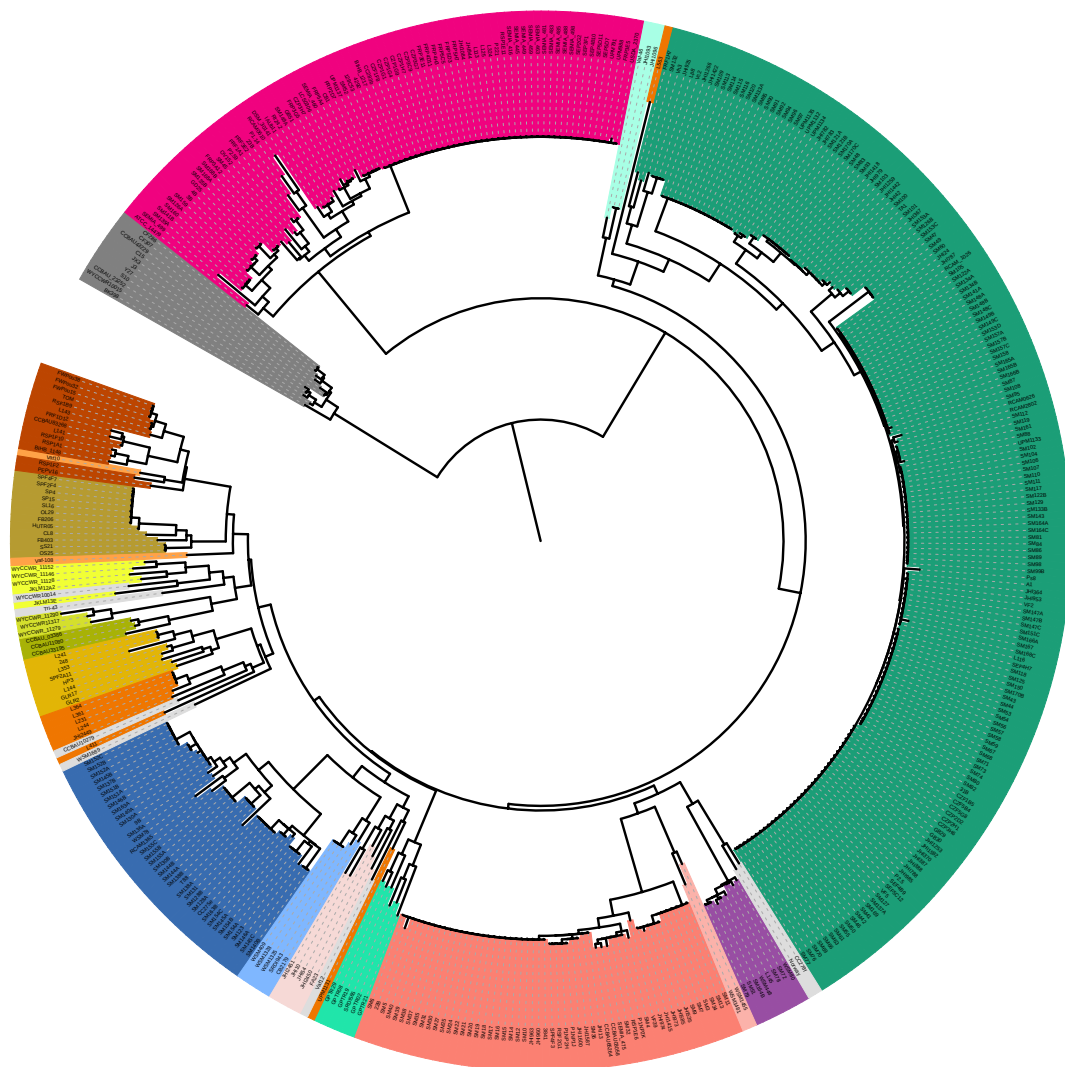
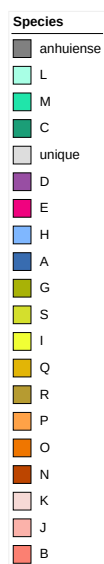


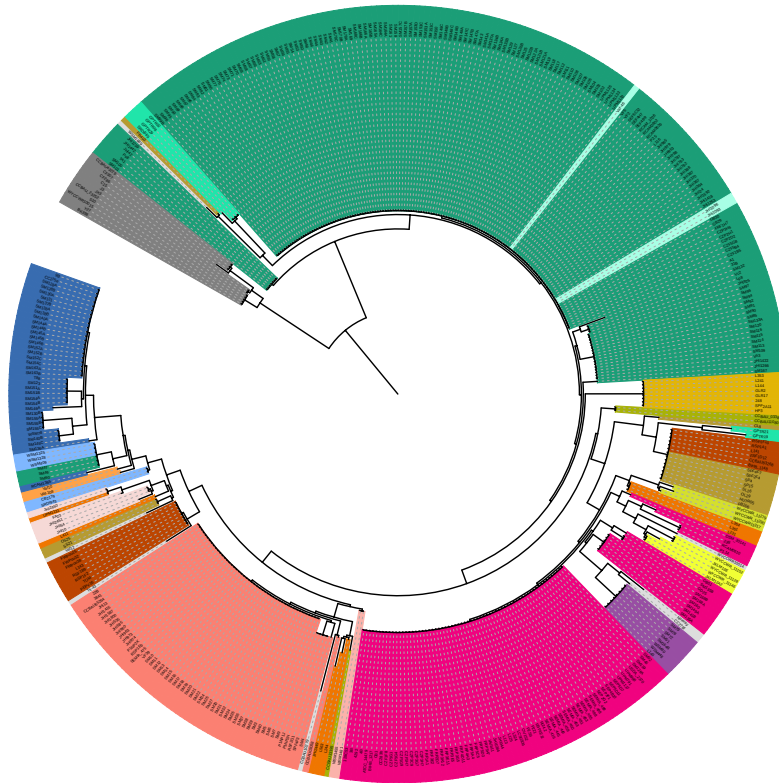
**Figure S1.** Phylogeny of the Rlc based on two sets of 60 core genes. a. Set A. b. Set B. The trees are rooted using *R. anhuiense* as the outgroup. The genospecies are indicated by coloured segments. Branches with >95% bootstrap support are black; those with lower support are coloured as indicated in the legend.

Tree scale: 0.01



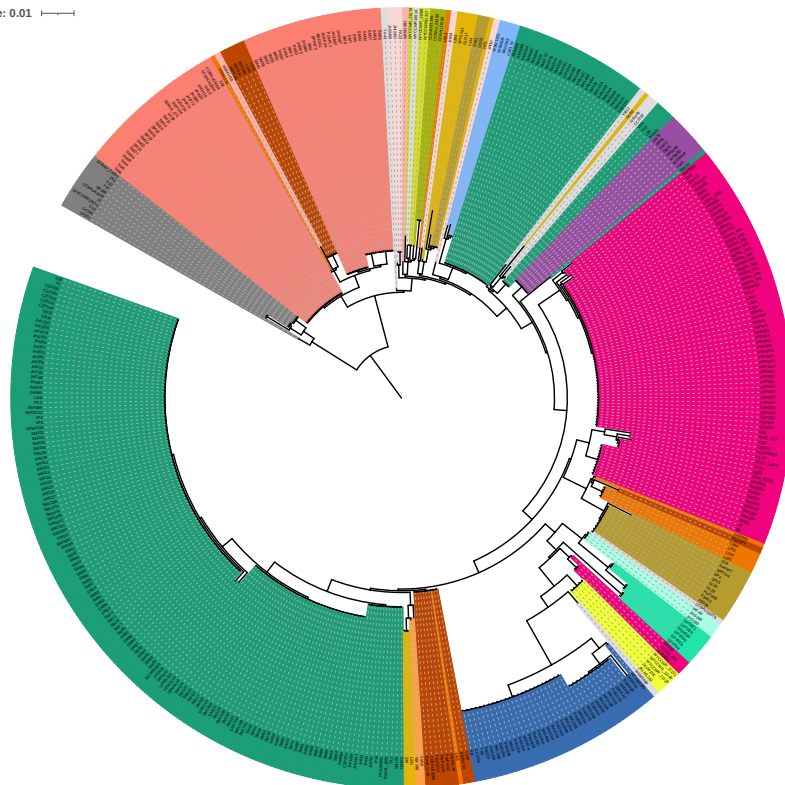
**Figure S2.** Phylogeny of concatenated *atpD-gyrB-recA* amplicon sequences. The sequences are those that would be amplified by widely-used PCR primers. The genospecies are indicated by coloured segments.

Tree scale: 0.01



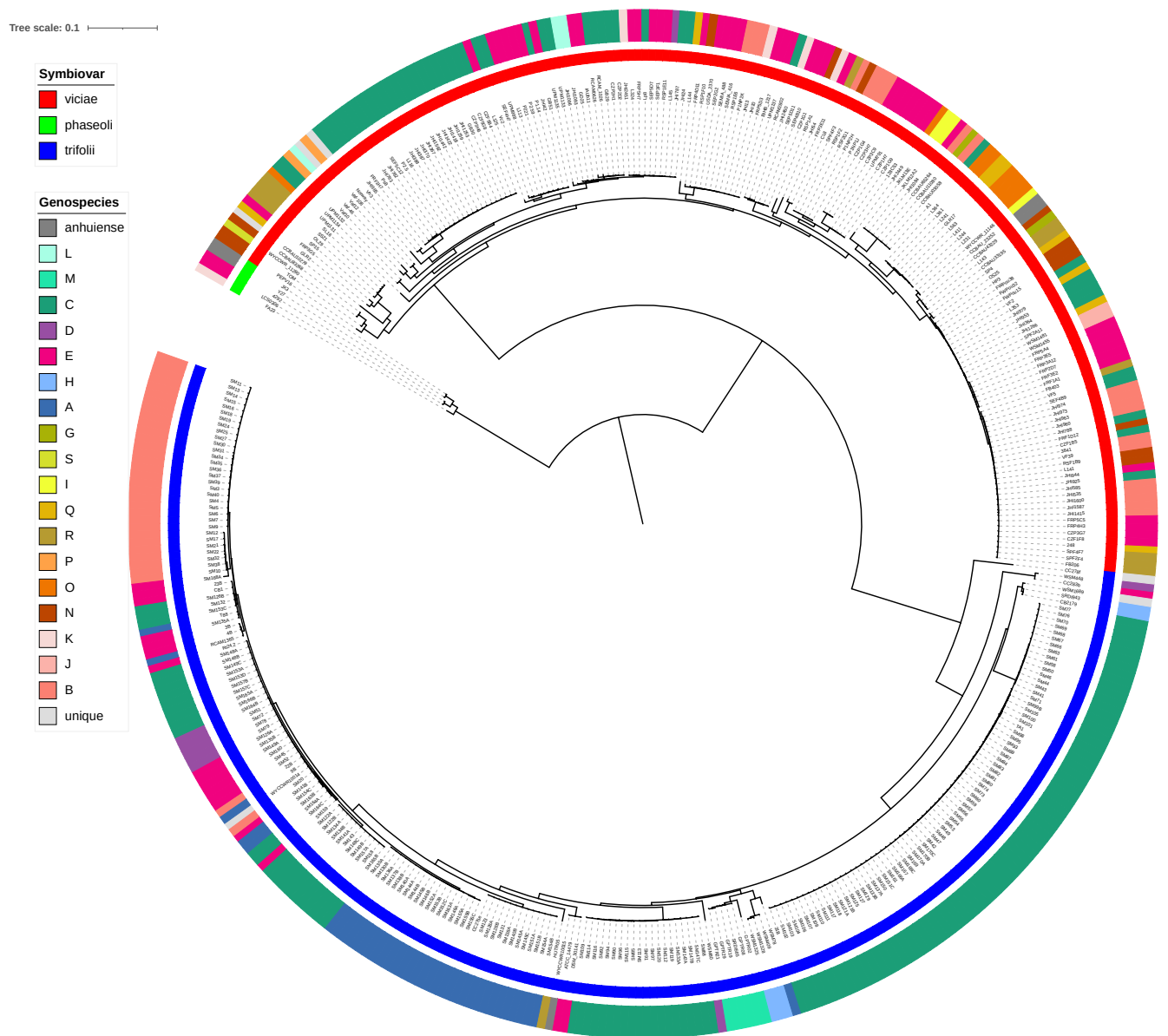
a

Tree scale: 0.01



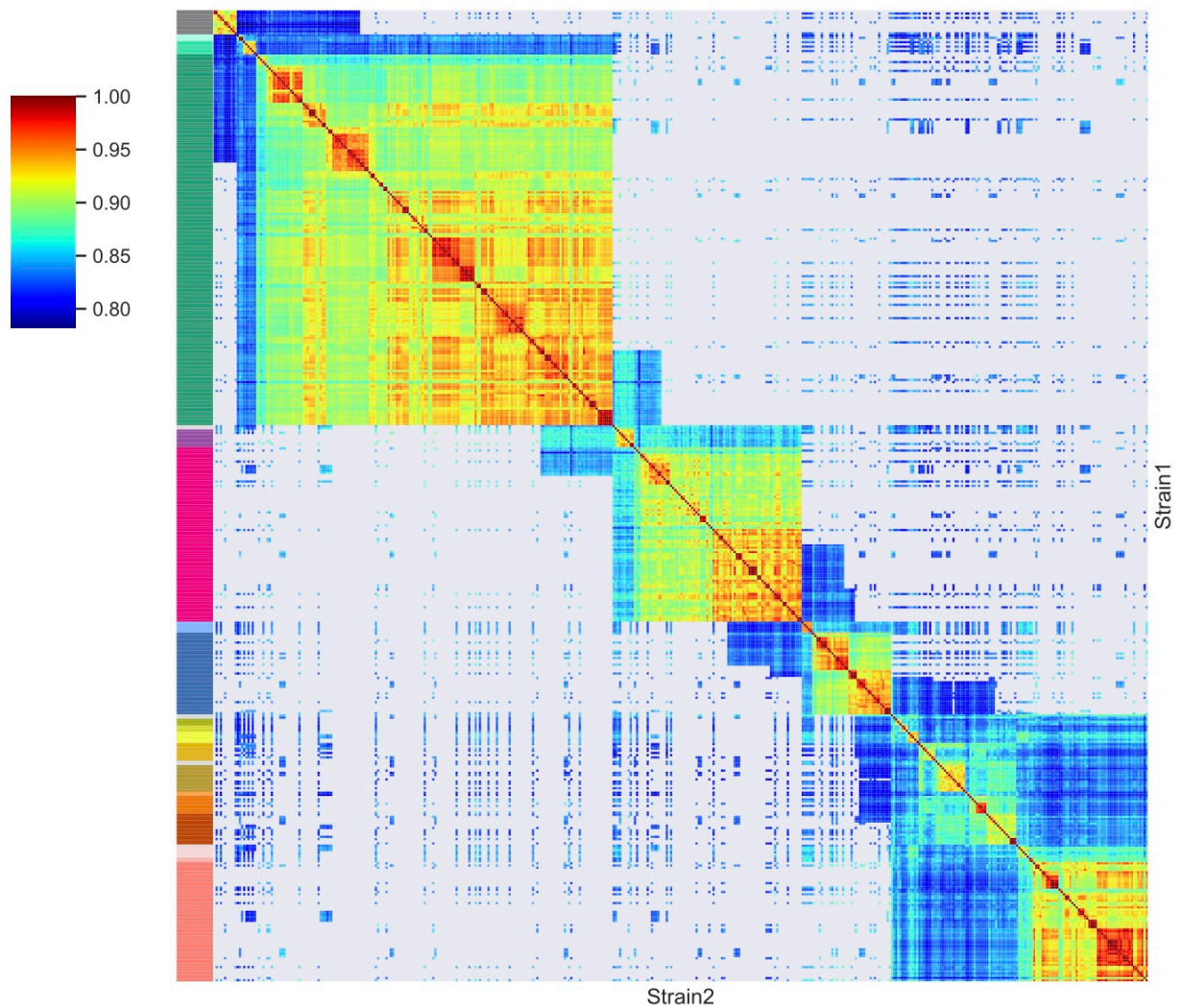
b

**Figure S3.** Phylogeny of two housekeeping genes, based on short sequences that could be amplified for high-throughput sequencing [74]. a. *recA* (251 bp). b. *rpoB* (254 bp). Sectors are shaded to indicate genospecies.



**Figure S4.** Phylogeny of *nodC*. The inner ring shows the separate clades of the three symbiovars. The outer circle shows the genospecies of each strain.





**Figure S5.** Sharing of accessory genes between strains. Pairwise sharing index, based on the number of orthogroups shared between two strains, normalised by total orthogroup content, for selected comparisons. The genospecies of each strain is indicated in the left bar.