Supplementary Information Summary

The Supplementary Information contains the **Supplementary Text,** which describes the source of *A. lacustris*' name, the **Supplementary Methods,** describing the details of all performed phylogenetic analyses, and also the **Supplementary Figures S1-S6** and **Supplementary Tables S1-S6**, which provide the full, uncollapsed phylogenies for all phylogenetic analyses performed in this study**.**

Supplementary Text

The name '*Candidatus* Arcanobacter lacustris' has been derived from the Latin words "arcanus", meaning mysterious or secret, and "lacustris", adjective from "lacus", the lake. The name reflects the mysterious nature of this bacterium (the uncertainty in lifestyle and low abundance) and the type of environment the single cell was isolated from.

Supplementary Methods

Detailed phylogenetic analyses

Panorthologs

The genome and corresponding protein coding sequences (CDSs) from 20 Rickettsiales and 4 non-Rickettsiales Alphaproteobacteria were downloaded from NCBI RefSeq (accession numbers: Supplementary Table 6). Sequences from these 24 genomes and from *A. lacustris* were clustered in ortholog clusters as described in (Guy *et al.*, 2013), These clusters are referred to as 'rickCOGs'.

From the 12,197 resulting clusters, 64 panorthologs (i.e. genes present in exactly one copy in all taxa) were extracted ('64-panorthologs' dataset). In addition, a further 65 rickCOGs were extracted, which are panorthologs in all taxa but missing in *A. lacustris* and added to the previous dataset ('129-panorthologs' dataset).

All panorthologs were subsequently aligned individually and trimmed as described above. All 129 clusters were annotated by using the aligned panorthologs as a query in a PSI-BLAST search versus Swiss-Prot. Alignments were concatenated into a final alignment consisting of 40,298 and 16,095 sites for the '129-panorthologs' and the '64-panorthologs' datasets, respectively. Genome phylogenies were then inferred with maximum likelihood and Bayesian methods as described above. Finally, for the '64-panorthologs' dataset, maximum-likelihood support for the Bayesian phylogeny was inferred by running 100 rapid bootstraps with RAxML.

Small subunit ribosomal DNA

From the SILVA database (Quast *et al.*, 2012), 64 representative Rickettsiales and free-living Alphaproteobacteria SSU rRNA sequences were retrieved. Sequences were selected so that as many Rickettsiales taxonomic groups as possible were represented by named strains and met the following criteria: sequence length ≥ 1200 bp, sequence quality ≥ 90, alignment quality ≥ 90 and pintail quality ≥ 90. The *A. lacustris* SSU rDNA sequence, predicted by RNAmmer 1.2, was added and the resulting dataset was aligned with SINA v1.2.11 (Pruesse *et al.*, 2012) using all sequences classified as Alphaproteobacteria as reference, trimmed and used to infer maximum likelihood phylogeny.

Exploration of *A. lacustris* phylogenetic diversity

The *A. lacustris* SSU sequence was used as a query in a standard BLASTN (Altschul *et al.*, 1990) search of the NCBI-nt database. Based on a subsequent distance tree analysis, the twelve most closely related sequences were retrieved and added to the SSU dataset. Finally, 65 sequence reads retrieved from the SRA (34) and VAMPs (31) databases (see Materials and Methods) with ≥ 95% sequence identity were clustered into 10 OTUs with UCLUST 1.2.22 (Edgar, 2010), using a 97% sequence identity threshold. Seed sequences were extracted and subsequently checked for primer adapter sequences with BLASTN search of the NCBI UniVec databases. Primer adapters were clipped if necessary and the clipped sequences were added to the dataset. To this initial dataset consisting of twelve most closely related BLASTN hits, seven SRA reads and three VAMPS reads, the 64 Rickettsiales used in the SSU rDNA phylogeny were added. The dataset was then aligned with SINA v1.2.11, using all sequences classified as Alphaproteobacteria in the SILVA database as reference. The alignment was then trimmed and an initial maximum likelihood phylogeny was inferred. From the obtained phylogeny, five BLASTN hits that were too distant from *A. lacustris* were manually removed. The remaining sequences were then realigned with SINA, trimmed, and a maximum likelihood phylogeny was inferred. All internal branches with ≤ 30 bootstrap support were collapsed with Newick-Utilities 1.6 (Junier & Zdobnov, 2010).

Flagella

Protein sequences of 14 core flagellar genes in 63 bacteria (Sassera *et al.*, 2011) were complemented with orthologs from *A. lacustris* and several newly sequenced Alphaproteobacteria (*Holospora undulata* HU1, *Candidatus* Odyssella thessalonicensis L13, *Kordiimonas gwangyangensis* DSM 19435, *Micavibrio aeruginosavorus* ARL-13, *Polymorphum gilvum* SL003B-26A1, *Kiloniella laminariae* DSM 19542 and *Magnetococcus marinus* MC-1; for accession numbers see Supplementary Table 1). Orthologs were identified in three steps. First, a PSI-BLAST search against the proteomes of *A. lacustris* and the new Alphaproteobacteria was done using the gene alignments from Sassera et al. as a query. All core flagellar genes had hits to all new taxa had hits, except *A. lacustris* (8 hits out of 14 genes), and *Holospora* (2 hits out of 14 genes). Because only two genes (*fliI* and *motB*) had hits in *Holospora, Holospora* was omitted from any further analyses. Hits were added to the base set and single gene phylogenies were inferred with FastTreeMP 2.1 (Price *et al.*, 2010). Finally, orthologs were distinguished from distant paralogs by manual inspection of the trees. In cases where orthologs in the new Alphaproteobacteria could not be clearly identified from the single gene trees, identification was done based on gene synteny and NCBI RefSeq annotation.

The final dataset comprised 70 taxa, with *Polymorphum gilvum* and *Bradyrhizobium japonicum* having each two complete sets of genes. Two concatenated alignments were prepared: one contained all 14 core flagella (with six flagella missing in *A. lacustris*), and another only contained the eight core flagella that were present in *A. lacustris*. Both alignments were subsequently trimmed and used to infer maximum likelihood and Bayesian phylogenies.

ATP/ADP-translocase

The ATP/ADP translocase experimentally characterized in *Rickettsia prowazekii* Madrid E (Tlc1) (Audia & Winkler, 2006) was used as a query for a BLASTP search to identify homologs in *A. lacustris* and for a PSI-BLAST search with one iteration to identify homologs in the nr database. PSI-BLAST hits from a selected set of taxa were extracted and pooled with the *A. lacustris* homolog. Taxa selection was done so that the number of sequences was significantly reduced, but all diverse major taxonomic groups (Rickettsiales, Rhizobiales, Chlamydiales, Deltaproteobacteria, and Gammaproteobacteria) and their subclades were still represented. Sequences were aligned and the alignment was visually inspected with Seaview (Gouy *et al.*, 2010). Poorly aligned and highly diverging sequences (based on their branch lengths in a neighbor joining tree) were manually removed. The remaining alignment was subsequently trimmed and used to calculate a BI phylogeny.

T4SS

Protein sequences from components of type IVA classified T4SS in *R. prowazekii* Madrid E, *R. bellii* RML369-C, *O. tsutsugamushi* Ikeda, *N. risticii* Illinois, *Wolbachia* endosymbiont of *Drosophila melanogaster, Wolbachia* endosymbiont TRS of *Brugia malayi, E. canis* Jake, *E. chaffeensis* Arkansas, *A. phagocytophilum* HZ, *A. marginale* Florida, *A. tumefaciens* C58 plasmid Ti, *X. citri* plasmid pXcB, *B. henselae* Houston-1 and *L. pneumophila* Lens were downloaded from the SecReT4 database (Bi *et al.*, 2013)(accession numbers: Supplementary Table 1). Taxon selection was guided by Figure 2 of (Gillespie *et al.*, 2010). Gene clusters were created for *virB4, virB8, virB9, virB10* and *virB11* based on SecReT4 annotation, complemented with corresponding homologs in *A. lacustris* and used to infer single gene ML phylogenies. Paralogs were distinguished from orthologs by manual inspection of the resulting trees and evaluation of gene synteny and removed. The resulting orthologous groups were then used to prepare a trimmed concatenated alignment from which ML phylogenies were inferred.

Supplementary References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**:403–410.

Audia JP, Winkler HH. (2006). Study of the Five Rickettsia prowazekii Proteins Annotated as ATP/ADP Translocases (Tlc): Only Tlc1 Transports ATP/ADP, While Tlc4 and Tlc5 Transport Other Ribonucleotides. *J Bacteriol* **188**:6261–6268.

Bi D, Liu L, Tai C, Deng Z, Rajakumar K, Ou H-Y. (2013). SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res* **41**:D660–D665.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**:2460–2461.

Gillespie JJ, Brayton KA, Williams KP, Diaz MAQ, Brown WC, Azad AF, *et al.* (2010). Phylogenomics Reveals a Diverse Rickettsiales Type IV Secretion System. *Infect Immun* **78**:1809–1823.

Gouy M, Guindon S, Gascuel O. (2010). SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* **27**:221–224.

Guy L, Nystedt B, Toft C, Zaremba-Niedzwiedzka K, Berglund EC, Granberg F, *et al.* (2013). A Gene Transfer Agent and a Dynamic Repertoire of Secretion Systems Hold the Keys to the Explosive Radiation of the Emerging Pathogen Bartonella. *PLoS Genet* **9**:e1003393.

Junier T, Zdobnov EM. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* **26**:1669–1670.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**:e9490.

Pruesse E, Peplies J, Glöckner FO. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**:1823–1829.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**:D590–D596.

Sassera D, Lo N, Epis S, D'Auria G, Montagna M, Comandatore F, *et al.* (2011). Phylogenomic Evidence for the Presence of a Flagellum and cbb3 Oxidase in the Free-Living Mitochondrial Ancestor. *Mol Biol Evol* **28**:3285–3296.

Supplementary Table Titles

Supplementary Table S1

Annotation of the 129 Rickettsiales panorthologous marker genes that were used for the phylogenomic analyses and completeness estimate and occurrence in the A. lacustris single cell draft genome. *Provided as Excel file.*

Supplementary Table S2

Annotated rickCOGs either unique to *A. lacustris* or uniquely shared with Rickettsiaceae, Anaplasmataceae, 'Midichloriaceae' or 'Holosporaceae'. Ortholog clusters mentioned in the main text are given in bold italics and colored according to putative function: mobile element (green), heme metabolism (brown), toxin-antitoxin systems (purple), chemotaxis (blue) and phage related proteins (black). *Provided as Excel file*

Supplementary Table S3

List of putative T4SS effector proteins in *A. lacustris* based on BLASTP versus SecReT4 effectors and experimentally verified effectors databases. *Provided as Excel file*

Supplementary Table S4

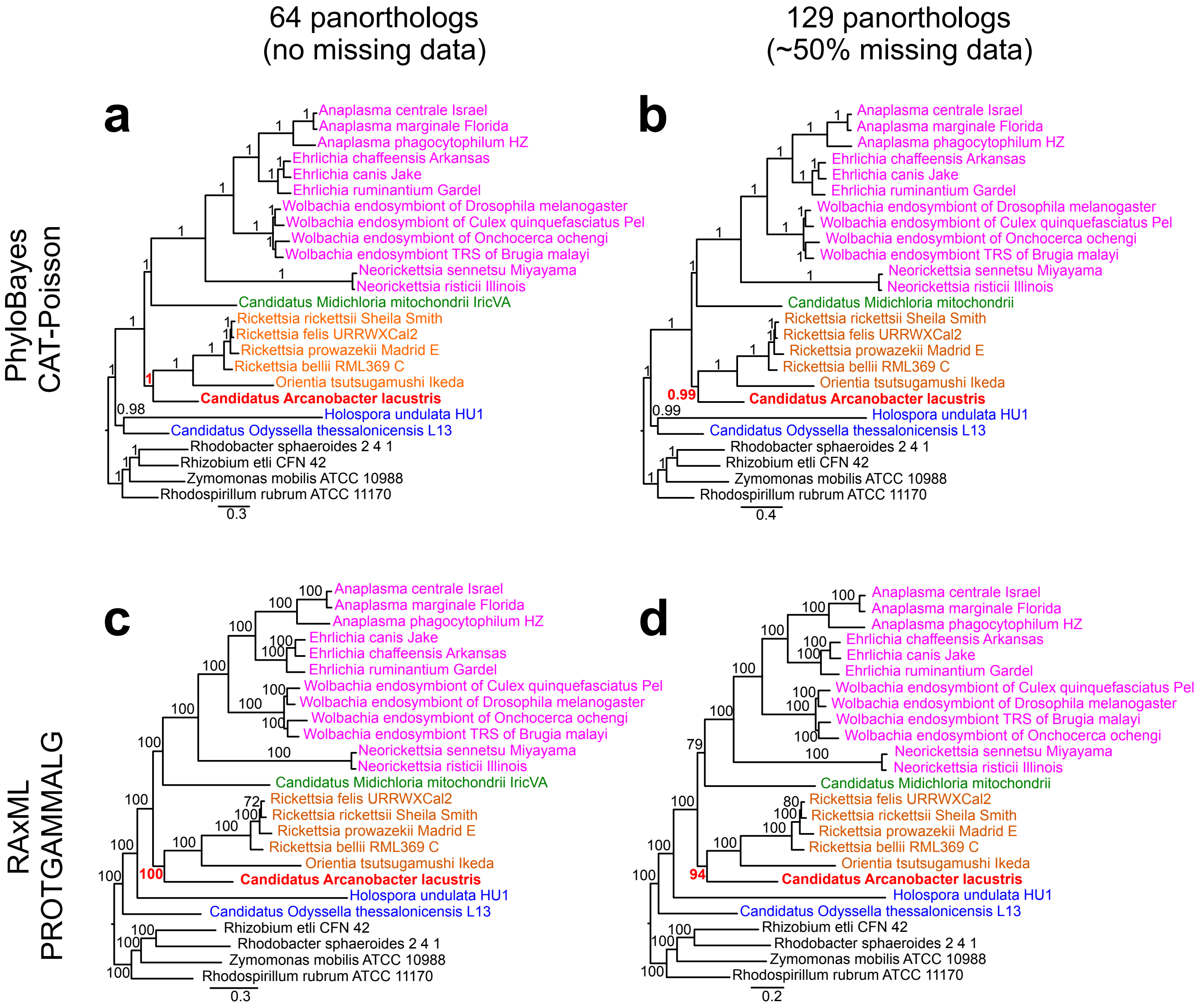
Accessions, databases and types of environment for all *A. lacustris* related sequences that were incorporated in Figure 6 and Supplementary Figure 6. *Provided as Excel file*

Supplementary Table S5

Accession numbers, project titles and type of environment for the SSU amplicon datasets available in SRA and VAMPS that contained reads with high identity to *A. lacustris* SSU. *Provided as Excel file*

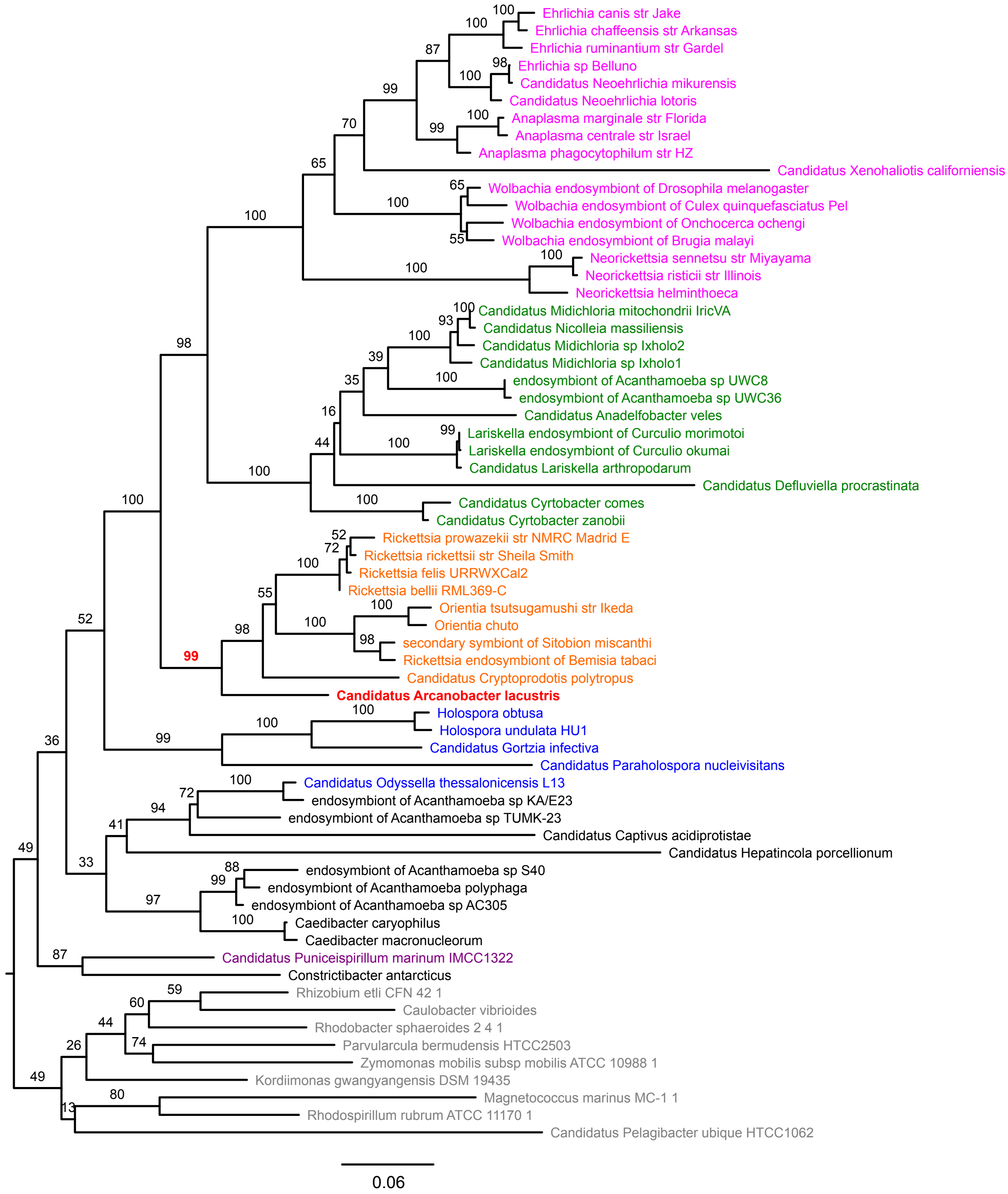
Supplementary Table S6

Accession numbers of publicly available genomes used in this study. *Provided as Excel file*


64 panorthologs
(no missing data)
129 panorthologs
(~50% missing data)
PhyloBayes
CAT-Poisson
RAxML
PROTGAMMALG
a
b
c
d
Anaplasma centrale Israel
Anaplasma marginale Florida
Anaplasma phagocytophilum HZ
Ehrlichia chaffeensis Arkansas
Ehrlichia canis Jake
Ehrlichia ruminantium Gardel
Wolbachia endosymbiont of Drosophila melanogaster
Wolbachia endosymbiont of Culex quinquefasciatus Pel
Wolbachia endosymbiont of Onchocerca ochengi
Wolbachia endosymbiont TRS of Brugia malayi
Neorickettsia sennetsu Miyayama
Neorickettsia risticii Illinois
Candidatus Midichloria mitochondrii IricVA
Candidatus Midichloria mitochondrii
Rickettsia rickettsii Sheila Smith
Rickettsia felis URRWXCal2
Rickettsia prowazekii Madrid E
Rickettsia bellii RML369 C
Orientia tsutsugamushi Ikeda
Candidatus Arcanobacter lacustris
Holospora undulata HU1
Candidatus Odyssella thessalonicensis L13
Rhodobacter sphaeroides 2 4 1
Rhizobium etli CFN 42
Zymomonas mobilis ATCC 10988
Rhodospirillum rubrum ATCC 11170
0.98
0.99
1
100
72
80
79
94
0.3
0.4
0.3
0.2

Supplementary Figure S1

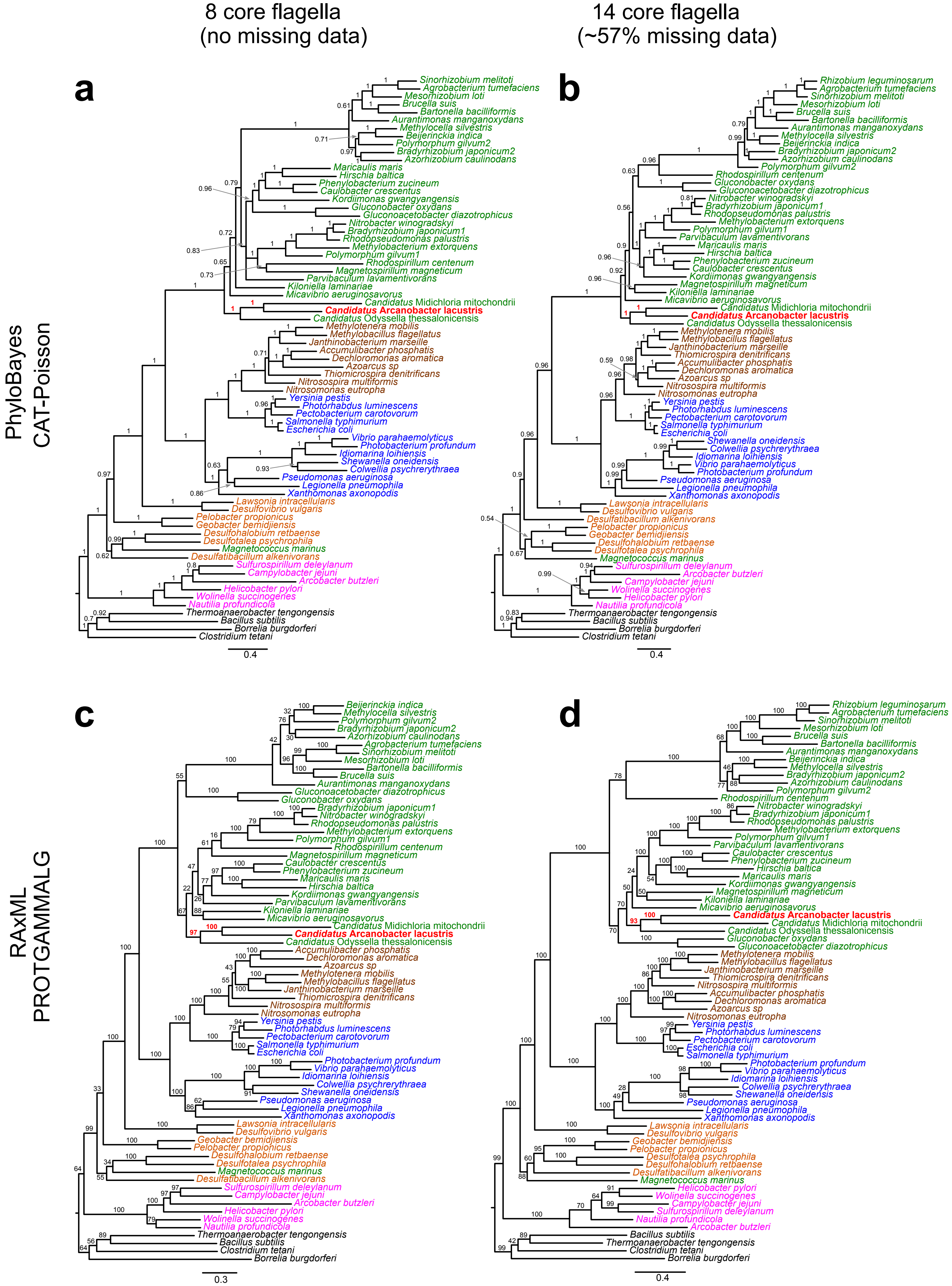
Robustness of phylogenetic placement of *A. lacustris* to tree inference method and missing data in concatenated panorthologs alignments. Phylogenies were either inferred with PhyloBayes (model: CAT-Poisson) **(a,b)** or RAxML (model: Γ+LG) **(c,d)** from either the concatenated alignment of 64 panorthologs (no missing data) **(a,c)** or the concatenated alignment of 129 panorthologs (~50% missing data) **(b,d)**. Taxa are colored according to their taxonomic affiliation at the family level: Anaplasmataceae: pink, 'Midichloriaceae': green, Rickettsiaceae: orange and 'Holosporaceae': blue. Outgroup taxa are in black. *A. lacustris* and branch support values of interest are shown in bold red.



Ehrlichia canis str Jake
Ehrlichia chaffeensis str Arkansas
Ehrlichia ruminantium str Gardel
Ehrlichia sp Belluno
Candidatus Neoehrlichia mikurensis
Candidatus Neoehrlichia lotoris
Anaplasma marginale str Florida
Anaplasma centrale str Israel
Anaplasma phagocytophilum str HZ
Candidatus Xenohaliotis californiensis
Wolbachia endosymbiont of Drosophila melanogaster
Wolbachia endosymbiont of Culex quinquefasciatus Pel
Wolbachia endosymbiont of Onchocerca ochengi
Wolbachia endosymbiont of Brugia malayi
Neorickettsia sennetsu str Miyayama
Neorickettsia risticii str Illinois
Neorickettsia helminthoeca
Candidatus Midichloria mitochondrii IricVA
Candidatus Nicolleia massiliensis
Candidatus Midichloria sp Ixholo2
Candidatus Midichloria sp Ixholo1
endosymbiont of Acanthamoeba sp UWC8
endosymbiont of Acanthamoeba sp UWC36
Candidatus Anadelfobacter veles
Lariskella endosymbiont of Curculio morimotoi
Lariskella endosymbiont of Curculio okumai
Candidatus Lariskella arthropodarum
Candidatus Defluviella procrastinata
Candidatus Cyrtobacter comes
Candidatus Cyrtobacter zanobii
Rickettsia prowazekii str NMRC Madrid E
Rickettsia rickettsii str Sheila Smith
Rickettsia felis URRWXCal2
Rickettsia bellii RML369-C
Orientia tsutsugamushi str Ikeda
Orientia chuto
secondary symbiont of Sitobion miscanthi
Rickettsia endosymbiont of Bemisia tabaci
Candidatus Cryptoprodotis polytropus
Candidatus Arcanobacter lacustris
Holospora obtusa
Holospora undulata HU1
Candidatus Gortzia infectiva
Candidatus Paraholospora nucleivisitans
Candidatus Odyssella thessalonicensis L13
endosymbiont of Acanthamoeba sp KA/E23
endosymbiont of Acanthamoeba sp TUMK-23
Candidatus Captivus acidiprotistae
Candidatus Hepatincola porcellionum
endosymbiont of Acanthamoeba sp S40
endosymbiont of Acanthamoeba polyphaga
endosymbiont of Acanthamoeba sp AC305
Caedibacter caryophilus
Caedibacter macronucleorum
Candidatus Puniceispirillum marinum IMCC1322
Constrictibacter antarcticus
Rhizobium etli CFN 42 1
Caulobacter vibrioides
Rhodobacter sphaeroides 2 4 1
Parvularcula bermudensis HTCC2503
Zymomonas mobilis subsp mobilis ATCC 10988 1
Kordiimonas gwangyangensis DSM 19435
Magnetococcus marinus MC-1 1
Rhodospirillum rubrum ATCC 11170 1
Candidatus Pelagibacter ubique HTCC1062
100
100
87
98
100
99
99
100
70
65
65
100
55
100
100
100
98
100
93
100
39
100
35
16
44
99
100
100
100
100
52
72
100
55
100
100
98
98
99
52
100
99
100
72
100
94
41
36
88
99
33
97
100
49
87
59
60
44
74
26
49
80
13
0.06

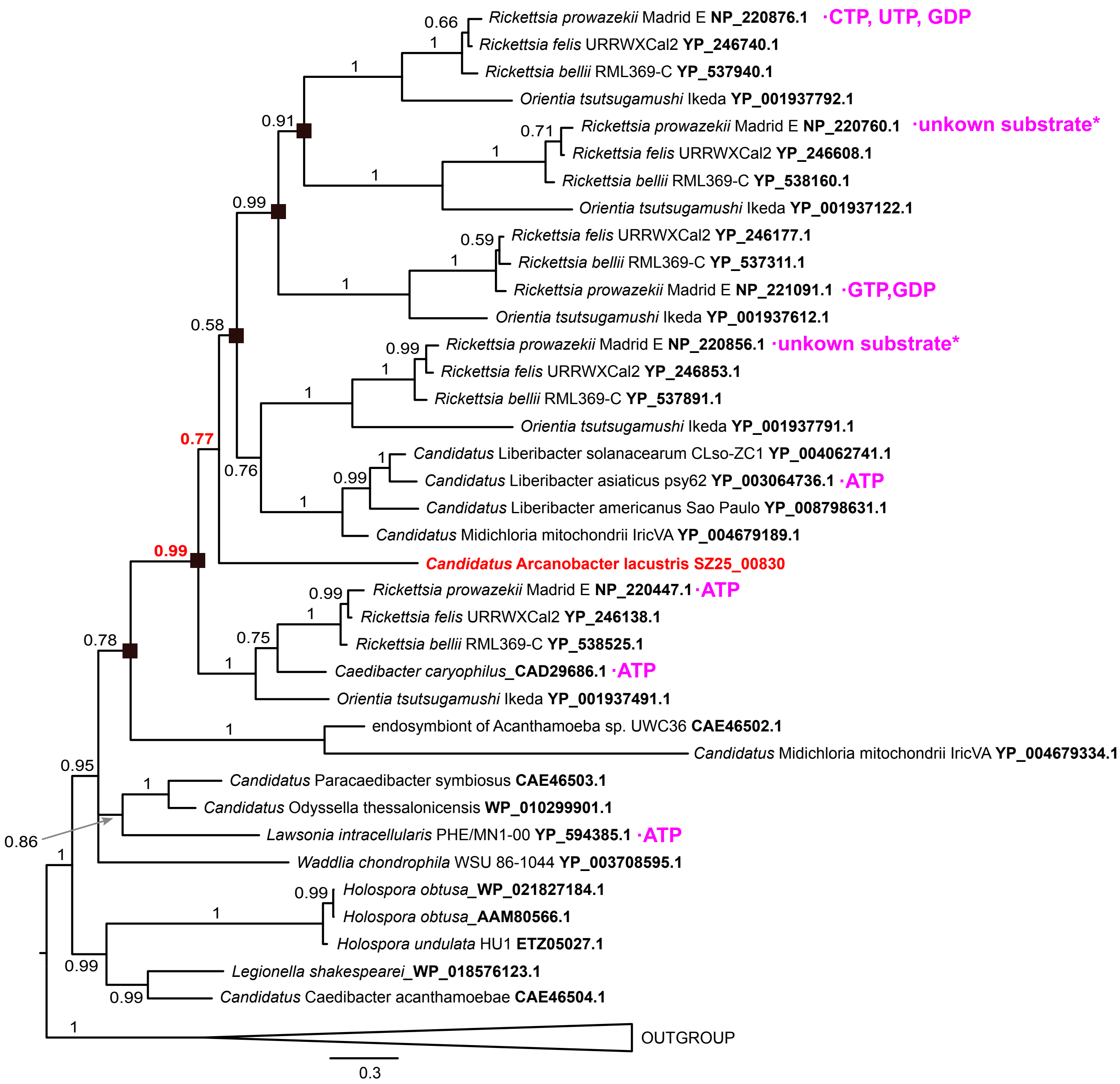
Supplementary Figure S2

Phylogenetic tree based on SSU rRNA gene. Tree topology and bootstrap support values were inferred with RAxML (model: Γ+LG). This is the full tree of the phylogenetic analysis shown in Figure 2b. The *A. lacustris* leaf and branch support values on branches of interest have been marked with bold red. Other taxa names were colored as follows: Anaplasmataceae: pink, 'Midichloriaceae': green, Rickettsiaceae: orange, 'Holosporaceae': blue, SAR116: purple, 'Free-living Alphaproteobacteria': grey.



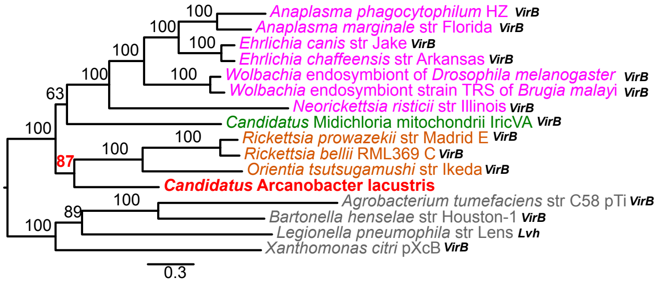
Supplementary Figure S3

Robustness of phylogenetic analyses to tree inference method and missing data in concatenated core flagellar alignments. Phylogenies were either inferred with PhyloBayes (model: CAT-Poisson) **(a,b)** or RAxML (model: Γ+LG) **(c,d)** from either the concatenated alignment of 8 core flagella **(a,c)** or the concatenated alignment of 14 core flagella **(b,d)**. Taxa are colored according to their taxonomic affiliation at the class level: Alphaproteobacteria: green, Betaproteobacteria: brown, Gammaproteobacteria: blue, Deltaproteobacteria: orange and Epsilonbacteria: pink. Outgroup taxa are in black. *A. lacustris* and branch support values of interest are shown in bold red.


0.66
Rickettsia prowazekii Madrid E NP_220876.1 ·CTP, UTP, GDP
1
Rickettsia felis URRWXCal2 YP_246740.1
1
Rickettsia bellii RML369-C YP_537940.1
Orientia tsutsugamushi Ikeda YP_001937792.1
0.91
0.71
Rickettsia prowazekii Madrid E NP_220760.1 ·unkown substrate*
1
Rickettsia felis URRWXCal2 YP_246608.1
1
Rickettsia bellii RML369-C YP_538160.1
Orientia tsutsugamushi Ikeda YP_001937122.1
0.99
0.59
Rickettsia felis URRWXCal2 YP_246177.1
1
Rickettsia bellii RML369-C YP_537311.1
1
Rickettsia prowazekii Madrid E NP_221091.1 ·GTP,GDP
Orientia tsutsugamushi Ikeda YP_001937612.1
0.58
0.99
Rickettsia prowazekii Madrid E NP_220856.1 ·unkown substrate*
1
Rickettsia felis URRWXCal2 YP_246853.1
1
Rickettsia bellii RML369-C YP_537891.1
Orientia tsutsugamushi Ikeda YP_001937791.1
0.77
1
Candidatus Liberibacter solanacearum CLso-ZC1 YP_004062741.1
0.76
0.99
Candidatus Liberibacter asiaticus psy62 YP_003064736.1 ·ATP
1
Candidatus Liberibacter americanus Sao Paulo YP_008798631.1
Candidatus Midichloria mitochondrii IricVA YP_004679189.1
0.99
Candidatus Arcanobacter lacustris SZ25_00830
0.99
Rickettsia prowazekii Madrid E NP_220447.1 ·ATP
1
Rickettsia felis URRWXCal2 YP_246138.1
0.78
0.75
Rickettsia bellii RML369-C YP_538525.1
1
Caedibacter caryophilus_CAD29686.1 ·ATP
Orientia tsutsugamushi Ikeda YP_001937491.1
endosymbiont of Acanthamoeba sp. UWC36 CAE46502.1
1
Candidatus Midichloria mitochondrii IricVA YP_004679334.1
0.95
1
Candidatus Paracaedibacter symbiosus CAE46503.1
Candidatus Odyssella thessalonicensis WP_010299901.1
0.86
Lawsonia intracellularis PHE/MN1-00 YP_594385.1 ·ATP
1
Waddlia chondrophila WSU 86-1044 YP_003708595.1
0.99
Holospora obtusa_WP_021827184.1
1
Holospora obtusa_AAM80566.1
Holospora undulata HU1 ETZ05027.1
0.99
Legionella shakespearei_WP_018576123.1
0.99
Candidatus Caedibacter acanthamoebae CAE46504.1
1
OUTGROUP
0.3

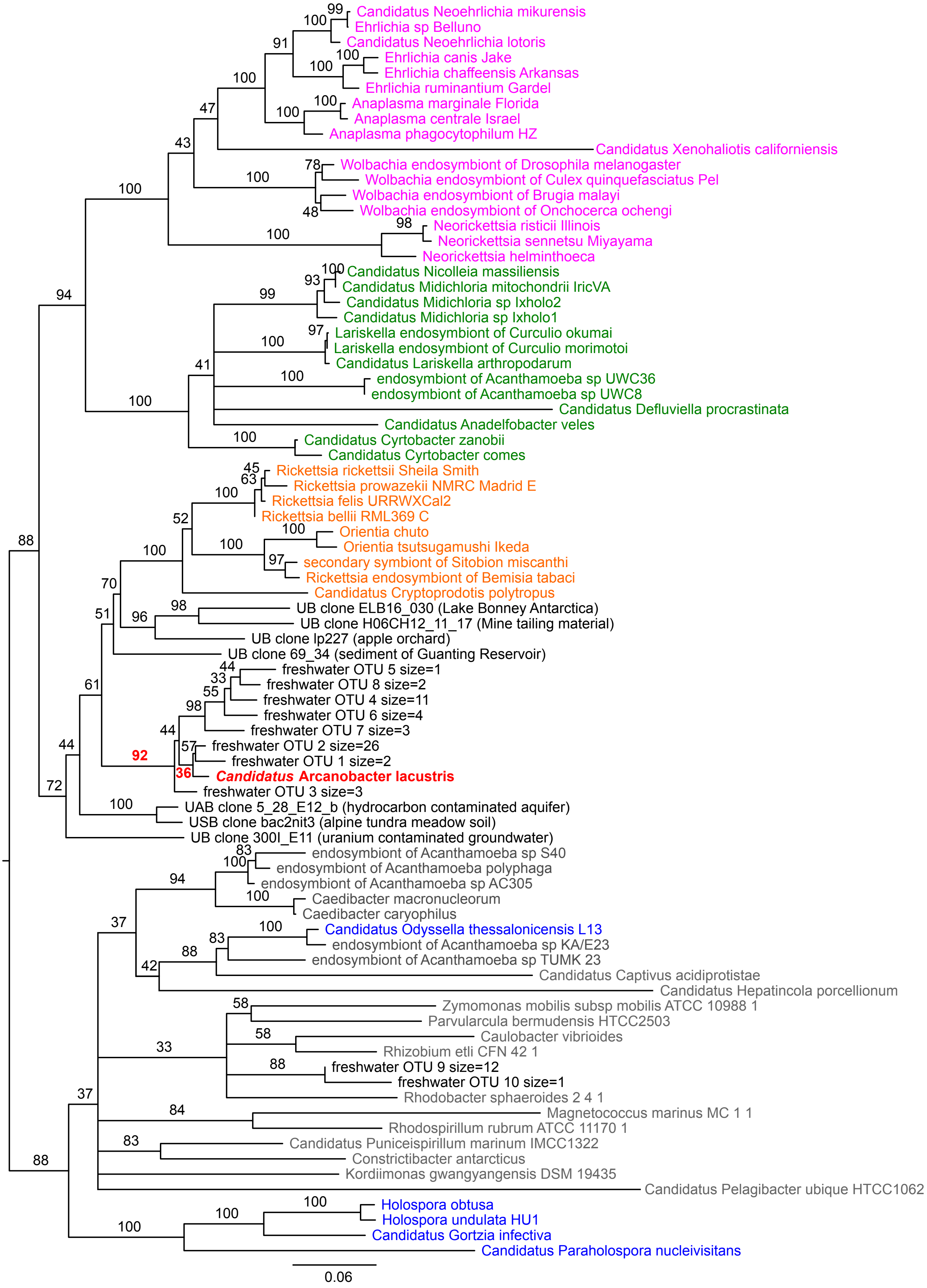
Supplementary Figure S4

Phylogenetic analysis of bacterial ATP,ADP-translocase proteins. Tree was inferred with PhyloBayes (model: CAT-Poisson). Inferred gene duplication events are indicated with black squares on the respective nodes. Accession numbers are given in bold. For proteins that have been functionally characterized, the substrate is given in bold pink. (*) Substrates for these proteins are not known, but shown not to be ATP, ADP, AMP, CTP, CMP, GTP, GDP, GMP, UTP or UMP. *A. lacustris* and branch support values of interest are shown in bold red. All Chlamydiales homologs are used as outgroup and have been collapsed to improve readability.


100
100
100
100
100
63
100
100
100
87
89
100
100
Anaplasma phagocytophilum HZ VirB
Anaplasma marginale str Florida VirB
Ehrlichia canis str Jake VirB
Ehrlichia chaffeensis str Arkansas VirB
Wolbachia endosymbiont of Drosophila melanogaster VirB
Wolbachia endosymbiont strain TRS of Brugia malayi VirB
Neorickettsia risticii str Illinois VirB
Candidatus Midichloria mitochondrii IricVA VirB
Rickettsia prowazekii str Madrid E VirB
Rickettsia bellii RML369 C VirB
Orientia tsutsugamushi str Ikeda VirB
Candidatus Arcanobacter lacustris
Agrobacterium tumefaciens str C58 pTi VirB
Bartonella henselae str Houston-1 VirB
Legionella pneumophila str Lens Lvh
Xanthomonas citri pXcB VirB
0.3

Supplementary Figure S5

Phylogenetic analysis based on the concatenated alignment of conserved T4SS proteins VirB4, VirB8, VirB9, VirB10 and VirB11. Tree was inferred with RAxML (model: Γ+LG). Taxa are colored according to their taxonomic affiliation at the family level: Anaplasmataceae: pink, 'Midichloriaceae': green and Rickettsiaceae: orange. Outgroup taxa are in grey. Names of the taxa specific T4SS, based on SecReT4 classification, are given in bold italics. *A. lacustris* and branch support values of interest are shown in bold red.


99
100
Candidatus Neoehrlichia mikurensis
Ehrlichia sp Belluno
91
Candidatus Neoehrlichia lotoris
100
Ehrlichia canis Jake
100
Ehrlichia chaffeensis Arkansas
100
Ehrlichia ruminantium Gardel
47
100
Anaplasma marginale Florida
100
Anaplasma centrale Israel
43
Anaplasma phagocytophilum HZ
Candidatus Xenohaliotis californiensis
78
Wolbachia endosymbiont of Drosophila melanogaster
100
100
Wolbachia endosymbiont of Culex quinquefasciatus Pel
Wolbachia endosymbiont of Brugia malayi
48
Wolbachia endosymbiont of Onchocerca ochengi
98
Neorickettsia risticii Illinois
100
Neorickettsia sennetsu Miyayama
Neorickettsia helminthoeca
100
Candidatus Nicolleia massiliensis
94
93
Candidatus Midichloria mitochondrii IricVA
99
Candidatus Midichloria sp Ixholo2
Candidatus Midichloria sp Ixholo1
97
Lariskella endosymbiont of Curculio okumai
100
Lariskella endosymbiont of Curculio morimotoi
41
Candidatus Lariskella arthropodarum
100
endosymbiont of Acanthamoeba sp UWC36
endosymbiont of Acanthamoeba sp UWC8
100
Candidatus Defluviella procrastinata
Candidatus Anadelfobacter veles
100
Candidatus Cyrtobacter zanobii
Candidatus Cyrtobacter comes
45
Rickettsia rickettsii Sheila Smith
63
Rickettsia prowazekii NMRC Madrid E
100
Rickettsia felis URRWXCal2
52
Rickettsia bellii RML369 C
100
Orientia chuto
88
100
Orientia tsutsugamushi Ikeda
100
97
secondary symbiont of Sitobion miscanthi
70
Rickettsia endosymbiont of Bemisia tabaci
Candidatus Cryptoprodotis polytropus
98
UB clone ELB16_030 (Lake Bonney Antarctica)
51
96
UB clone H06CH12_11_17 (Mine tailing material)
UB clone lp227 (apple orchard)
UB clone 69_34 (sediment of Guanting Reservoir)
44
freshwater OTU 5 size=1
33
freshwater OTU 8 size=2
61
55
freshwater OTU 4 size=11
98
freshwater OTU 6 size=4
44
freshwater OTU 7 size=3
57
freshwater OTU 2 size=26
92
freshwater OTU 1 size=2
44
36
Candidatus Arcanobacter lacustris
72
freshwater OTU 3 size=3
100
UAB clone 5_28_E12_b (hydrocarbon contaminated aquifer)
USB clone bac2nit3 (alpine tundra meadow soil)
UB clone 300I_E11 (uranium contaminated groundwater)
83
endosymbiont of Acanthamoeba sp S40
100
endosymbiont of Acanthamoeba polyphaga
94
endosymbiont of Acanthamoeba sp AC305
100
Caedibacter macronucleorum
Caedibacter caryophilus
37
100
Candidatus Odyssella thessalonicensis L13
83
endosymbiont of Acanthamoeba sp KA/E23
88
endosymbiont of Acanthamoeba sp TUMK 23
42
Candidatus Captivus acidiprotistae
Candidatus Hepatincola porcellionum
58
Zymomonas mobilis subsp mobilis ATCC 10988 1
Parvularcula bermudensis HTCC2503
58
Caulobacter vibrioides
33
Rhizobium etli CFN 42 1
88
freshwater OTU 9 size=12
freshwater OTU 10 size=1
37
Rhodobacter sphaeroides 2 4 1
84
Magnetococcus marinus MC 1 1
Rhodospirillum rubrum ATCC 11170 1
83
Candidatus Puniceispirillum marinum IMCC1322
88
Constrictibacter antarcticus
Kordiimonas gwangyangensis DSM 19435
Candidatus Pelagibacter ubique HTCC1062
100
Holospora obtusa
100
Holospora undulata HU1
100
Candidatus Gortzia infectiva
Candidatus Paraholospora nucleivisitans
0.06

Supplementary Figure S6

Phylogenetic tree based on SSU rRNA gene including *A. lacustris* related reads identified in the lake metagenomes, SRA and VAMPS SSU amplicon datasets and related sequences in NCBI-nt. This is the full tree of the phylogenetic analysis shown in Figure 6. Internal branches with bootstrap support less than 30 were collapsed. Taxa are colored according to their taxonomic affiliation at the family level; Anaplasmataceae: pink, 'Midichloriaceae': green, Rickettsiaceae: orange and 'Holosporaceae': blue. Outgroup taxa are grey. The *A. lacustris* leaf and branch support values on branch of interests have been marked with bold red. For NCBI-nt sequences, their environmental origin is stated between parentheses.