Supplementary material: biosynthetic pathways of amino acids

This section supplements the main text describing the procedures and results regarding the reconstruction of the biosynthetic pathways for amino acids in *Deianiraea*, and the phylogenetic analyses of the genes involved.

The prediction of the function of each gene putatively involved in the pathways was verified in detail by manual inspection of blastp results against NCBI databases, and conserved domain search results (Marchler-Bauer et al. 2015). The pathways were reconstructed manually with the reference of Biocyc (Caspi et al. 2016), Pathway tools (Karp et al. 2015), and KEGG (Kanehisa et al. 2016), which were also used for comparing with other bacteria. In particular, the presence in other *Rickettsiales* (excluding samples coming from metagenomic projects) of all the genes involved in amino acid biosynthesis (both those present in *Deianiraea* and those absent) was further directly evaluated by taxon-specific blastp searches. In order to consider a biosynthetic pathway present in a given organism, two main criteria were applied: the gene coding to the last enzymatic step, giving the final compound, must be present, and half (or more) of the genes specific of the pathway should be present.

Further analyses were conducted on the five biosynthetic pathways, responsible for the synthesis of eight amino acids, which were identified as exclusive of *Deianiraea* respect to all known *Rickettsiales*. In particular, phylogenetic analyses were employed to evaluate the evolutionary origin of those genes and the possibility of horizontal gene transfer (HGT). In order to identify the orthologs, the predicted proteins from 2644 complete bacterial genomes were downloaded from NCBI. For each *Deianiraea* protein analysed, the respective best blastp hit on each organismal protein set was identified. In order to filter out paralogs, multiple strategies were employed. First, preliminary screening of the best hits from the NCBI annotation was performed, followed by examination of preliminary single gene trees (data not shown). Additionally, in the cases when *Deianiraea* itself possessed two (or more) paralogous genes, the respective best hits on

each organism were compared, and only the hit with highest identity was retained. Finally, all the *Deianiraea* genes that were found to have orthologs in an insufficient and not taxonomically representative set of organisms (in particular for what concerned the presence in other *Alphaproteobacteria*) were discarded from the analysis. Subsequently, the retained genes of each pathway were treated separately. For each pathway, a representative selection of proteobacterial organisms plus an outgroup was selected for the analysis. Then, each ortholog was aligned with MUSCLE and polished with Gblocks, and a concatenated alignment of all orthologs was produced. For each alignment, ML (maximum likelihood) phylogenetic analyses were performed with RAxML, after selection of the best model using ProtTest. For comparison, the respective organismal phylogenies were inferred, by selecting single copy orthogroups with OrthoFinder (Emms and Kelly 2015), which were aligned, polished and concatenated before ML and BI phylogenies, as described above. The Shimodaira-Hasegawa (SH) test was performed with RAxML, in order to compare the phylogeny of each pathway with the respective organismal phylogeny.

In the following sections, a detailed report is presented on the inferred reconstruction of the five metabolic pathways exclusive of *Deianiraea* among *Rickettsiales,* followed by the respective phylogenetic results.

Aromatic amino acids synthesis

Biosynthesis of aromatic amino acids (tryptophan, phenylalanine, tyrosine) in bacteria initially involves the shikimate pathway. This pathway, starting from erythrose 4-phosphate (an intermediate of the pentose-phosphate pathway) and phosphoenolpyruvate (and requiring another phosphoenolpyruvate at the 6th step), leads to chorismate through 7 enzymatic steps (Pittard and Yang 2008; Bender 2012). In turn, chorismate leads to the synthesis of the aromatic amino acids and other aromatic compounds. The gene sequence of the enzyme catalysing the first step of the

shikimate pathway (3-deoxy-7-phosphoheptulonate synthase) is atypical respect to all *Proteobacteria*. Only a limited number of significant blastp hits was found, and always with very low identity (Supplementary material 15), with sequences assigned to rare bacterial phyla retrieved in metagenomic samples (Anantharam et al. 2016; Lawson et al. 2017; Probst et al. 2018; Brown C.T et al. unpublished: [KKT47881.1]).

Additionally, a peculiar gene fusion of enzymatic activities was retrieved (Deia_00424), namely for the genes encoding in the order the enzymes of the 5th, 4th and 3rd of the shikimate pathway (respectively, shikimate kinase AroL, shikimate dehydrogenase AroE, and 3-dehydroquinate dehydratase AroD). As a matter of fact, several different combinations of gene fusions involved in the shikimate pathway are common in different evolutionary lineages (Bender 2012). Interestingly, the 3-dehydroquinate dehydratase is more similar to type I enzymes found in *Gammaproteobacteria* and *Betaproteobacteria*, rather than to type II, typical in other *Alphaproteobacteria*. The gene for the 6th step of the pathway (3-phosphoshikimate 1-carboxyvinyltransferase) is present also in two *Anaplasmataceae*, namely "*Ca*. Neoehrlichia lotoris" (Daugherty, S.C et al. unpublished: [LANX01000001]) and "*Ca*. Xenolissoclinum pacificiensis" (Kwan and Schmidt 2013).

The biosynthetic pathway for tryptophan immediately deviates after chorismate (five genes for six enzymatic steps in *Escherichia coli*). This pathway is absent in *Deianiraea* and in all other *Rickettsiales*: only a single gene (anthranilate phosphoribosyltransferase TrpD) was found in the *Rickettsiales* bacterium Ac37b (Felsheim, R.F et al. unpublished. [NZ_CP009217]).

The synthesis of both tyrosine and phenylalanine involves the conversion of chorismate into prephenate by a chorismate mutase. The following step is specific for each amino acid, involving prephenate dehydratase for phenylalanine, and prephenate/cyclohexenyl dehydrogenase for tyrosine. Similarly to other *Alphaproteobacteria*, in *Deianiraea* these three enzymatic activities are encoded by three distinct genes, differently from the other *Proteobacteria* classes, presenting two genes, in which different isoforms of chorismate mutase are fused with the each of the other two

enzymes. The last step is the amination of both tyrosine and phenylalanine, performed by a generalist amino acid aminotransferase (e.g. Deia_00080, Deia_00820, Deia_00902).

Phylogenetic analyses: Used eight genes (operationally subdividing Deia_00424 into three distinct genes according to the homology-predicted enzymatic activities) and 71 organisms, 1,250 sites. For the phylogenomics, used 59 ortholog genes and 10,126 sites.

Figure 13.1 Maximum likelihood tree of the eight concatenated genes involved in the aromatic amino acids biosynthesis with the LG+I+G+F substitution model with 100 bootstrap pseudo-replicates. The position of *Deianiraea* genes is unsupported. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values.

Figure 13.2 Reference maximum likelihood phylogenomic tree of the organisms employed for phylogenesis in **13.1** with the LG+I+G substitution model with 100 bootstrap pseudo-replicates. The five proteobacterial classes (as indicated in the figure) are all monophyletic and highly supported, including the positioning of *Deianiraea* within *Alphaproteobacteria*, which is consistent with the phylogenomic analyses. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values.

Figure 13.1

Scale bar: 0.3

- 88 Erythrobacter flavus CP022528
- 100 Croceicoccus naphthovorans CP011770
- 100 Sphingopyxis granuli CP012199
- 100 Sphingorhabdus flavimaris CP022336
- 97 Sphingobium baderi CP013264
- 100 Paracoccus yeei CP020442
- 100 Rhodovulum sulfidophilum AP014800
- 56 Thioclava nitratireducens CP019437
- 79 Loktanella vestfoldensis CP021431
- 48 Kozakia baliensis CP014681
- 100 Acetobacter pasteurianus HF677570
- 68 Azospirillum lipoferum FQ311868
- 86 Haematospirillum jordaniae CP014525
- 23 Caulobacter henricii CP013002
- 32 Ochrobactrum pseudogrignonense CP015775
- 100 Mesorhizobium sp. CP018171
- 99 Bosea vaviloviae CP017147
- 78 Methylobacterium oryzae CP003811
- 58 Oligotropha carboxidovorans CP002821
- 100 Bradyrhizobium lablabi LT670844
- 89 Labilithrix luteola CP012333
- Psychromonas sp. CP004404
- 100 Pseudoalteromonas haloplanktis CR954246
- 97 Shewanella baltica CP001252
- 36 Oceanimonas sp. CP003171
- 69 Aeromonas hydrophila CP007518
- 80 Vibrio cholerae CP001235
- 72 Pasteurella multocida CP003022
- 100 Haemophilus influenzae CP007805
- 96 Erwinia amylovora FN434113
- 100 Yersinia pestis CP000305
- 63 Escherichia coli CP014270
- 66 Klebsiella pneumoniae CP011985
- 79 Kosakonia cowanii CP019445
- 56 **"*Candidatus* Deianiraea vastatrix"**
- 67 Vogesella sp. LT607802
- 99 Neisseria gonorrhoeae CP001050
- 100 Nitrosomonas ureae CP013341
- 94 Thauera sp. CP001281
- 53 Bordetella pertussis CP002695
- 47 Variovorax paradoxus CP001635
- 54 Rhizobacter gummiphilus CP015118
- 100 Comamonas aquatica CP016603
- 100 Alicycliphilus denitrificans CP002449
- 100 Acidovorax ebreus CP001392
- 28 Polynucleobacter duraquae CP007501
- 100 Oxalobacter formigenes CP019430
- 31 Herbaspirillum seropedicae CP002039
- 100 Ralstonia solanacearum FP885906
- 100 Cupriavidus basilensis CP010536
- 62 Pandoraea faecigallinarum CP011807
- 87 Burkholderia pseudomallei CP003781
- 67 Dyella thiooxydans CP014841
- 78 Acidihalobacter ferrooxidans CP019434
- 91 Methylomonas denitrificans CP014476
- 100 Microbulbifer thermotolerans CP014864
- 100 Pseudomonas aeruginosa CP007147
- 100 Halomonas sp. CP022286
- 100 Kushneria marisflavi CP021358
- Helicobacter cinaedi AP012344
- 100 Campylobacter coli CP007181
- 80 Sulfurospirillum sp. CP021416
- Geoalkalibacter subterraneus CP010311
- Desulfuromonas soudanensis CP010802
- Geobacter sulfurreducens CP002031
- 42 Anoxybacillus flavithermus CP000922
- 45 Geobacillus subterraneus CP014342
- Bacillus megaterium CP003017
- Kurthia sp. CP013217
- 52 Salimicrobium jeotgali CP011361
- 100 Virgibacillus halodenitrificans CP017962

Outgroup

Branch support values: 72, 67, 67, 100, 100, 86, 100, 100, 52

Figure 13.1

Figure 13.2

0.2

Paracoccus yeei CP020442
Thioclava nitratireducens CP019437
Loktanella vestfoldensis CP021431
Rhodovulum sulfidophilum AP014800
Caulobacter henricii CP013002
Ochrobactrum pseudogrignonense CP015775
Mesorhizobium sp.CP018171
Methylobacterium oryzae CP003811
Bosea vaviloviae CP017147
Oligotropha carboxidovorans CP002821
Bradyrhizobium lablabi LT670844
Sphingobium baderi CP013264
Sphingopyxis granuli CP012199
Sphingorhabdus flavimaris CP022336
Erythrobacter flavus CP022528
Croceicoccus naphthovorans CP011770
Haematospirillum jordaniae CP014525
Azospirillum lipoferum FQ311868
Kozakia baliensis CP014681
Acetobacter pasteurianus HF677570
"*Candidatus* Deianiraea vastatrix"

Alphaproteobacteria

Alicycliphilus denitrificans CP002449
Acidovorax ebreus CP001392
Comamonas aquatica CP016603
Variovorax paradoxus CP001635
Rhizobacter gummiphilus CP015118
Bordetella pertussis CP002695
Cupriavidus basilensis CP010536
Ralstonia solanacearum FP885906
Polynucleobacter duraquae CP007501
Pandoraea faecigallinarum CP011807
Burkholderia pseudomallei CP003781
Herbaspirillum seropedicae CP002039
Oxalobacter formigenes CP019430
Thauera sp. CP001281
Neisseria gonorrhoeae CP001050
Vogesella sp. LT607802
Nitrosomonas ureae CP013341

Betaproteobacteria

Kosakonia cowanii CP019445
Klebsiella pneumoniae CP011985
Escherichia coli CP014270
Erwinia amylovora FN434113
Yersinia pestis CP000305
Haemophilus influenzae CP007805
Pasteurella multocida CP003022
Vibrio cholerae CP001235
Aeromonas hydrophila CP007518
Oceanimonas sp CP003171
Pseudoalteromonas haloplanktis CR954246
Shewanella baltica CP001252
Psychromonas sp CP004404
Pseudomonas aeruginosa CP007147
Halomonas sp. CP022286
Kushneria marisflavi CP021358
Microbulbifer thermotolerans CP014864
Methylomonas denitrificans CP014476
Acidihalobacter ferrooxidans CP019434
Dyella thiooxydans CP014841

Gammaproteobacteria

Geobacter sulfurreducens CP002031
Desulfuromonas soudanensis CP010802
Geoalkalibacter subterraneus CP010311
Labilithrix luteola CP012333

Deltaproteobacteria

Helicobacter cinaedi AP012344
Campylobacter coli CP007181
Sulfurospirillum sp. CP021416

Epsilonproteobacteria

Salimicrobium jeotgali CP011361
Virgibacillus halodenitrificans CP017962
Kurthia sp. CP013217
Bacillus megaterium CP003017
Geobacillus subterraneus CP014342
Anoxybacillus flavithermus CP000922

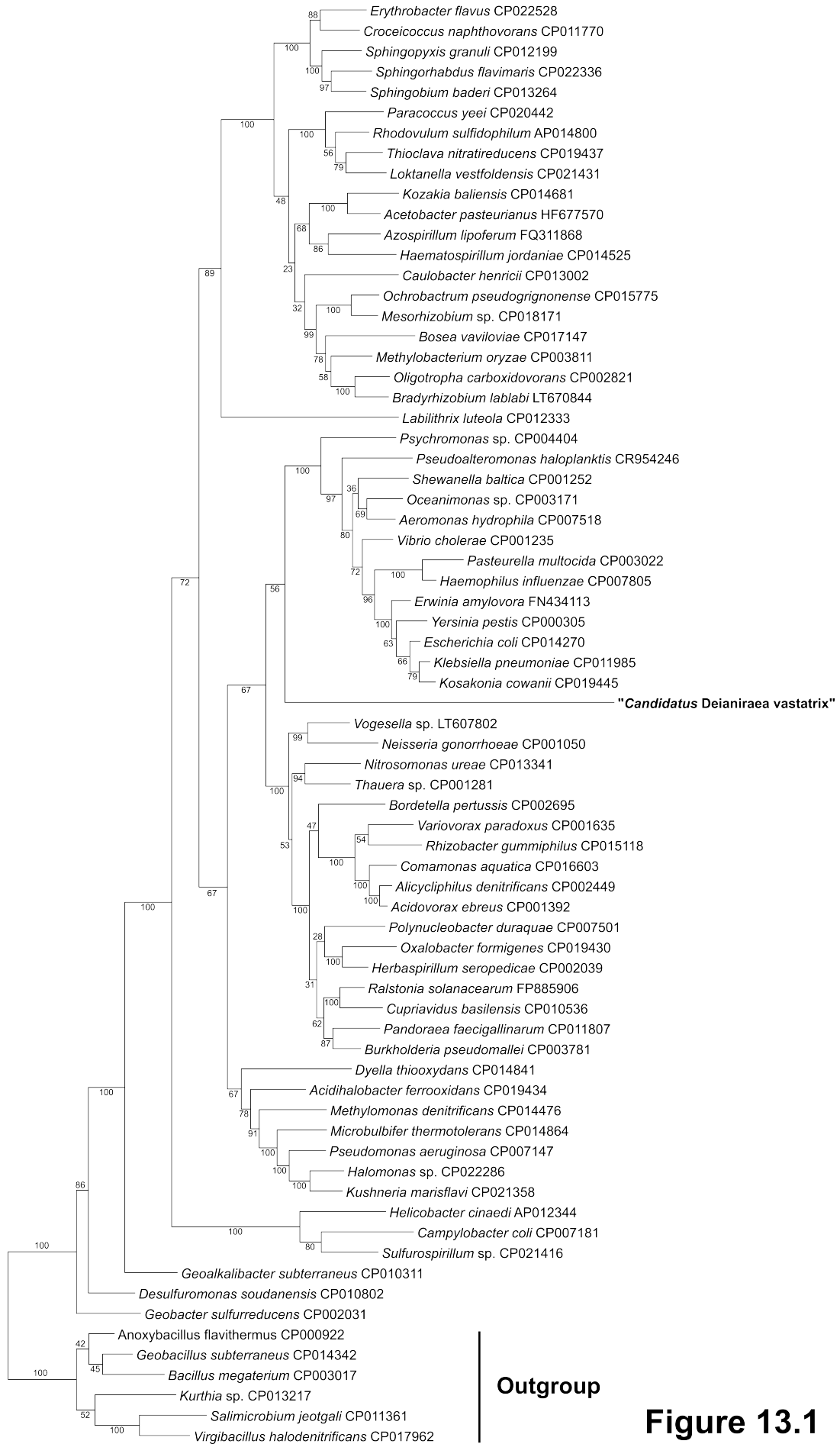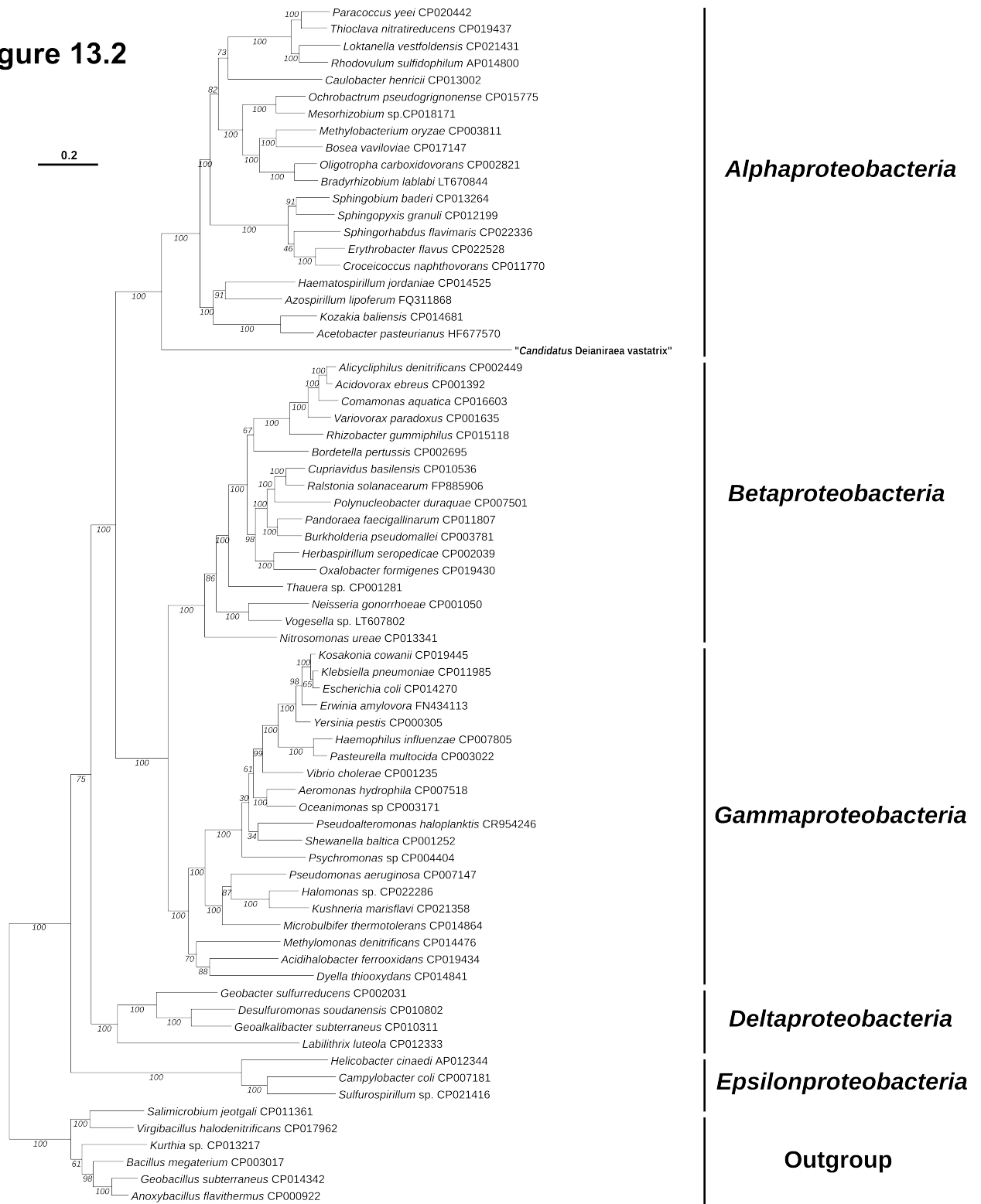Outgroup

Branched chain amino acids synthesis

The biosynthesis of the three branched chain amino acids (valine, leucine, isoleucine) involves a number of shared enzymes (Salmon et al. 2006). Typically, 3-methyloxybutanoate, common intermediate of leucine and valine, is obtained in three enzymatic steps from two pyruvate molecules. Two isozymes working in the first step were found in *Deianiraea* (Deia_00067 acetolactate/acetohydroxybutanoate synthase large subunit; Deia_01102 acetohydroxybutanoate/acetolactate synthase large subunit). A putative homologue is present in "*Occidentia massiliensis*" (*Rickettsiaceae*) (Mediannikov et al. 2014). Another ORF (Deia_1007) with similarities to the accessory regulatory subunit of acetolactate synthase was identified, but, since the alignment involved only partial length of database reference sequences (data not shown), it was judged non-functional and not further considered. The gene for the second step of the pathway (ketol-acid reductoisomerase) is also present in few *Anaplasmataceae*, namely "*Ca.* Neoehrlichia lotoris" (Daugherty, S.C et al. unpublished [LANX01000001]) and *Anaplasma* spp. (Dark et al. 2009).

For the synthesis of valine from 3-methyloxybutanoate, a single additional step is needed, exerted by the branched-chain aminotransferase. Contrarily to other genes of the pathway, this is present in several *Rickettsiales*, having a likely a broader generalist amino acid aminotransferase activity, or possibly acting on the biosynthetic intermediates derived from hosts.

For leucine biosynthesis from 3-methyloxybutanoate, three specific enzymatic reactions are necessary, involving four proteins (two enzymes and a dimeric complex formed by the other two proteins), plus a fourth terminal amination step performed by the same branched-chain aminotransferase as for valine.

The most common bacterial biosynthetic pathway for isoleucine starts from the deamination of threonine, to produce 2-oxobutanoate. This compound is then processed by the same sequential four enzymes as for valine (including terminal aminotransferase), to obtain isoleucine. The specific gene
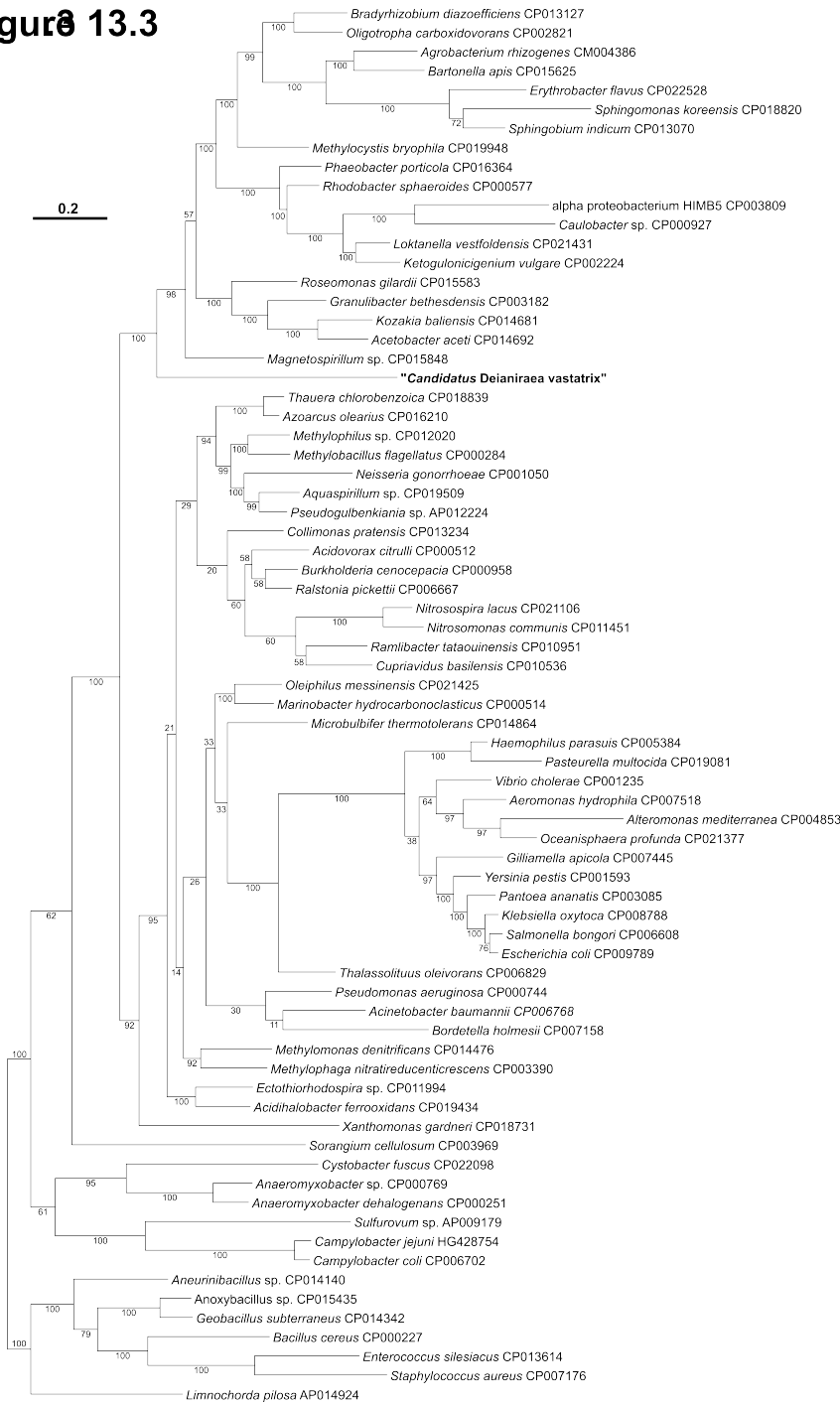
for the initial step (threonine deaminase IlvA) is not present in *Deianiraea*, differently from other *Alphaproteobacteria*. A detailed inspection allowed to determine that ORF Deia_00356, initially flagged as an additional homologue of the leucine synthesis protein 2-isopropylmalate synthase, displays higher homology and similar domain architecture to citramalate synthase (data not shown). Citramalate synthase is involved in an alternative pathway for the synthesis of isoleucine found in a growing number of bacteria (Xu et al. 2004; Risso et al. 2008), including *Alphaproteobacteria* (Tang et al. 2009; McKinlay et al. 2010). Starting from pyruvate and acetyl-CoA, citramalate is formed. This compound is then processed by the two terminal leucine synthesis enzymes (excluding the aminotransferase) to obtain 2-oxobutanoate, which can then enter the canonical pathway. As possible additional alternative, methionine can also be a source of 2-oxobutyrate, through the hydrolytic action of methionine gamma-synthase/lyase.

Phylogenetic analysis: Used eight genes and 73 organisms, 2,674 sites. For the phylogenomics, used 42 ortholog genes and 6,630 sites.

Figure 13.3 Maximum likelihood tree of the eight concatenated genes involved in the branched amino acids biosynthesis with the LG+I+G substitution model with 100 bootstrap pseudo-replicates. The position of *Deianiraea* genes is consistent with organismal phylogeny in **16.7** with high support. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values.

Figure 13.4 Reference maximum likelihood phylogenomic tree of the organisms employed for phylogenesis in **13.3** with the LG+I+G substitution model with 100 bootstrap pseudo-replicates. The five proteobacterial classes (as indicated in the figure) are all monophyletic and supported, including the positioning of *Deianiraea* within *Alphaproteobacteria*, which is consistent with the phylogenomic analyses. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values**.**

Figure 13.3

Figure 13.3

Figure 13.4

Erythrobacter flavus CP022528
Sphingobium indicum CP013070
Sphingomonas koreensis CP018820
Rhodobacter sphaeroides CP000577
Phaeobacter porticola CP016364
Loktanella vestfoldensis CP021431
Ketogulonicigenium vulgare CP002224
Caulobacter sp. CP000927
Bartonella apis CP015625
Agrobacterium rhizogenes CM004386
Methylocystis bryophila CP019948
Oligotropha carboxidovorans CP002821
Bradyrhizobium diazoefficiens CP013127
Magnetospirillum sp. CP015848
Roseomonas gilardii CP015583
Granulibacter bethesdensis CP003182
Kozakia baliensis CP014681
Acetobacter aceti CP014692
"*Candidatus* Deianiraea vastatrix"
alpha proteobacterium HIMB5 CP003809

Alphaproteobacteria

Aquaspirillum sp. CP019509
Neisseria gonorrhoeae CP001050
Pseudogulbenkiania sp. AP012224
Thauera chlorobenzoica CP018839
Azoarcus olearius CP016210
Bordetella holmesii CP007158
Ramlibacter tataouinensis CP010951
Acidovorax citrulli CP000512
Collimonas pratensis CP013234
Burkholderia cenocepacia CP000958
Ralstonia pickettii CP006667
Cupriavidus basilensis CP010536
Nitrosomonas communis CP011451
Nitrosospira lacus CP021106
Methylophilus sp. CP012020
Methylobacillus flagellatus CP000284

Betaproteobacteria

Klebsiella oxytoca CP008788
Escherichia coli CP009789
Salmonella bongori CP006608
Pantoea ananatis CP003085
Yersinia pestis CP001593
Gilliamella apicola CP007445
Pasteurella multocida CP019081
Haemophilus parasuis CP005384
Vibrio cholerae CP001235
Aeromonas hydrophila CP007518
Oceanisphaera profunda CP021377
Alteromonas mediterranea CP004853
Acinetobacter baumannii CP006768
Pseudomonas aeruginosa CP000744
Microbulbifer thermotolerans CP014864
Thalassolituus oleivorans CP006829
Marinobacter hydrocarbonoclasticus CP000514
Oleiphilus messinensis CP021425
Ectothiorhodospira sp. CP011994
Acidihalobacter ferrooxidans CP019434
Methylophaga nitratireducenticrescens CP003390
Methylomonas denitrificans CP014476
Xanthomonas gardneri CP018731

Gammaproteobacteria

Sulfurovum sp. AP009179
Campylobacter jejuni HG428754
Campylobacter coli CP006702

Epsilonproteobacteria

Cystobacter fuscus CP022098
Anaeromyxobacter sp. CP000769
Anaeromyxobacter dehalogenans CP000251
Sorangium cellulosum CP003969

Deltaproteobacteria

Limnochorda pilosa AP014924
Aneurinibacillus sp. CP014140
Enterococcus silesiacus CP013614
Staphylococcus aureus CP007176
Bacillus cereus CP000227
Anoxybacillus sp. CP015435
Geobacillus subterraneus CP014342

Outgroup

0.2

Cysteine synthesis

In most bacteria cysteine biosynthesis is exerted by a two-step pathway, starting from serine. In the *Deianiraea* genome only the second enzyme is encoded (Deia_00050: O-acetylserine sulfhydrylase/cysteine synthase A CysK). Thus, its substrate(s) acetyl-serine (or phosphoserine) should be either directly imported from *Paramecium*, or produced by an unidentified acetyl-(or phosho)-transferase acting on serine (instead of the canonical serine acetyltransferase/kinase, absent in *Deianiraea*). Alternatively, cysteine might also be produced from methionine via the reversed trans-sulfuration pathway (**methionine synthesis** pathway) (Bender 2012; Makino et al. 2016)

Phylogenetic analyses: Used one gene and 69 organisms, 261 sites. For the phylogenomics, used 49 ortholog genes and 6,009 sites.

Figure 13.5 Maximum likelihood tree of the gene involved in the cysteine biosynthesis with the LG+I+G substitution model with 100 bootstrap pseudo-replicates. The position of *Deianiraea* genes and the main proteobacterial classes are unsupported. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values.

Figure 13.6 Reference maximum likelihood phylogenomic tree of the organisms employed for phylogenesis in **13.5** and with the LG+I+G substitution model with 100 bootstrap pseudo-replicates. The five proteobacterial classes (as indicated in the figure) are all monophyletic and highly supported, including the positioning of *Deianiraea* within *Alphaproteobacteria*, which is consistent with the phylogenomic analyses. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values.
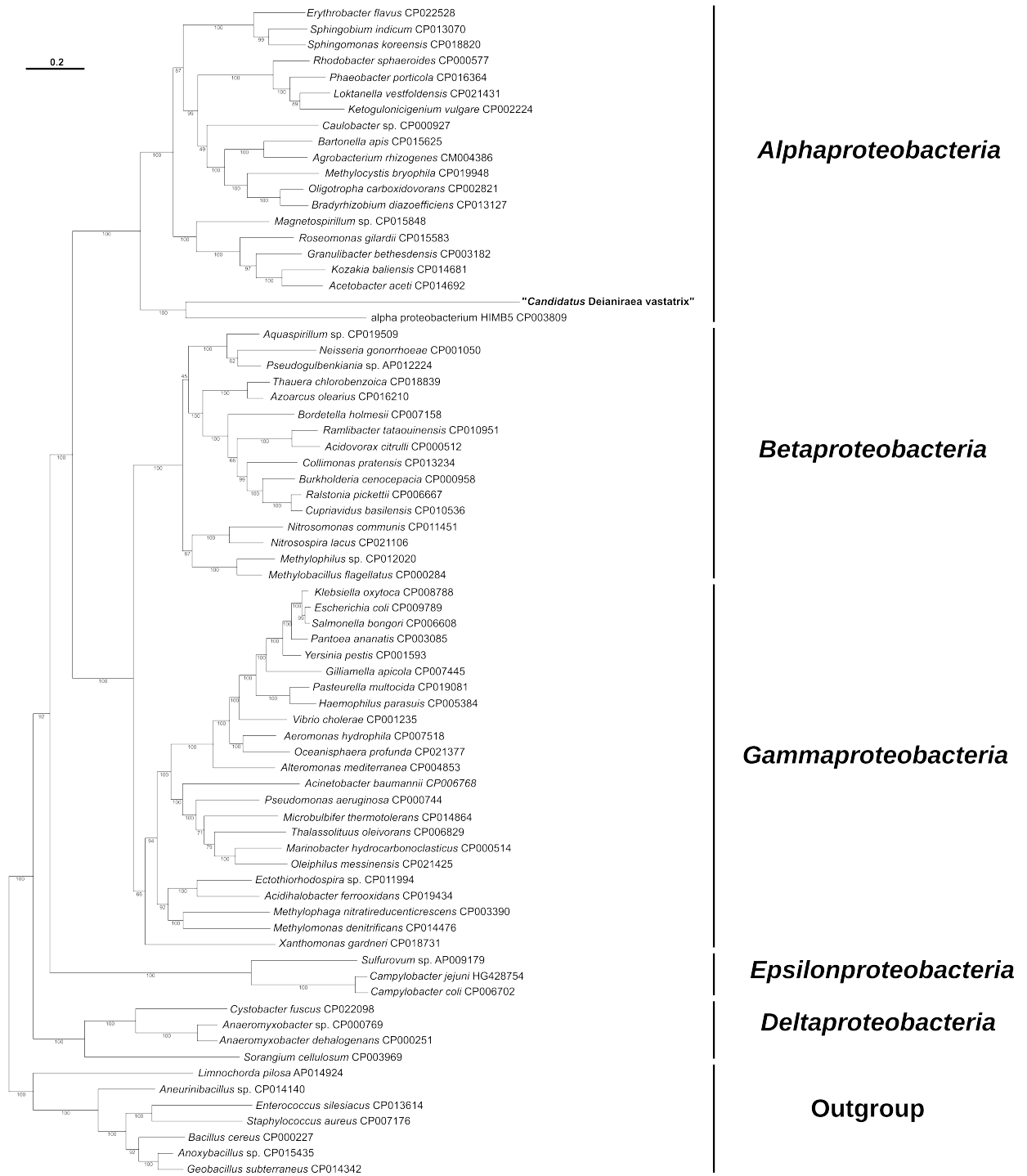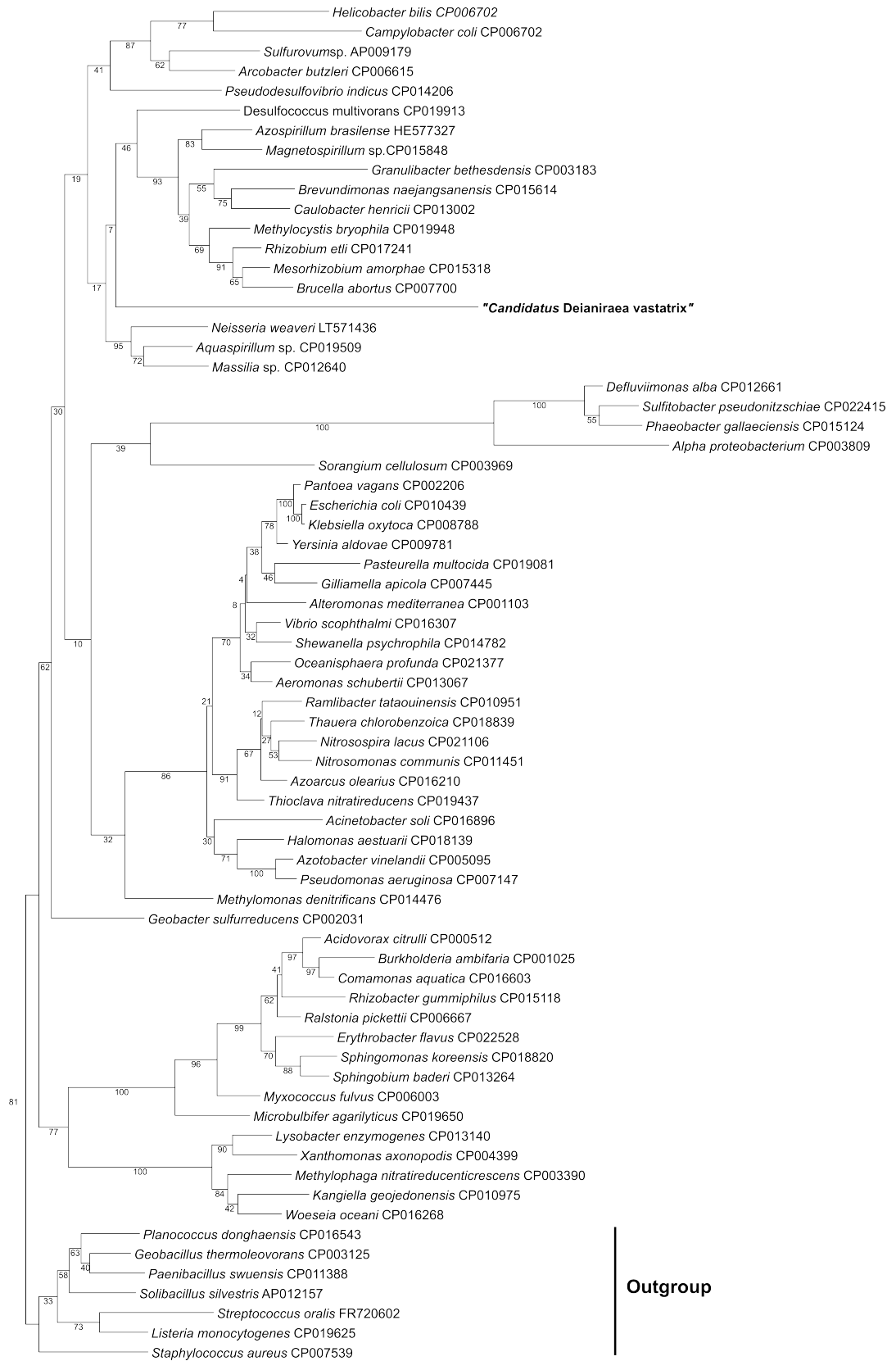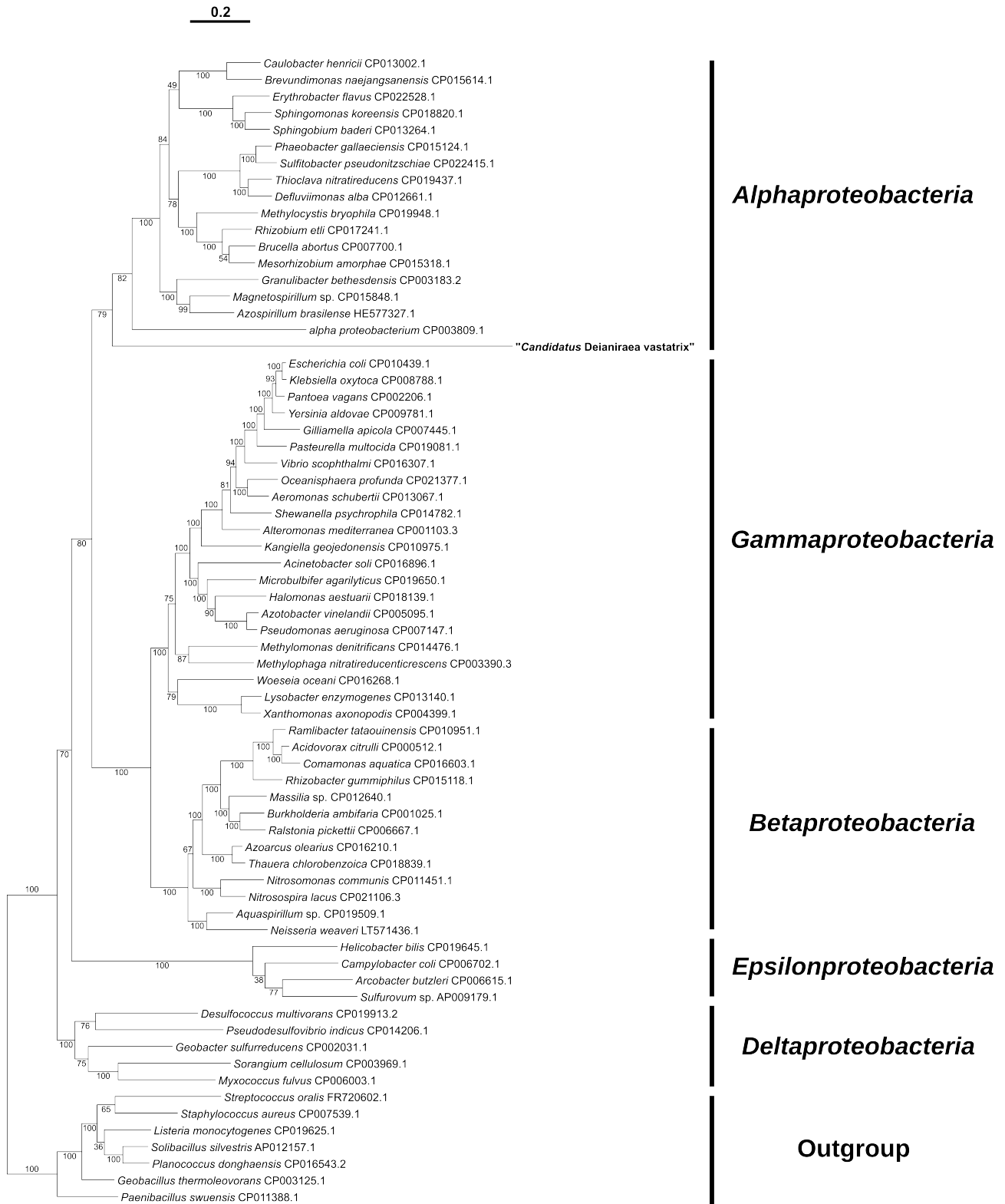
Figure 13.5

0.1

Helicobacter bilis CP006702
Campylobacter coli CP006702
77
87
Sulfurovum sp. AP009179
62
Arcobacter butzleri CP006615
41
Pseudodesulfovibrio indicus CP014206
19
Desulfococcus multivorans CP019913
46
Azospirillum brasilense HE577327
83
Magnetospirillum sp. CP015848
93
Granulibacter bethesdensis CP003183
55
Brevundimonas naejangsanensis CP015614
75
Caulobacter henricii CP013002
39
7
Methylocystis bryophila CP019948
69
Rhizobium etli CP017241
91
Mesorhizobium amorphae CP015318
65
Brucella abortus CP007700
17
"Candidatus Deianiraea vastatrix"
Neisseria weaveri LT571436
95
Aquaspirillum sp. CP019509
72
Massilia sp. CP012640
30
Defluviimonas alba CP012661
100
Sulfitobacter pseudonitzschiae CP022415
55
Phaeobacter gallaeciensis CP015124
100
Alpha proteobacterium CP003809
39
Sorangium cellulosum CP003969
Pantoea vagans CP002206
100
Escherichia coli CP010439
100
78
Klebsiella oxytoca CP008788
Yersinia aldovae CP009781
38
Pasteurella multocida CP019081
46
4
Gilliamella apicola CP007445
8
Alteromonas mediterranea CP001103
10
Vibrio scophthalmi CP016307
32
70
Shewanella psychrophila CP014782
Oceanisphaera profunda CP021377
34
Aeromonas schubertii CP013067
21
Ramlibacter tataouinensis CP010951
12
Thauera chlorobenzoica CP018839
27
Nitrosospira lacus CP021106
53
67
Nitrosomonas communis CP011451
86
91
Azoarcus olearius CP016210
Thioclava nitratireducens CP019437
Acinetobacter soli CP016896
30
Halomonas aestuarii CP018139
71
Azotobacter vinelandii CP005095
100
32
Pseudomonas aeruginosa CP007147
62
Methylomonas denitrificans CP014476
Geobacter sulfurreducens CP002031
Acidovorax citrulli CP000512
97
Burkholderia ambifaria CP001025
41
97
Comamonas aquatica CP016603
62
Rhizobacter gummiphilus CP015118
99
Ralstonia pickettii CP006667
Erythrobacter flavus CP022528
70
Sphingomonas koreensis CP018820
96
88
Sphingobium baderi CP013264
100
Myxococcus fulvus CP006003
Microbulbifer agarilyticus CP019650
81
77
Lysobacter enzymogenes CP013140
90
Xanthomonas axonopodis CP004399
100
Methylophaga nitratireducenticrescens CP003390
84
Kangiella geojedonensis CP010975
42
Woeseia oceani CP016268
Planococcus donghaensis CP016543
63
Geobacillus thermoleovorans CP003125
40
58
Paenibacillus swuensis CP011388
Solibacillus silvestris AP012157
33
Streptococcus oralis FR720602
73
Listeria monocytogenes CP019625
Staphylococcus aureus CP007539

Outgroup

Figure 13.6

0.2

Caulobacter henricii CP013002.1
Brevundimonas naejangsanensis CP015614.1
Erythrobacter flavus CP022528.1
Sphingomonas koreensis CP018820.1
Sphingobium baderi CP013264.1
Phaeobacter gallaeciensis CP015124.1
Sulfitobacter pseudonitzschiae CP022415.1
Thioclava nitratireducens CP019437.1
Defluviimonas alba CP012661.1
Methylocystis bryophila CP019948.1
Rhizobium etli CP017241.1
Brucella abortus CP007700.1
Mesorhizobium amorphae CP015318.1
Granulibacter bethesdensis CP003183.2
Magnetospirillum sp. CP015848.1
Azospirillum brasilense HE577327.1
alpha proteobacterium CP003809.1
"*Candidatus* Deianiraea vastatrix"

Alphaproteobacteria

Escherichia coli CP010439.1
Klebsiella oxytoca CP008788.1
Pantoea vagans CP002206.1
Yersinia aldovae CP009781.1
Gilliamella apicola CP007445.1
Pasteurella multocida CP019081.1
Vibrio scophthalmi CP016307.1
Oceanisphaera profunda CP021377.1
Aeromonas schubertii CP013067.1
Shewanella psychrophila CP014782.1
Alteromonas mediterranea CP001103.3
Kangiella geojedonensis CP010975.1
Acinetobacter soli CP016896.1
Microbulbifer agarilyticus CP019650.1
Halomonas aestuarii CP018139.1
Azotobacter vinelandii CP005095.1
Pseudomonas aeruginosa CP007147.1
Methylomonas denitrificans CP014476.1
Methylophaga nitratireducenticrescens CP003390.3
Woeseia oceani CP016268.1
Lysobacter enzymogenes CP013140.1
Xanthomonas axonopodis CP004399.1

Gammaproteobacteria

Ramlibacter tataouinensis CP010951.1
Acidovorax citrulli CP000512.1
Comamonas aquatica CP016603.1
Rhizobacter gummiphilus CP015118.1
Massilia sp. CP012640.1
Burkholderia ambifaria CP001025.1
Ralstonia pickettii CP006667.1
Azoarcus olearius CP016210.1
Thauera chlorobenzoica CP018839.1
Nitrosomonas communis CP011451.1
Nitrosospira lacus CP021106.3
Aquaspirillum sp. CP019509.1
Neisseria weaveri LT571436.1

Betaproteobacteria

Helicobacter bilis CP019645.1
Campylobacter coli CP006702.1
Arcobacter butzleri CP006615.1
Sulfurovum sp. AP009179.1

Epsilonproteobacteria

Desulfococcus multivorans CP019913.2
Pseudodesulfovibrio indicus CP014206.1
Geobacter sulfurreducens CP002031.1
Sorangium cellulosum CP003969.1
Myxococcus fulvus CP006003.1

Deltaproteobacteria

Streptococcus oralis FR720602.1
Staphylococcus aureus CP007539.1
Listeria monocytogenes CP019625.1
Solibacillus silvestris AP012157.1
Planococcus donghaensis CP016543.2
Geobacillus thermoleovorans CP003125.1
Paenibacillus swuensis CP011388.1

Outgroup

Histidine synthesis

Biosynthesis of histidine in bacteria is a multi-step pathway, involving ten enzymatic reactions, starting from 5-phosphoribosyl diphosphate, in turn obtained from the pentose phosphate intermediate ribose-5-phosphate via the ribose-phosphate diphosphokinase (Bender 2012; Winkler and Ramos-Montañez 2009). Depending on the organism, the number of genes involved in the pathway typically varies from eight to ten, given that alternative combinations of fused or separated of enzymatic activities for two pair of genes (regarding steps $2^{nd}+3^{rd}$ and $6^{th}+8^{th}$) might occur. *Deianiraea* possesses eight genes, with the two fused combinations. In particular, the presence of the bifunctional fused imidazoleglycerol-phosphate dehydratase/histidinol-phosphatase HisB ($6^{th}+8^{th}$ steps) is peculiar. Indeed, almost all *Alphaproteobacteria* possess separated genes for those enzymatic activities (with possible exceptions (Tully et al. 2017)), while the fused HisB is more common in other proteobacterial lineages, such as *Gammaproteobacteria* and *Epsilonproteobacteria*. Among *Rickettsiales*, the gene for the enzyme catalysing the 7th step (Deia_00820 Histidinol-phosphate aminotransferase) finds a counterpart in *Rickettsiales* bacterium Ac37b (Felsheim, R.F et al. unpublished. [NZ_CP009217]), and possibly in "*Candidatus* Rickettsia asemboensis" (Jima et al. 2015).

Phylogenetic analyses Used nine genes (operationally subdividing HisB gene into two distinct genes according to the homology-predicted enzymatic activities) and 72 organisms, 1,123 sites. For the phylogenomics, used 56 ortholog genes and 7,917 sites.

Figure 13.7 Maximum likelihood tree of the nine concatenated genes involved in the histidine biosynthesis with the LG+I+G+F substitution model with 100 bootstrap pseudo-replicates. The *Deianiraea* genes are very far-related to any other organism. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values

Figure 13.8 Reference maximum likelihood phylogenomic tree of the organisms employed for

phylogenesis in **13.7** with the LG+I+G substitution model with 100 bootstrap pseudo-replicates. The five proteobacterial classes (as indicated in the figure) are all monophyletic and highly supported, including the positioning of *Deianiraea* within *Alphaproteobacteria*, which is consistent with the phylogenomic analyses. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values.
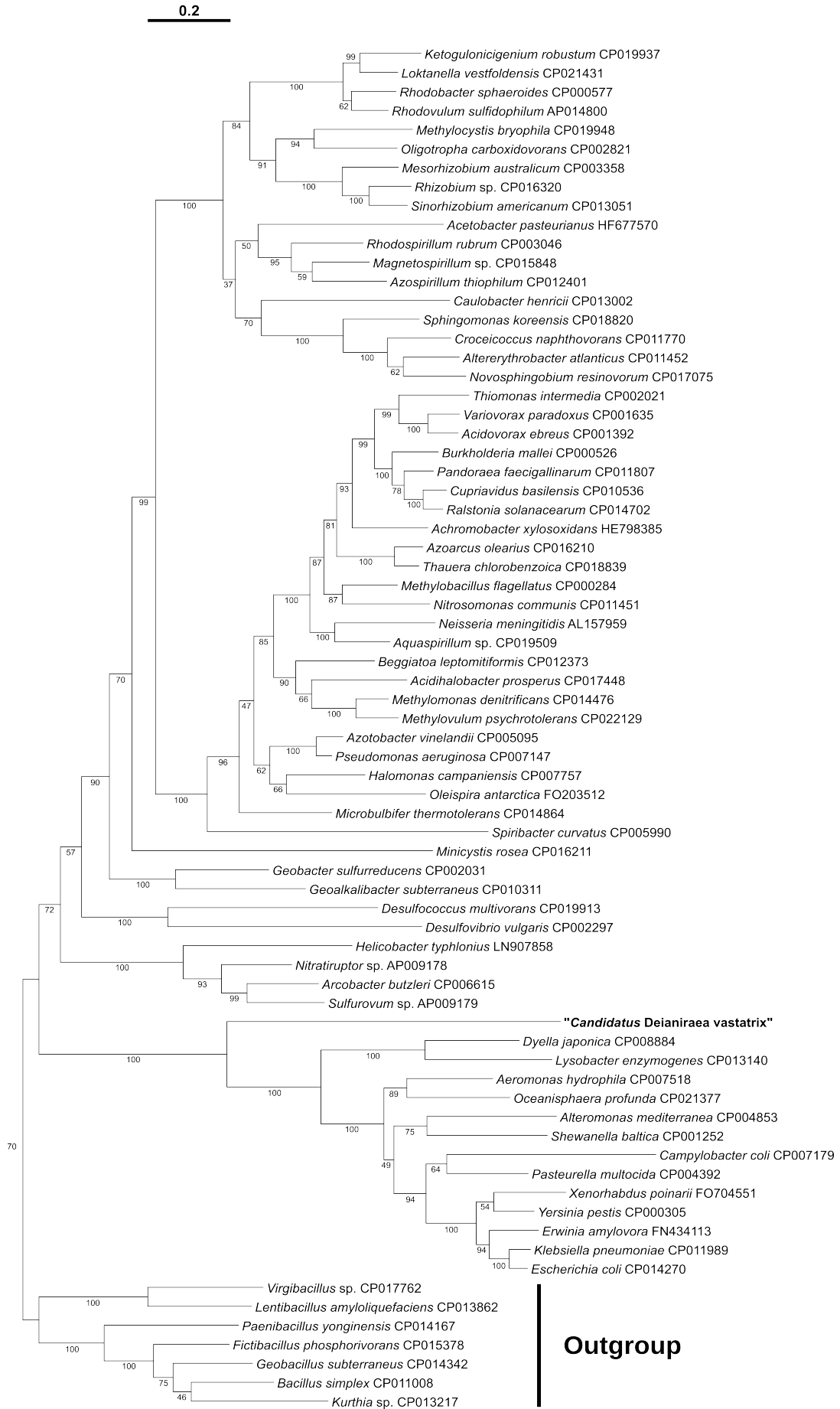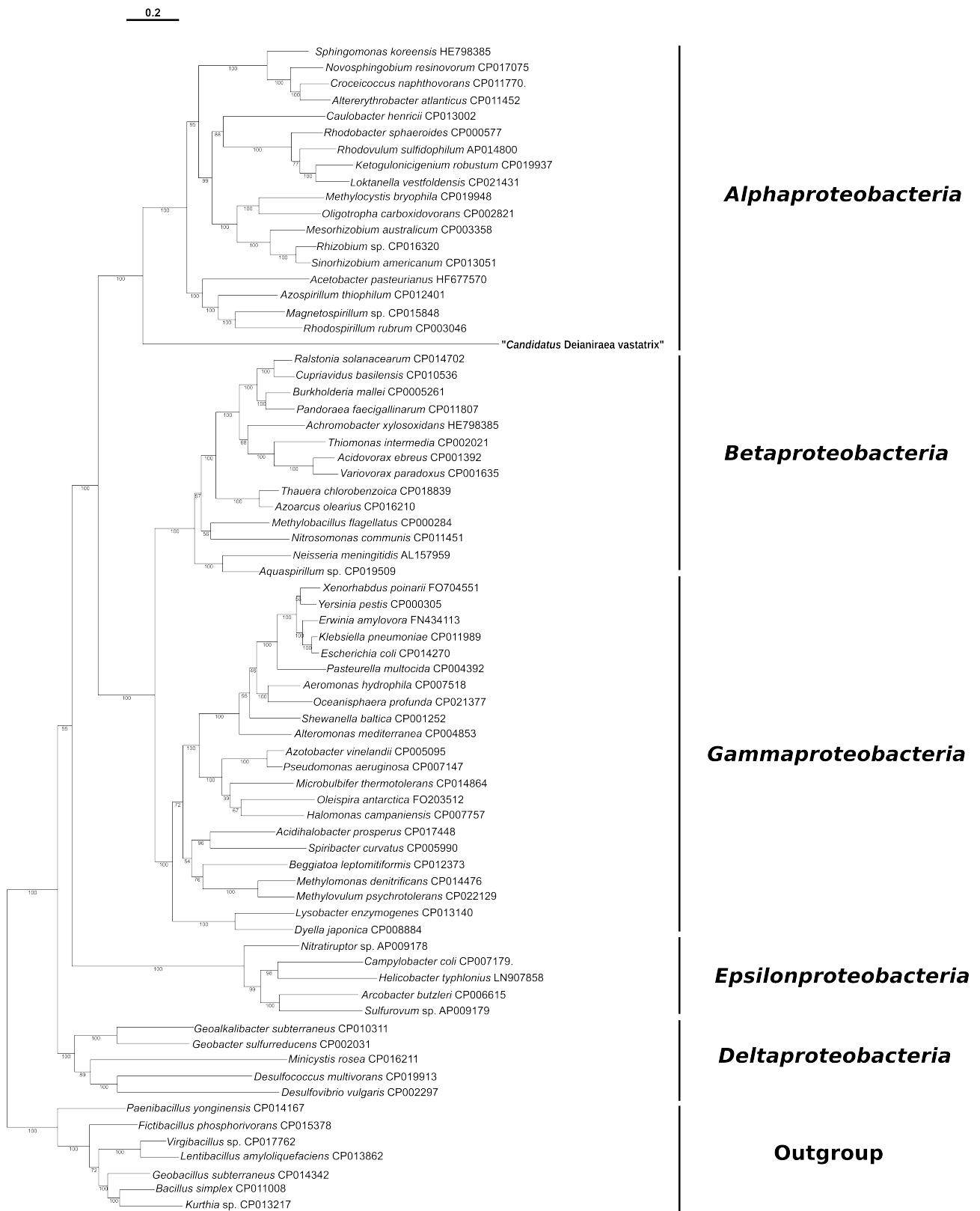
Figure 13.7

0.2

99 — *Ketogulonicigenium robustum* CP019937
— *Loktanella vestfoldensis* CP021431
100
62 — *Rhodobacter sphaeroides* CP000577
— *Rhodovulum sulfidophilum* AP014800
84
94 — *Methylocystis bryophila* CP019948
— *Oligotropha carboxidovorans* CP002821
91
100 — *Mesorhizobium australicum* CP003358
100 — *Rhizobium* sp. CP016320
100 — *Sinorhizobium americanum* CP013051
100
— *Acetobacter pasteurianus* HF677570
50 — *Rhodospirillum rubrum* CP003046
95 — *Magnetospirillum* sp. CP015848
37 59 — *Azospirillum thiophilum* CP012401
— *Caulobacter henricii* CP013002
70 — *Sphingomonas koreensis* CP018820
100 — *Croceicoccus naphthovorans* CP011770
100 — *Altererythrobacter atlanticus* CP011452
62 — *Novosphingobium resinovorum* CP017075
— *Thiomonas intermedia* CP002021
99 — *Variovorax paradoxus* CP001635
100 — *Acidovorax ebreus* CP001392
99 — *Burkholderia mallei* CP000526
100 — *Pandoraea faecigallinarum* CP011807
78 — *Cupriavidus basilensis* CP010536
93 100 — *Ralstonia solanacearum* CP014702
81 — *Achromobacter xylosoxidans* HE798385
87 — *Azoarcus olearius* CP016210
100 — *Thauera chlorobenzoica* CP018839
100 — *Methylobacillus flagellatus* CP000284
87 — *Nitrosomonas communis* CP011451
85 — *Neisseria meningitidis* AL157959
100 — *Aquaspirillum* sp. CP019509
— *Beggiatoa leptomitiformis* CP012373
90 — *Acidihalobacter prosperus* CP017448
47 66 — *Methylomonas denitrificans* CP014476
100 — *Methylovulum psychrotolerans* CP022129
100 — *Azotobacter vinelandii* CP005095
96 62 — *Pseudomonas aeruginosa* CP007147
90 — *Halomonas campaniensis* CP007757
100 66 — *Oleispira antarctica* FO203512
— *Microbulbifer thermotolerans* CP014864
— *Spiribacter curvatus* CP005990
57 — *Minicystis rosea* CP016211
72 — *Geobacter sulfurreducens* CP002031
100 — *Geoalkalibacter subterraneus* CP010311
— *Desulfococcus multivorans* CP019913
100 — *Desulfovibrio vulgaris* CP002297
— *Helicobacter typhlonius* LN907858
100 — *Nitratiruptor* sp. AP009178
93 — *Arcobacter butzleri* CP006615
99 — *Sulfurovum* sp. AP009179
"*Candidatus* Deianiraea vastatrix"
— *Dyella japonica* CP008884
100 — *Lysobacter enzymogenes* CP013140
100 — *Aeromonas hydrophila* CP007518
89 — *Oceanisphaera profunda* CP021377
100 75 — *Alteromonas mediterranea* CP004853
— *Shewanella baltica* CP001252
49 — *Campylobacter coli* CP007179
64 — *Pasteurella multocida* CP004392
94 — *Xenorhabdus poinarii* FO704551
54 — *Yersinia pestis* CP000305
100 — *Erwinia amylovora* FN434113
94 — *Klebsiella pneumoniae* CP011989
100 — *Escherichia coli* CP014270
70 — *Virgibacillus* sp. CP017762
100 — *Lentibacillus amyloliquefaciens* CP013862
— *Paenibacillus yonginensis* CP014167
100 — *Fictibacillus phosphorivorans* CP015378
100 — *Geobacillus subterraneus* CP014342
75 — *Bacillus simplex* CP011008
46 — *Kurthia* sp. CP013217

Outgroup

Figure 13.8

Methionine synthesis and metabolism

In bacteria (including *Deianiraea*), three enzymatic steps preliminary to the synthesis of methionine lead from aspartate to homoserine (Bender 2012). The first difference of *Deianiraea* respect to other *Rickettsiales* resides in the last step of this pathway, catalysed by the homoserine dehydrogenase. This enzymatic activity is indeed absent in all other *Rickettsiales* except "*Ca.* Arcanobacter lacustris" (Martijn et al. 2015), in which homoserine is likely the precursor of threonine, which cannot be synthesised by any other *Rickettsiales* (including *Deianiraea*). In *Deianiraea*, the homoserine dehydrogenase activity is fused to aspartate kinase (catalysing the first step of homoserine synthesis), similarly to many *Gammaproteobacteria* and *Deltaproteobacteria*, and differently respect to most *Alphaproteobacteria*, possessing separate genes for the two enzymatic activities.

The bacterial pathway specific for methionine biosynthesis from homoserine involves three to four steps (Ferla and Patrick 2014; Bender 2012). The first step is always the activation of homoserine, mostly through acylation from a CoA carrier. This reaction can be catalysed alternatively catalysed by distinct enzymes (MetA or MetX), depending on the phylogenetic lineage. While among *Alphaproteobacteria* the two genes are almost equally abundant (Ferla and Patrick 2014), *Deianiraea* lacks both. Considering that it possesses the enzymatic potential for the following steps, most likely it is able to circumvent such apparent deficiency either by direct import of activated homoserine, or by producing it autonomously through a different reaction. Interestingly, according to blastp search, another ORF (Deia_00649) finds some homology to proteins annotated as "Methionine biosynthesis protein MetW", mostly belonging to other *Alphaproteobacteria* (Supplementary material 15). This protein was suggested to be involved in the acylation of homoserine, possibly in cooperation with MetX (Andersen et al. 1998; Alaminos and Ramos 2001). Thus, ORF Deia_00649 was provisionally included in the *Deianiraea* gene set for methionine biosynthesis, while further analyses would be obviously necessary to clarify this point.

Further synthetic steps involve the sulfurylation of the activated homoserine, to produce homocysteine. Two alternative strategies are employed by bacteria, either a two-step trans-sulfurylation route, catalysed by cystathionine gamma-synthase (MetB) and cystathionine beta-lyase (MetC) using cysteine a sulfhydryl group donor, or a direct sulfurylation from free hydrogen sulfide, catalysed by one among two possible O-acetylhomoserine thiolases (MetY or MetZ). The four proteins (MetB, MetC, MetY, MetZ) are all homologous, and the apparently redundant presence of multiple alternative pathways in the same bacterium is frequent. Two complete genes of belonging to this homology group were identified in *Deianiraea* (Deia_00032, Deia_00704), plus an additional much shorter ORF (Deia_00017), which was thus not considered in the analysis. According to sequence identity, the complete genes were putatively assigned to bacterial MetC and MetY, respectively. Both of these forms are common in *Alphaproteobacteria*. The presence of MetC gene in absence of MetB might be explained by hypothesising a possible bifunctional activity, involving also the gamma-synthase activity (Ferla and Patrick 2014). Among *Rickettsiales*, putative MetC homologues were found also in some *Anaplasmataceae*, in particular several *Wolbachia* (e.g. Foster et al. 2005), "*Ca*. Neoehrlichia lotoris" (Daugherty, S.C et al. unpublished: [LANX01000001]), and few *Anaplasma* species (*A. marginale* and *A. centrale*) (Dark et al. 2009; Herndon et al. 2010).

The last biosynthetic step is the methylation of homocysteine, to obtain methionine. Two distinct homocysteine transmethylases can catalyse this reaction, and the two forms might coexist in the same organism. *Deianiraea* has two different ORFs (Deia_00662, Deia_00705) belonging to the same homology group, the cobalamin-independent homocysteine transmethylase MetE (methyl group donor is 5-methyltetrahydropteroyl-tri-L-glutamate), differently than several other *Alphaproteobacteria*, which more frequently display the cobalamin-dependent form.

Another ORF (Deia_00702), encoding the 5,10-methylenetetrahydrofolate reductase MetF, was taken into account in the analysis. Though not directly involved in methionine biosynthesis, it is as well related to methionine metabolism, and it is absent in all other *Rickettsiales*, thus it was

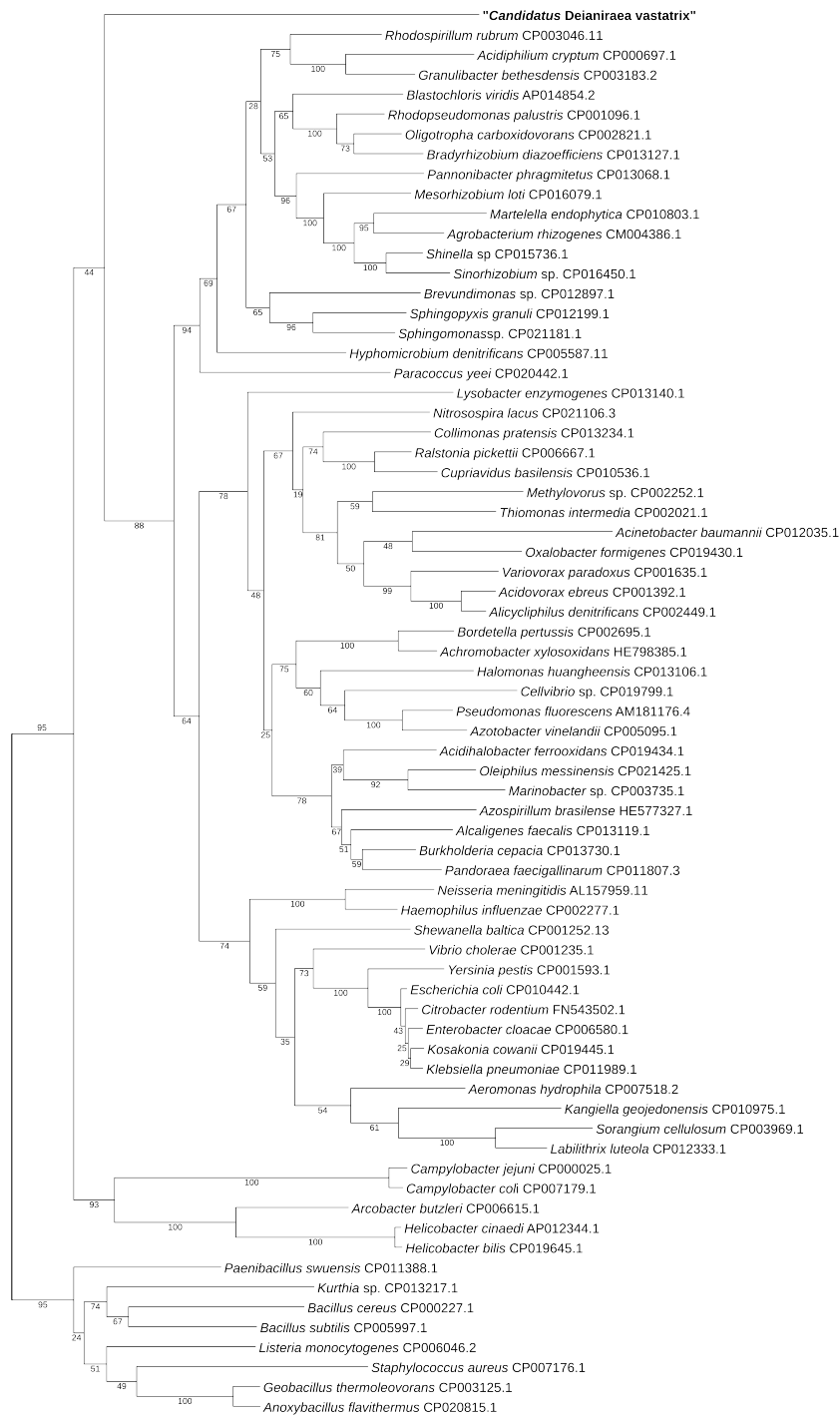considered putatively relevant, and was included in the analysis.

Phylogenetic analyses: Used three genes and 70 organisms, 1,071 sites. For the phylogenomics, used 53 ortholog genes and 9,246 sites.

Figure 13.9 Maximum likelihood tree of the three concatenated genes involved in the methionine biosynthesis with the LG+I+G+F substitution model with 100 bootstrap pseudo-replicates. The position of *Deianiraea* genes is unsupported. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values.

Figure 13.10 Reference maximum likelihood phylogenomic tree of the organisms employed for phylogenesis in **13.9** with the LG+I+G substitution model with 100 bootstrap pseudo-replicates. The five proteobacterial classes (as indicated in the figure) are all monophyletic and highly supported, including the positioning of *Deianiraea* within *Alphaproteobacteria*, which is consistent with the phylogenomic analyses. Scale bar stands for estimated sequence divergence. Number on branches stand for bootstrap values.
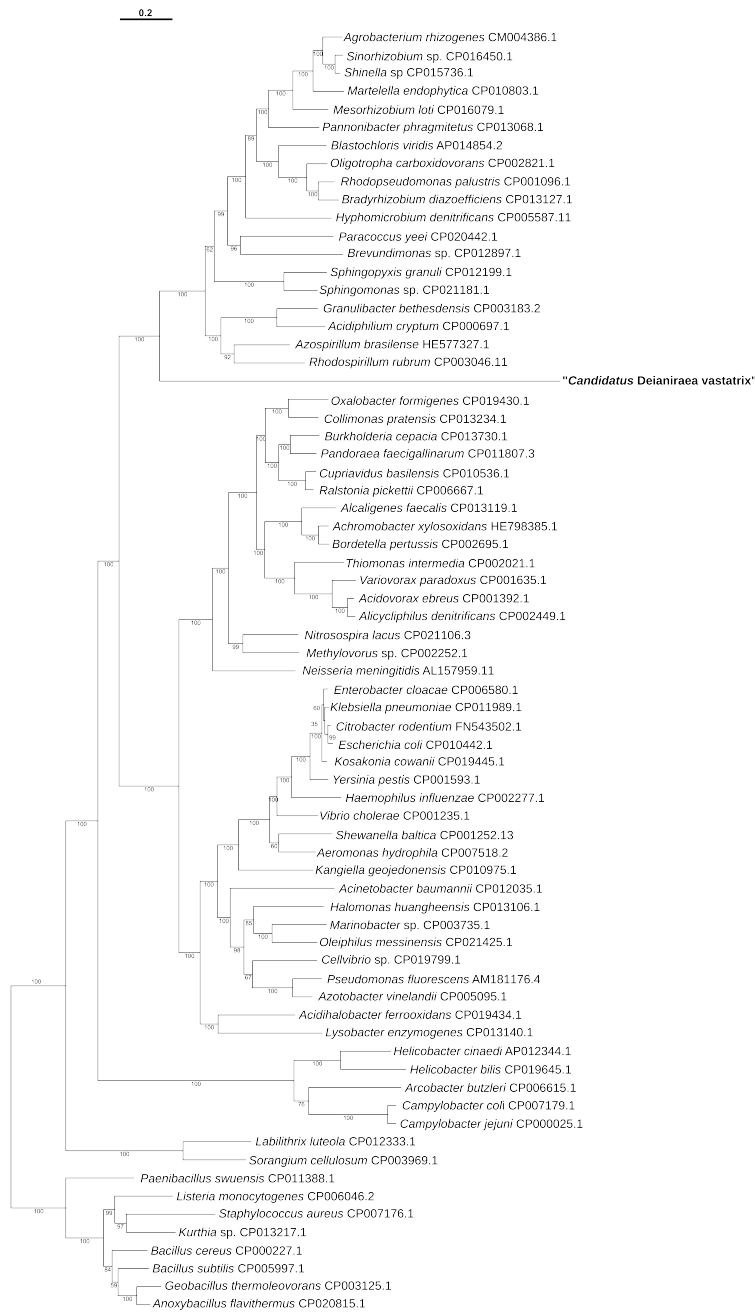
Figure 13.9

0.2

"*Candidatus* Deianiraea vastatrix"
- *Rhodospirillum rubrum* CP003046.11
  - 75
    - *Acidiphilium cryptum* CP000697.1
    - *Granulibacter bethesdensis* CP003183.2 — 100
  - *Blastochloris viridis* AP014854.2
  - 28 / 65
    - *Rhodopseudomonas palustris* CP001096.1
    - *Oligotropha carboxidovorans* CP002821.1 — 100
    - *Bradyrhizobium diazoefficiens* CP013127.1 — 73
  - 53
    - *Pannonibacter phragmitetus* CP013068.1
    - *Mesorhizobium loti* CP016079.1 — 96
    - 100
      - *Martelella endophytica* CP010803.1
      - *Agrobacterium rhizogenes* CM004386.1 — 95
      - *Shinella* sp CP015736.1 — 100
      - *Sinorhizobium* sp. CP016450.1 — 100
  - 67 / 69
    - *Brevundimonas* sp. CP012897.1
    - 65
      - *Sphingopyxis granuli* CP012199.1
      - *Sphingomonas*sp. CP021181.1 — 96
  - 94
    - *Hyphomicrobium denitrificans* CP005587.11
    - *Paracoccus yeei* CP020442.1
- 44
  - *Lysobacter enzymogenes* CP013140.1
  - *Nitrosospira lacus* CP021106.3
  - 67 / 74
    - *Collimonas pratensis* CP013234.1
    - *Ralstonia pickettii* CP006667.1 — 100
    - *Cupriavidus basilensis* CP010536.1
  - 19
    - *Methylovorus* sp. CP002252.1
    - *Thiomonas intermedia* CP002021.1 — 59
  - 81
    - *Acinetobacter baumannii* CP012035.1
    - *Oxalobacter formigenes* CP019430.1 — 48
    - 50
      - *Variovorax paradoxus* CP001635.1
      - *Acidovorax ebreus* CP001392.1 — 99
      - *Alicycliphilus denitrificans* CP002449.1 — 100
  - 78 / 48
    - *Bordetella pertussis* CP002695.1
    - *Achromobacter xylosoxidans* HE798385.1 — 100
    - 75
      - *Halomonas huangheensis* CP013106.1
      - 60
        - *Cellvibrio* sp. CP019799.1
        - *Pseudomonas fluorescens* AM181176.4 — 64
        - *Azotobacter vinelandii* CP005095.1 — 100
  - 25
    - *Acidihalobacter ferrooxidans* CP019434.1
    - 39
      - *Oleiphilus messinensis* CP021425.1
      - *Marinobacter* sp. CP003735.1 — 92
  - 78
    - *Azospirillum brasilense* HE577327.1
    - 67
      - *Alcaligenes faecalis* CP013119.1
      - 51
        - *Burkholderia cepacia* CP013730.1
        - *Pandoraea faecigallinarum* CP011807.3 — 59
- 88 / 64
  - *Neisseria meningitidis* AL157959.11
  - *Haemophilus influenzae* CP002277.1 — 100
  - 74
    - *Shewanella baltica* CP001252.13
    - *Vibrio cholerae* CP001235.1
    - 59 / 73
      - *Yersinia pestis* CP001593.1
      - 100
        - *Escherichia coli* CP010442.1
        - *Citrobacter rodentium* FN543502.1 — 100
        - 43
          - *Enterobacter cloacae* CP006580.1
          - *Kosakonia cowanii* CP019445.1 — 25
          - *Klebsiella pneumoniae* CP011989.1 — 29
    - 35
      - *Aeromonas hydrophila* CP007518.2
      - 54
        - *Kangiella geojedonensis* CP010975.1
        - 61
          - *Sorangium cellulosum* CP003969.1
          - *Labilithrix luteola* CP012333.1 — 100
- 95
  - *Campylobacter jejuni* CP000025.1
  - *Campylobacter coli* CP007179.1 — 100
  - 93
    - *Arcobacter butzleri* CP006615.1
    - 100
      - *Helicobacter cinaedi* AP012344.1
      - *Helicobacter bilis* CP019645.1 — 100
- *Paenibacillus swuensis* CP011388.1
- 95
  - *Kurthia* sp. CP013217.1
  - 74
    - *Bacillus cereus* CP000227.1
    - *Bacillus subtilis* CP005997.1 — 67
  - 24
    - *Listeria monocytogenes* CP006046.2
    - *Staphylococcus aureus* CP007176.1 — 51
  - 49
    - *Geobacillus thermoleovorans* CP003125.1
    - *Anoxybacillus flavithermus* CP020815.1 — 100

Outgroup

Figure 13.10

Alphaproteobacteria

Betaproteobacteria

Gammaproteobacteria

Epsilonproteobacteria

Deltaproteobacteria

Outgroup

Conclusions

Summing up, *Deianiraea* possesses the distinctive ability to synthesise eight amino acids more than any other known *Rickettsiales* bacterium. The respective biosynthetic enzymes can be grouped into five pathways. For each pathway, no evidence of recent HGT was found. Indeed, all the genes are highly divergent from any other organism (mostly <50% identity), and no relevant trend of higher similarity to very distantly related organisms was observed, since the identity values respect to best hits on the entire nr protein database, on *Proteobacteria*, and *Alphaproteobacteria* (which would be the closest relatives in the absence of HGT) are all comparable (Supplementary material 15). Moreover, for no pathway the combined result of phylogenetic analyses and GC content and CAI deviation tests showed a clear evidence of recent HGT. For what concerns phylogenetic analyses, in most cases, the whole pathway-specific dataset, lacked a sufficient phylogenetic signal, as confirmed by the highly significant deviations respect to organismal phylogenies in SH tests. Thus, it was not possible to reconstruct confidently the position of *Deianiraea* genes. Nevertheless, given that they were always found with a very long branch (consistent with sequence divergence), and that they were never found associated to a recent (family level or lower) group of bacteria, the conclusion that those genes did not undergo any recent HGT was indirectly confirmed. Therefore, the hypothesis that they could be inherited directly from *Rickettsiales* ancestor resulted fortified. Based on these lines of reasoning, further hypotheses on the evolutionary trends of *Deianiraea* and *Rickettsiales* in general were fittingly drawn.

References

Alaminos MM, Ramos JL. The methionine biosynthetic pathway from homoserine in *Pseudomonas putida* involves the metW, metX, metZ, metH and metE gene products. *Arch. Microbiol.* **176**, 151-154 (2001)

Andersen GL, Beattie GA, Lindow SE. Molecular characterization and sequence of a methionine biosynthetic locus from *Pseudomonas syringae*. *J. Bacteriol.* **180**, 4497-4507 (1998)

Anantharaman K *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Comm.* **7**, 13219 (2016)

Bender DA. *Amino Acid Metabolism*, Third Edition (2012) John Wiley & Sons, Ltd

Brown, C. T. *et al.* unpublished: [KKT47881.1]

Dark MJ *et al*. Conservation in the face of diversity: multistrain analysis of an intracellular bacterium. *BMC Genomics*. **10**, 16 (2009)

Daugherty SC *et al.* unpublished: [LANX01000001]

Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16(1): 157.

Felsheim RF *et al.* unpublished. [NZ_CP009217])

Ferla MP, Patrick WM. Bacterial methionine biosynthesis. *Microbiology* **60**, 1571-1584 (2014)

Foster J, *et al.* The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode *PLoS Biol.* **3**, e121 (2005)

Herndon DR, Palmer GH, Shkap V, Knowles DP Jr, Brayton KA. Complete genome sequence of *Anaplasma marginale* subsp. *centrale*. *Bacteriol*. **192**, 379-380 (2010)

Jima DD *et al.* Whole-genome sequence of "*Candidatus* Rickettsia asemboensis" strain NMRCii,

isolated from fleas of western Kenya. *Genome Announc.* **3**, pii: e00018-15 (2015)

Kwan JC, Schmidt EW. Bacterial endosymbiosis in a chordate host: long-term co-evolution and conservation of secondary metabolism. *PLoS One* **8**, e80822 (2013)

Lawson CE *et al.* Metabolic network analysis reveals microbial community interactions in anammox granules. *Nat. Comm.* **8**, 15416 (2017)

Makino Y *et al.* An archaeal ADP-dependent serine kinase involved in cysteine biosynthesis and serine metabolism *Nat. Comm.* **7**, 13446 (2016)

Marchler-Bauer A *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222-D226 (2015)

Martijn J *et al.* Single-cell genomics of a rare environmental alphaproteobacterium provides unique insights into *Rickettsiaceae* evolution. *ISME J.* **9**, 2373-2385 (2015)

McKinlay JB, Harwood CS. Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. *Proc. Nat. Acad. Sci. USA* **107**, 11669–11675 (2010)

Mediannikov O *et al.* High quality draft genome sequence and description of *Occidentia massiliensis* gen. nov., sp. nov., a new member of the family *Rickettsiaceae. Stand. Genomic Sci.* **9**, 9 (2014)

Pittard J, Yang J. Biosynthesis of the aromatic amino acids. *EcoSal Plus* **3**, (2008) doi:10.1128/ecosalplus.3.6.1.8.

Probst A.J *et al.* Depth-based differentiation of microbial function through sediment-hosted aquifers and enrichment of novel symbionts in the deep terrestrial subsurface. *Nat. Microbiol.* **3**, 328-336 (2018)

Risso C, Van Dien SJ, Orloff A, Lovley DR, Coppi MV. Elucidation of an alternate isoleucine biosynthesis pathway in *Geobacter sulfurreducens*. *J. Bacteriol.* **190**, 2266–2274 (2008)

Salmon KA, Yang CR, Hatfield GW. Biosynthesis and regulation of the branched-chain amino

acids. *EcoSal Plus* **2** (2006); doi:10.1128/ecosalplus.3.6.1.5.

Tang KH, Feng X, Tang YJ, Blankenship RE. Carbohydrate metabolism and carbon fixation in *Roseobacter denitrificans* och114. *PLoS One* **4**, e7233 (2009)

Tully BJ, Sachdeva R, Graham ED, Heidelberg JF. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ.* **5**, e3558 (2017)

Winkler ME, Ramos-Montañez S. Biosynthesis of histidine. *EcoSal Plus* **3** (2009) doi: 10.1128/ecosalplus.3.6.1.9.

Xu H *et al.* Isoleucine biosynthesis in *Leptospira interrogans* serotype lai strain 56601 proceeds via a threonine-independent pathway. *J. Bact.* **186**, 5400-5409 (2004)