions³⁸. Several different archaeal molecular mechanisms facilitating DNA transfer
413 have been described³⁹, providing plausible means for LGT between Bacteria and Asgard
414 archaea. Finally, frequent LGTs between *Psyn* and different bacterial species might
415 partially explain the apparent chimeric profiles of Asgard genomes⁴⁰, which seemingly
416 result from the frequent mixing of ancestral archaeal genes with laterally inherited
417 bacterial ones.

418 Regarding the eukaryogenesis process, the possible contribution of Asgard to the
419 eukaryotic folate metabolism remains elusive. Indeed, our results support the existence
420 of multiple and independent lateral gene transfers from diverse bacteria to Asgard,
421 blurring an exact scenario. Nevertheless, the ThyA family of thymidylate synthases
422 appears ancestral in Asgard considering the phylogenetic positioning of the Asgard
423 sequences deeply nested in the bacterial phylogenetic tree. Our analyses also indicate that
424 eukaryotic ThyA does not derive from Asgard archaea but also has a bacterial origin. Both
425 scenarios suggest that the present-day thymidylate metabolism in Eukarya and Asgard
426 archaea originated in bacteria. It was recently proposed that the contribution of Asgard
427 to the present eukaryotic gene repertoire is very limited⁴¹. However, widespread recent
428 LGTs events from bacteria to Asgard displacing the original archaeal genes would

429 underestimate the contribution of the current day Asgard to the eukaryotic gene
430 repertoire. In turn, our study on the thymidylate metabolism might represent only the tip
431 of an iceberg of functional bacterial genes in Asgard genomes.

432 Taken together, our data support the idea that Asgard archaea harbor highly
433 mosaic genomes that have received many bacterial genes, even for central metabolic
434 processes such as the folate metabolism. We have also experimentally demonstrated that
435 some of the transferred *Psyn* genes, like *Psyn thyX*, are fully functional *in vivo* and *in vitro*.
436 More general studies on Asgard gene evolution are now needed to better appreciate the
437 pervasiveness of interdomain LGT, as well as the origin of the eukaryotic gene repertoire
438 in this fascinating group of prokaryotes.

439 **Materials and Methods**

440 **Bioinformatics methods and genome-wide analyses.** The complete genome sequence
441 available for the *Candidatus* Prometheoarchaeum syntrophicum strain MK-D1 (*Psyn*)
442 chromosome (NZ_CP042905.1) was used for bioinformatics analyses. This genome is
443 4,278 Mb with a GC content of 31.2%. RPKM values for gene expression were determined
444 using the Geneious Prime® 2021.1.1 (Build 2021-03-12). Metabolic reconstructions were
445 performed using the RAST server and KEGG databases. These analyses were
446 complemented with highly sensitive manual similarity searches using HHpred⁴².

447 **Structural alignments and structure modelling.** PROfile Multiple Alignment with
448 predicted Local Structures and 3D constraints (PROMALS3D) was used to align the *Psyn*
449 ThyX sequence with the *Thermotoga (T.) maritima* ThyX structure (PDB 5CHP).
450 PROMALS3D (<http://prodata.swmed.edu/promals3d/promals3d.php>) aligns multiple
451 protein sequences and/or structures, with enhanced information from database searches,

452 secondary structure prediction, and 3D structures. Input sequences and structures used
453 FASTA and PDB formats, respectively. (red: alpha-helix, blue: beta-strand).

454 A model of the *Psyn* ThyX homotetramer was constructed with the protein
455 structure modeling program Modeller using PDB structures 1O26, 3N0B, and 6J61 as
456 templates²¹. Ten initial models were constructed and the best one was chosen using
457 molpdf and DOPE score, and by evaluating the stereochemical quality with PROCHECK. A
458 structural superposition of the model with PDB structure 3GT9 using the Chimera
459 software²² allowed the addition of FAD, dUMP and folate molecules. To improve the
460 structural model obtained, molecular dynamics simulations were performed with
461 CHARMM⁴³ and Namd⁴⁴ molecular mechanics softwares.

462 **Automated prediction of LGT.** Automated analyses of potential LGT events were
463 performed using a HGTector²³. This method first performs Diamond all-against-all
464 similarity searches of each protein-coding sequence of *Psyn*. The NCBI genome database
465 was downloaded and compiled (on August 2021) on a local machine. The output of the
466 similarity searches was passed to the analysis program HGTector where the taxonomy of
467 the self-group was defined as *Candidatus* Prometheoarchaeum syntrophicum (NCBI
468 TaxID: 2594042). The close-group was automatically inferred by HGTector as the
469 “superkingdom Archaea” (TaxID: 2157) that contains 1062 taxa in the reference database.
470 From the statistical distribution of close and distal hits, HGTector predicted a list of
471 potentially transferred genes. The output of the program is a scatter plot, where the
472 horizontal axis represents the close score, the vertical axis represents the distal score, and
473 potential LGT genes are colored in yellow. We slightly modified the source code which
474 was made publicly available at: <https://github.com/cgneo/neoHGT>.

475 **Phylogenetic analyses.** For phylogenetic analyses, we constructed a database with a
476 backbone of 183 *thyX* sequences that cover all the major bacterial and archaeal phyla
477 derived from a previous study¹⁰. These were submitted to BLASTP, with an *e*-value cut-
478 off of 1e-05, against the *Psyn* genome, as well as against the 140 metagenomic bins larger
479 than 2Mb assigned to Asgard archaea publicly available in the Genbank database. We used
480 similar approaches with the other folate enzymes with 355 ThyA, 53 FTHFS, 40 MetH, 58
481 PurH, 70 Fola, 85 Fold, 33 MTHFR and 49 MTHFS sequences to find homologous
482 sequences in Asgard genomes. Homologs found in Asgard genomes were then used as
483 seeds for reciprocal BLAST searches against an NR database (15/01/2021 version) to find
484 the best homolog outside Asgards. Identical sequences (same identifying code or 100%
485 sequence similarity) were finally removed to generate the final dataset.

486 For analysis of the genomic environment of the *Psyn thyX* gene, we collected all the
487 genes located 15 kb downstream and 15 kb upstream of the *Psyn thyX* genes and searched
488 for homologs with BLASTP against an NR database (15/01/2021 version). The first
489 BLASTP hit was retrieved, as well as the percentages of the archaeal sequences in the
490 taxonomy profile of the first top 100 hits.

491 Phylogenetic analyses were performed using the obtained protein sequence data
492 sets (see above) that were aligned using MAFFT v7.388 with default settings⁴⁵. Identical
493 sequences belonging to the same phylum or genus were removed to retain a single
494 representative sequence. Ambiguously aligned sites were removed using trimAl
495 v1.4.rev15⁴⁶ with the “-automated1” and “-phylip” options. The phylogenetic position of
496 the Asgard sequences was then inferred with PhyML v. 1.8.1⁴⁷ using the best substitution
497 models as determined by Smart Model Selection⁴⁸, and 1000 bootstrapped data sets were
498 used to estimate the statistical confidences of the nodes. Trees were then visualized using

499 the FigTree software (<https://github.com/rambaut/figtree/releases>). Raw data including
500 BLAST outputs, table with the accession numbers of each Asgard sequences, alignment
501 files, alignments used for the phylogenies, Newick tree files and additional information
502 can be downloaded at <http://gofile.me/2ppPR/HvmT893hn>.

503 **Design and synthesis of *Psyn ThyX* expression plasmid.** pTwist-*Psyn-ThyX* plasmid
504 (3370bp), carrying the synthetic *Psyn thyX* gene of 897bp was synthesized by Twist
505 Biosciences (<https://www.twistbioscience.com>) and confirmed by sequencing. In the
506 final design, the target gene *Psyn thyX* was placed under the control of a strong
507 bacteriophage T5 promotor carrying *lacO* sites. This construct also carried an appropriate
508 ribosome binding site, a lambda t0 transcriptional terminator, and the Amp^R gene. The
509 pTwist-*Psyn-ThyX* plasmid uses a p15A origin of replication and contains the codon-
510 optimized synthetic gene of *Psyn-ThyX* with an N-terminal 6xHis sequence.

511 **Genetic complementation tests.** The ability of *Psyn thyX* to permit thymidine-
512 independent growth of the *E. coli* thymidine-auxotroph FE013 strain ($\Delta thyA::aphA3 lacI$,
513 derived from wild type MG1655 strain)¹⁰ was scored after three days at room
514 temperature or at 37°C in the presence of 1mM IPTG using either minimal M9 (shown in
515 Figure 5) or thymidine-deprived enriched medium L⁺. Eight individual colonies all
516 demonstrating similar phenotypes in the absence of thymidine were tested.

517 **Protein purification.** *E. coli* FE013 cells carrying pTwist-*Psyn-ThyX* were grown at 37°C
518 and shaken at 150rpm in liquid Luria Bertani medium supplemented with ampicillin (final
519 100 µg/mL) until an absorbance of 600 nm around 0.7 was reached. Expression of
520 recombinant *Psyn ThyX* was induced by the addition of 0.5 mM IPTG (isopropyl-b-D-
521 thiogalactopyranoside) during 3 hours at 37°C. Cells were harvested by centrifugation at
522 6000 x g at 4°C for 30min before storage at -20°C. The *Psyn ThyX* recombinant protein

523 was purified on Protino Ni-TED column (Macherey-Nagel) as previously described⁴⁹.
524 Stepwise protein elution was performed with imidazole at 50mM, 100mM, 150mM and
525 250mM in a phosphate buffer (50mM Na₂HPO₄-NaH₂PO₄, pH8, 300mM NaCl). The 100mM
526 imidazole fractions containing the *Psyn* ThyX enzyme were pooled, buffer exchanged on
527 Econo-Pac PD-10 Columns (Bio-Rad Laboratories, Hercules, CA) with phosphate buffer
528 (50mM Na₂HPO₄-NaH₂PO₄, pH8, 300mM NaCl), concentrated to a final concentration of
529 4μM and stored at -20°C in the same phosphate buffer complemented with glycerol (10%
530 v/v).

531 The purified proteins were analyzed on 12% TGX Precast Gels (Bio-Rad
532 Laboratories, Hercules, CA) followed by Coomassie Brilliant Blue staining. For
533 immunoblotting detection, after protein transfer, the nitrocellulose membranes (Bio-
534 Rad) were treated with 20% blocking buffer (Li-Cor Biosciences, Lincoln, US) followed by
535 incubation with anti-His-tag mouse primary antibody (dilution 1/5000, Bio-Rad) and
536 revelation with a IRDye 800CW Goat anti-mouse IgG secondary antibody (dilution
537 1/10000, Li-Cor Biosciences) using a ChemiDoc™ Touch Imaging System (Bio-Rad).

538 **Enzymatic activity tests.** The NADPH oxidase assay consists in measuring the conversion
539 of NADPH to NADP⁺ via the decrease in absorbance at 340 nm. During the test, in 96-well
540 plates, one hundred microlitres of the standard reaction mixture (50mM Buffer Na₂HPO₄-
541 NaH₂PO₄ pH 8, NaCl 300 mM, MgCl₂ 2 mM, FAD 50 μM, β-mercaptoethanol 1.43 mM, dUMP
542 20 μM, NADPH 750 μM, glycerol 8%, 0.4 μM of purified *Psyn*ThyX) were incubated in a
543 Chameleon II microplate reader (Hidex) at 25°C. Measurements took place over 15
544 minutes and were done in triplicates. The enzyme-free reaction was used as a negative
545 control.

546 Tests of ThyX inhibition were performed by incubation of molecules at 100 or
547 200 μ M with *PsynThyX* enzyme (0.4 μ M) in the standard reaction mixture, without
548 NADPH, for 10 min at 25°C before starting measurements by automatically injecting
549 NADPH. The molecules “2716” and C8C1 were solubilized in dimethylsulfoxide (DMSO)
550 and used at 1% final concentration of DMSO during the test. % of inhibition was calculated
551 using the following equation: $((V_o - V_i)/V_o) * 100$; V_o and V_i are, respectively, the initial
552 rates of the reaction without or with addition of molecule to the assay.

553 **References**

- 554 1. Iyer LM, *et al.* Quod erat demonstrandum? The mystery of experimental validation
555 of apparently erroneous computational analyses of protein sequences. *Genome*
556 *Biol* **2**, RESEARCH0051 (2001).
557
- 558 2. Myllykallio H, Lipowski G, Leduc D, Filee J, Forterre P, Liebl U. An alternative flavin-
559 dependent mechanism for thymidylate synthesis. *Science* **297**, 105-107 (2002).
560
- 561 3. Giladi M, Bitan-Banin G, Mevarech M, Ortenberg R. Genetic evidence for a novel
562 thymidylate synthase in the halophilic archaeon *Halobacterium salinarum* and in
563 *Campylobacter jejuni*. *FEMS Microbiol Lett* **216**, 105-109 (2002).
564
- 565 4. Myllykallio H, Leduc D, Filee J, Liebl U. Life without dihydrofolate reductase Fola.
566 *Trends Microbiol* **11**, 220-223 (2003).
567
- 568 5. Leduc D, *et al.* Functional evidence for active site location of tetrameric
569 thymidylate synthase X at the interphase of three monomers. *Proc Natl Acad Sci U*
570 *SA* **101**, 7252-7257 (2004).
571
- 572 6. Carreras CW, Santi DV. The catalytic mechanism and structure of thymidylate
573 synthase. *Annu Rev Biochem* **64**, 721-762 (1995).
574
- 575 7. Schober AF, *et al.* A Two-Enzyme Adaptive Unit within Bacterial Folate
576 Metabolism. *Cell Rep* **27**, 3359-3370 e3357 (2019).
577
- 578 8. Escartin F, Skouloubris S, Liebl U, Myllykallio H. Flavin-dependent thymidylate
579 synthase X limits chromosomal DNA replication. *Proc Natl Acad Sci U S A* **105**,
580 9948-9952 (2008).
581
- 582 9. Myllykallio H, Sournia P, Heliou A, Liebl U. Unique Features and Anti-microbial
583 Targeting of Folate- and Flavin-Dependent Methyltransferases Required for
584 Accurate Maintenance of Genetic Information. *Front Microbiol* **9**, 918 (2018).
585

- 586 10. Stern A, Mayrose I, Penn O, Shaul S, Gophna U, Pupko T. An evolutionary analysis
587 of lateral gene transfer in thymidylate synthase enzymes. *Syst Biol* **59**, 212-225
588 (2010).
589
- 590 11. Giovannoni SJ, *et al.* Genome streamlining in a cosmopolitan oceanic bacterium.
591 *Science* **309**, 1242-1245 (2005).
592
- 593 12. Arnoriaga-Rodriguez M, *et al.* Obesity-associated deficits in inhibitory control are
594 phenocopied to mice through gut microbiota changes in one-carbon and aromatic
595 amino acids metabolic pathways. *Gut*, (2021).
596
- 597 13. Krone UE, McFarlan SC, Hogenkamp HP. Purification and partial characterization
598 of a putative thymidylate synthase from *Methanobacterium thermoautotrophicum*.
599 *Eur J Biochem* **220**, 789-794 (1994).
600
- 601 14. Nyce GW, White RH. dTMP biosynthesis in Archaea. *J Bacteriol* **178**, 914-916
602 (1996).
603
- 604 15. Maden BE. Tetrahydrofolate and tetrahydromethanopterin compared:
605 functionally distinct carriers in C1 metabolism. *Biochem J* **350 Pt 3**, 609-629
606 (2000).
607
- 608 16. Adam PS, Borrel G, Gribaldo S. An archaeal origin of the Wood-Ljungdahl H4MPT
609 branch and the emergence of bacterial methylotrophy. *Nat Microbiol* **4**, 2155-2163
610 (2019).
611
- 612 17. Zaremba-Niedzwiedzka K, *et al.* Asgard archaea illuminate the origin of eukaryotic
613 cellular complexity. *Nature* **541**, 353-358 (2017).
614
- 615 18. Imachi H, *et al.* Isolation of an archaeon at the prokaryote-eukaryote interface.
616 *Nature* **577**, 519-525 (2020).
617
- 618 19. Liu Y, *et al.* Expanded diversity of Asgard archaea and their relationships with
619 eukaryotes. *Nature* **593**, 553-557 (2021).
620
- 621 20. Williams TA, Cox CJ, Foster PG, Szollosi GJ, Embley TM. Phylogenomics provides
622 robust support for a two-domains tree of life. *Nat Ecol Evol* **4**, 138-147 (2020).
623
- 624 21. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Curr*
625 *Protoc Protein Sci* **86**, 291-2937 (2016).
626
- 627 22. Pettersen EF, *et al.* UCSF Chimera--a visualization system for exploratory research
628 and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
629
- 630 23. Zhu Q, Kosoy M, Dittmar K. HGTector: an automated method facilitating genome-
631 wide discovery of putative horizontal gene transfers. *BMC Genomics* **15**, 717
632 (2014).
633

- 634 24. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for
635 Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*
636 **428**, 726-731 (2016).
637
- 638 25. Graziani S, *et al.* Functional analysis of FAD-dependent thymidylate synthase ThyX
639 from *Paramecium bursaria* Chlorella virus-1. *J Biol Chem* **279**, 54340-54347
640 (2004).
641
- 642 26. Graziani S, *et al.* Catalytic mechanism and structure of viral flavin-dependent
643 thymidylate synthase ThyX. *J Biol Chem* **281**, 24048-24057 (2006).
644
- 645 27. Abu El Asrar R, *et al.* Discovery of a new *Mycobacterium tuberculosis* thymidylate
646 synthase X inhibitor with a unique inhibition profile. *Biochem Pharmacol* **135**, 69-
647 78 (2017).
648
- 649 28. Modranka J, *et al.* Synthesis and Structure-Activity Relationship Studies of
650 Benzo[b][1,4]oxazin-3(4H)-one Analogues as Inhibitors of *Mycobacterial*
651 Thymidylate Synthase X. *ChemMedChem* **14**, 645-662 (2019).
652
- 653 29. Galperin MY, Koonin EV. Sources of systematic error in functional annotation of
654 genomes: domain rearrangement, non-orthologous gene displacement and operon
655 disruption. *In Silico Biol* **1**, 55-67 (1998).
656
- 657 30. Shane B, Stokstad EL. Transport and metabolism of folates by bacteria. *J Biol Chem*
658 **250**, 2243-2253 (1975).
659
- 660 31. Graber JR, Breznak JA. Folate cross-feeding supports symbiotic homoacetogenic
661 spirochetes. *Appl Environ Microbiol* **71**, 1883-1889 (2005).
662
- 663 32. Leduc D, *et al.* Flavin-dependent thymidylate synthase ThyX activity: implications
664 for the folate cycle in bacteria. *J Bacteriol* **189**, 8537-8545 (2007).
665
- 666 33. Soto-Martin EC, *et al.* Vitamin Biosynthesis by Human Gut Butyrate-Producing
667 Bacteria and Cross-Feeding in Synthetic Microbial Communities. *mBio* **11**, (2020).
668
- 669 34. Gogarten JP, Doolittle WF, Lawrence JG. Prokaryotic evolution in light of gene
670 transfer. *Mol Biol Evol* **19**, 2226-2238 (2002).
671
- 672 35. Pal C, Papp B, Lercher MJ. Adaptive evolution of bacterial metabolic networks by
673 horizontal gene transfer. *Nat Genet* **37**, 1372-1375 (2005).
674
- 675 36. Spang A, *et al.* Complex archaea that bridge the gap between prokaryotes and
676 eukaryotes. *Nature* **521**, 173-179 (2015).
677
- 678 37. Deschamps P, Zivanovic Y, Moreira D, Rodriguez-Valera F, Lopez-Garcia P.
679 Pangenome evidence for extensive interdomain horizontal transfer affecting
680 lineage core and shell genes in uncultured planktonic thaumarchaeota and
681 euryarchaeota. *Genome Biol Evol* **6**, 1549-1563 (2014).
682

- 683 38. Polz MF, Alm EJ, Hanage WP. Horizontal gene transfer and the evolution of
684 bacterial and archaeal population structure. *Trends Genet* **29**, 170-175 (2013).
685
- 686 39. Wagner A, *et al.* Mechanisms of gene flow in archaea. *Nat Rev Microbiol* **15**, 492-
687 501 (2017).
688
- 689 40. Garg SG, *et al.* Anomalous Phylogenetic Behavior of Ribosomal Proteins in
690 Metagenome-Assembled Asgard Archaea. *Genome Biol Evol* **13**, (2021).
691
- 692 41. Knopp M, Stockhorst S, van der Giezen M, Garg SG, Gould SB. The Asgard Archaeal-
693 Unique Contribution to Protein Families of the Eukaryotic Common Ancestor Was
694 0.3. *Genome Biol Evol* **13**, (2021).
695
- 696 42. Zimmermann L, *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit
697 with a New HHpred Server at its Core. *J Mol Biol* **430**, 2237-2243 (2018).
698
- 699 43. Brooks BR, *et al.* CHARMM: the biomolecular simulation program. *J Comput Chem*
700 **30**, 1545-1614 (2009).
701
- 702 44. Phillips JC, *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**,
703 1781-1802 (2005).
704
- 705 45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
706 improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).
707
- 708 46. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated
709 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-
710 1973 (2009).
711
- 712 47. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New
713 algorithms and methods to estimate maximum-likelihood phylogenies: assessing
714 the performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010).
715
- 716 48. Lefort V, Longueville JE, Gascuel O. SMS: Smart Model Selection in PhyML. *Mol Biol*
717 *Evol* **34**, 2422-2424 (2017).
718
- 719 49. Ulmer JE, Boum Y, Thouvenel CD, Myllykallio H, Sibley CH. Functional analysis of
720 the Mycobacterium tuberculosis FAD-dependent thymidylate synthase, ThyX,
721 reveals new amino acid residues contributing to an extended ThyX motif. *J*
722 *Bacteriol* **190**, 2056-2064 (2008).
723

724 **Acknowledgements.** We thank CNRS, INSERM, and E. Polytechnique for their financial
725 support. X.L is a student of the Bachelor of Science program of E. Polytechnique.

726 **Author contributions.** J.F., H.B., U.L, and H.M. conceived the study and wrote the
727 manuscript with contributions from all other authors. J.F. and H.M performed

728 phylogenetic and bioinformatics analyses. H.B and L.M provided biochemical data and
729 U.L and H.M performed genetic experiments. J-C.L. provided a structural model and Z.L
730 performed HGTector analyses. All authors read and commented on the manuscript.

731 **Competing interests.** The authors declare no competing interests.

732 **Additional information.** Supplementary information for this manuscript includes the
733 supplementary figures 1-9 and the supplementary table 1.

734

735 **Figure legends:**

736 **Fig. 1 a** ThyA- and ThyX-dependent folate cycles. ThyA catalyzes the methylation of dUMP
737 to dTMP, leading to the formation of H₂folate that is subsequently reduced by Fola (left
738 panel). The flavoenzyme ThyX uses methylene from CH₂H₄folate and acquires the
739 reducing hydride from NADPH. CH₂H₄folate functions only as a carbon source, resulting
740 in H₄folate at the end of the catalytic cycle (right panel). Serine hydroxymethyltransferase
741 GlyA is universally present. **b** *Psyn thyX* is located between the nucleotides 765,745 and
742 764,847 on the genome of *Psyn* (NZ_CP042905, 4 427 796 bp). The *Psyn thyX* gene is
743 surrounded by the *nif3* and DUF2095 genes. For details, see text. Structure-based
744 sequence alignment of *Psyn* ThyX with *Thermotoga maritima* ThyX (corresponding to PDB
745 structure 5CHP) is also shown. Functionally important residues are indicated above the
746 alignment. Asterisks refer to folate binding, stars to nucleotide-binding and filled circles
747 to flavin binding residues. **c** Structural model for the *Psyn* ThyX homotetramer (left panel)
748 using PDB structures 1O26, 3N0B and 6J61 as templates. Superposition of the model with
749 PDB structure 3GT9 allowed the addition of FAD (cyan), dUMP (yellow), and folate
750 (green) molecules (middle panel) to the structural model. The substrate and co-factor
751 configuration is highlighted in the right panel.

752 **Fig. 2.** Heatmap for phyletic distribution of different folate-dependent enzymes in
753 different phyla of Asgard archaea. The red color indicates the presence of the entire gene,
754 orange designates a partial copy. *Psyn* is indicated with an asterisk and its gene
755 distribution is framed in black. The (meta)genomic assembly and gene names are in
756 [supplementary table 1](#) (sheet “distribution of folate genes”).

757 **Fig. 3.** Automated prediction of gene transfers between *Psyn* and bacteria. **a** Plot
758 indicating the distributions of “close” and “distal” scores revealed by similarity searches.
759 The scatter plot was created by the HGTector. Potentially transferred genes (n=149) are
760 indicated in yellow, whereas the other coloured labels refer to *Psyn* folate-dependent
761 enzymes (for annotation, see [supplementary Figs. 2 and 3](#)). **b** Distribution of the
762 functional groups for potentially transferred *Psyn* genes obtained using the BlastKOALA
763 annotation²⁴.

764 **Fig. 4** Unrooted Maximum likelihood phylogenetic tree of ThyX (198 sequences,
765 alignment of 157 amino acids with the LG+G+I phylogenetic model) and ThyA (390
766 sequences, alignment of 121 amino acids with the LG+G+I model) with a focus on Asgard

767 proteins. The obtained ThyX tree indicates multiple lateral gene transfers from bacteria
768 to *Psyn* and other Asgard genomes. By contrast, the ThyA tree refutes the occurrence of
769 recent and multiple lateral gene transfers from bacteria to Asgard genomes. Asgard
770 sequences are indicated in red, other archaea in blue, bacteria in black, and eukarya in
771 green. The scale bar indicates the average number of substitutions per site. Bootstrap
772 values are proportional to the size of the circles on the branches. . The complete trees are
773 available in the supplementary data.

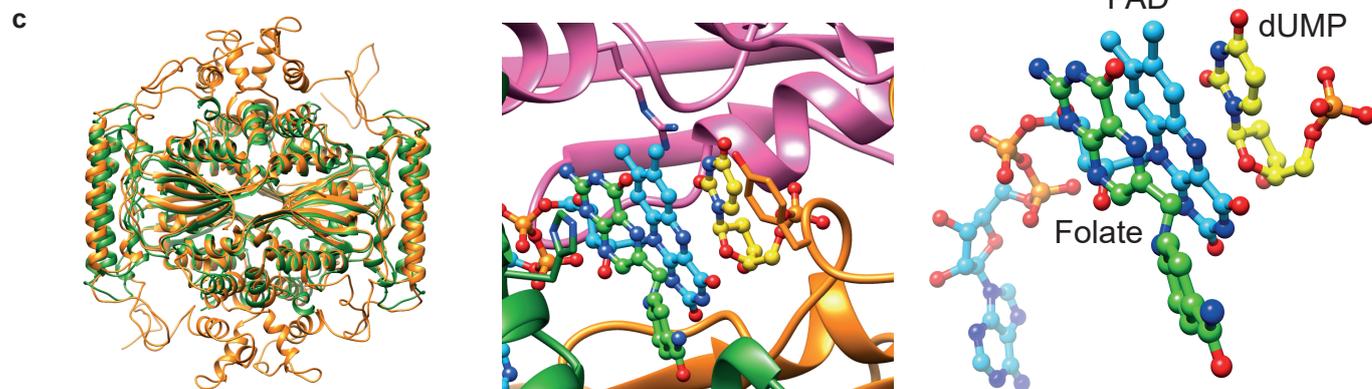
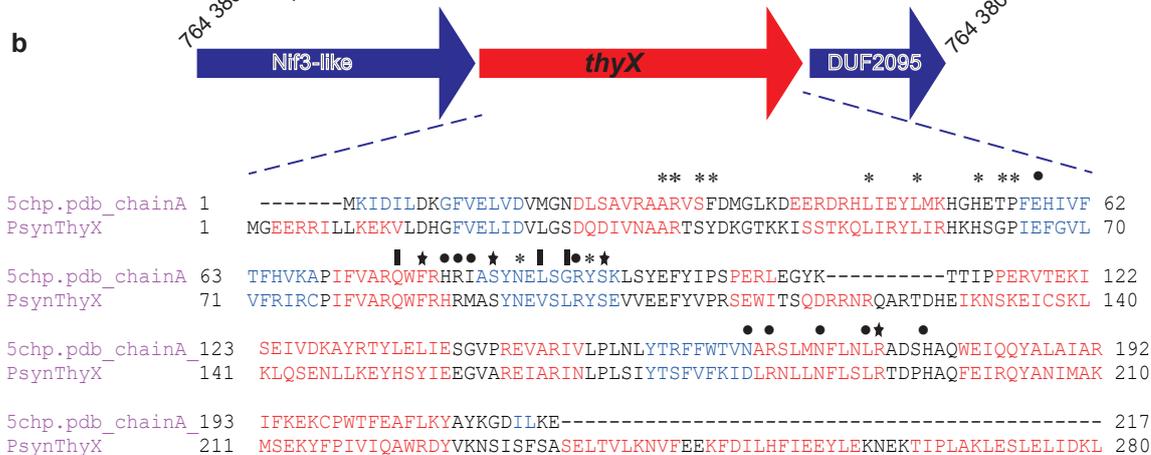
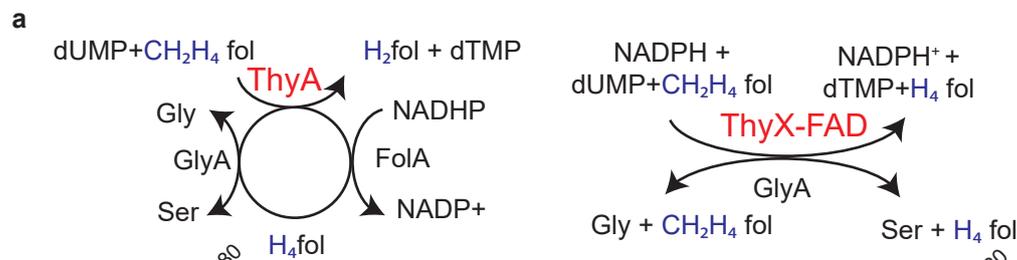
774 **Fig. 5. a** Map for the synthetic plasmid pTwist-*Psyn*-ThyX. Transcription of the *Psyn thyX*
775 gene is under the control of a synthetic T5 promoter carrying *lacO*. An ampicillin
776 resistance marker and a p15A replication origin are integrated, as well as a lambda *tO*
777 terminator. **b** Interkingdom complementation assay using *Psyn thyx*. The ability of *Psyn*
778 *thyX* to permit thymidine-independent growth of *E. coli* thymidine-auxotroph FE013⁸
779 (Δ *thyA*, *lacI*) was scored after three days at 37°C in the presence of 1mM IPTG using
780 minimal M9 medium. 1 and 2 correspond to two independent clones of *Psyn thyX*/FE013
781 whereas 3 and 4 are negative (no insert) and positive (*C. trachomatics thyX*) controls,
782 respectively. The plate on the left contains IPTG, but no thymidine, whereas the plate on
783 the right contains thymidine, but no IPTG.

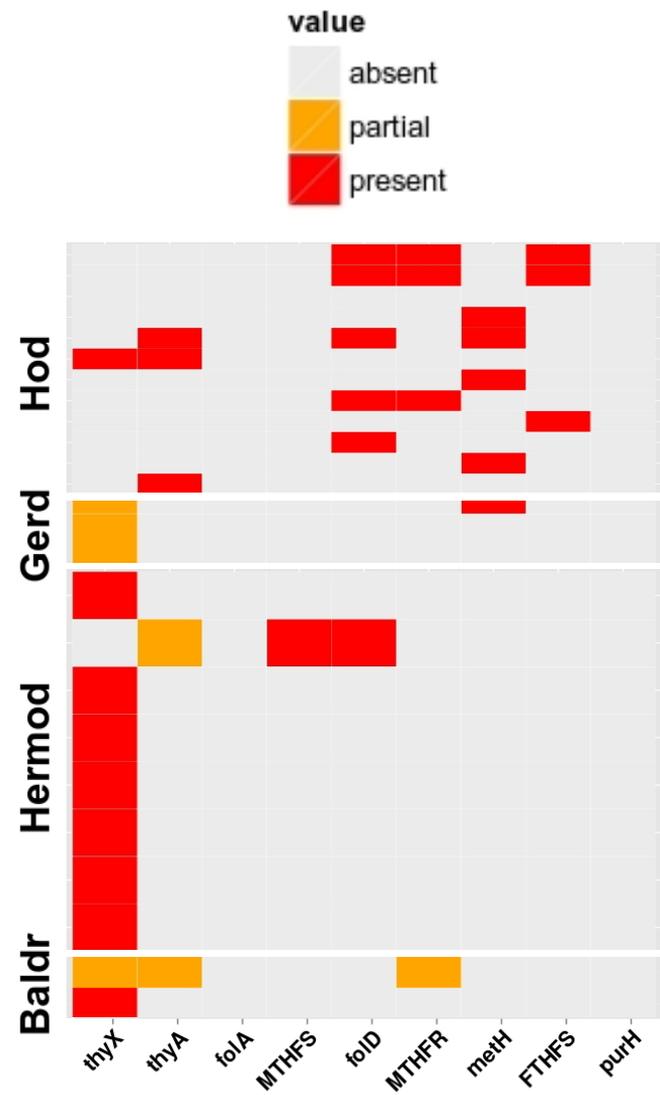
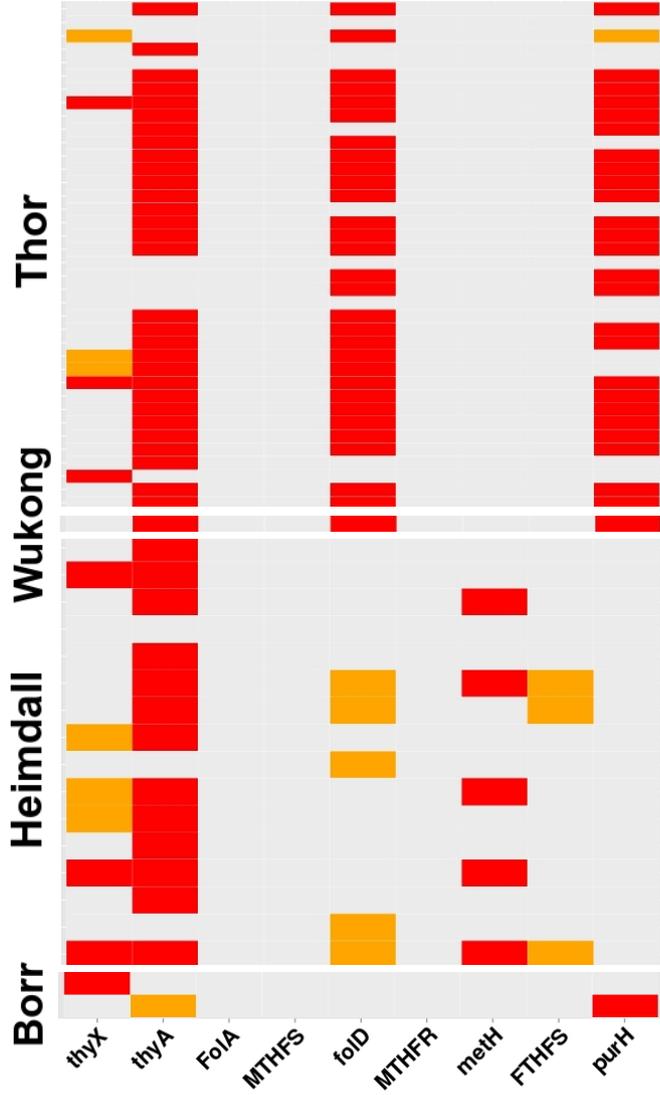
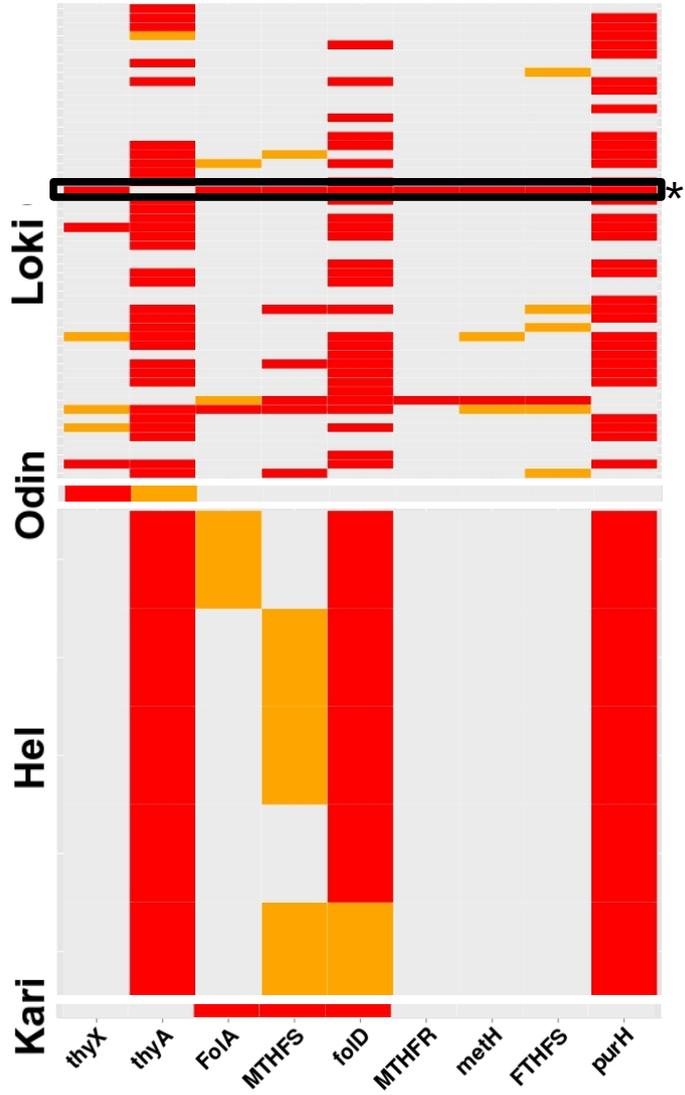
784 **Fig. 6.** Biochemical characterization and inhibition analyses of *Psyn* ThyX. **a** 12% TGX™
785 PAGE of *Psyn* ThyX purified by one-step affinity chromatography. 0.5, 1, and 2 micrograms
786 of purified protein were loaded from the left to the right. MW: molecular marker. **b**
787 Western immunoblot of IPTG-induced and non-induced whole cell lysates of *E. coli* FE013
788 using anti-His monoclonal antibodies. A single band with an apparent molecular mass of
789 ~37 kDa is revealed. T⁺: control lane (*Pyrococcus abyssi* RNA ligase). **c** Representative
790 NADPH oxidation activity curves of *Psyn* ThyX as measured following the absorption
791 change at 340 nm. The control without enzyme is also shown. Assay mixtures contained
792 0.4 μM enzyme, 50 μM FAD, 20 μM dUMP, and 750 μM NADPH. **d** Oxidation rate of NADPH
793 under saturating dUMP concentrations. The affinity of dUMP for *Psyn* ThyX was
794 determined from the hyperbolic saturation curve. **e** Evaluation of inhibiting properties of
795 H₄folate (200 μM), CH₂H₄folate (200 μM), and the folate analog '2716' (100 μM) ^{23, 24}.
796 Molecule '2716' is solubilized in DMSO. Controls: without molecule and in presence of 1%
797 DMSO. **f** Percentage of inhibition calculated for the three folate derivatives or inhibitors.
798 Standard deviations (S.D) of three independent measurements are also shown.

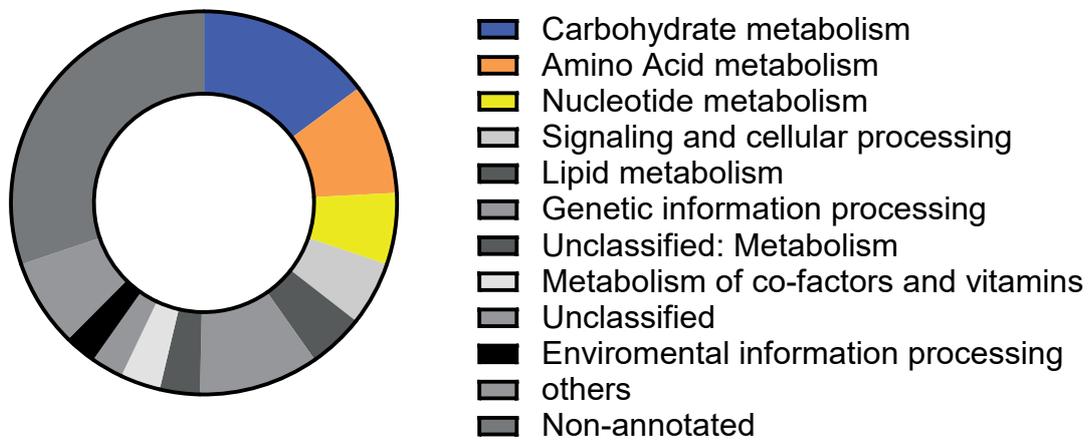
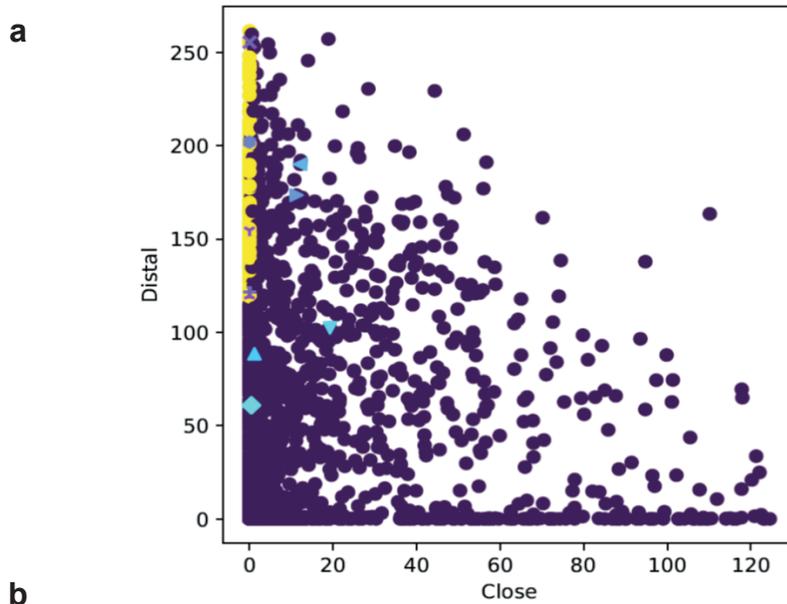
799 **Fig. 7.** Unrooted Maximum likelihood phylogenetic tree of the folate-related proteins :
800 FoaA (83 sequences, alignment of 153 amino acids with the CpREV+G phylogenetic
801 model), FoaD (130 sequences, alignment of 244 amino acids with the BLOSUM62+G
802 model), MTHFS (56 sequences, alignment of 149 amino acids with the LG+G model),
803 FTHFS (79 sequences, alignment of 509 amino acids with the BLOSUM62+G model), PurH
804 (94 sequences, alignment of 437 amino acids with the BLOSUM62+G model), MetH (47
805 sequences, alignment of 131 amino acids with the BLOSUM62+G model) and MTHFR (53
806 sequences, alignment of 280 amino acids with the BLOSUM62+G model) with a focus on
807 Asgard proteins. The obtained trees indicate multiple lateral gene transfers from bacteria
808 to *Psyn* and other Asgard genomes for most folate genes. Asgard FoaD sequences form a
809 monophyletic group with the other Archaeal sequences refuting gene transfer
810 occurrences. Asgard sequences are indicated in red and *Psyn* with red arrows, other
811 archaea are in blue, bacteria in black, and Eukarya in green. The scale bar indicates the
812 average number of substitutions per site. Bootstrap values are proportional to the size of
813 the circles on the branches. The complete trees are available in the supplementary data.

814 **Fig. 8. a** Automated and manual procedures were used for metabolic reconstruction of the
815 complete *Psyn* folate-dependent metabolism. Abbreviations not explained in the figure
816 are DHF, dihydrofolate; THF, tetrahydrofolate; IMP, inosine monophosphate. **b** Functional
817 connections between the folate-dependent *Psyn* enzymes (indicated in yellow) as
818 suggested by the genomic context analyses. Purple and gray refer to genes functionally
819 linked with the folate-dependent enzymes. For additional discussion, see text.

820

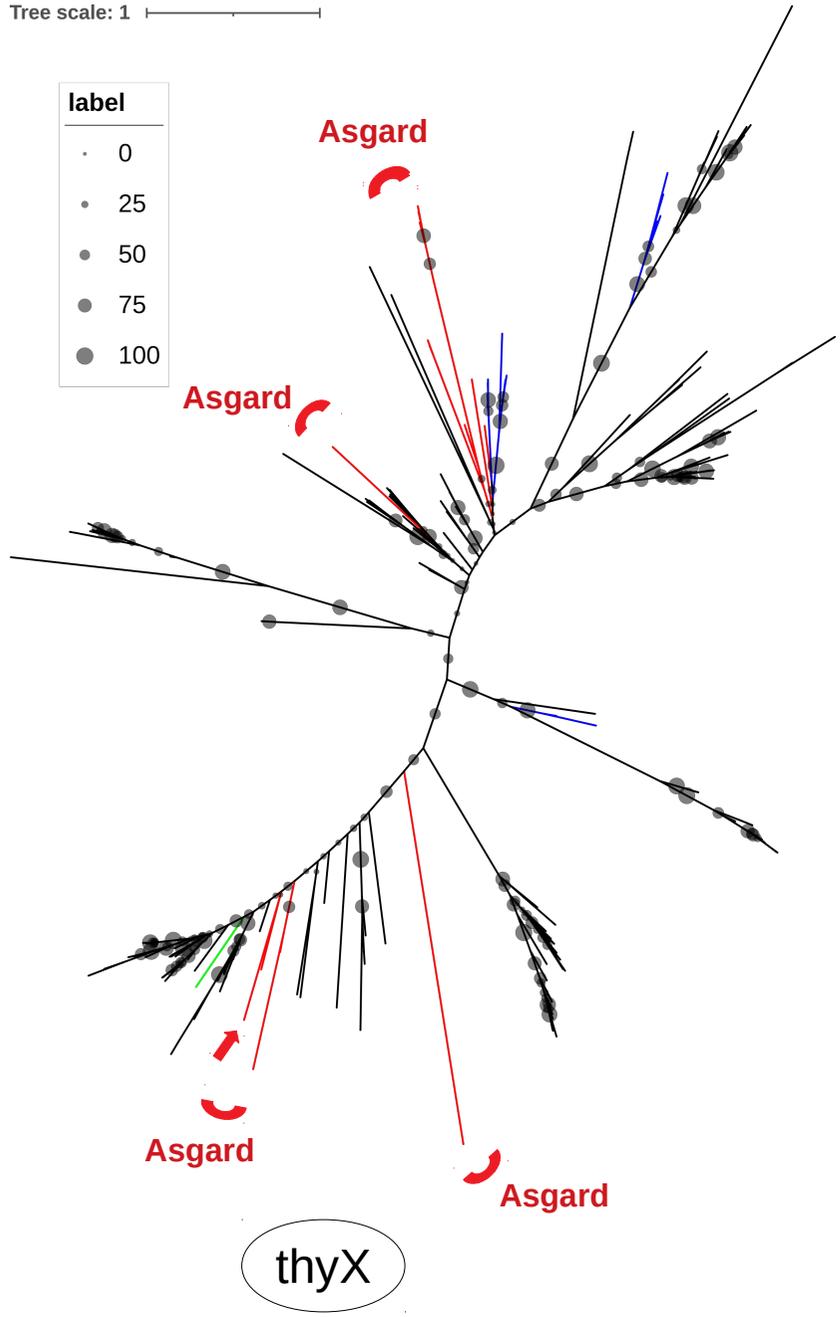






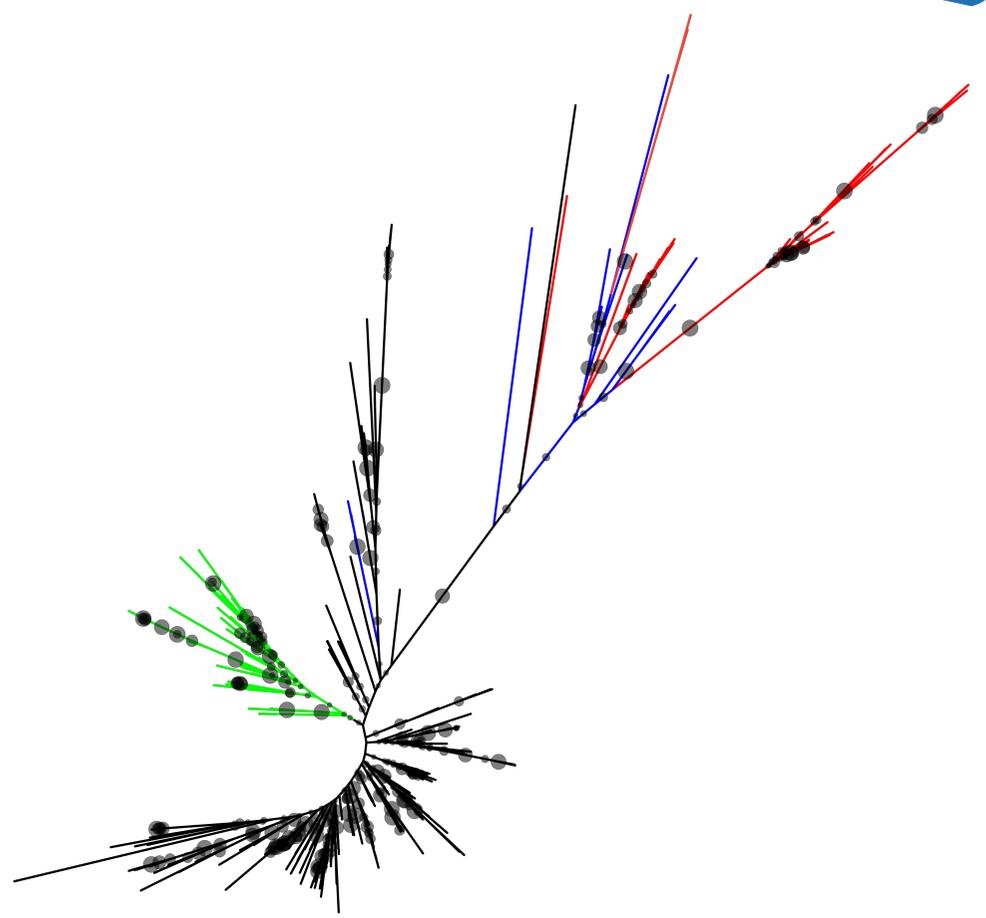
Tree scale: 1

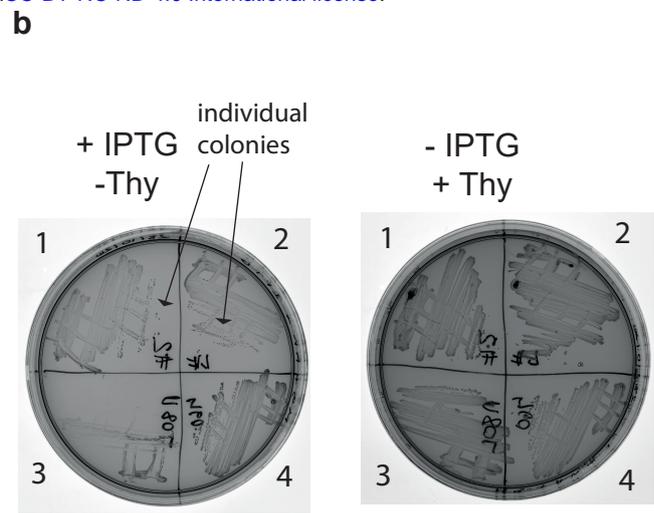
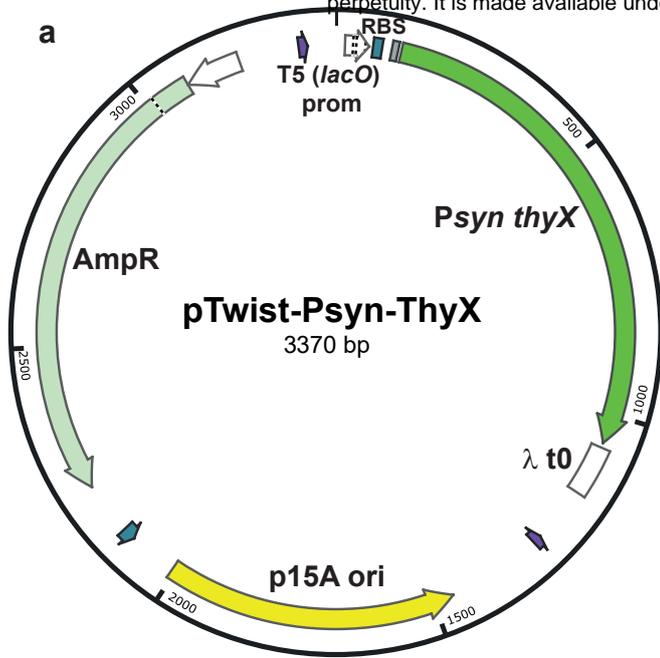
label
· 0
· 25
● 50
● 75
● 100

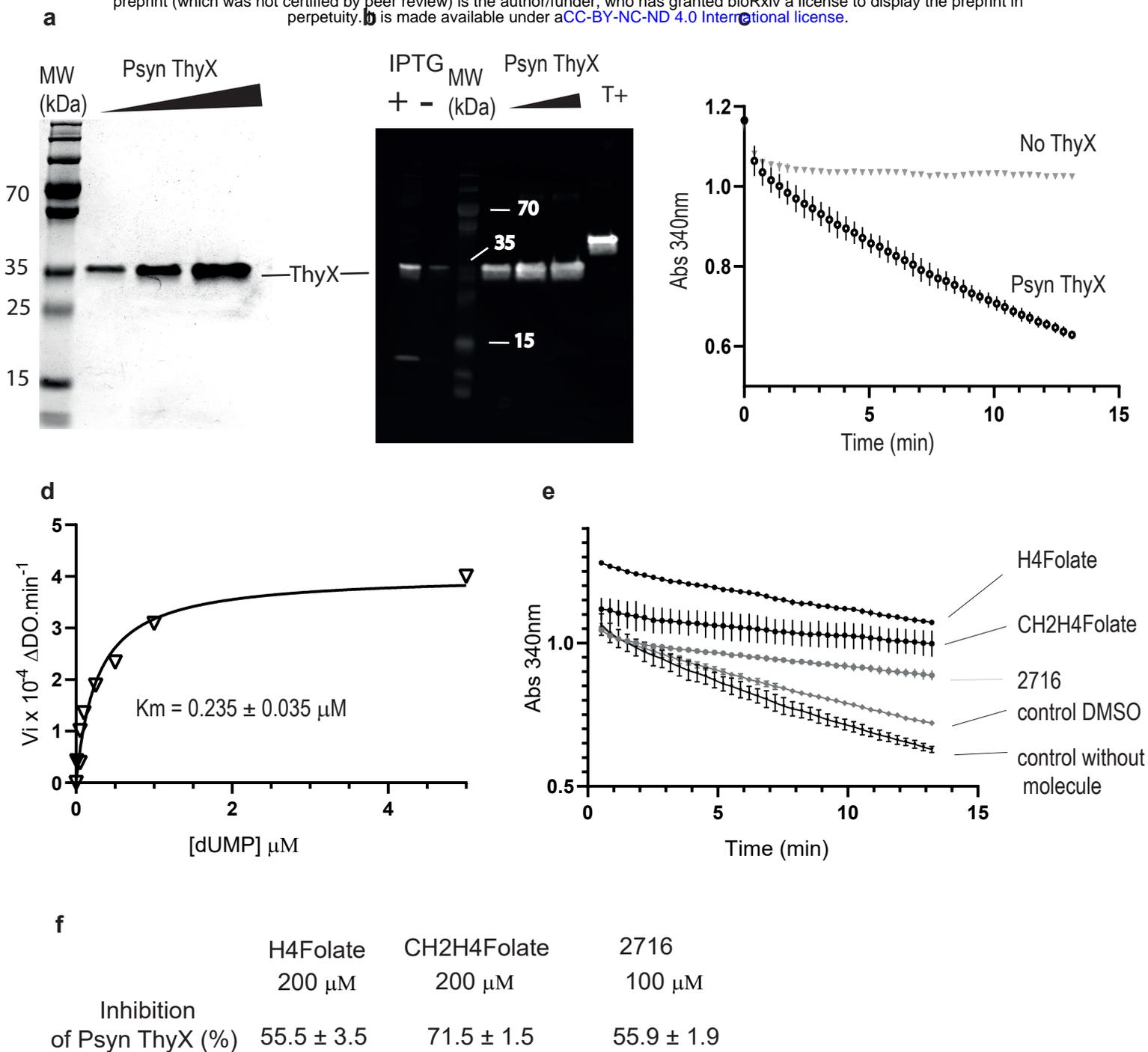


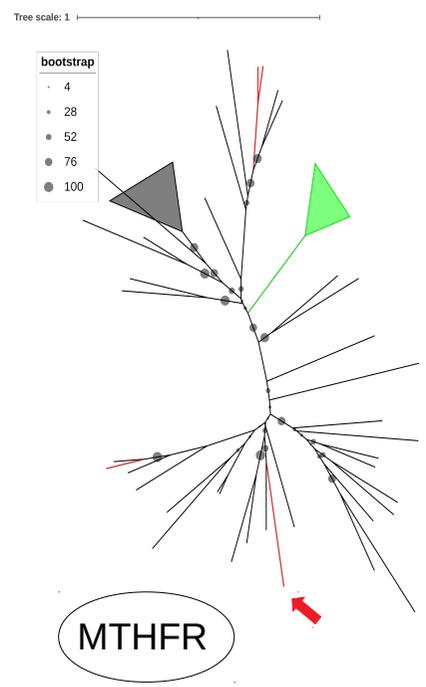
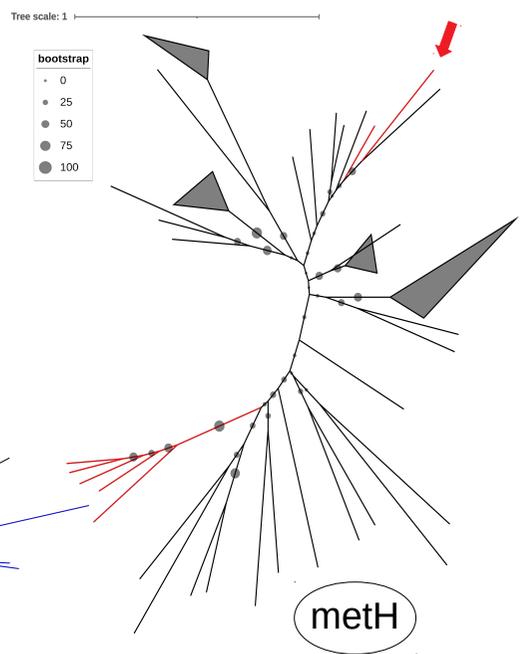
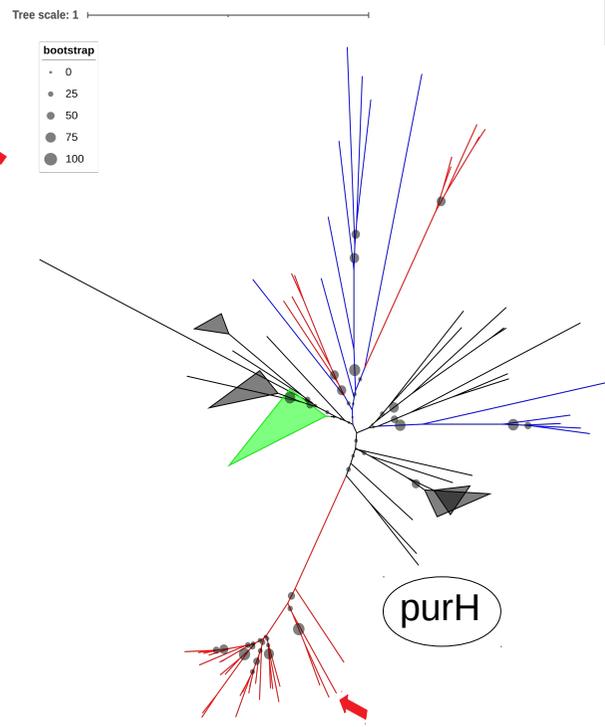
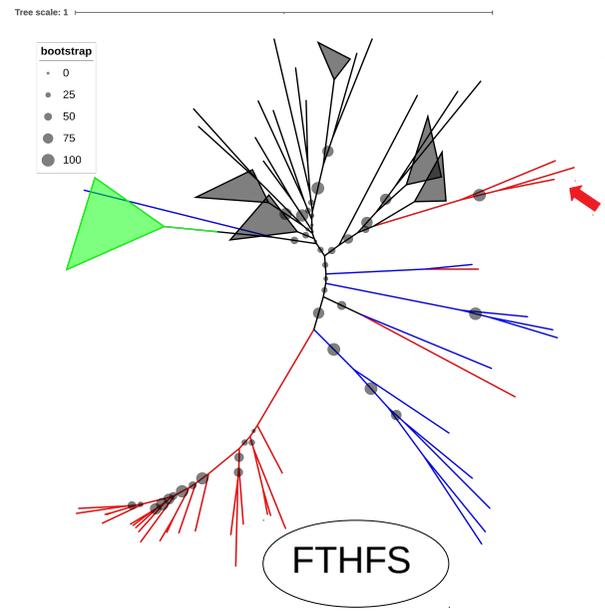
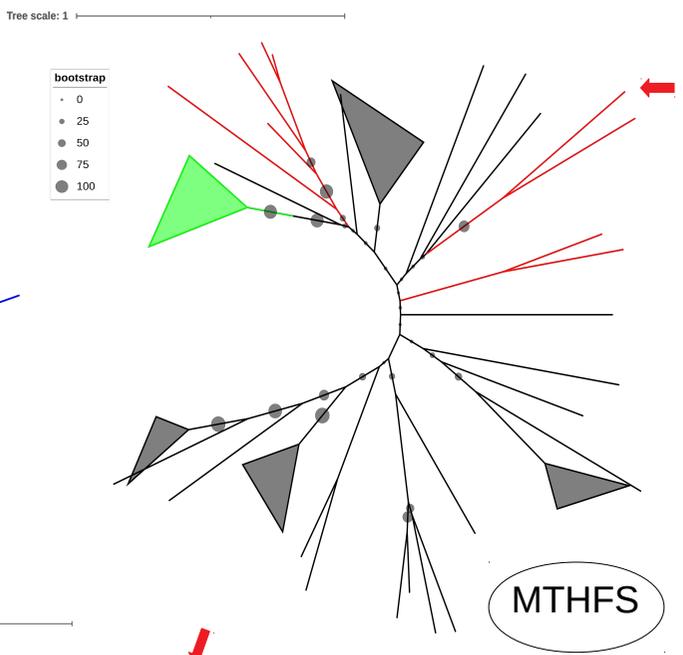
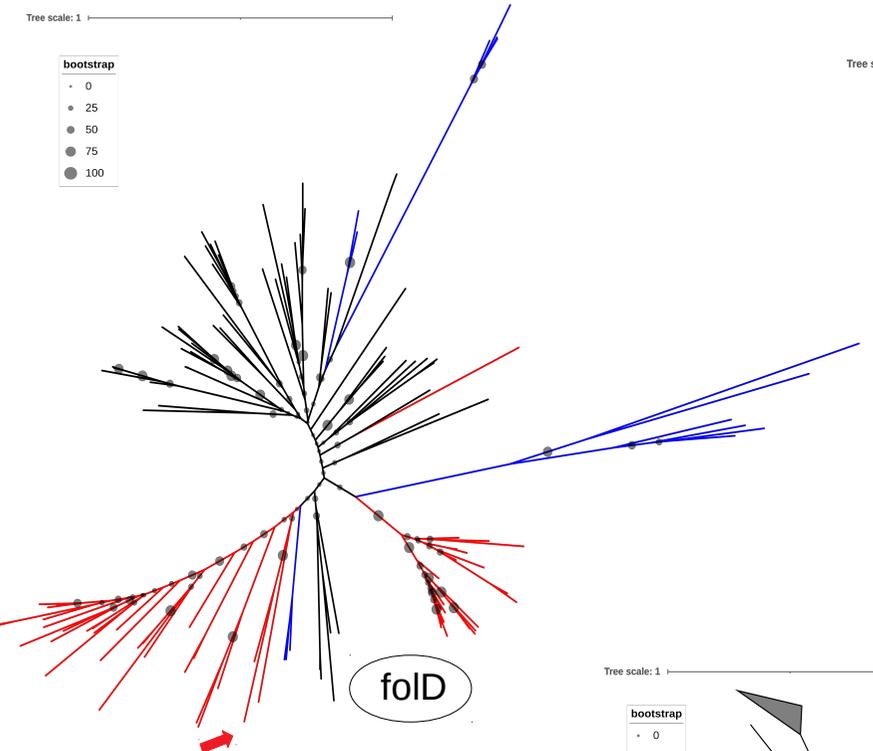
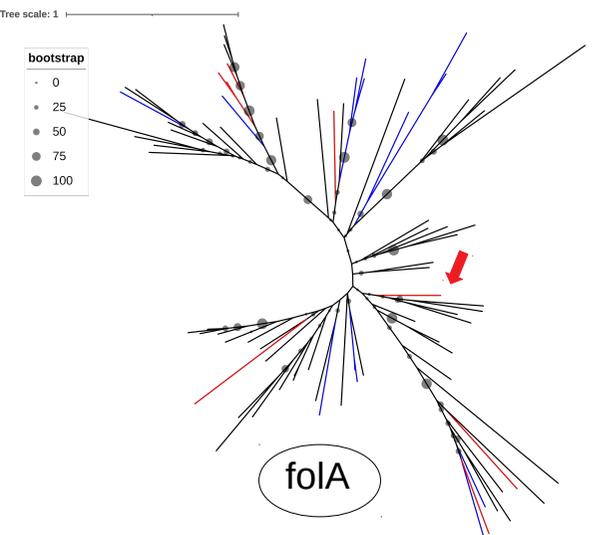
thyA

Asgard + other Archaea

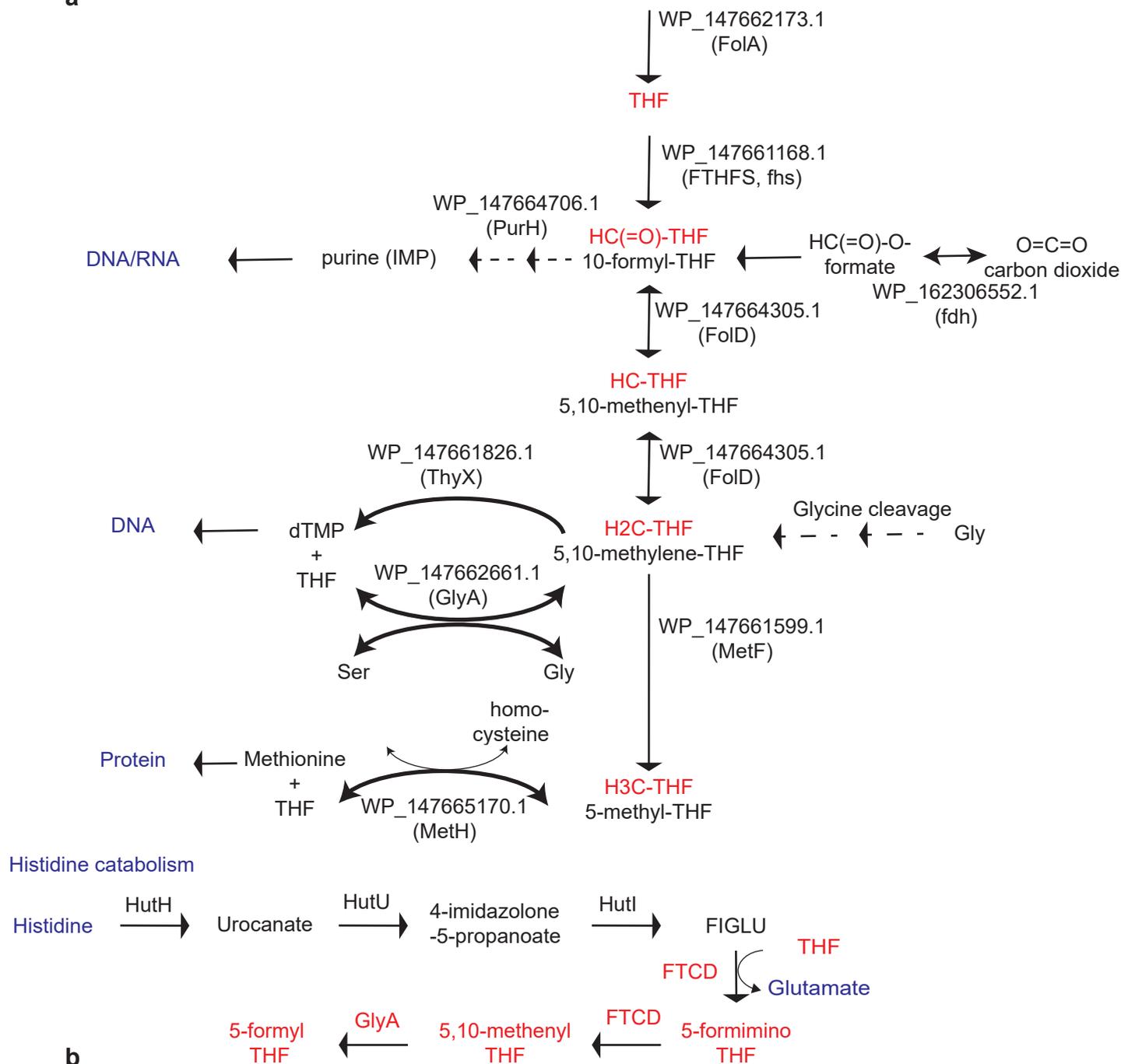








a



b

