

A comparative study demonstrates vertical inheritance and
maintenance of Arsenic Resistance Metabolism in
eukaryotes

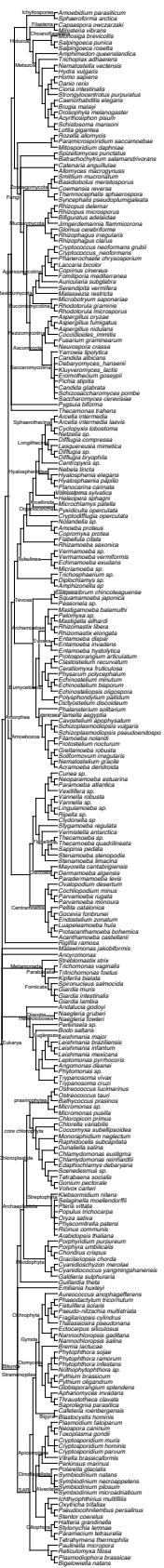
Running title : Arsenic Resistance Metabolism in eukaryotes

Ribeiro Giulia.M.¹ and Lahr Daniel.J.G.^{1,a}

¹Department of Zoology, Institute of Biosciences, University of São Paulo, Rua do Matão, Tv. 14, n101, São Paulo, SP - 05508-090, Brazil

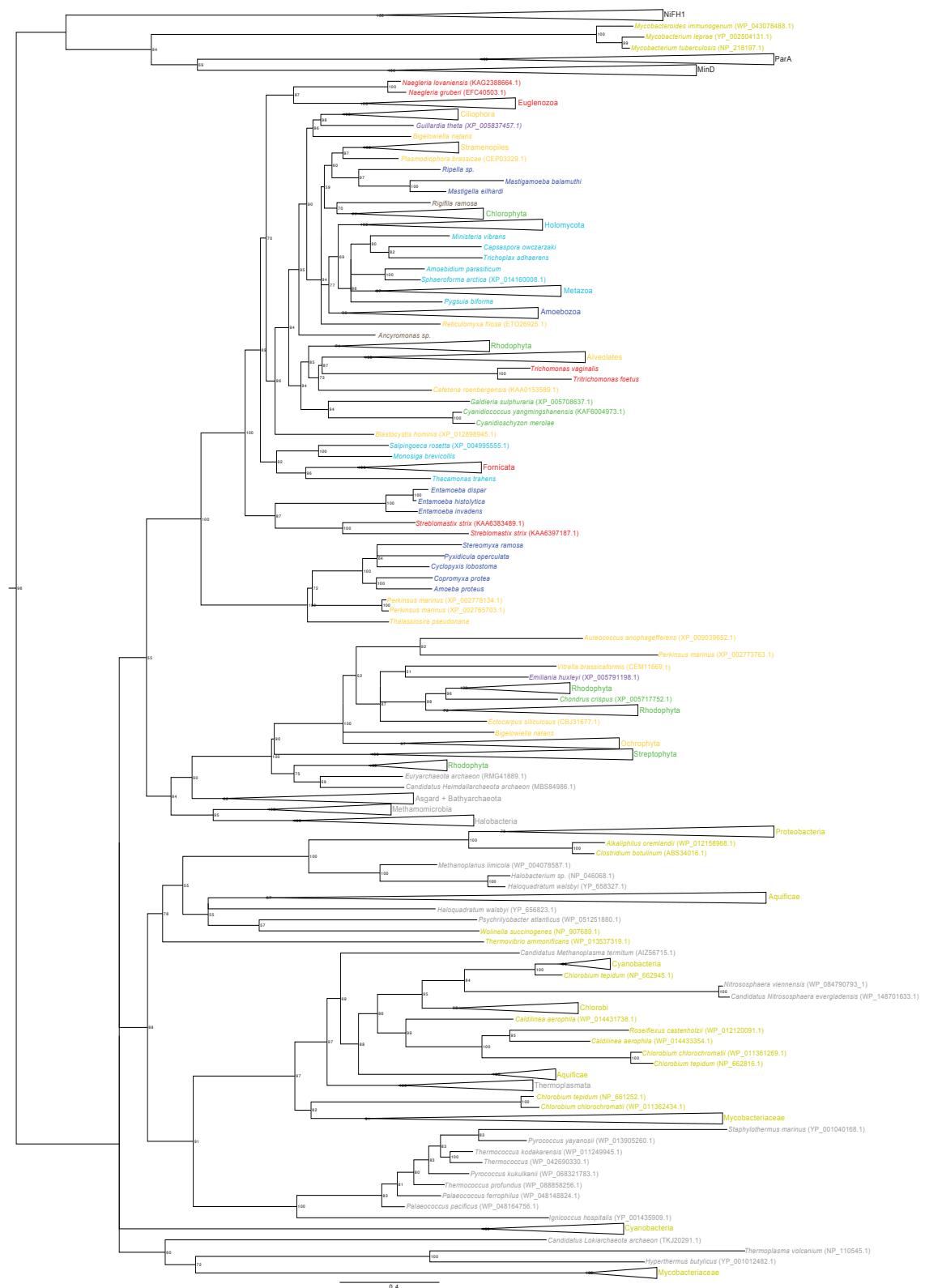
a-Corresponding author: Daniel J.G. Lahr, Rua do Matão, Travessa 14, 101, Cidade, São Paulo, SP, Brazil, 05588-090, Telephone number: +55 11 3091-0948, e-mail: dlahr@ib.usp.br

June 17, 2021



Supplementary Figure S1: Reference eukaryotic species tree (continue next page)

Other notes about Supplementary Figure S1: Reference eukaryotic species tree. The species tree is a compilation of the phylogenetic literature of major groups of eukaryotes that determined the relationships between the taxa.



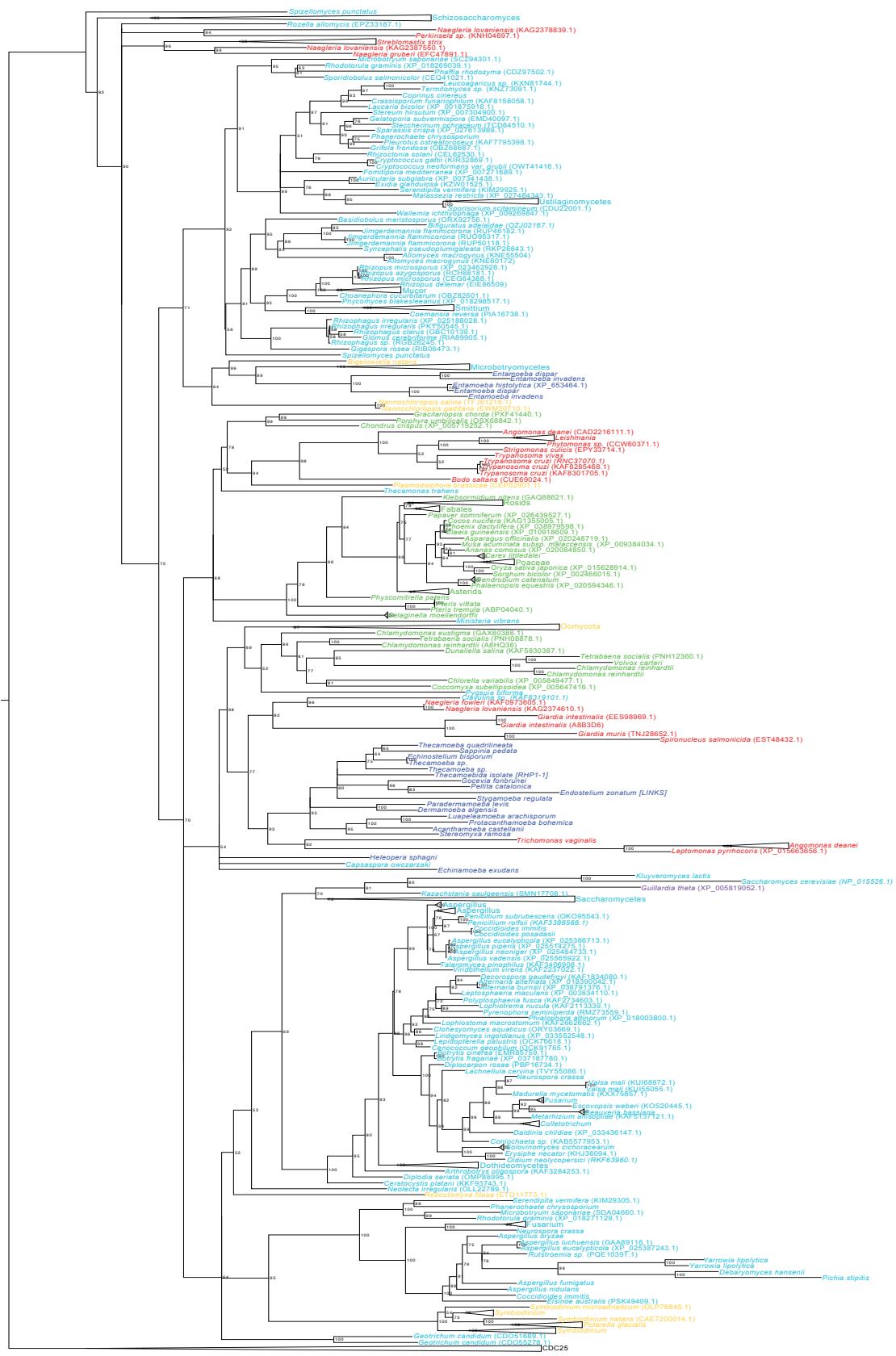
Supplementary Figure S2: Phylogenetic reconstruction of Arsenical pump ATPase ArsA/GET3 (ARSA) and external groups of the P-loop NTPase superfamily (IPR027417). (continue next page)

Other notes about Supplementary Figure S2: Phylogenetic reconstruction of Arsenical pump ATPase ArsA/GET3 (ARSA) and external groups of the P-loop NTPase superfamily (IPR027417). The alignment has 501 taxa and 270 amino acids. The tree used the substitution model LG + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes.



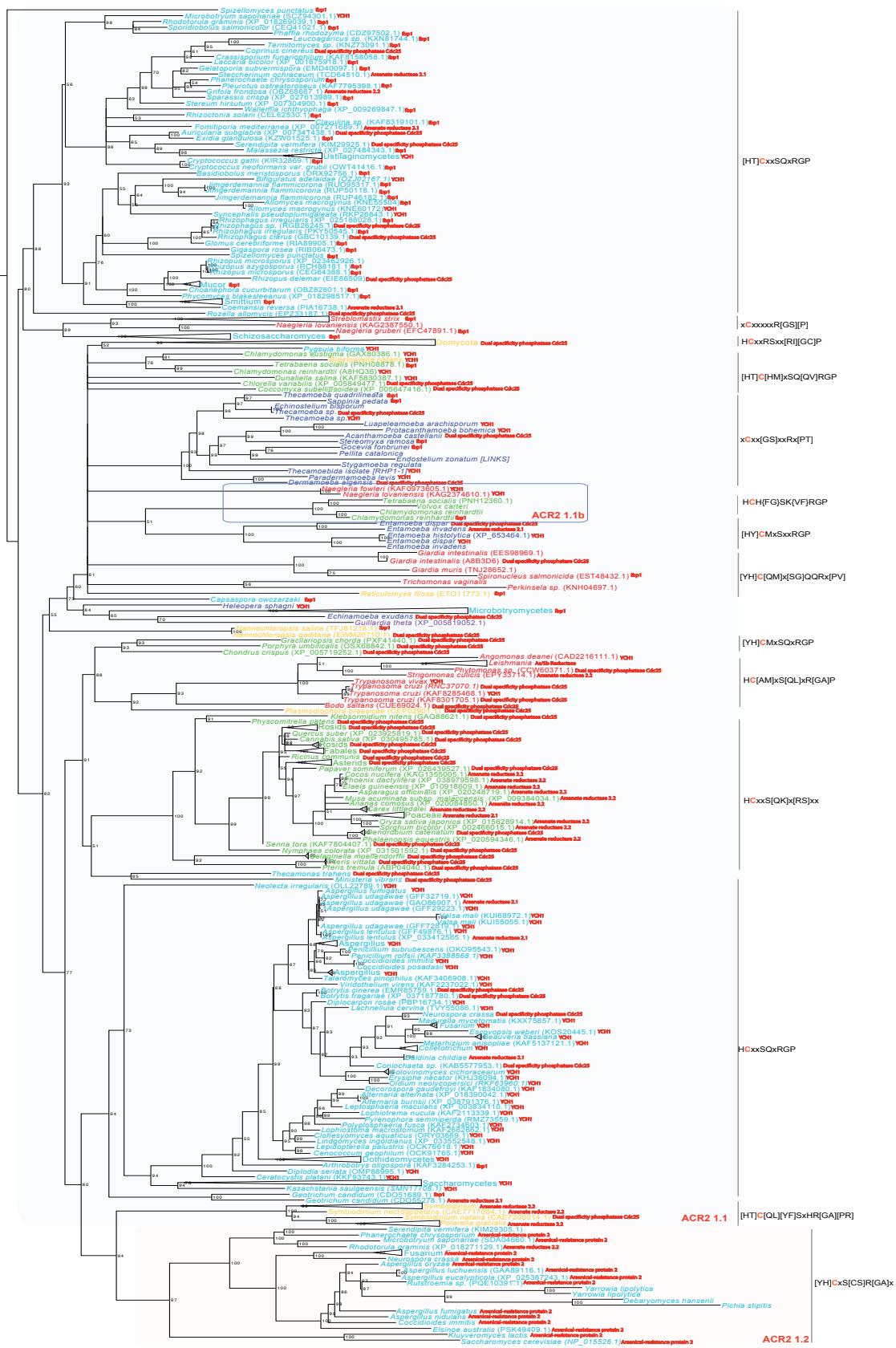
Supplementary Figure S3: Phylogenetic reconstruction of Arsenical pump ATPase ArsA/GET3 (ARSA). (continue next page)

Other notes about Supplementary Figure S3: Phylogenetic reconstruction of Arsenical pump ATPase ArsA/GET3 (ARSA). The alignment has 472 taxa and 294 amino acids. The tree used the substitution model LG + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes. Protein domain organization is indicated at right. Groups are shaded and sub-groups are indicated in a box. In addition, the names of the groups and sub-groups are shown at the right of the corresponding clades.



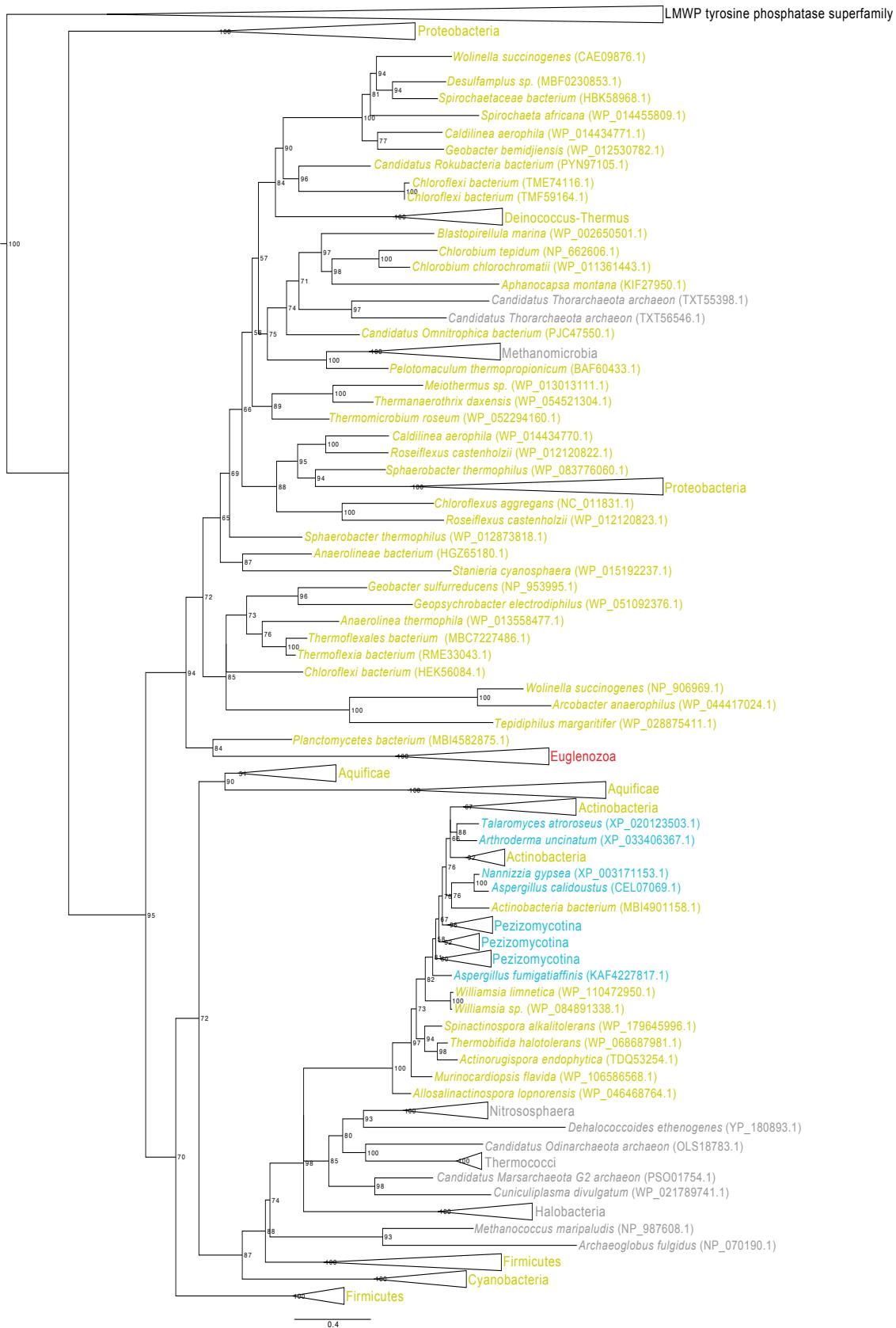
Supplementary Figure S4: Phylogenetic reconstruction of Arsenate reductase 2 (ACR2).
(continue next page)

Other notes about Supplementary Figure S4: Phylogenetic reconstruction of Arsenate reductase 2 (ACR2) and external group Cell cycle control phosphatase (CDC25). The alignment has 340 taxa and 124 amino acids. The tree used the substitution model LG + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes.



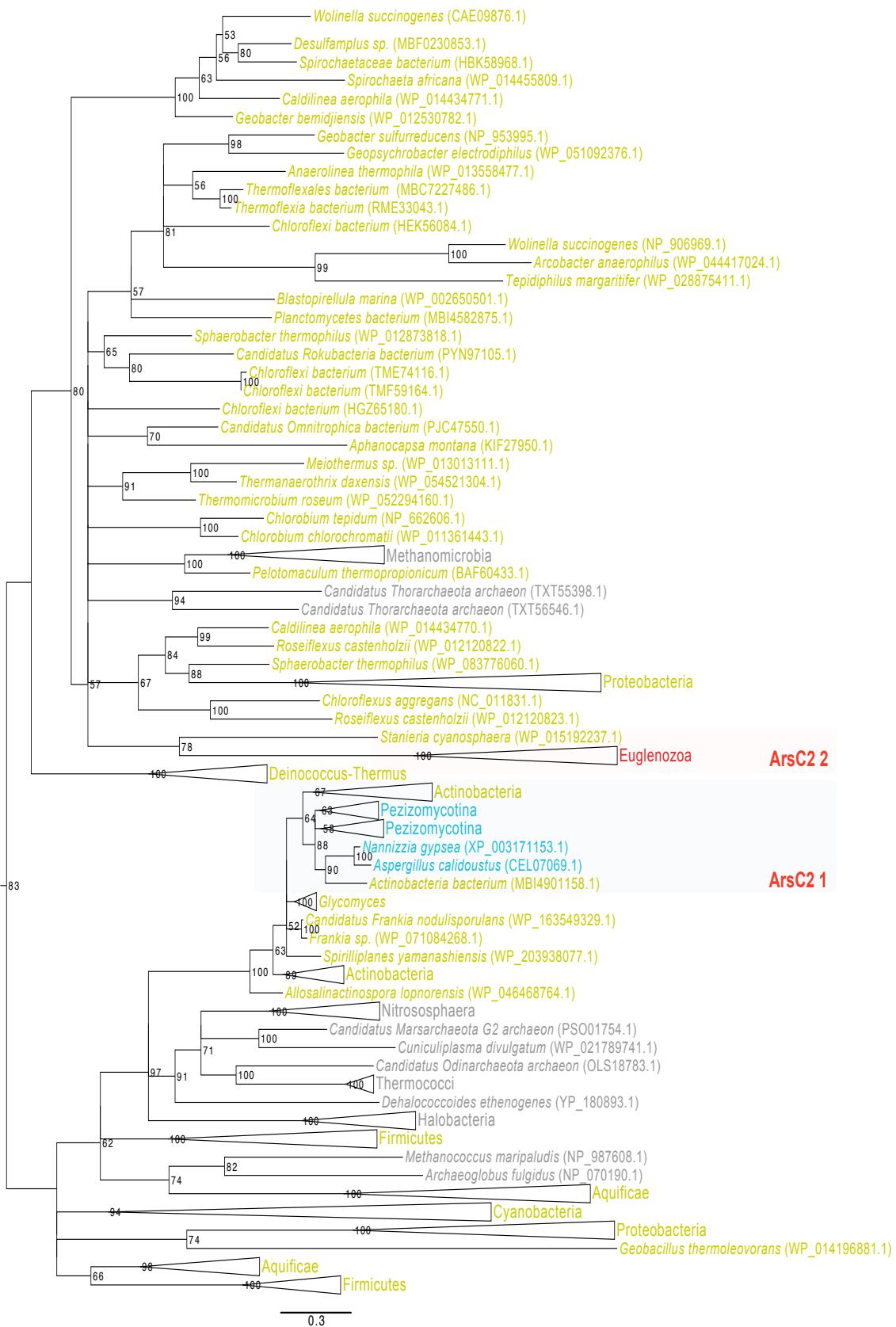
Supplementary Figure S5: Phylogenetic reconstruction of Arsenate reductase 2 (ACR2).
 (continue next page)

Other notes about Supplementary Figure S5: Phylogenetic reconstruction of Arsenate reductase 2 (ACR2) and external group Cell cycle control phosphatase (CDC25). The alignment has 340 taxa and 124 amino acids. The tree used the substitution model LG + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes. Sequence annotation is next to terminal label. Conserved domain is indicated at left, and functional motifs of the main groups of the tree are at right. Groups are shaded and sub-groups are indicated in a box. In addition, the names of the groups and sub-groups are shown at the right of the corresponding clades.



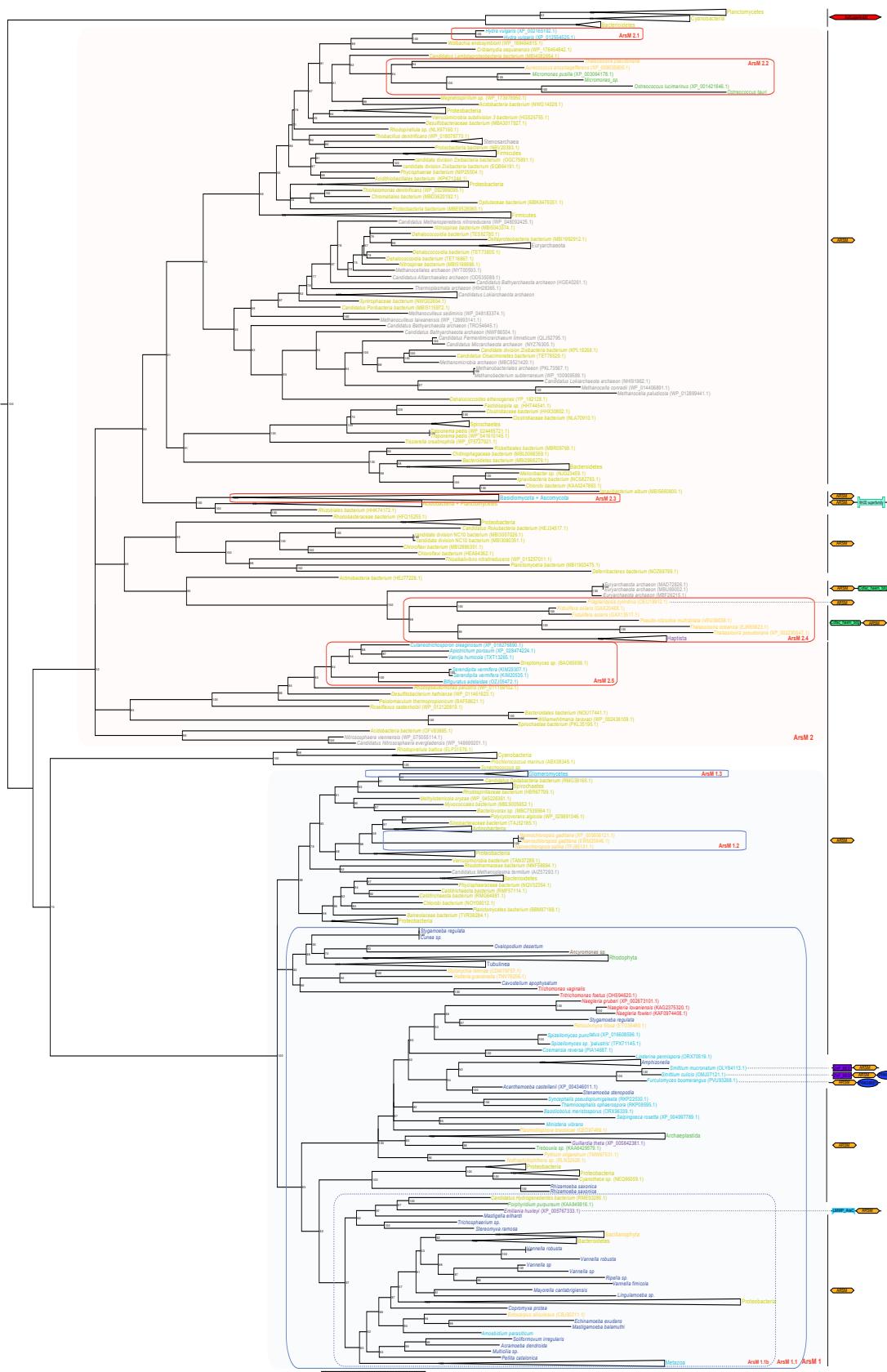
Supplementary Figure S6: Phylogenetic reconstruction of Arsenate reductase, glutathione/glutaredoxin type (ARSC2). (continue next page)

Other notes about Supplementary Figure S6: Phylogenetic reconstruction of Arsenate reductase, glutathione/glutaredoxin type (ARSC2) and external group Low molecular weight protein tyrosine phosphatase (IPR017867). The alignment has 253 taxa and 130 amino acids. The tree used the substitution model LG + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes.



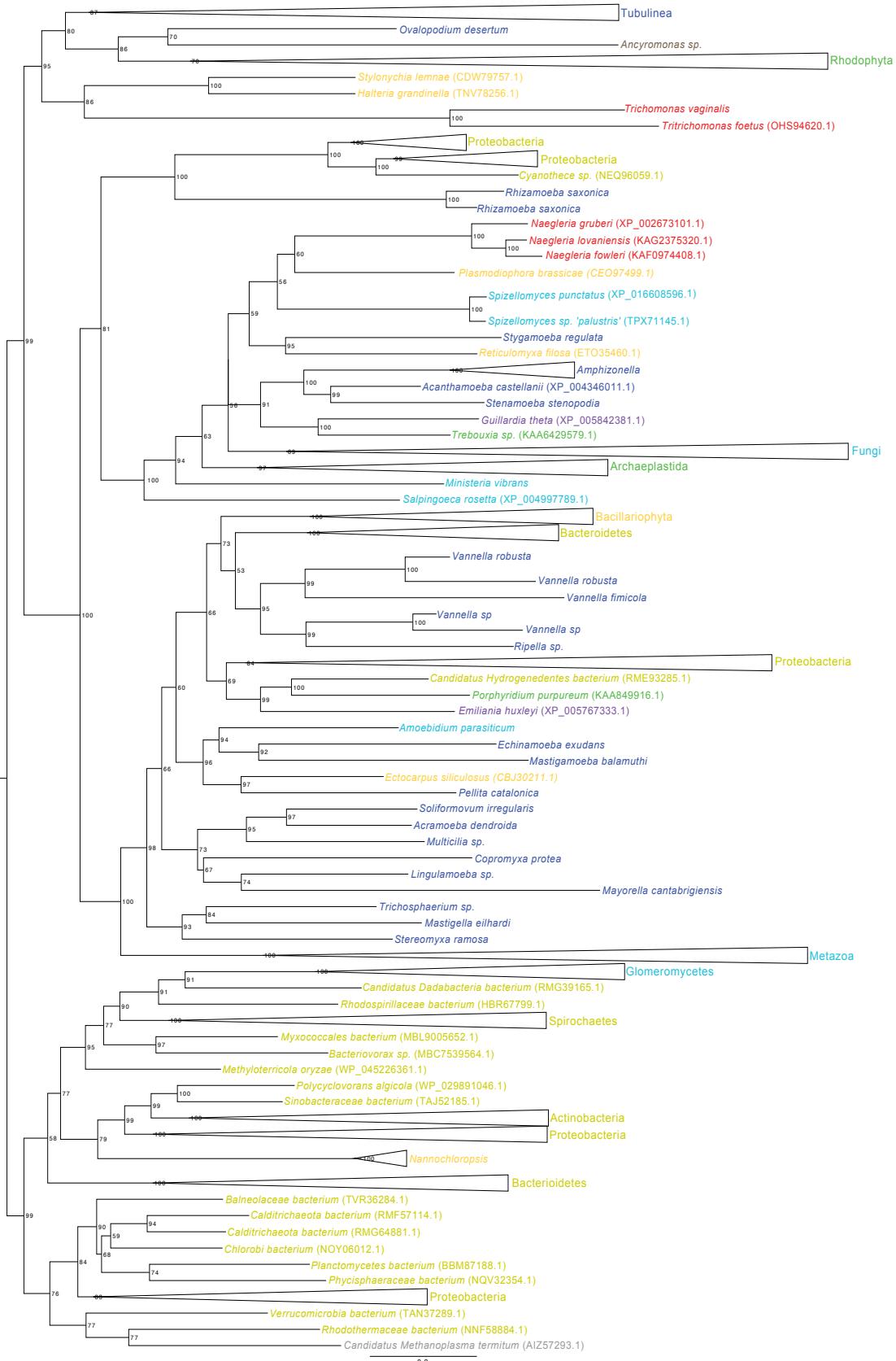
Supplementary Figure S7: Phylogenetic reconstruction of Arsenate reductase, glutathione/glutaredoxin type (ARSC2). (continue next page)

Other notes about Supplementary Figure S7: Phylogenetic reconstruction of Arsenate reductase, glutathione/glutaredoxin type (ARSC2). The alignment has 239 taxa and 129 amino acids. The tree used the substitution model LG + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes. Clusters of eukaryotic sequences are shaded. In addition, the names of the groups of eukaryotes are shown at the right of the corresponding clades.



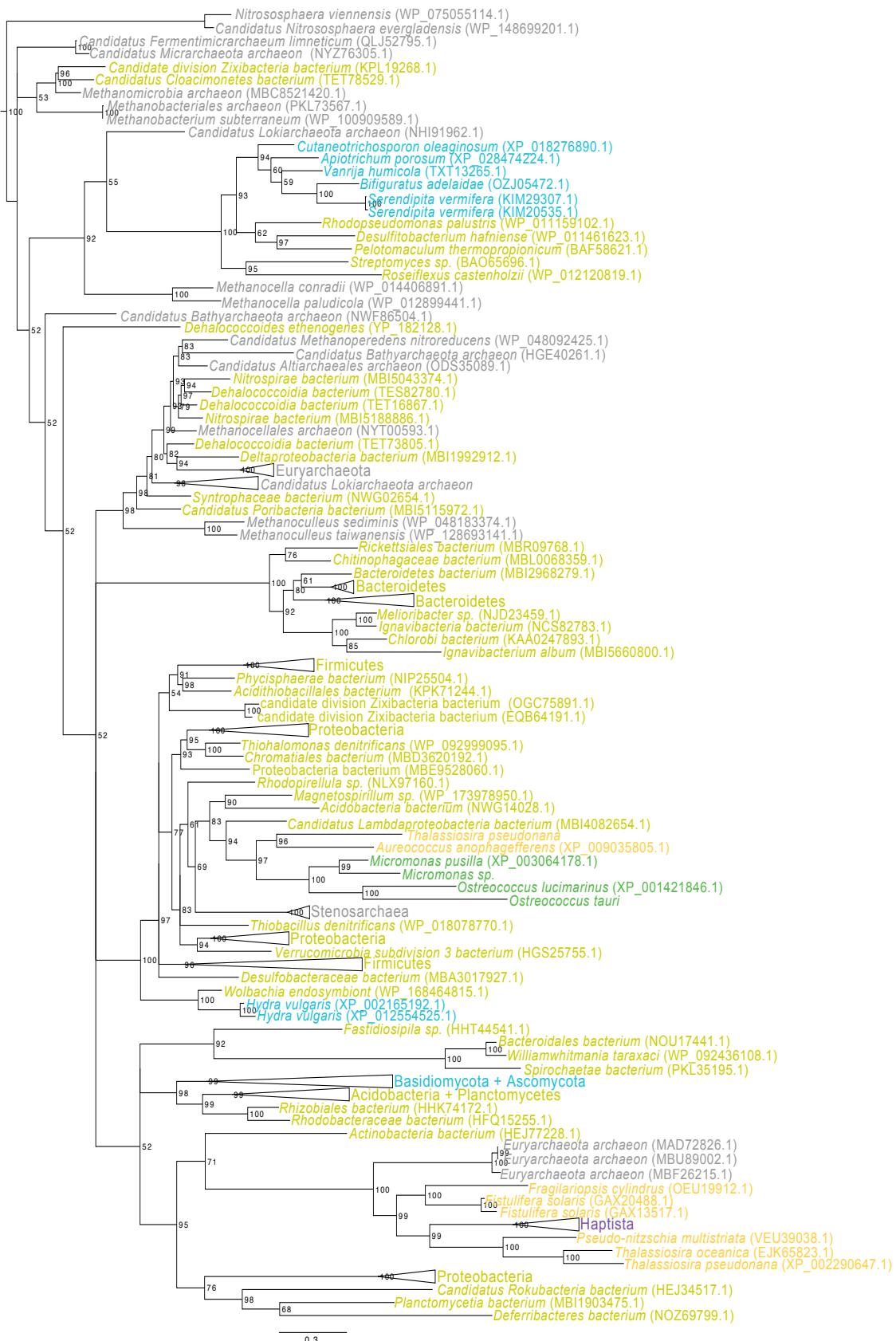
Supplementary Figure S8: Phylogenetic reconstruction of Arsenite methyltransferase (ARSM). (continue next page)

Other notes about Supplementary Figure S8: Phylogenetic reconstruction of Arsenite methyl-transferase (ARSM). The alignment has 632 taxa and 229 amino acids. The tree used the substitution model LG + I + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes. Protein domain organization is indicated at right. Groups that contains eukaryotic sequences are shaded, sub-groups are indicated in a box. In addition, the names of the groups and sub-groups of eukaryotes are shown at the right of the corresponding clades.



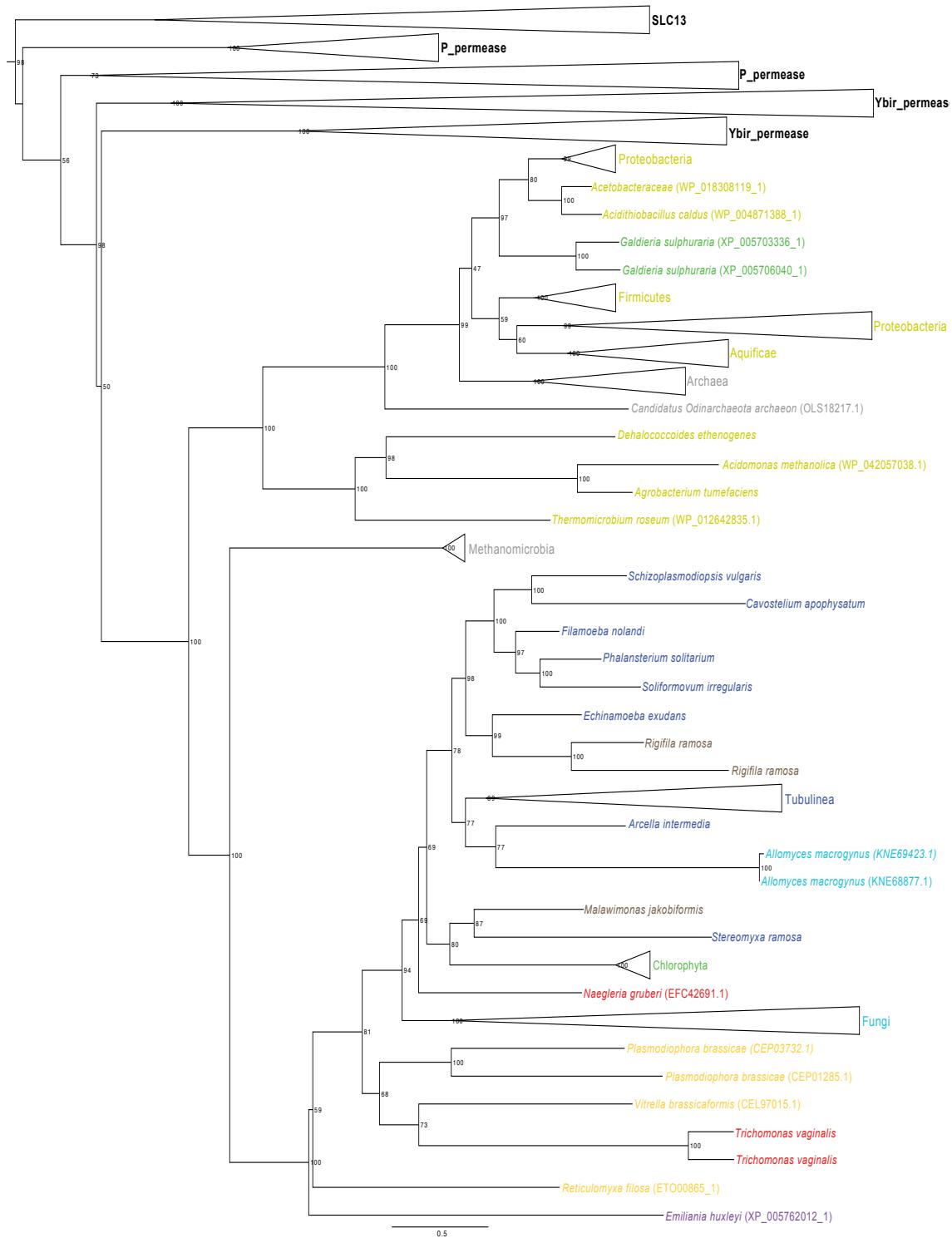
Supplementary Figure S9: Phylogenetic reconstruction of Arsenite methyltransferase (ARSM). (continue next page)

Other notes about Supplementary Figure S9: Phylogenetic reconstruction of Arsenite methyltransferase Group 1 (ARSM 1). The alignment has 345 taxa and 343 amino acids. The tree used the substitution model LG + I + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes.



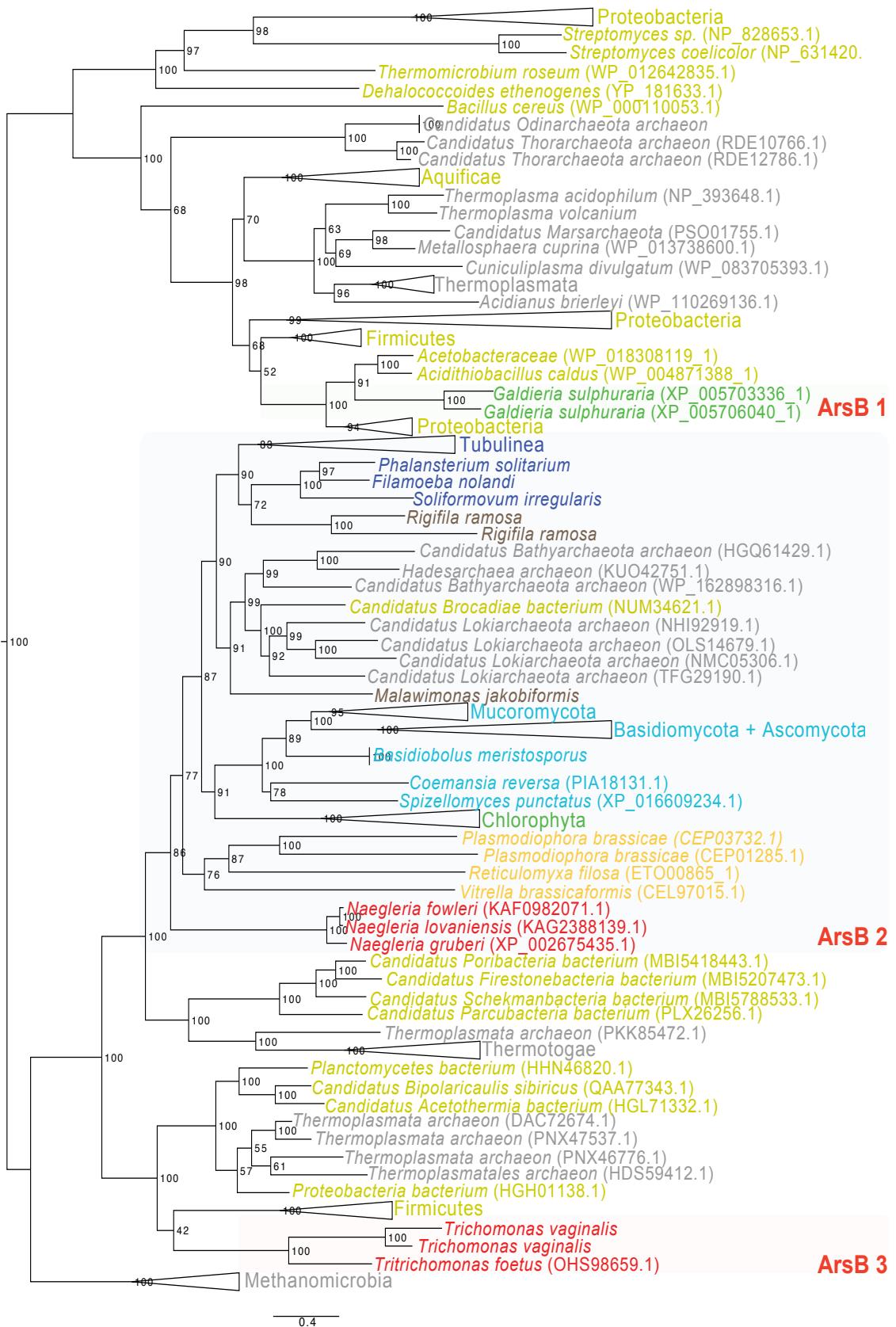
Supplementary Figure S10: Phylogenetic reconstruction of Arsenite methyltransferase (ARSM). (continue next page)

Other notes about Supplementary Figure S10: Phylogenetic reconstruction of Arsenite methyltransferase Group 2 (ARSM 2). The alignment has 234 taxa and 245 amino acids. The tree used the substitution model LG + I + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes.



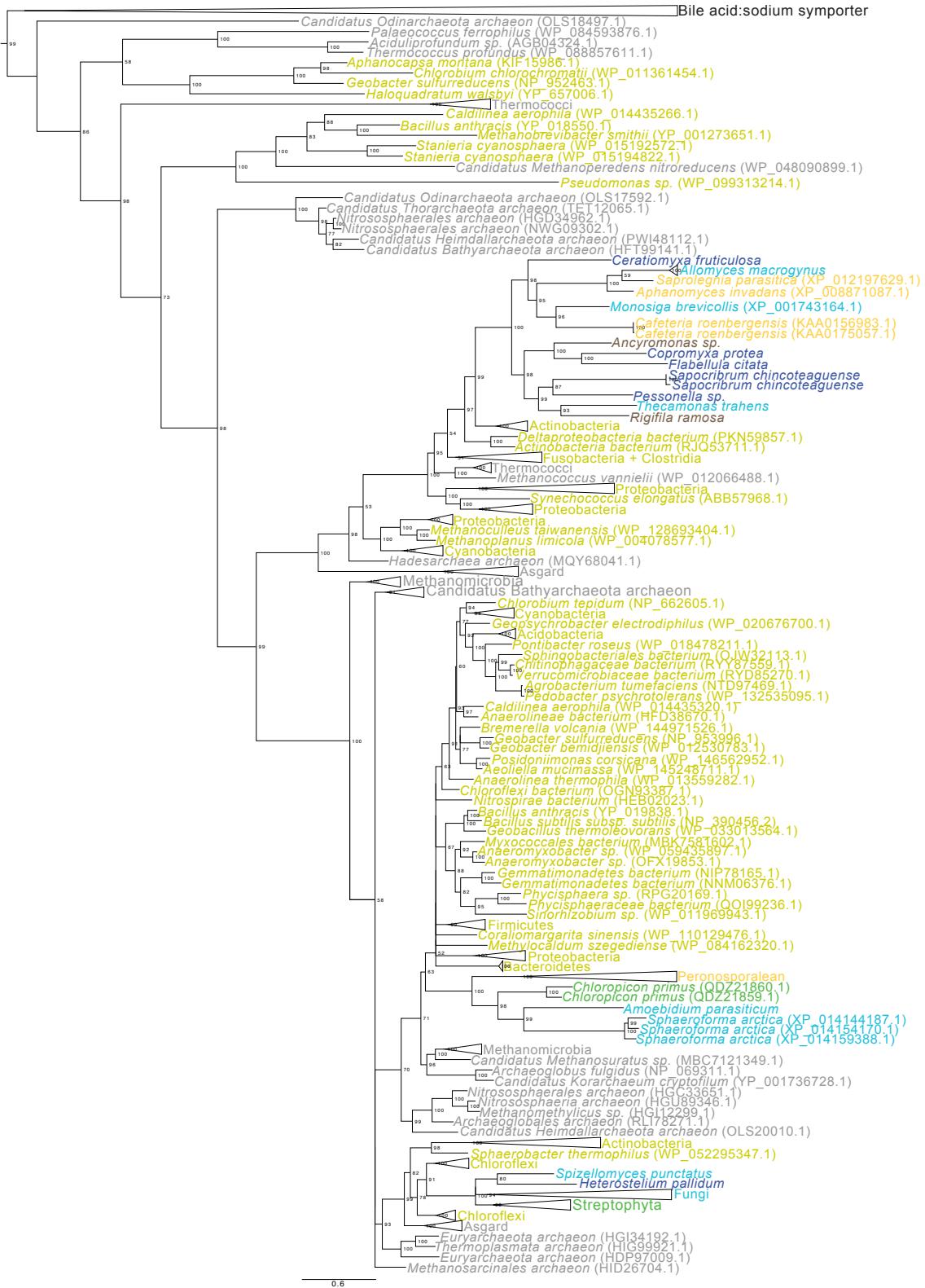
Supplementary Figure S11: Phylogenetic reconstruction of Arsenical pump membrane protein (ARSB) with external groups of ArsB/NhaD superfamily. (continue next page)

Other notes about Supplementary Figure S11: Preliminary phylogenetic reconstruction of the Arsenical pump membrane protein (ARSB) with external groups. The alignment has 167 taxa and 385 amino acids. The tree used the substitution model LG + F + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes. External protein families are collapsed and colored in black.



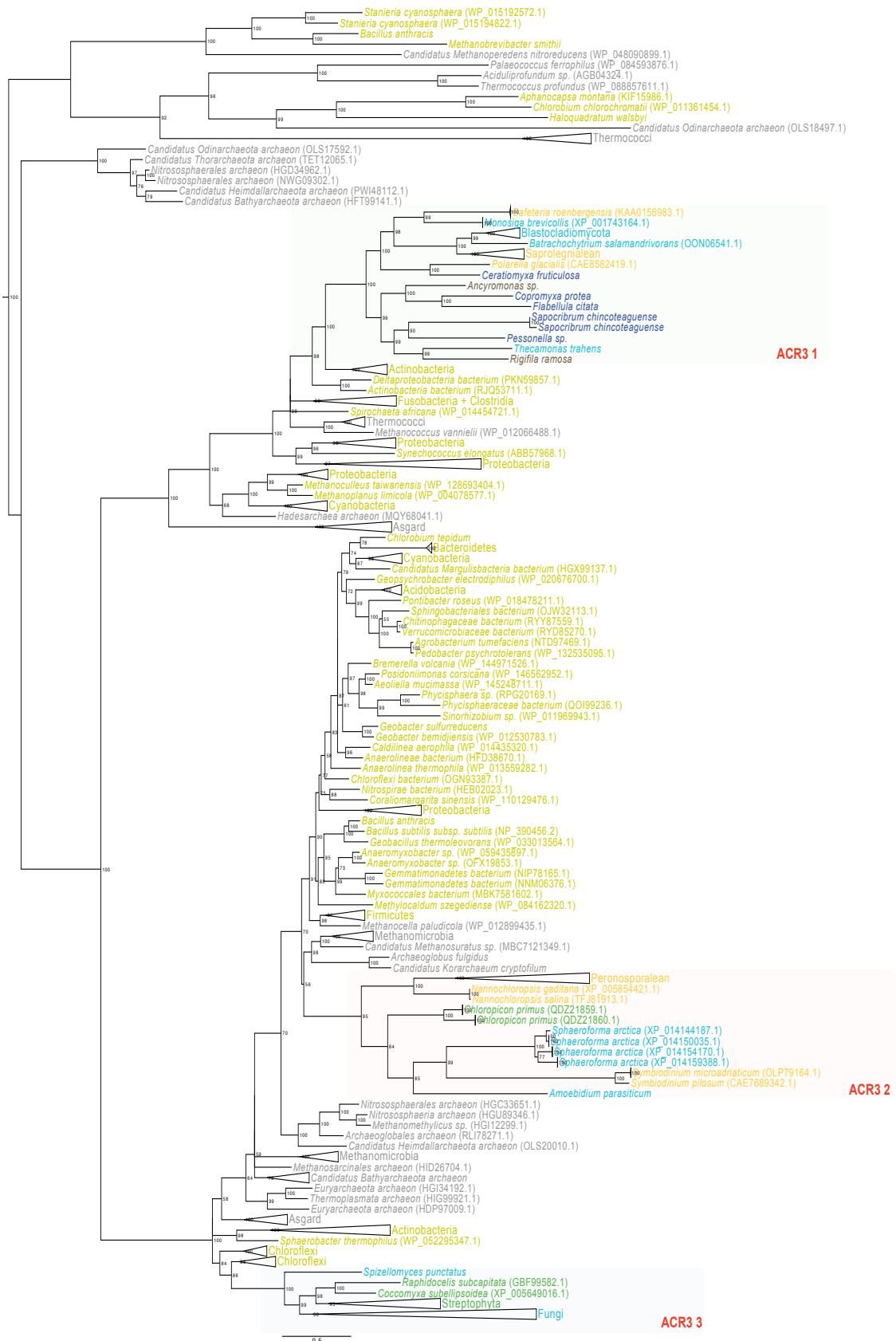
Supplementary Figure S12: Phylogenetic reconstruction of Arsenical pump membrane protein (ARSB). (continue next page)

Other notes about Supplementary Figure S12: Phylogenetic reconstruction of the Arsenical pump membrane protein (ARSB). The alignment has 180 taxa and 419 amino acids. The tree used the substitution model LG + F + I + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes. Groups that contains eukaryotic sequences are shaded. In addition, the names of the groups are shown at the right of the corresponding clades.



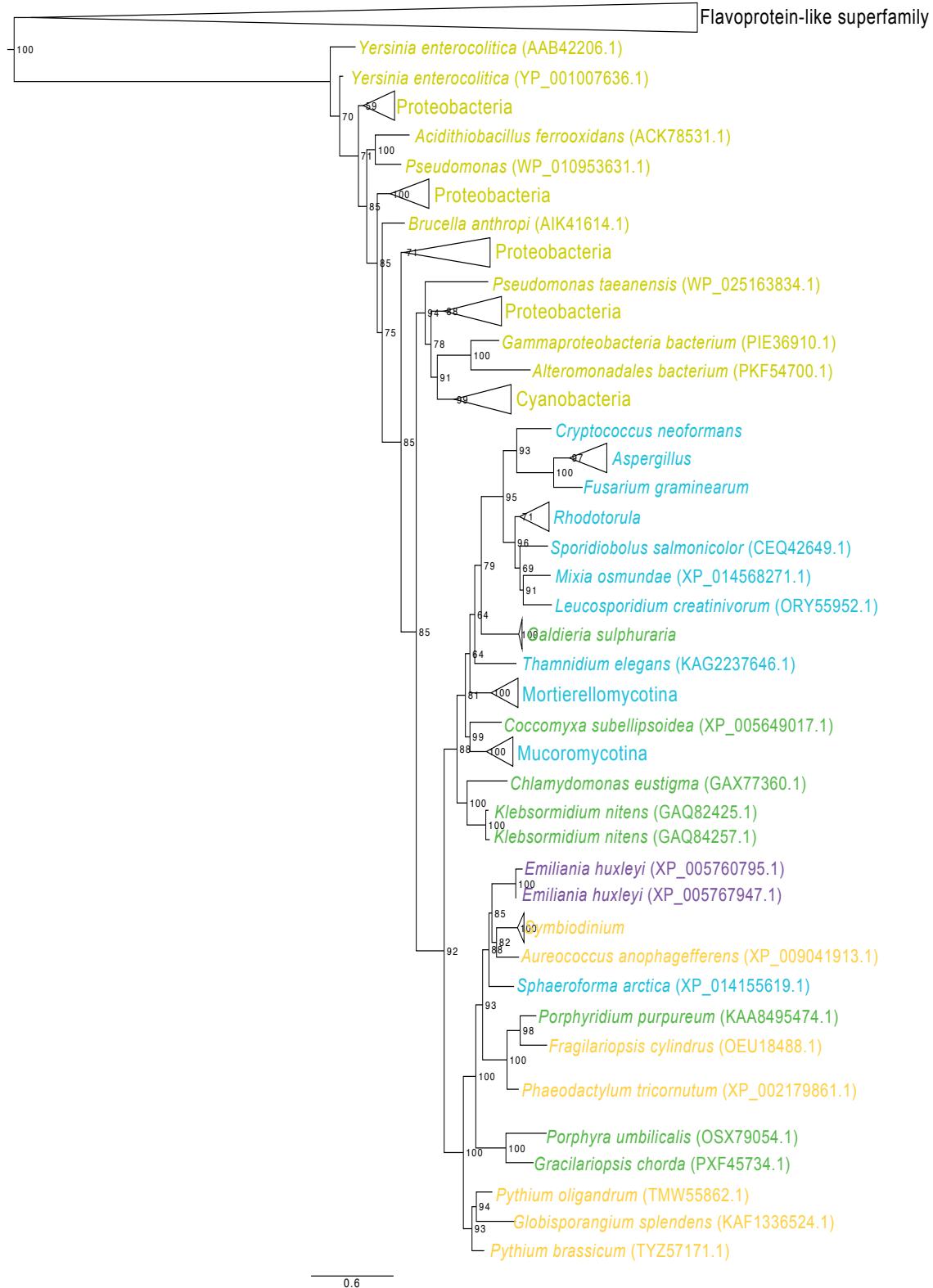
Supplementary Figure S13: Phylogenetic reconstruction of Arsenical resistance protein (ACR3) with external groups Bile acid:sodium symporter family. (continue next page)

Other notes about Supplementary Figure S13: Phylogenetic reconstruction of Arsenical resistance protein (ACR3) with external groups Bile acid:sodium symporter family. The alignment has 277 taxa and 325 amino acids. The tree used the substitution model LG + G4. Branch-length is indicated at the left edge. Numbers at the clades represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes.



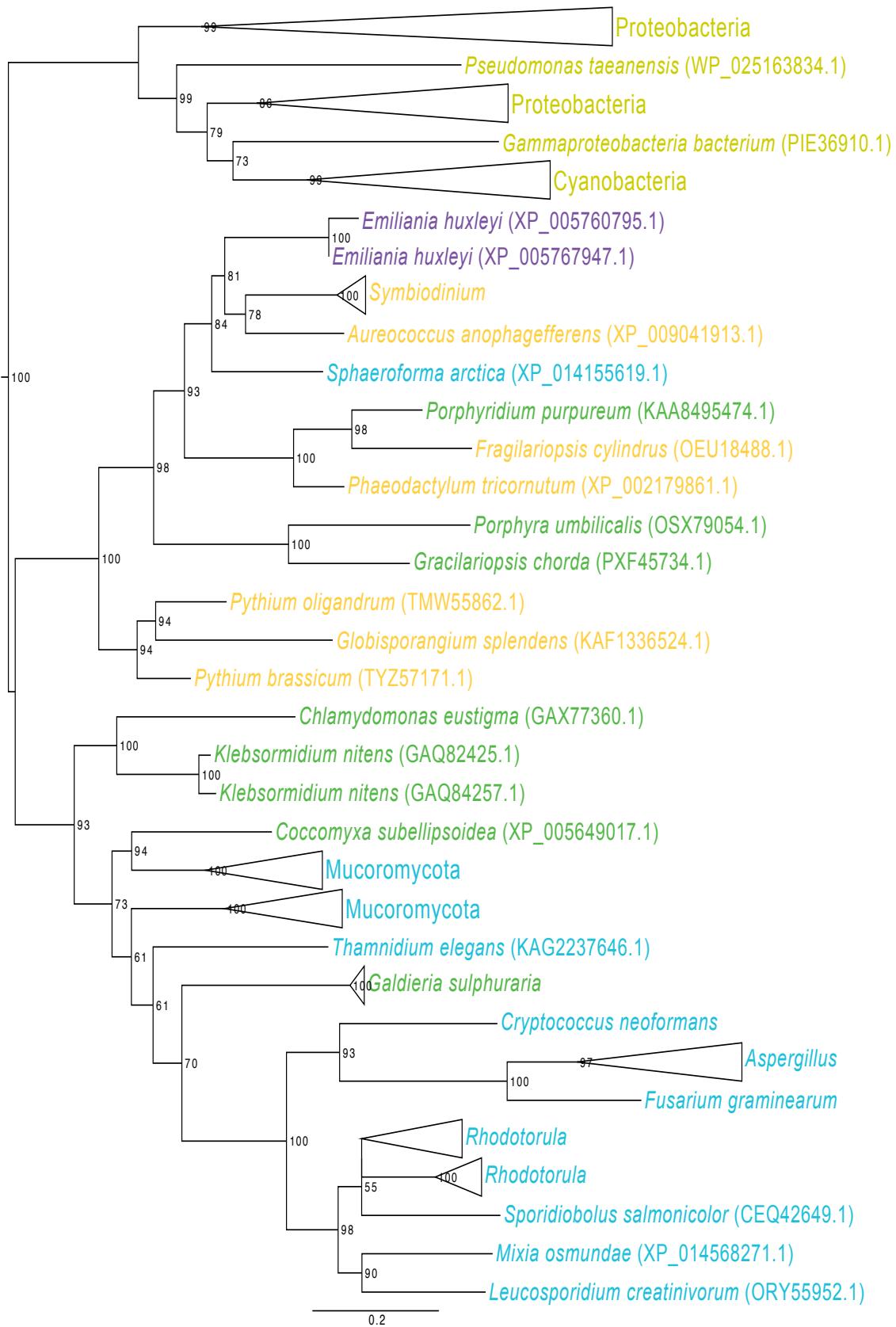
Supplementary Figure S14: Phylogenetic reconstruction of Arsenical resistance protein (ACR3). (continue next page)

Other notes about Supplementary Figure S14: Phylogenetic reconstruction of Arsenical resistance protein (ACR3). The alignment has 302 taxa and 337 amino acids. The tree used the substitution model LG + G4. Branch-length is indicated at the left edge. Numbers at the clades represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes. Clusters of eukaryotic sequences are shaded. In addition, the names of the groups of eukaryotes are shown at the right of the corresponding clades.



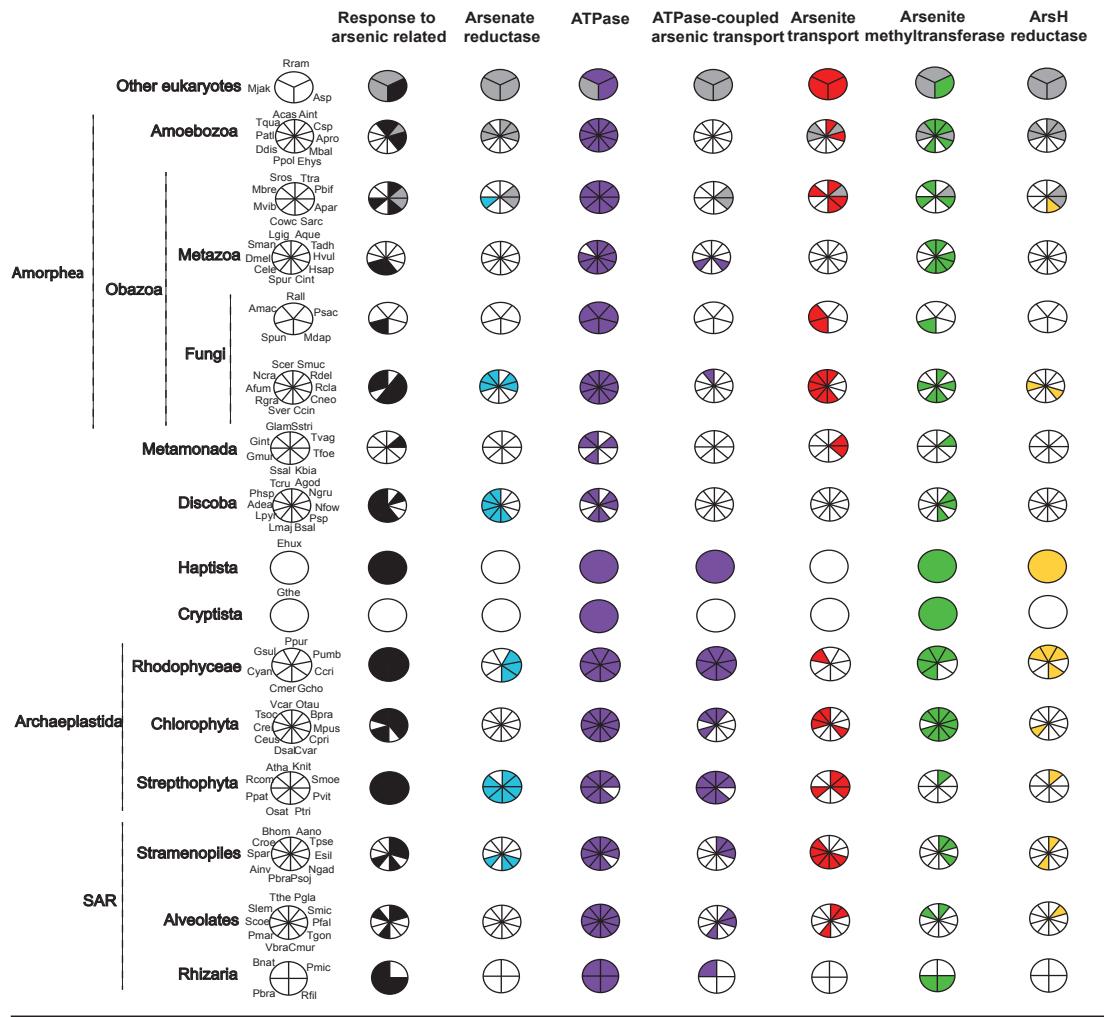
Supplementary Figure S15: Phylogenetic reconstruction of Arsenate resistance ArsH (ARSH) with external group of Flavoprotein-like superfamily (IPR029039). (continue next page)

Other notes about Supplementary Figure S15: Phylogenetic reconstruction of Arsenate resistance ArsH (ARSH) with external group of Flavoprotein-like superfamily (IPR029039). The alignment has 184 taxa and 216 amino acids. The tree used the substitution model LG + I + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes.



Supplementary Figure S16: Phylogenetic reconstruction of Arsenate resistance ArsH (ARSH). (continue next page)

Other notes about Supplementary Figure S16: Phylogenetic reconstruction of Arsenate resistance ArsH (ARSH). The alignment has 169 taxa and 215 amino acids. The tree used the substitution model LG + I + G4. Branch-length is indicated at the left edge. Numbers at the clades nodes represent percent of ultra-fast bootstrap support (UFBoot) calculated with 1000 runs. Only clades with a minimum 50% UFBoot are shown. The colors of terminal labels indicate the domain or great group of the sequence: Dark yellow - Bacteria, Gray - Archaea, Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes.



Other eukaryotes: Rram - <i>Rigifila ramosa</i> ; Mjak - <i>Malawimonas jakobiformis</i> ; Asp - <i>Ancyromonas sp.</i> ;	Dmel - <i>Drosophila melanogaster</i> ; Sman - <i>Schistosoma mansoni</i> ; Lig - <i>Lottia gigantea</i> ;	Psp - <i>Perkinsela sp.</i> ; Bsal - <i>Bodo saltans</i> ;	Smoe - <i>Selaginella moellendorffii</i> ;
Amoebozoa: Aint - <i>Arcella intermedia</i> ; Csp - <i>Centropyxis sp.</i> ; Apro - <i>Amoeba proteus</i> ; Mbal - <i>Mastigamoeba balamuthi</i> ; Ehys - <i>Entamoeba histolytica</i> ; Ppol - <i>Physarum polycephalum</i> ; Ddis - <i>Dictyostelium discoideum</i> ; Patl - <i>Paramoeba atlantica</i> ; Tqua - <i>Thecamoeba quadrilineata</i> ; Acas - <i>Acanthamoeba castellanii</i> ;	Rall - <i>Zoellera allomyces</i> ; Psac - <i>Paramicroporidium saccamoebae</i> ; Mdap - <i>Mitosporidium daphniae</i> ; Spun - <i>Spizellomyces punctatus</i> ; Amae - <i>Allomyces macrogyrus</i> ; Smuc - <i>Smittium mucronatum</i> ; Rdel - <i>Rhizopus delemar</i> ; Rcla - <i>Rhizophagus clarus</i> ; Ccne - <i>Cryptococcus neoformans</i> ; Ccni - <i>Coprinopsis cinerea</i> ; Sver - <i>Serendipita vermifera</i> ; Rgra - <i>Rhodotorula graminis</i> ; Afum - <i>Aspergillus fumigatus</i> ; Ncra - <i>Neurospora crassa</i> ; Scer - <i>Saccharomyces cerevisiae</i> ;	Psal - <i>Ostreococcus tauri</i> ; Bbra - <i>Bathycoccus prasinus</i> ; Mpri - <i>Micromonas pusilla</i> ; Cpri - <i>Chloropicon primus</i> ; Cvar - <i>Chlorella variabilis</i> ;	Ptri - <i>Populus trichocarpa</i> ;
Metazoa: Aque - <i>Amphimedon queenslandica</i> ; Tadh - <i>Trichoplax adhaerens</i> ; Hvul - <i>Hydra vulgaris</i> ; Hsap - <i>Homosapiens</i> ;	Sros - <i>Salpingoeca rosetta</i> ; Lig - <i>Amphimedon quadrivalvis</i> ; Mtbl - <i>Ministeria vibrans</i> ; Mbre - <i>Monosiga brevicollis</i> ; Sros - <i>Salpingoeca rosetta</i> ;	Dsal - <i>Dunaliella salina</i> ;	Osat - <i>Oryza sativa</i> ;
Metamonada: Aque - <i>Amphimedon queenslandica</i> ; Tadh - <i>Trichoplax adhaerens</i> ; Hvul - <i>Hydra vulgaris</i> ; Hsap - <i>Homosapiens</i> ;	Gint - <i>Giardia intestinalis</i> ; Glam - <i>Giardia lamblia</i> ;	Crei - <i>Chlamydomonas reinhardtii</i> ;	Ppat - <i>Phycomitrella patens</i> ;
Discoba: Ago - <i>Amoebeidium parasiticum</i> ;	Gint - <i>Giardia intestinalis</i> ;	Tsoc - <i>Tetrahymena socialis</i> ;	Rcom - <i>Ricinus communis</i> ;
Haptista: Gint - <i>Giardia intestinalis</i> ;	Gint - <i>Giardia intestinalis</i> ;	Vcar - <i>Volvox carteri</i> ;	Atha - <i>Arabidopsis thaliana</i> ;
Cryptista:	Gint - <i>Giardia intestinalis</i> ;	Streptophyta:	Stramenopiles:
Rhizarians:	Gint - <i>Giardia intestinalis</i> ;	Knit - <i>Perkinsus marinus</i> ;	Aano - <i>Aureococcus anophagefferens</i> ;
Rhodophyceae:	Gint - <i>Giardia intestinalis</i> ;	Dsal - <i>Dunaliella salina</i> ;	Tpse - <i>Thalassiosira pseudonana</i> ;
Chlorophyta:	Gint - <i>Giardia intestinalis</i> ;	Ceus - <i>Chlamydomonas eustigma</i> ;	Esl - <i>Ectocarpus siliculosus</i> ;
Streptophyta:	Gint - <i>Giardia intestinalis</i> ;	Crei - <i>Chlamydomonas reinhardtii</i> ;	Ngad - <i>Nannochloropsis gaditana</i> ;
Stramenopiles:	Gint - <i>Giardia intestinalis</i> ;	Vcar - <i>Volvox carteri</i> ;	Psoj - <i>Phytophthora sojae</i> ;
Alveolates:	Gint - <i>Giardia intestinalis</i> ;	Tsoc - <i>Tetrahymena socialis</i> ;	Pbra - <i>Pythium brassicace</i> ;
Rhizaria:	Gint - <i>Giardia intestinalis</i> ;	Vcar - <i>Volvox carteri</i> ;	Ainv - <i>Aphanomyces invadans</i> ;

Supplementary Figure S17: Distribution of the main arsenic resistance functions in major eukaryotic taxa. (continue next page)

Other notes about Supplementary Figure S17: The figure illustrates the presence and absence of the ontology functions related to arsenic metabolism for a subsample of 115 eukaryotic species. Ontologies are grouped by function type: Response to arsenic (GO:0046685, GO:0071722, GO:0010038); Arsenate reductase (GO:0008794, GO:0030611, GO:0030612); ATPase (GO:0016887); ATPase-coupled arsenic transport (GO:0015446, GO:0046685, GO:0071722, GO:0015700, GO:0046872, GO:0010038); Arsenite transport (GO:0015105, GO:0015103); Arsenite methyltransferase (GO:0030791, GO:0030792); ArsH reductase (GO:0010038, GO:0008753). Function type is indicated at the top. In addition, species are grouped by classification in eukaryotic great lineages. Unfilled and gray sectors represent absence in the genomes or transcriptomes, respectively. Colors represent eukaryotic great lineages (Light-blue - Obazoa, Dark-blue - Amoebozoa, Red - Metamonada and Discoba, Green - Archaeplastida, Light-yellow - SAR, Purple - Haptista and Cryptista, Brown - Other eukaryotes).

Supplementary Table S1: Results of approximately unbiased (AU) tests of alternative topologies for ACR2, ACR3, ARSB and ArsM gene trees. (continue next page)

Table S1: Results of approximately unbiased (AU) tests of alternative topologies for ACR2, ACR3, ARSB and ArsM gene trees. AU Topology Tests for a combination of topology constraints related to possible origin of arsenic resistance metabolism genes. P-AU represents the p-value of the topologies tested by the approximately unbiased (AU) criteria. Accepted trees are indicated in bold (> 0.05).

	Hypothesis tested	logL	p-AU
ACR2	Unconstrained tree	-43842.82091	0.786
	Fungi monophyletic	-43864.41782	0.489
	Metamonada monophyletic	-43857.63337	0.455
	Discoba monophyletic	-43897.62565	0.158
	Amoebozoa monophyletic	-43890.38004	0.172
	Archaeplastida monophyletic	-43870.00233	0.387
	SAR monophyletic	-43884.91836	0.193
ARSB	Fungi + Metamonada + Discoba + SAR + Archaeplastida + Amoebozoa monophyletic	-43980.45213	2.22E-019
	Unconstrained tree	-7,877,770,938	0.926
	Eukaryotic monophyletic	-7,927,148,607	0.000107
	ARSB 2 monophyletic	-7,879,461,563	0.076
	ARSB 1 + ARSB 2 monophyletic	-791,950,358	9.09e-06
	ARSB 1 + ARSB 3 monophyletic	-7,943,097,566	5.86e-05
ACR3 test (groups)	ARSB 2 + ARSB 3 monophyletic	-7,888,092,247	4.09e-06
	Unconstrained tree	-8,230,043,048	0.881
	Eukaryotic monophyletic	-8,264,627,396	8.85e-123
	ACR3 1 + ACR3 2 monophyletic	-8,256,999,753	2.26e-07
	ACR3 1 + ACR3 3 monophyletic	-8,261,617,522	8.48e-34
ACR3 test (classification)	ACR3 2 + ACR3 3 monophyletic	-8,234,837,386	0.119
	Unconstrained tree	-7,763,168,156	1
	Eukaryotic monophyletic	-7,796,133,918	3.01e-05
	Oomycetes monophyletic	-7,804,225,848	2.31e-07
	Archaeplastida monophyletic	-7,776,974,866	7.23e-05
ARSM1	Fungi monophyletic	-7,804,238,927	1.17e-77
	Unconstrained tree	-8,984,400,596	0.976
	Eukaryotic ARSM 1 monophyletic	-9,016,839,843	1.16e-75
	Eukaryotic ARSM 1.1 monophyletic	-9,007,446,498	0.00102
ARSM2	Eukaryotic ARSM 1.b monophyletic	-8,993,608,104	0.032
	Unconstrained tree	-5,172,209,384	1
	Eukaryotic ARSM 2 monophyletic	-5,219,695,659	3.04e-13
	ARSM 2.3 + ARSM 2.5 monophyletic	-5,187,353,866	0.000119
	ARSM 2.1 + ARSM 2.3 + ARSM 2.5 monophyletic	-5,260,039,797	1.57e-48
	ARSM 2.2 + ARSM 2.4 monophyletic	-5,187,273,107	9.66e-07

Supplementary Table S1: Results of approximately unbiased (AU) tests of alternative topologies for ACR2, ACR3, ARSB and ArsM gene trees. AU Topology Tests for a combination of topology constrains related to possible origin of arsenic resistance metabolism genes. P-AU represents the p-value of the topologies tested by the approximately unbiased (AU) criteria. Accepted trees are indicated in bold (> 0.05).

Additional data-set tables:

Supplementary Data-set S1: Taxa dataset from the three-domain database used in this work. Database of the study. We assembled a database of OrthoMCL, Genomic and Transcriptomic data from the three domains of life. Transcriptomic data are a compilation of the assembled dataset generated for [36, 37] from the BioProjects indicated. Genomic data are protein data from NCBI or Ensemble databases. We classified eukaryotic species based on [35]. The sources of the data are in the fourth column. We subsampled the database for Figure 2 and Supplementary Figure S17. Figure codes of the subsampled sequences are identified in the second column.

Supplementary Data-set S2: Query sequences of the arsenic resistance genes used in this study. Accession number of functional characterized arsenic resistance proteins in NCBI and Uniprot databases used as query in this study.

Supplementary Data-set S3: Gene ontology of arsenic resistance protein types. Each of the arsenic resistance genes generates protein types differentially defined between the lineages. From the BLAST2GO annotations, we compiled the BLAST definitions and the ontology of each of the protein types. Each of the sheets contains the results of the Blast2GO annotation.

Supplementary Data-set S4: Conserved domains annotation of CD-Search tool from NCBI. Each of the sheets contains the raw domain annotations from the NCBI tool for the sequences of each of the arsenic resistance genes.