

An integrated phylogenomic approach toward pinpointing the origin of mitochondria

Zhang Wang and Martin Wu*

Figure S1. A rooted RAxML ML tree made with the mitochondrial marker dataset of 47 Alphaproteobacteria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Bootstrap values for internal nodes are shown beside them.

Figure S2. A rooted Bayesian tree made with the mitochondrial marker dataset of 47 Alphaproteobacteria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Posterior probability values for internal nodes are shown beside them.

Figure S3. A rooted RAxML ML tree made with the nuclear marker dataset of 47 Alphaproteobacteria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Bootstrap values for internal nodes are shown beside them.

Figure S4. A rooted Bayesian tree made with the nuclear marker dataset of 47 Alphaproteobacteria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Posterior probability values for internal nodes are shown beside them.

Figure S5. A rooted RAxML ML tree made with the phylum-level marker dataset of 47 Alphaproteobacteria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Bootstrap values for internal nodes are shown beside them.

Figure S6. A rooted Bayesian tree made with the phylum-level marker dataset of 47 Alphaproteobacteria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Posterior probability values for internal nodes are shown beside them.

Figure S7. A rooted RAxML ML tree made with the mitochondrial marker dataset of 72 Alphaproteobacteria and 6 mitochondria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Bootstrap values for internal nodes are shown beside them. Mitochondria representatives are highlighted in red.

Figure S8. A rooted Bayesian tree made with the mitochondrial marker dataset of 72 Alphaproteobacteria and 6 mitochondria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Posterior probability

values for internal nodes are shown beside them. Mitochondria representatives are highlighted in red.

Figure S9. A rooted RAxML ML tree made with the nuclear marker dataset of 72 Alphaproteobacteria and 6 eukaryote representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Bootstrap values for internal nodes are shown beside them. Eukaryote representatives are highlighted in red.

Figure S10. A rooted Bayesian tree made with the Dayhoff6 recoded mitochondria marker dataset of 72 Alphaproteobacteria and 6 mitochondria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Posterior probability values for internal nodes are shown beside them. Mitochondria representatives are highlighted in red.

Figure S11. A rooted Bayesian tree made with the Dayhoff6 recoded nuclear marker dataset of 72 Alphaproteobacteria and 6 eukaryote representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Posterior probability values for internal nodes are shown beside them. Eukaryote representatives are highlighted in red.

Figure S12. A rooted RAxML ML tree made with the Dayhoff4 recoded mitochondria marker dataset of 72 Alphaproteobacteria and 6 mitochondria representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Bootstrap values for internal nodes are shown beside them. Mitochondria representatives are highlighted in red.

Figure S13. A rooted RAxML ML tree made with the Dayhoff4 recoded nuclear marker dataset of 72 Alphaproteobacteria and 6 eukaryote representatives. The tree was rooted using Beta and Gammaproteobacteria as the outgroup. Bootstrap values for internal nodes are shown beside them. Eukaryote representatives are highlighted in red.

Figure S14. A rooted Bayesian consensus tree made with the nuclear dataset of 72 Alphaproteobacteria and 22 eukaryotes (highlighted in red). The tree was rooted using Beta and Gammaproteobacteria as the outgroup. The posterior probability support values of the internal nodes are shown beside them.

Supplementary Table 1. List of 50 top-ranked splits in the spectral analysis of the datasets with and without 18 genomes. Long-branch lineages that violate the well established phylogenetic relationships (e.g., the monophyly of mitochondria or Rickettsiales) are highlighted in bold.

Supplementary Table 2. List of nuclear and mitochondrial marker genes sorted by aminoGC, chi_square and substitution rate respectively.

Supplementary Table 3. Overview of 30 eukaryotic lineages selected for identifying the mitochondria-derived nuclear genes.

Supplementary Table 4. List of 200 Alphaproteobacterial genomes used in the phylogenomic analysis.

Supplementary File 1. The distribution of the 29 mitochondria-derived nuclear genes among individual genomes in the concatenated sequence alignment. The presence of gene in each genome is denoted as '1' and highlighted in green. The absence of gene is denoted as '0' and highlighted in red. The alpha-proteobacteria, nuclear, and beta/gamma-proteobacteria genomes are highlighted in red, blue and black, and taxonomy showing at order, phylum and class levels respectively.

Supplementary File 2. The presence of paralogs of the 29 mitochondria-derived nuclear genes and 26 mitochondria-encoded genes among individual genomes. The single copy gene in each genome is denoted as '1' and highlighted in green. The absence of a gene in each genome is denoted as '0' and highlighted in red. The gene with species-specific duplications in each genome is denoted as "copy number+S" and highlighted in gold. The gene with non-species-specific duplications in each genome is denoted as "copy number+NS" and highlighted in pink. The alpha-proteobacteria, nuclear, and beta/gamma-proteobacteria genomes are highlighted in red, blue and black, and taxonomy showing at order, phylum and class levels respectively.

Supplementary File 3. Individual protein sequences of 26 mitochondria-encoded genes and 29 mitochondria-derived nuclear genes.

Supplementary File 4. Individual sequence alignments of 26 mitochondria-encoded genes and 29 mitochondria-derived nuclear genes.



Figure S1

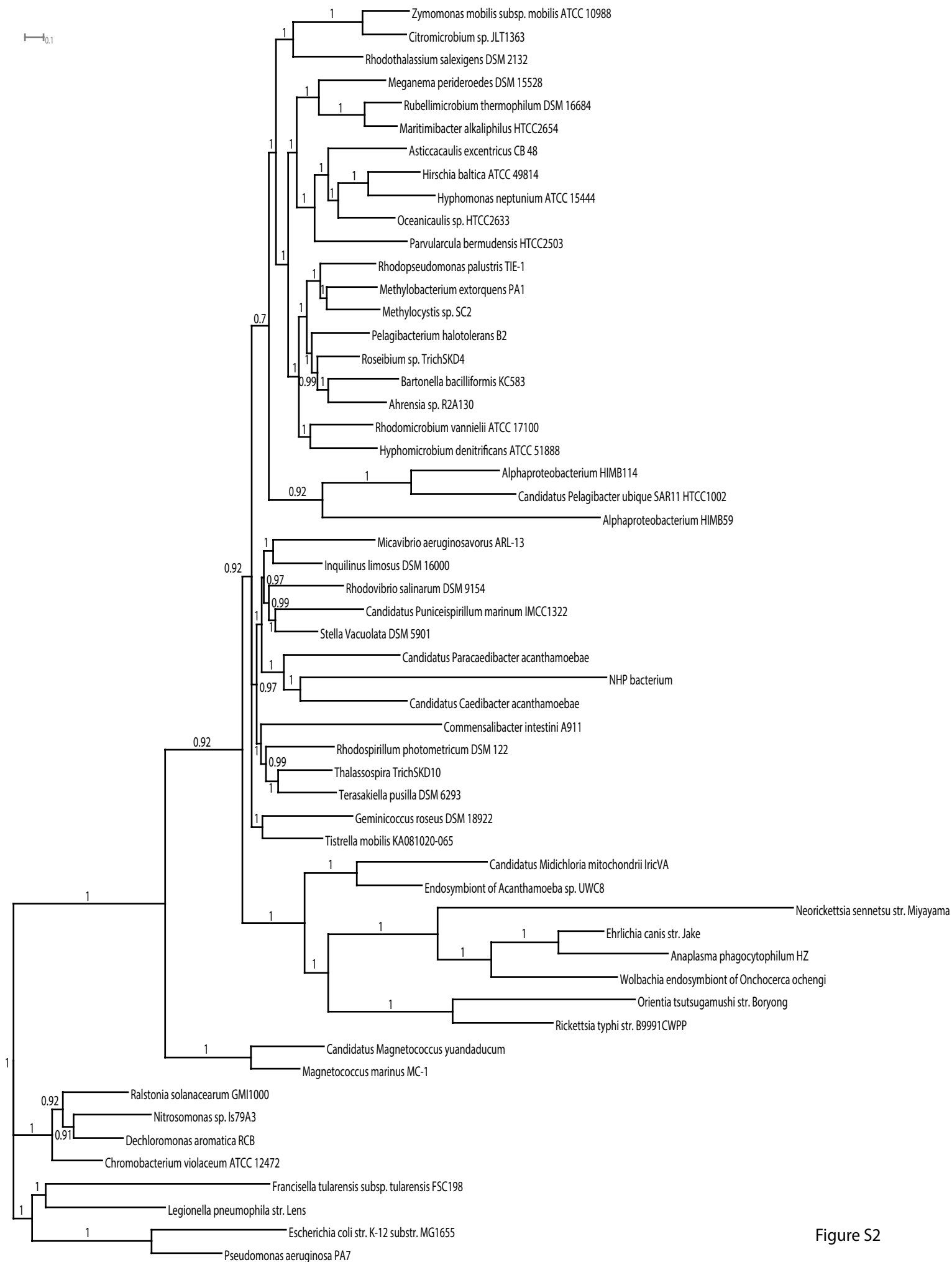


Figure S2

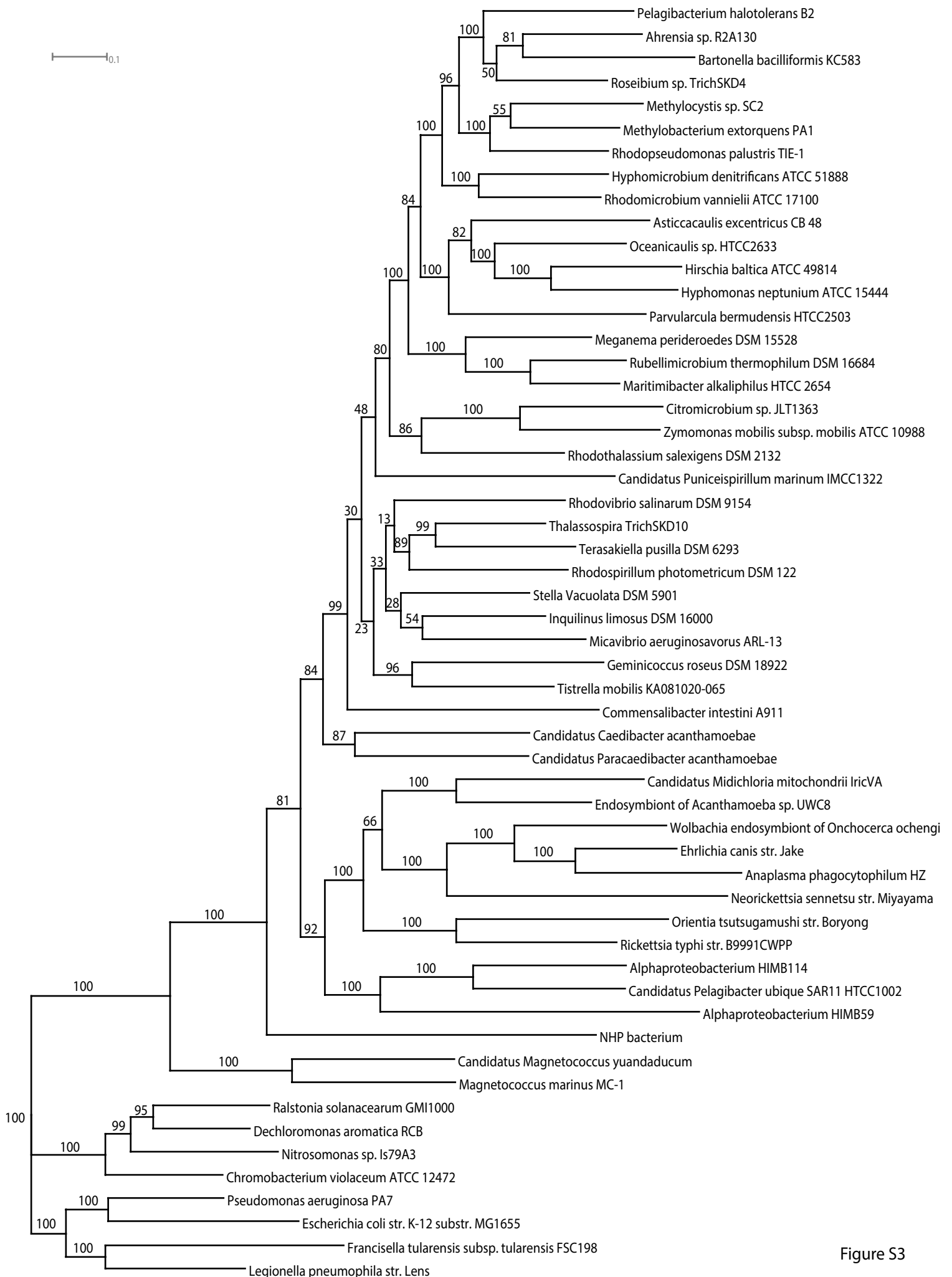


Figure S3

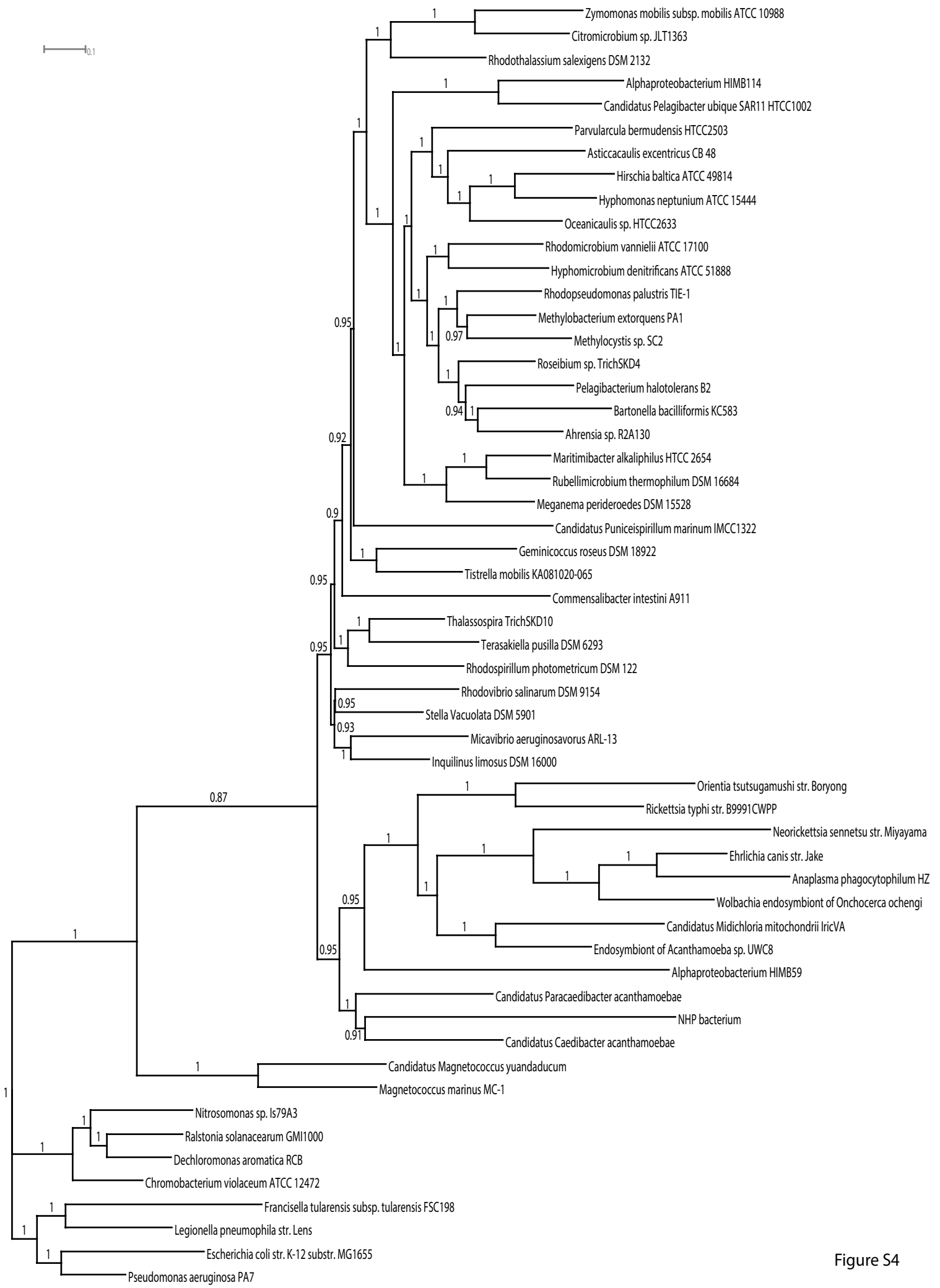


Figure S4

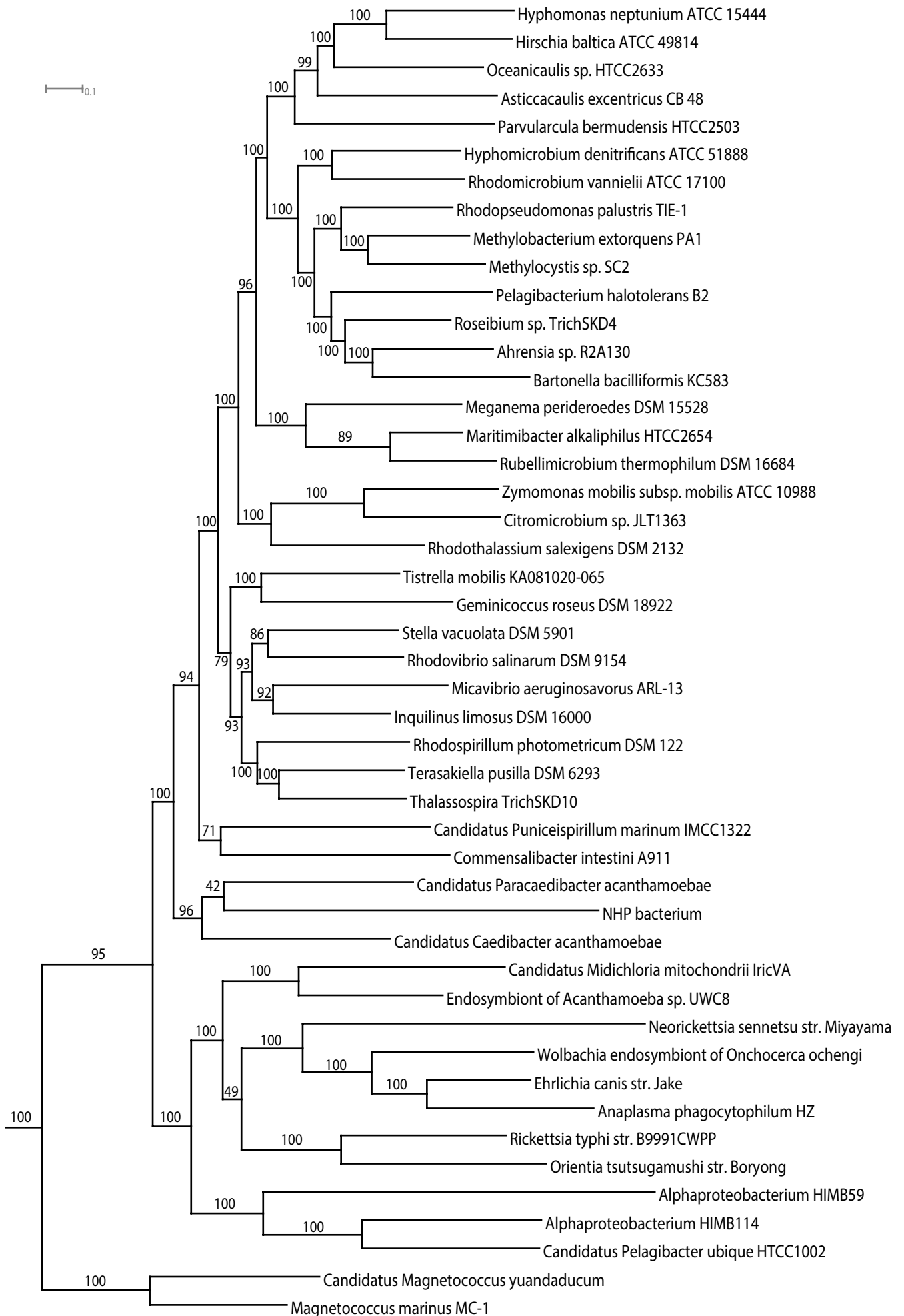


Figure S5

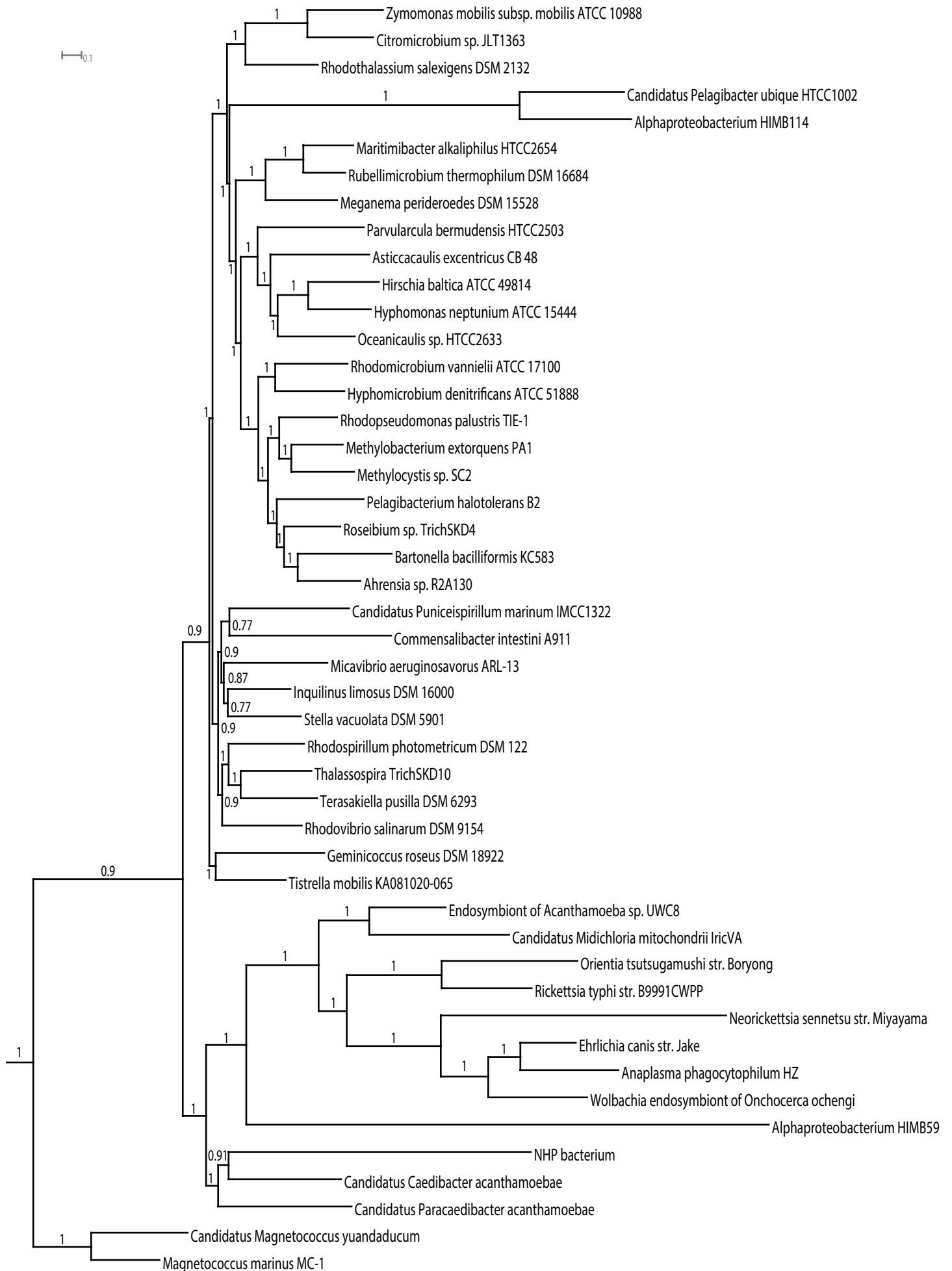


Figure S6



Figure S7

0.1



Figure S8

0.1

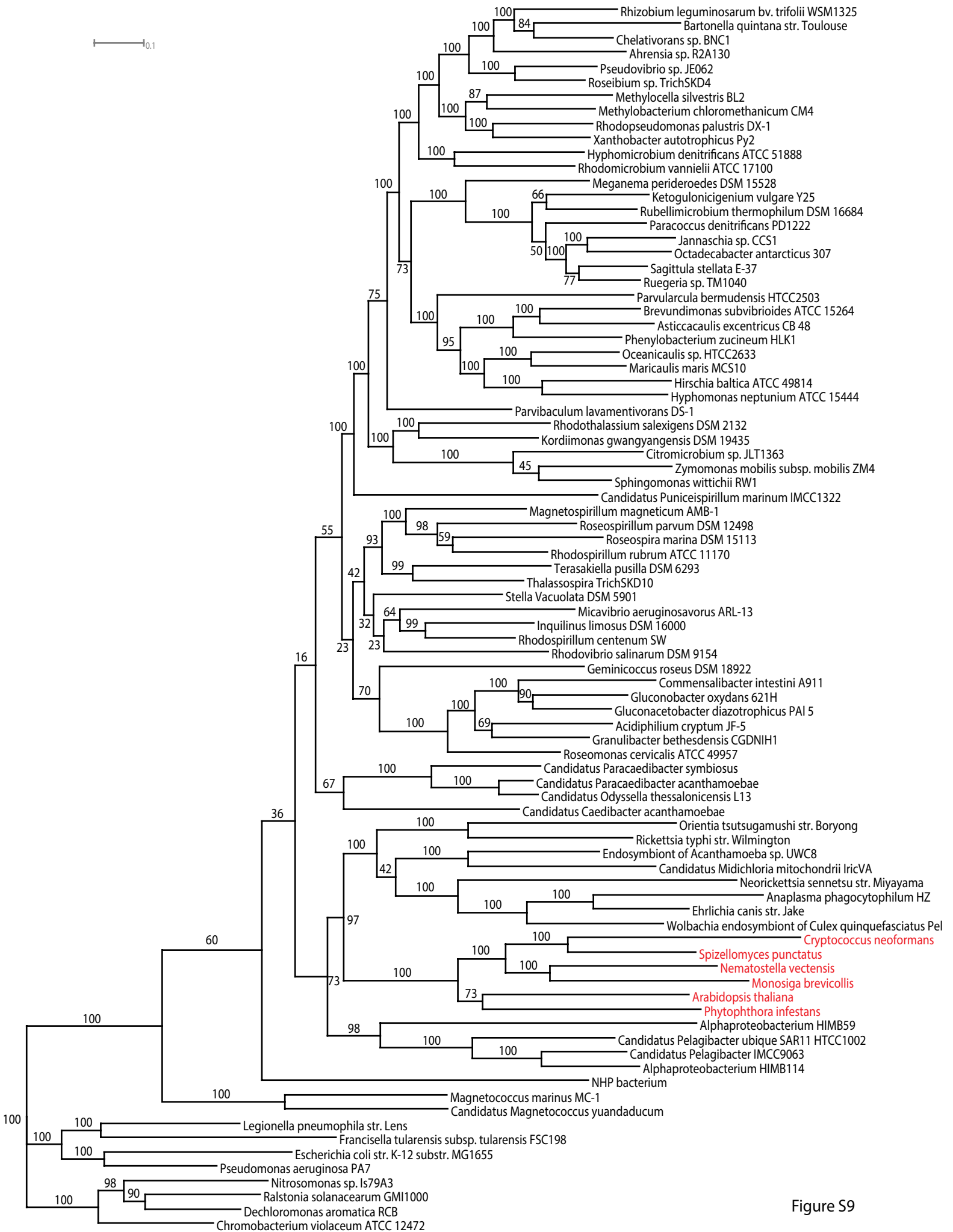


Figure S9

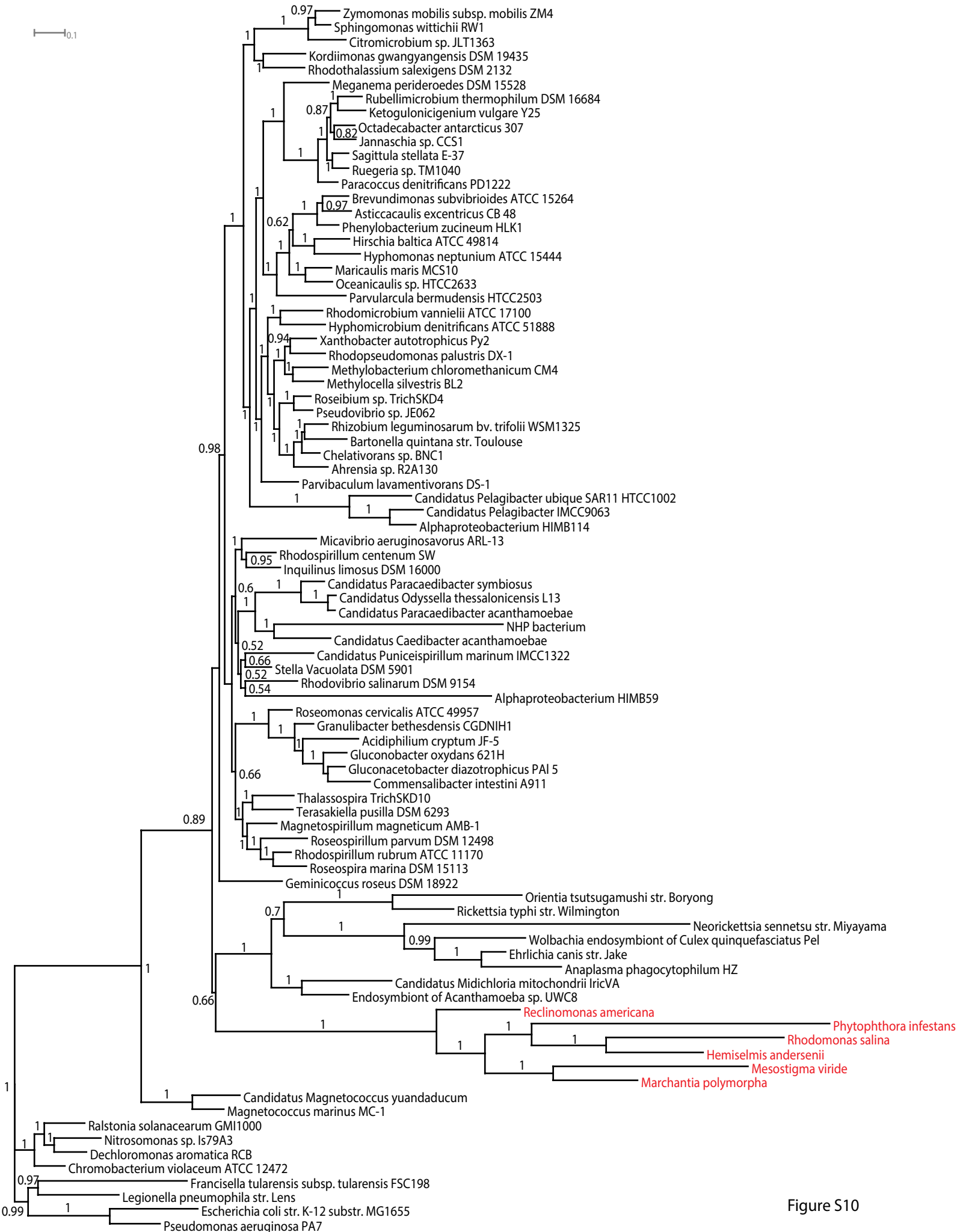


Figure S10

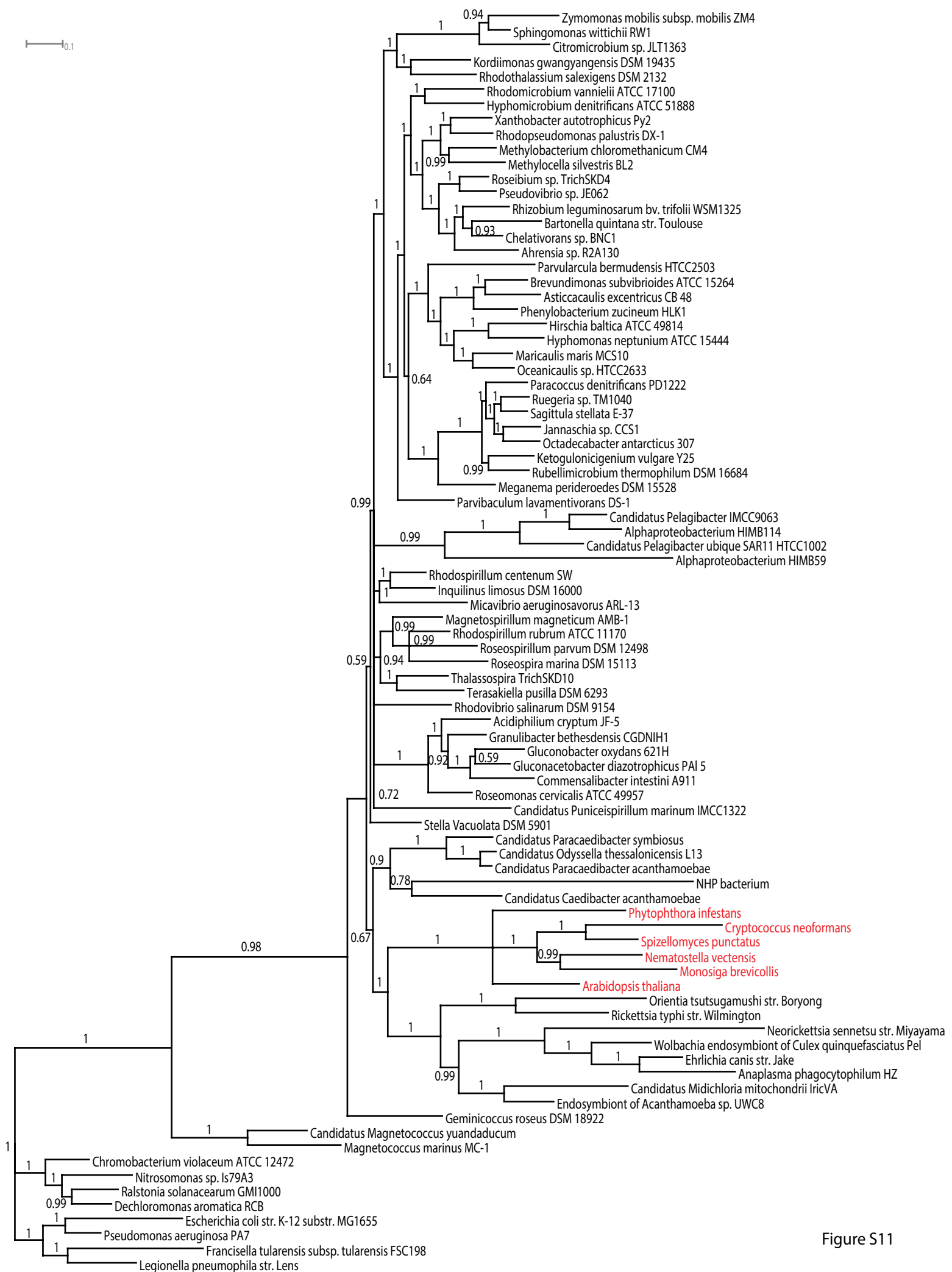


Figure S11

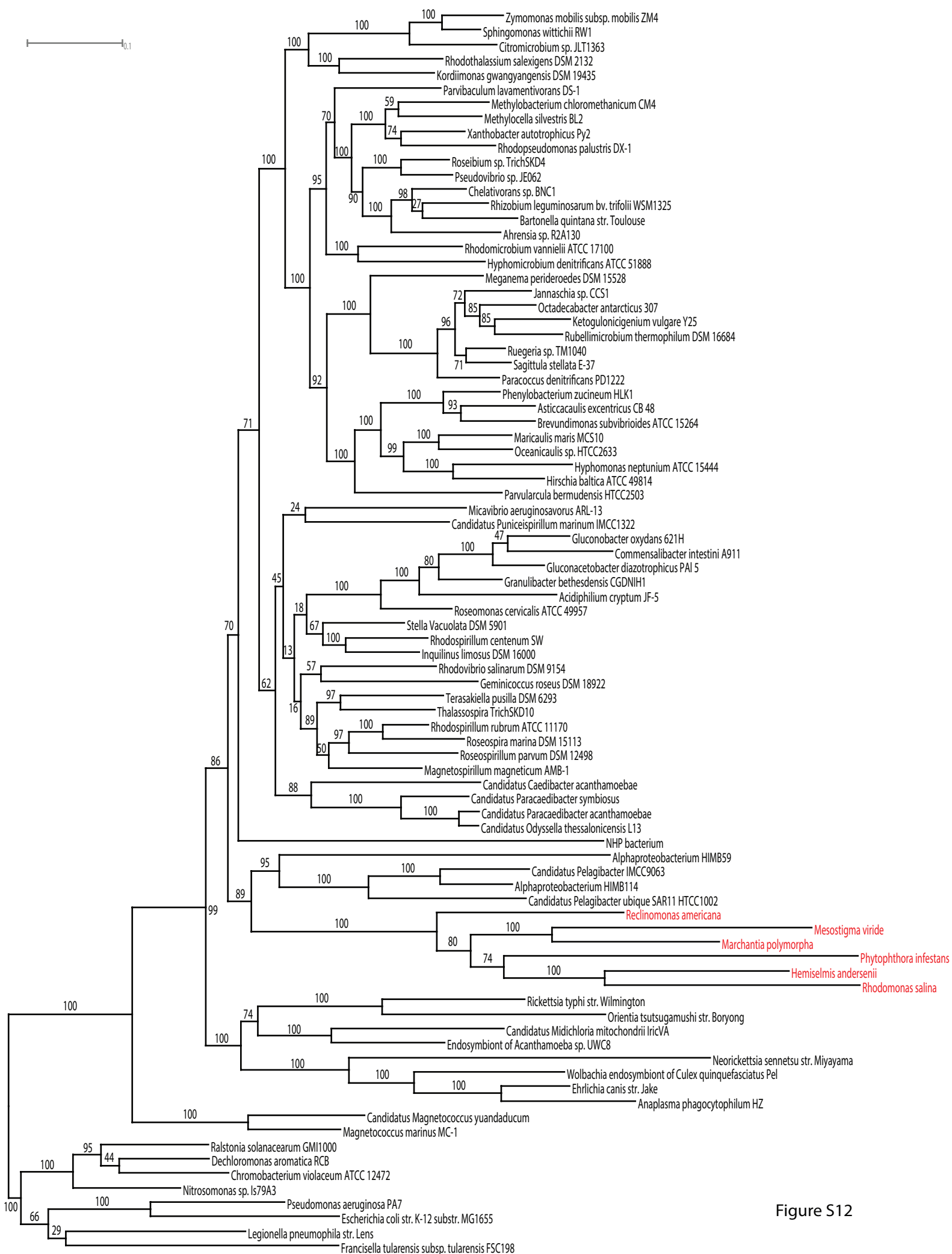


Figure S12

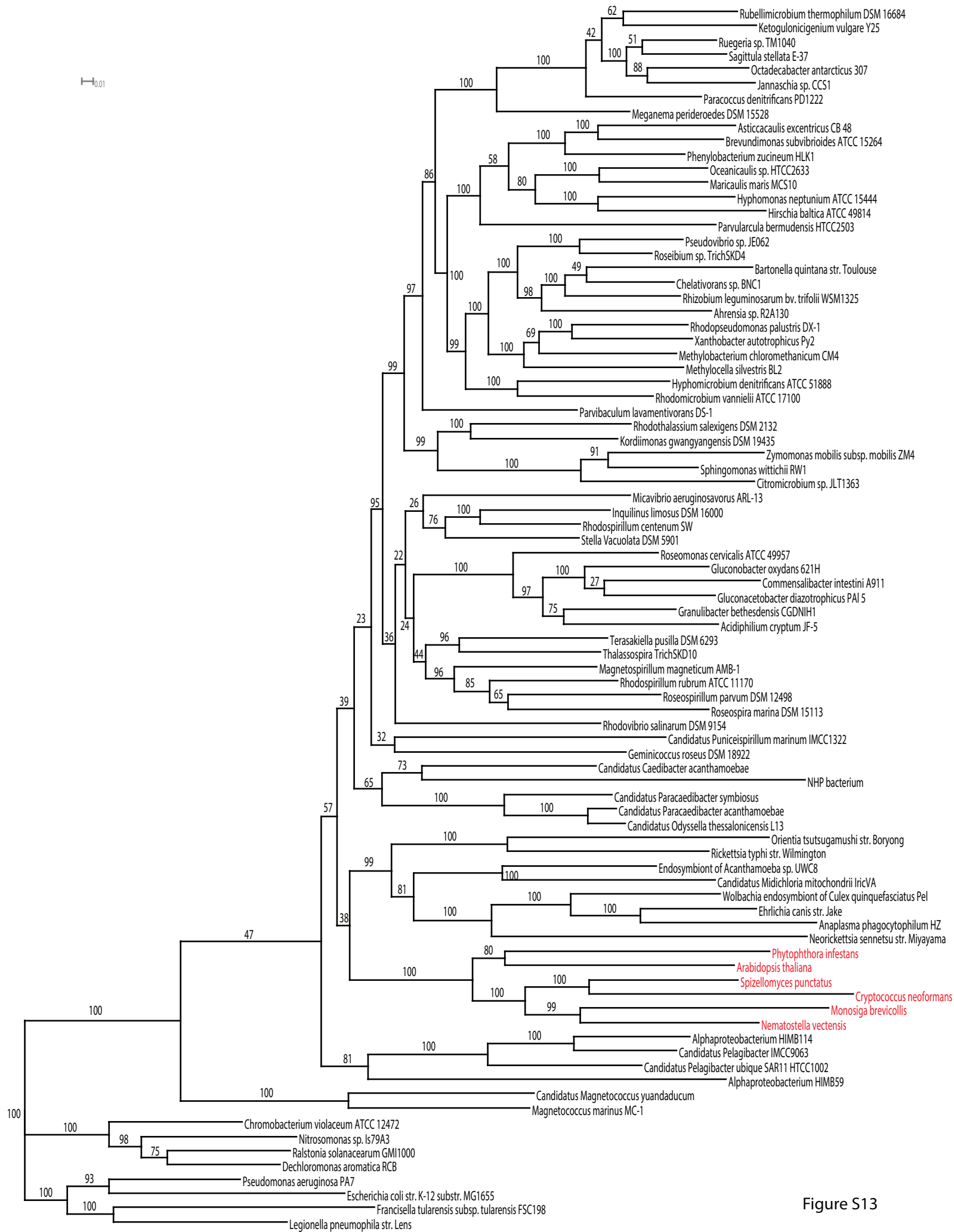


Figure S13

0.1



Figure S14