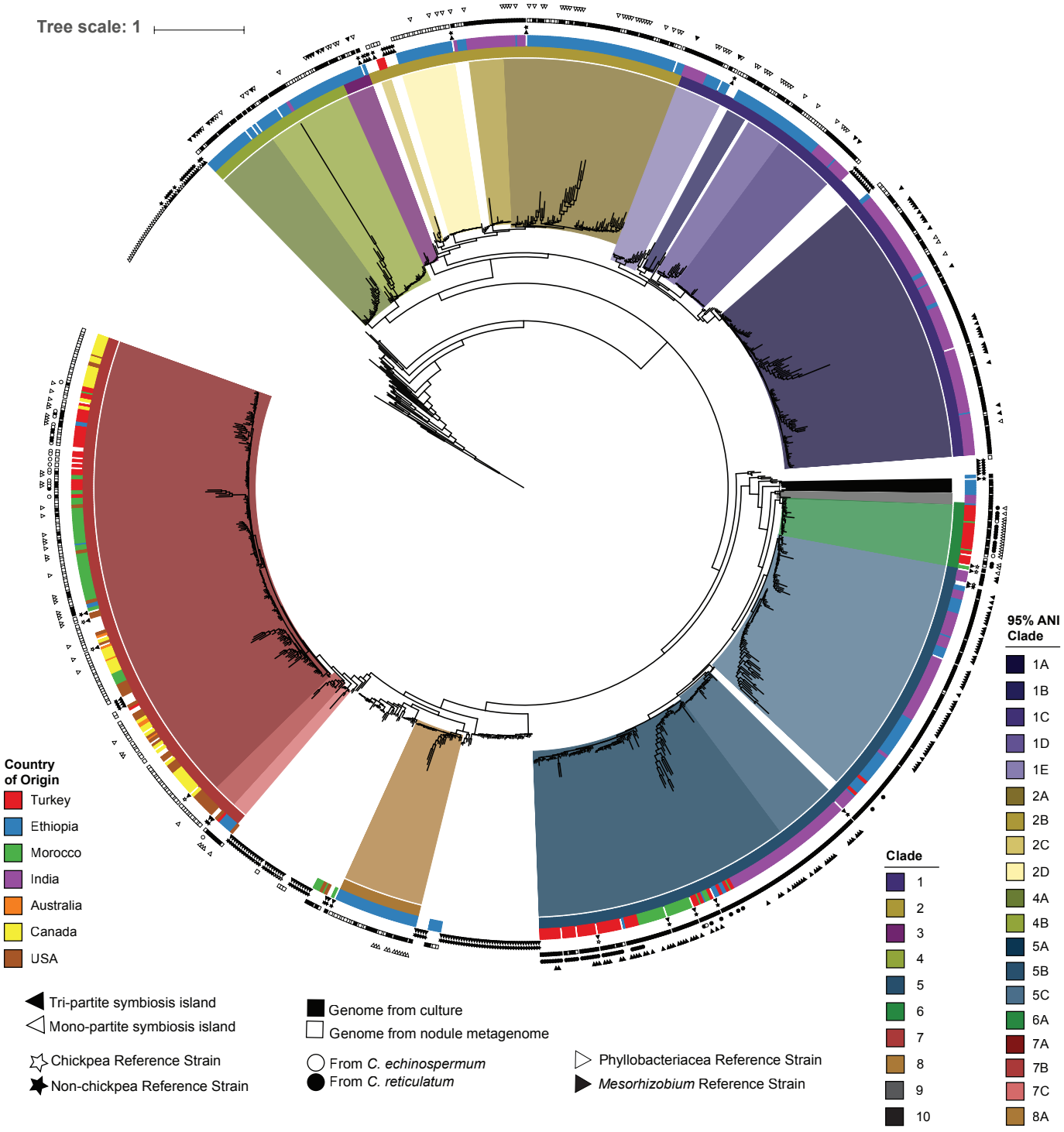


Fig S1: Phylogenetic tree generated from set of 400 universal single-copy marker genes annotated in genomes from *Mesorhizobium* cultures and root-nodule DNA extracts, as well as published reference genomes in the bacterial family Phyllobacteriaceae, with *Sinorhizobium melliloti* (in the family Rhizobiaceae) as an outgroup. Annotations are the same as in main-text **Fig. 1A**. The structure of the phylogeny demonstrates that genomes assembled from chickpea root-nodule DNA and bacterial isolates from chickpea nodules occur not only within the family Phyllobacteriaceae but specifically within the genus Mesorhizobium.

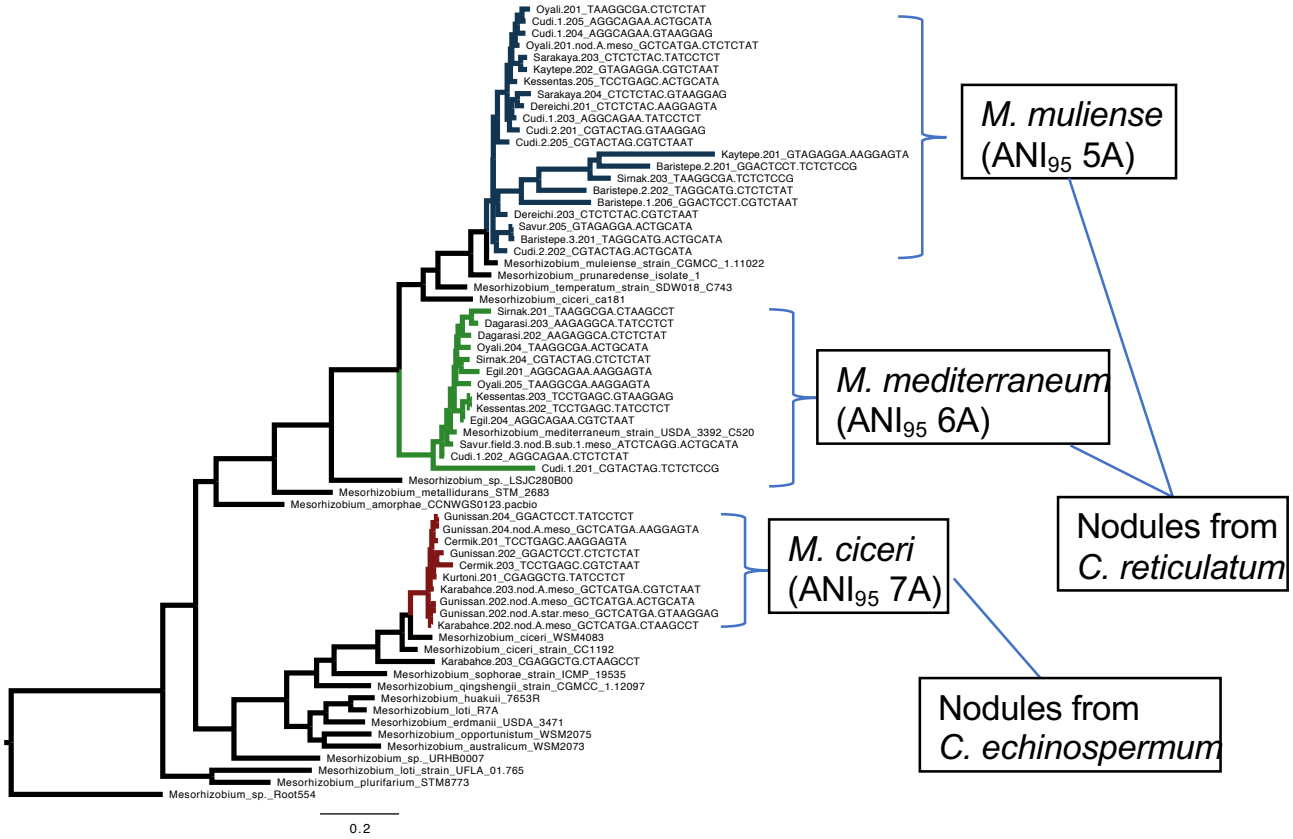


Fig S2: Phylogenetic tree generated from set of 400 universal single-copy marker genes annotated in genomes assembled from nodules collected from wild *C.* reticulatum and *C. echinospermum* in their native ranges in Southeastern Turkey, along with a minimal set of phylogenetically-representative set of *Mesorhizobium* reference genomes.

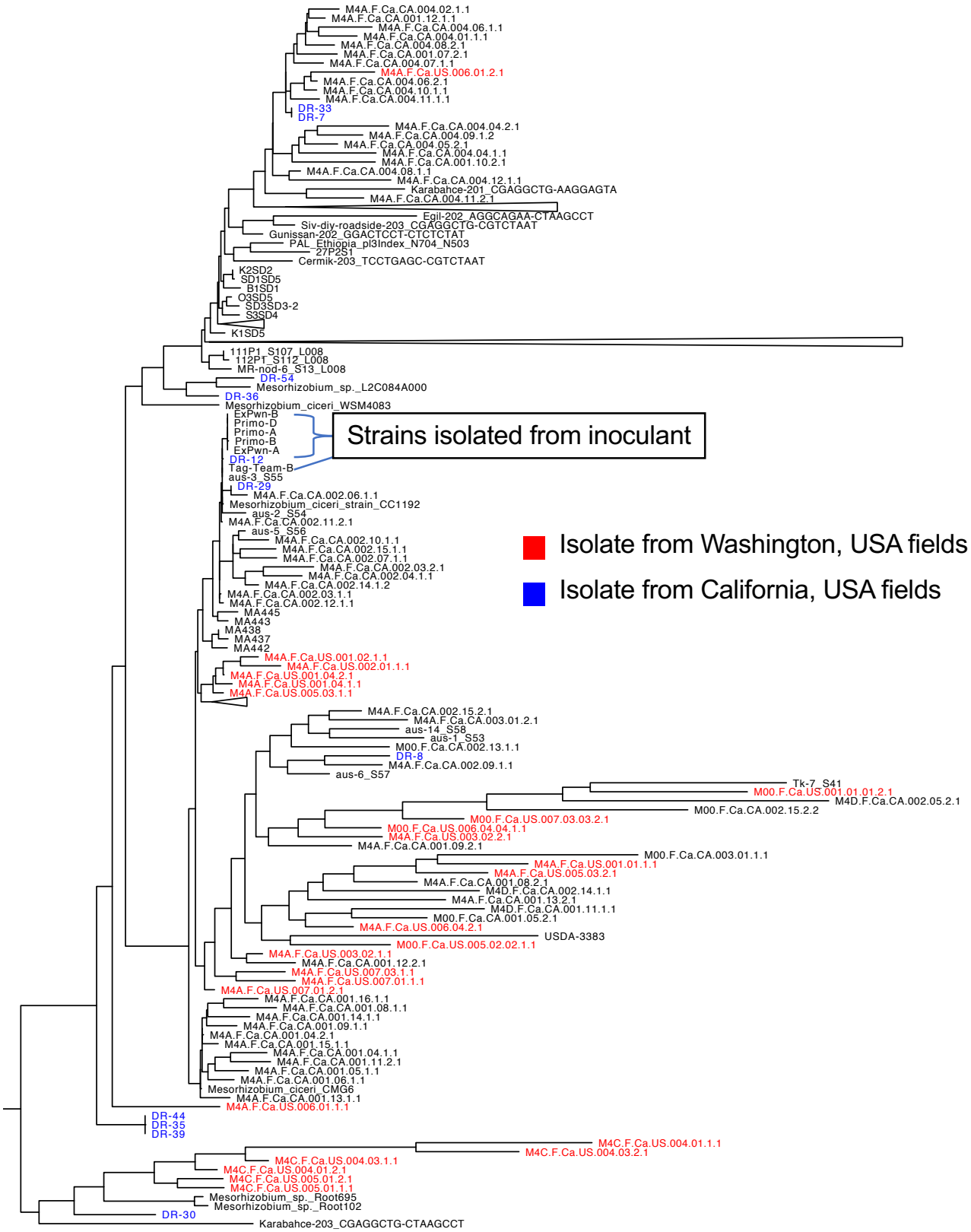


Fig S3: Phylogenetic tree generated from set of 400 universal single-copy marker genes annotated in genomes from clade 7A, which contains the type strain of *M. ciceri* as well as strains that nodulate *C. echinospermum* in its native range. Most of the leaves of this tree are from isolates cultured from nodules of cultivated chickpea grown in Canada and the United States (colored red from Washington state and blue from California). We obtained inoculant used for sampled fields from Washington and California and display the phylogenetic placement of strains obtained from this inoculant. Nodule-isolates from the fields are within the same ANI_{95} group as the inoculant strains, but are phylogenetically disparate.

Fig. S4A-C

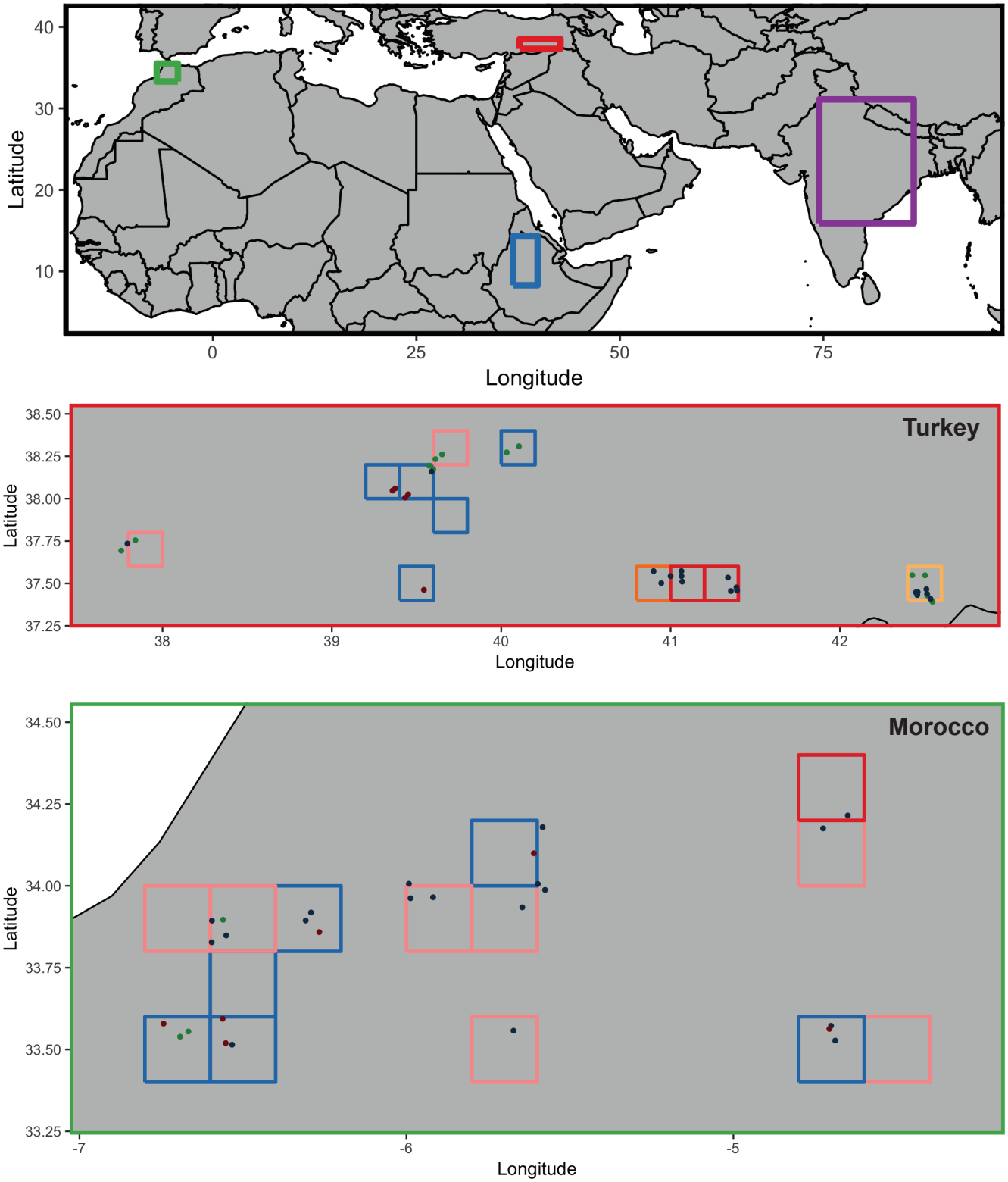


Fig. S4: 0.2 degree sampling grids displayed in geographic space and colored by beta-diversity cluster (**Fig 2B** dendrogram). Point colors represent ANI95 species groups for sample-collection points.

Fig. S4D

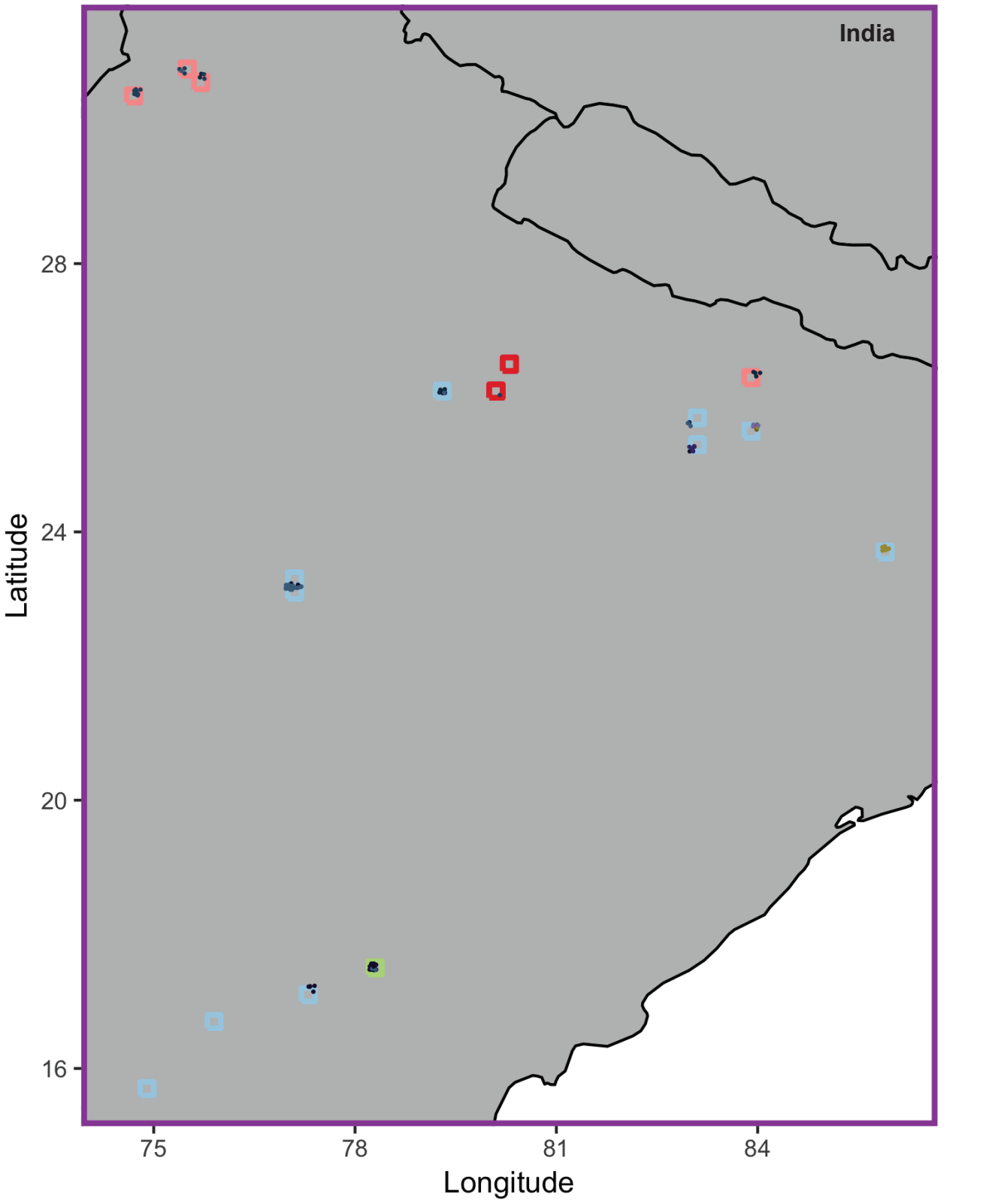

India
28
24
20
16
Latitude
75
78
81
84
Longitude
Phylogenetic-Diversity Cluster
A1
A3
B1
B3
B5
A2
A4
B2
B4
B6
95% ANI Clade
1A
1C
1E
2B
2D
4B
5B
6A
7B
8A
1B
1D
2A
2C
4A
5A
5C
7A
7C

Fig. S4E

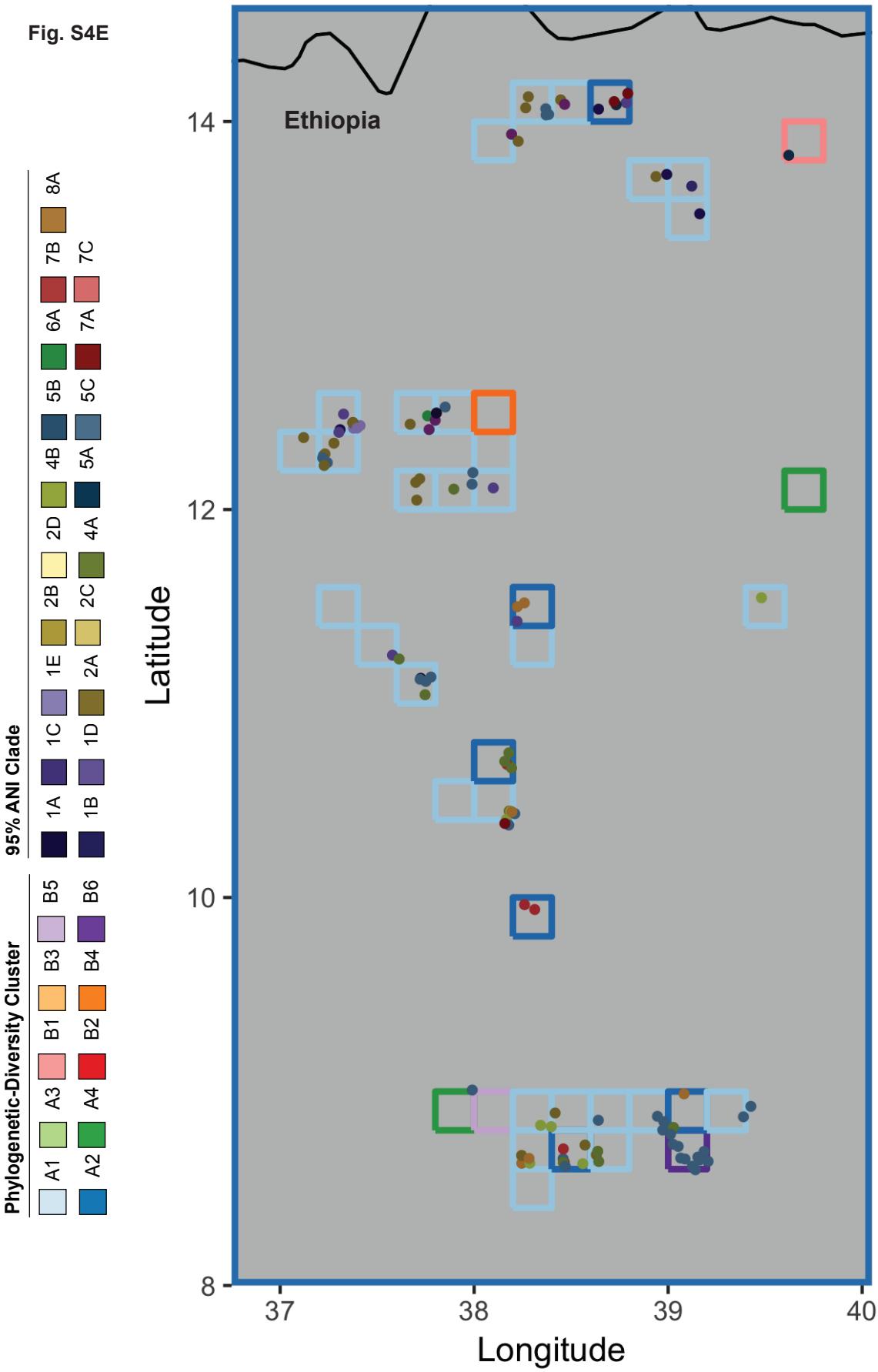

Ethiopia
Latitude
Longitude
14
12
10
8
37
38
39
40
95% ANI Clade
1A
1B
1C
1D
1E
2A
2B
2C
2D
4A
4B
5A
5B
5C
6A
7A
7B
7C
8A
Phylogenetic-Diversity Cluster
A1
A2
A3
A4
B1
B2
B3
B4
B5
B6

Fig S5A

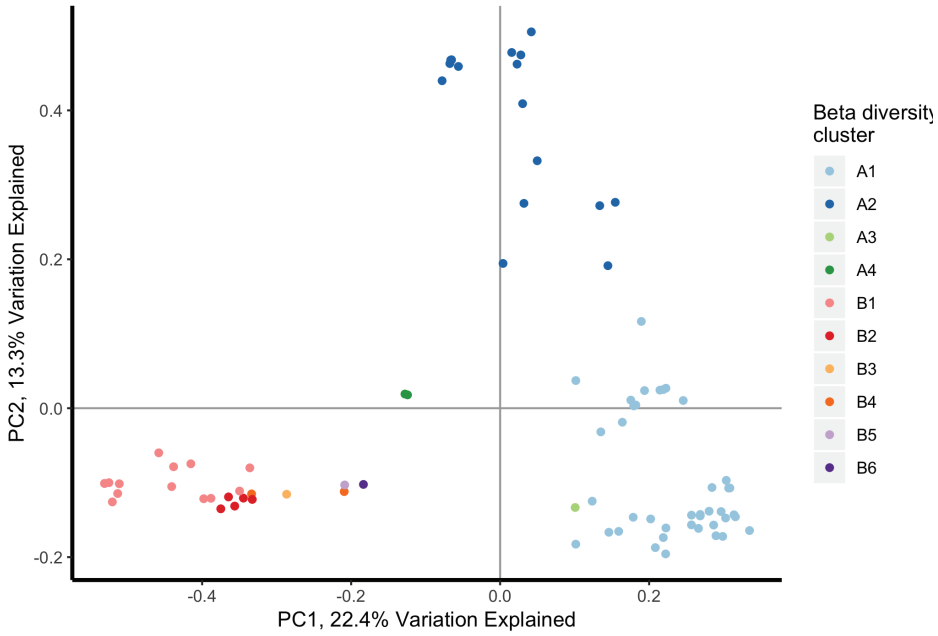


Fig S5B

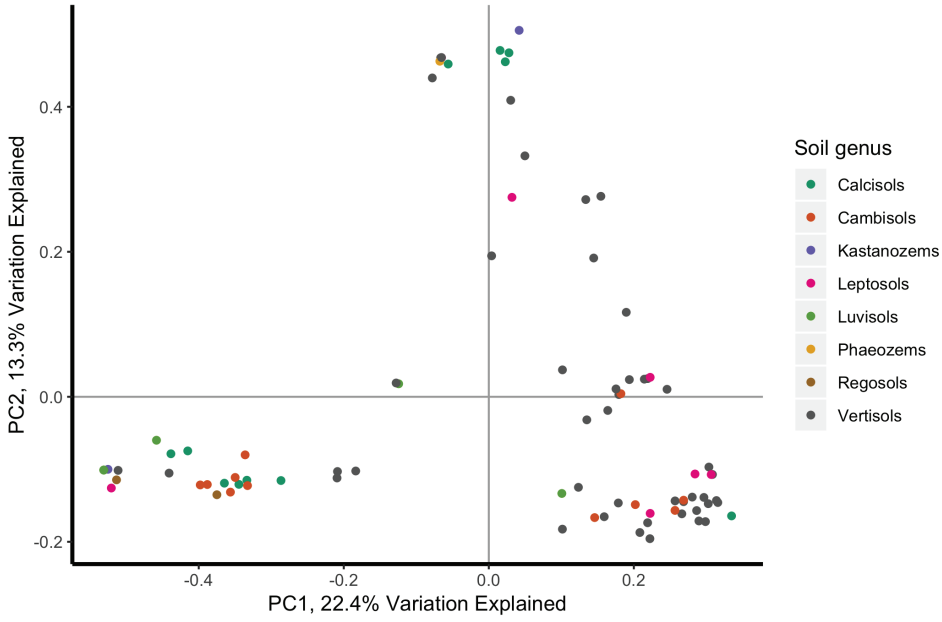


Fig S5: PCoA of 0.2-degree sampling grid cells ordinated by phylogenetic β-diversity, **(A)** colored as per clusters in 2B dendrogram and **(B)** colored by soil type.

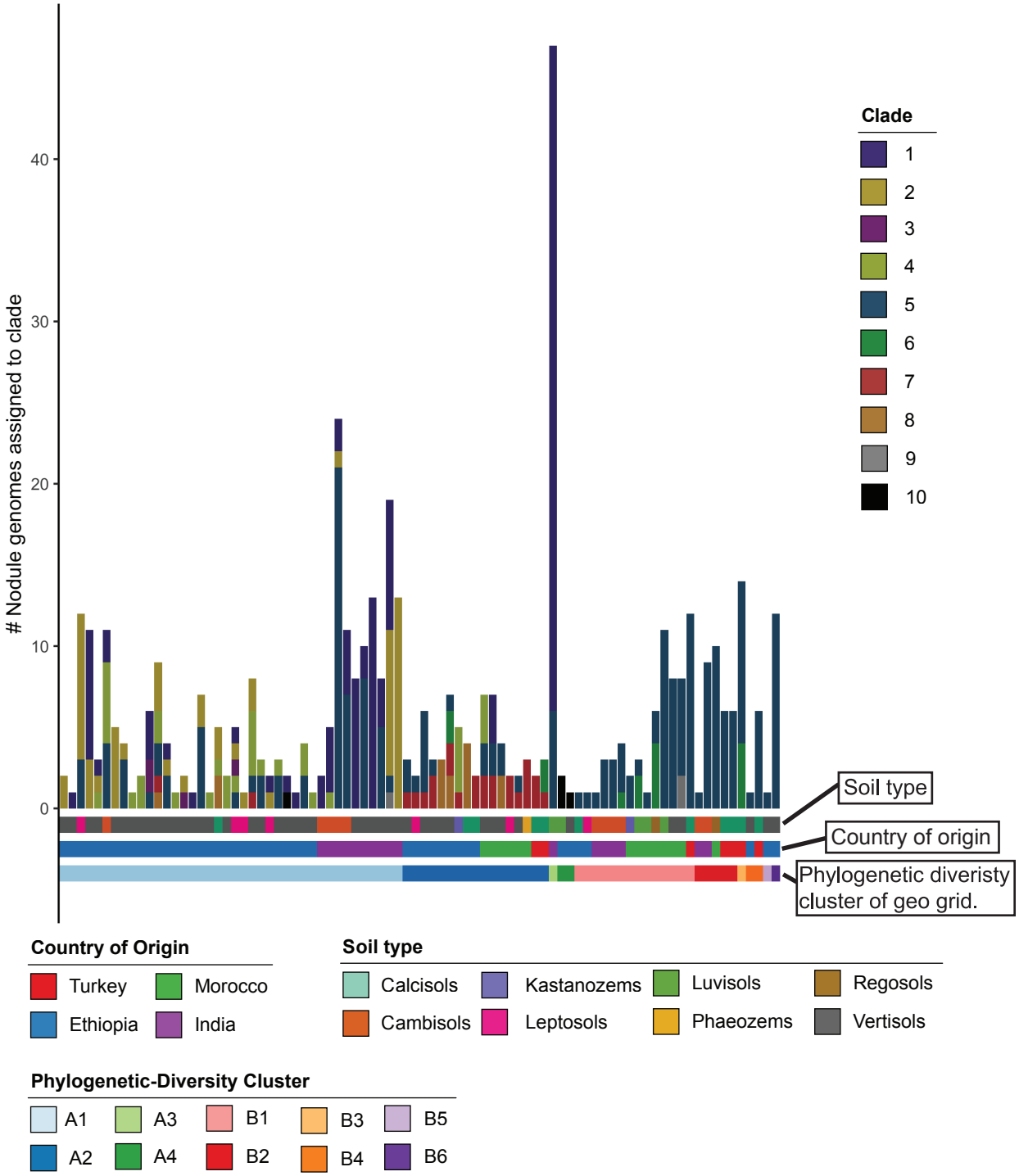


Fig S6: Stacked barchart depicting the phylogenetic composition of each geographic grid cell. Each bar represents an individual 0.2 degree x 0.2 degree cell, with the height of the bar representing the number of nodules sampled from that cell, and colored by the major Mesorhizobium clade each nodule was assigned to. Grid cells are ordered left to right by beta-diversity cluster as per Fig. 2B dendrogram, with colored annotations below representing the country that cell belongs to and the major soil-type inferred for that cell.

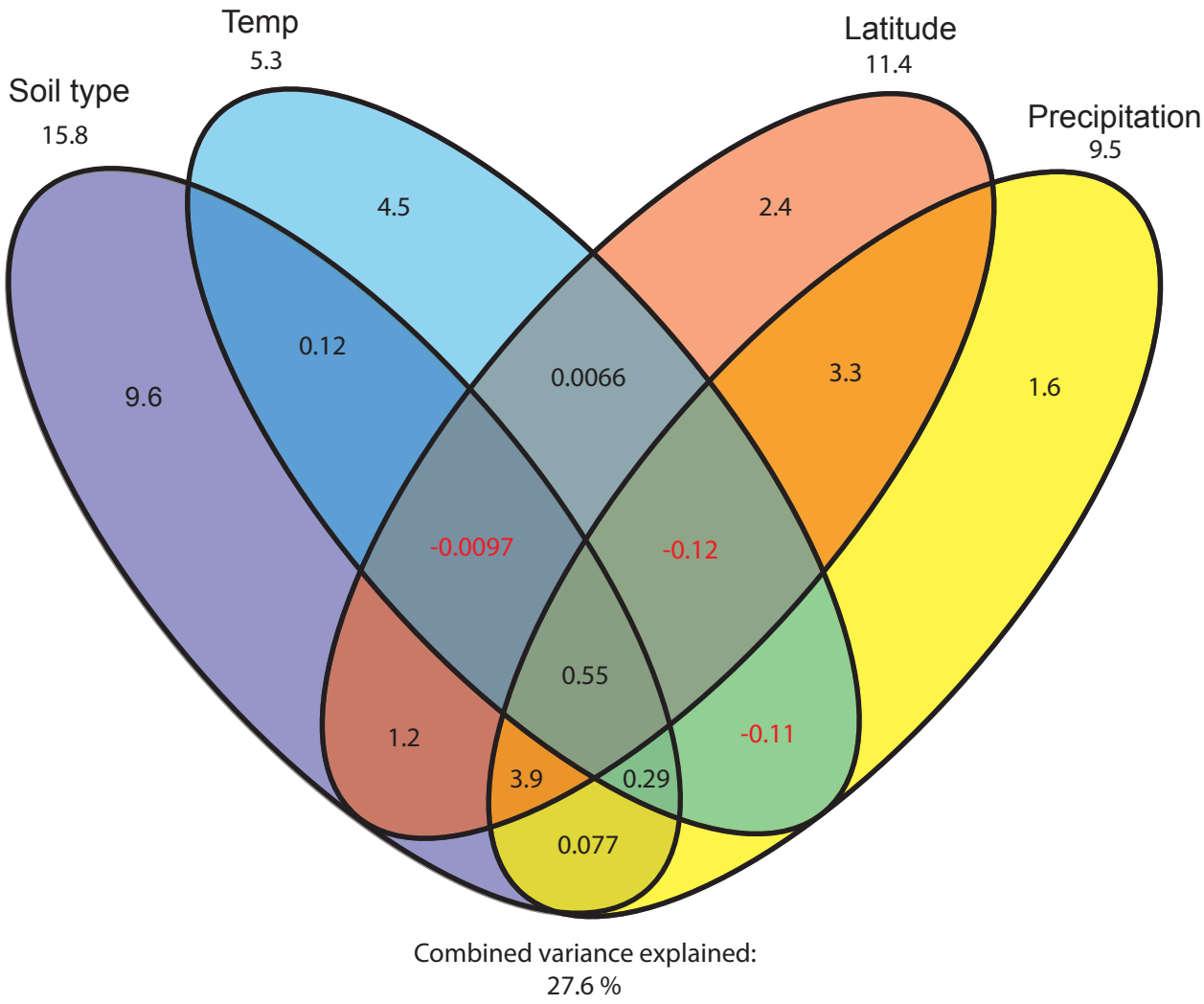


Fig S7: Venn diagram depicting the overlap in percent variation explained by the environmental variables found to be significantly explanatory in our forward selection analysis, for the observed ordination of geographic grids by Mesorhizobium phylogenetic community similarity (as per dataset S3).

Fig S8A

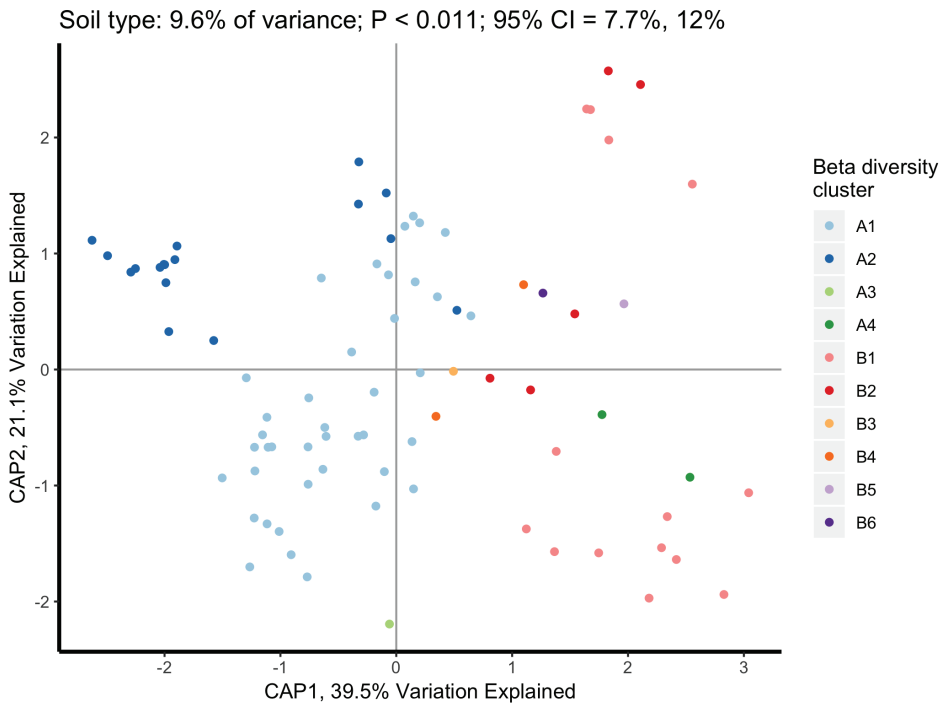


Fig S8B

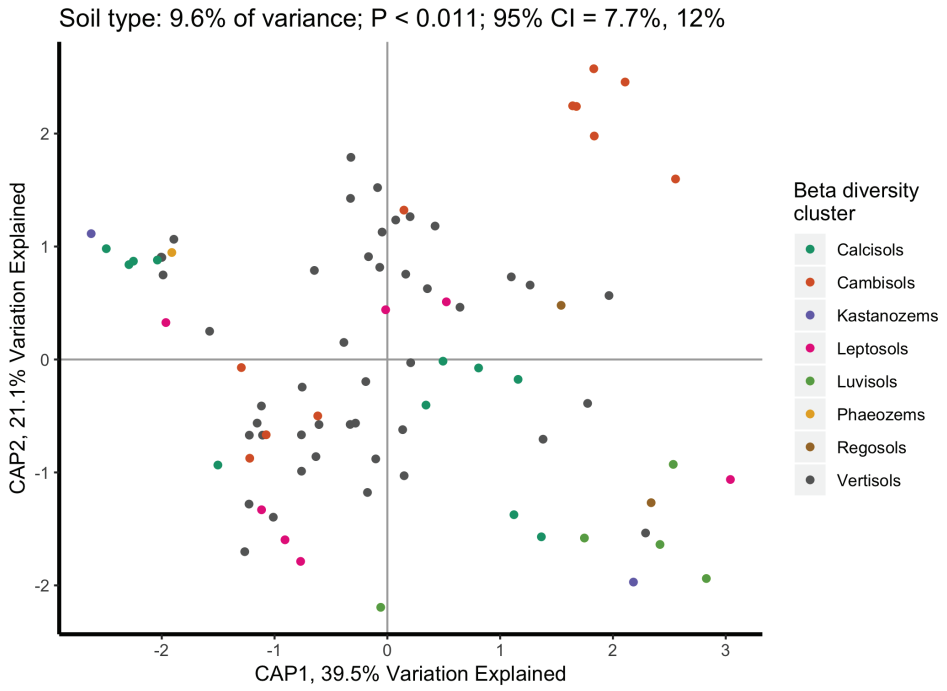


Fig S8: Constrained analysis of principal coordinates whereby ordination of geographic sampling grids by Mesorhizobium phylogenetic community dissimilarity is constrained by soil type, conditioned on mean annual temperature, mean annual precipitation, and latitude. **(A)** Colored by biodiveristy cluster from Fig. 2B. **(B)** Colored by soil type.

Fig. S9A-C

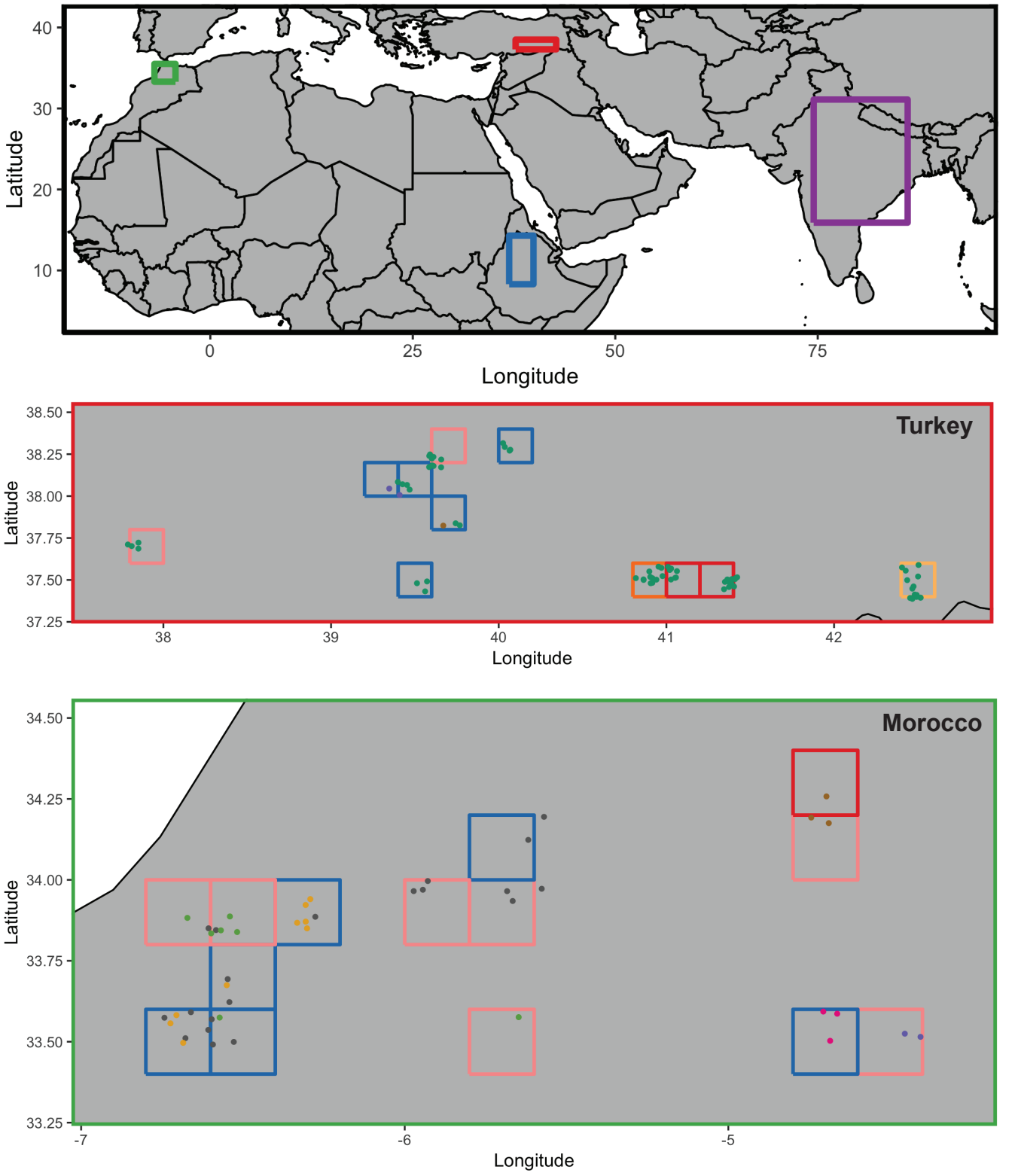
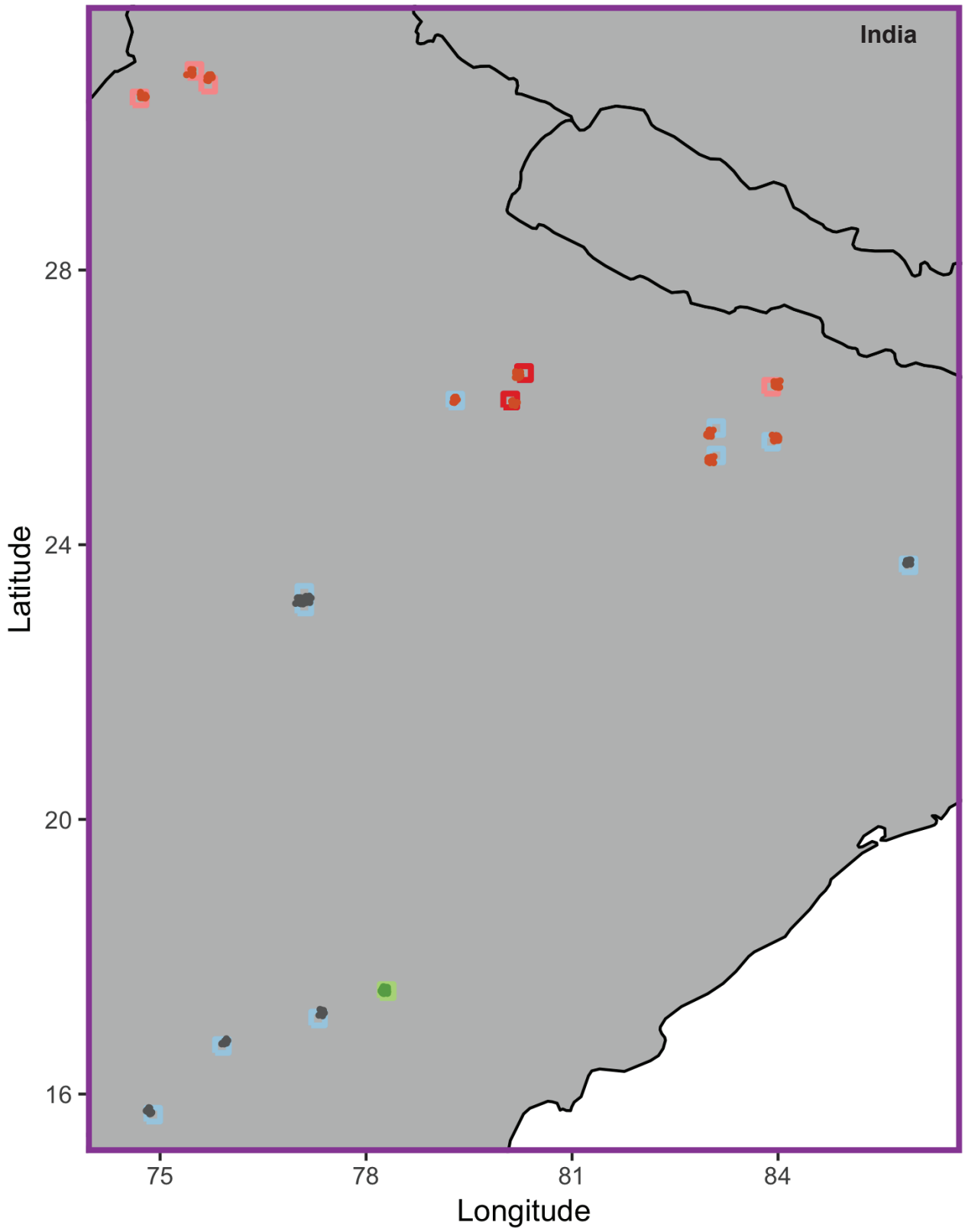


Fig. S9: 0.2 degree sampling grids displayed in geographic space and colored by beta-diversity cluster (**Fig 2B** dendrogram). Point colors represent FAO soil classification accessed through the International Soil Reference and Information Center SoilGrids250 dataset for sample-collection points.

Fig. S9D



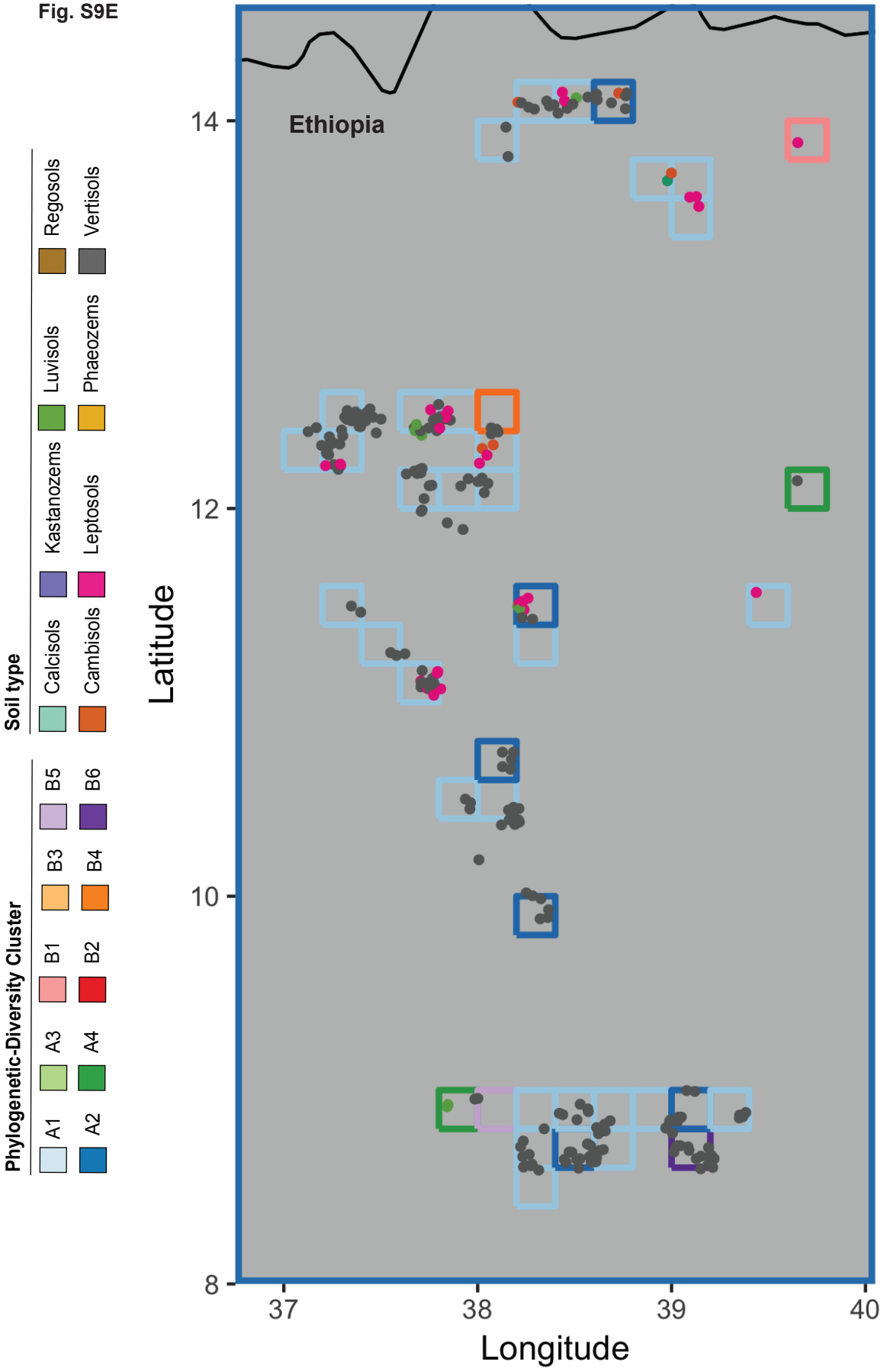
Phylogenetic-Diversity Cluster

A1 A3 B1 B3 B5
A2 A4 B2 B4 B6

Soil type

Calcisols Kastanozems Luvisols Regosols
Cambisols Leptosols Phaeozems Vertisols

Fig. S9E


Soil type
Calcisols
Kastanozems
Luvisols
Regosols
Cambisols
Leptosols
Phaeozems
Vertisols
Phylogenetic-Diversity Cluster
A1
A2
A3
A4
B1
B2
B3
B4
B5
B6
Ethiopia
Latitude
Longitude
14
12
10
8
37
38
39
40

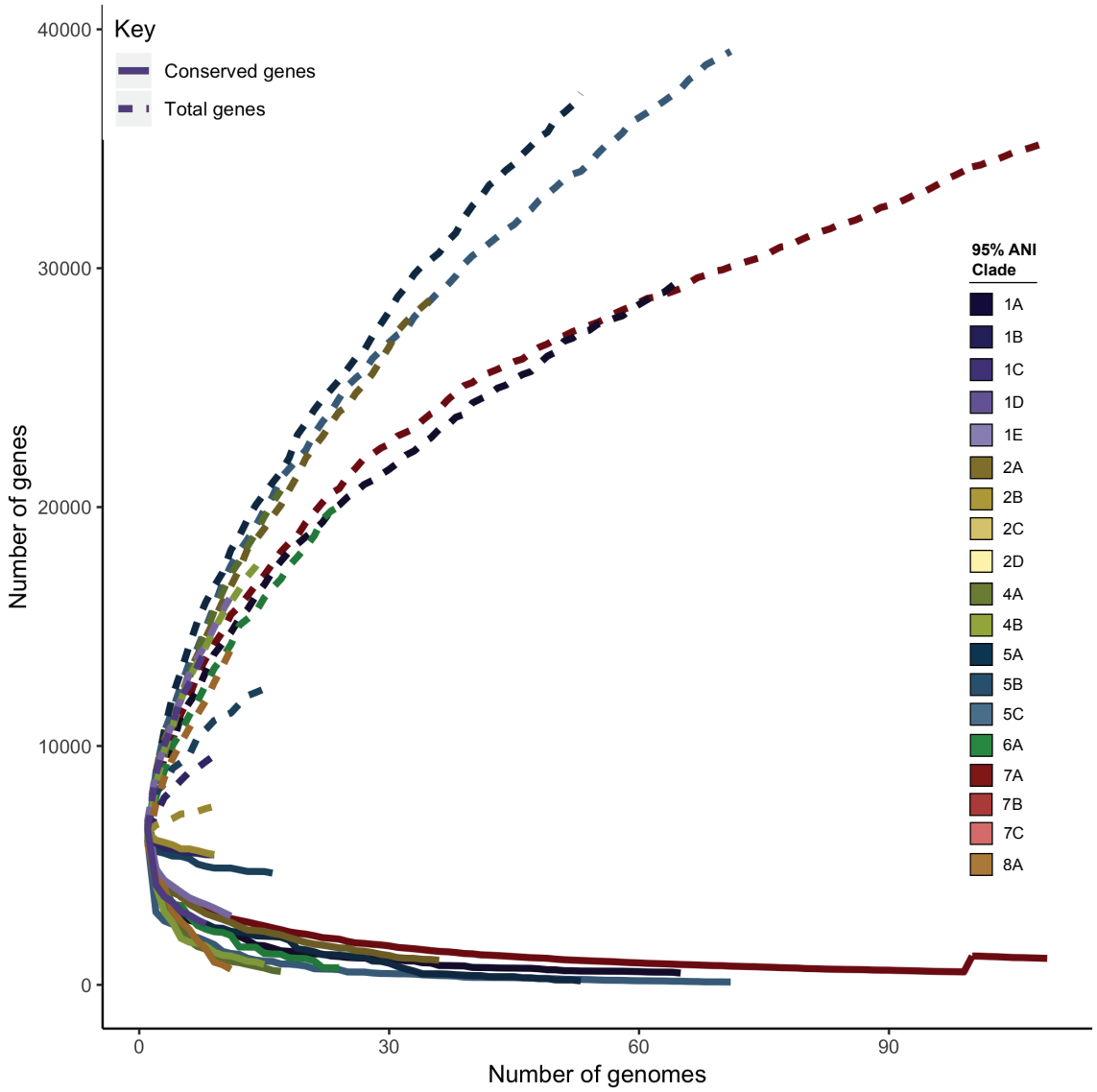


Fig. S10: Pangenome gene accumulation curves for each 95% ANI group. Lines depict the average number of genes (core or accessory) present across rarefied genomes, with 10 replications, as the number of genomes increases.

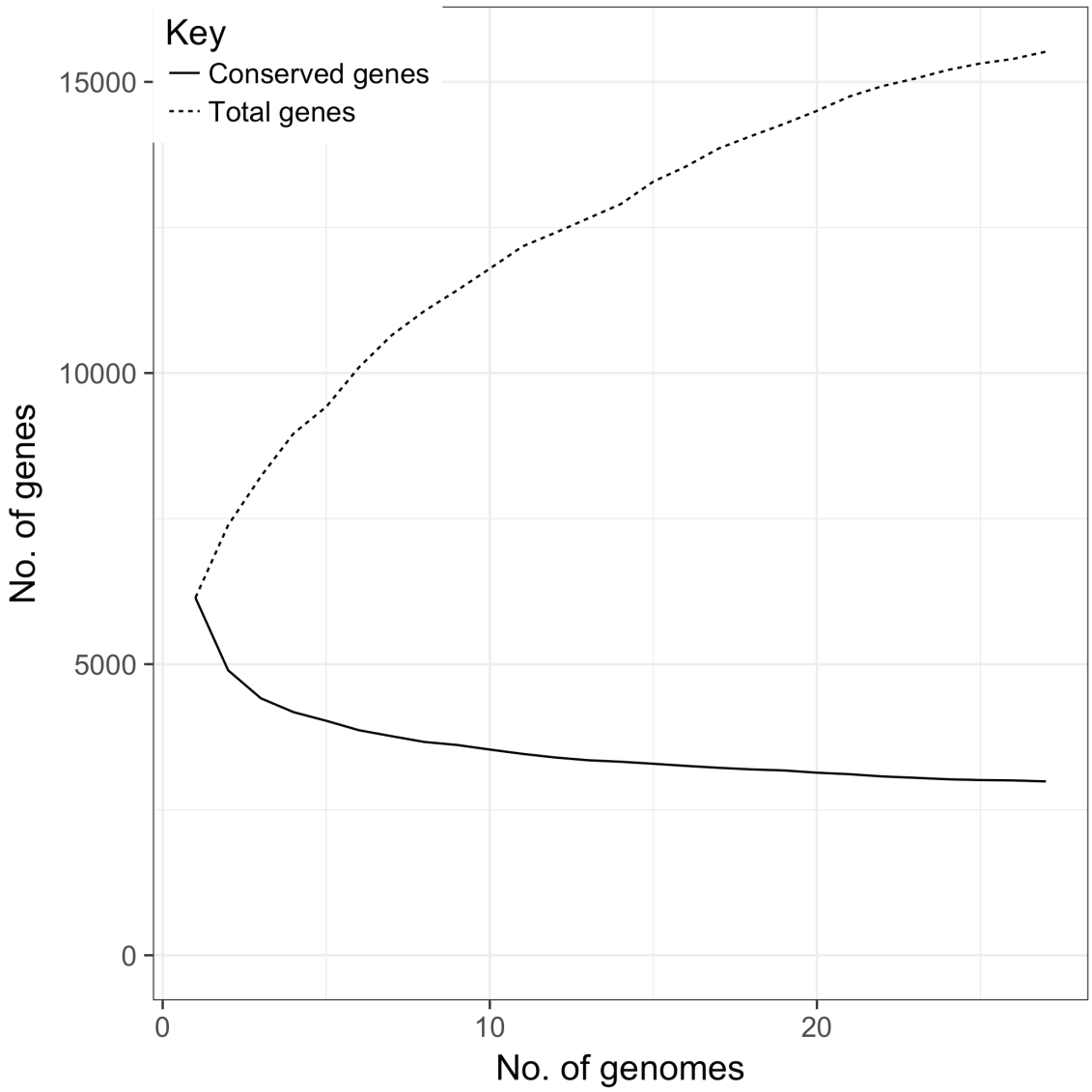


Fig. S11: Pangenome gene accumulation curve for 27 high-quality genomes from ANI95 group 1A from a single field in Southern India. The curve highlights the striking gene presence- absence diversity of even narrowly-defined bacterial lineages in circumscribed areas, as genomes from this single group from this single field appear to share less than half of their genes (~3000) with the whole group and contain greater than 15,000 orthologous groups of genes across the 27 genomes (with no indication that gene-level diversity will stabilize without much greater sampling).

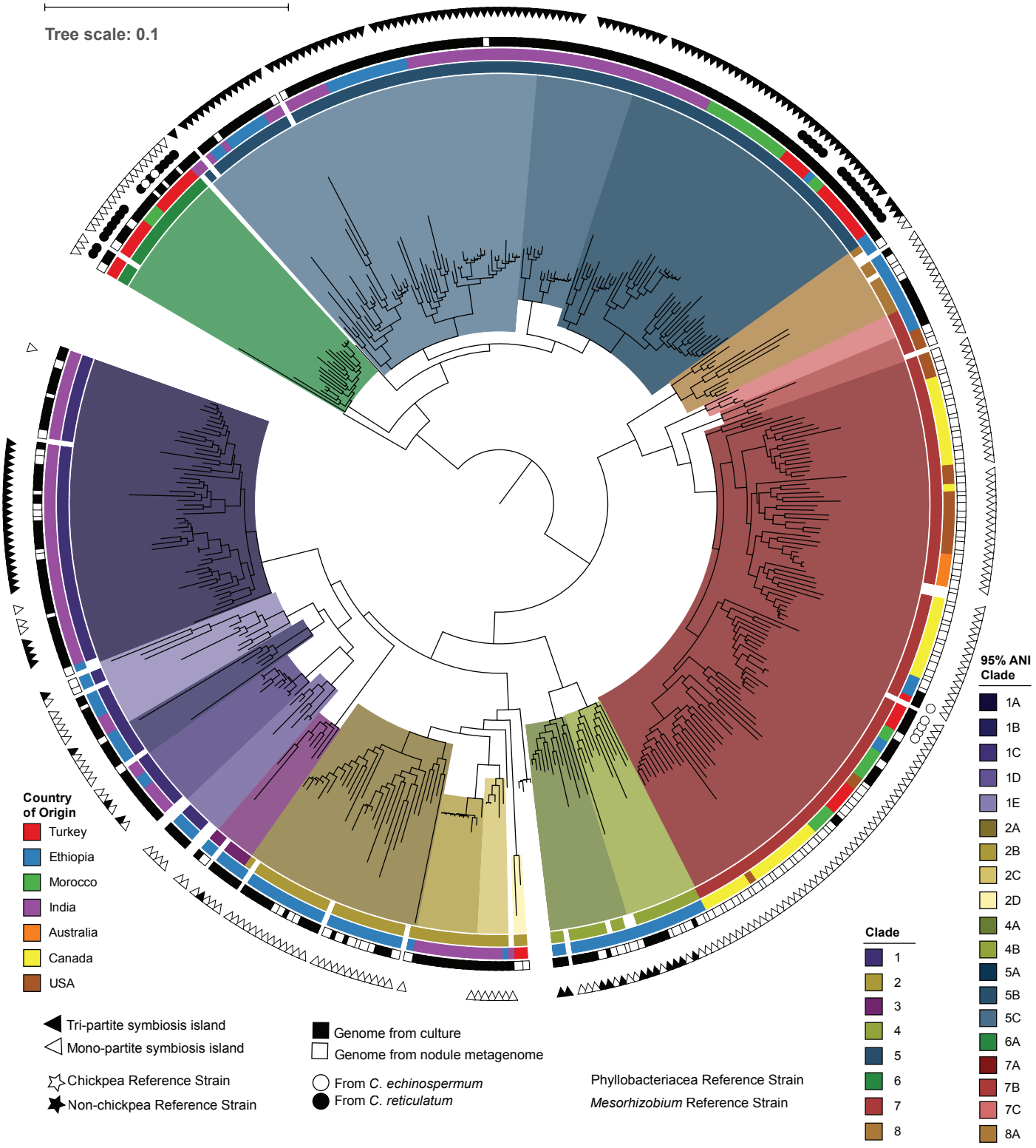


Fig. S12: Maximum-likelihood phylogeny inferred from the the pres- ence/absence of each gene in the genus-wide pangenome. Order and color of annotation rings as per fig 1A.

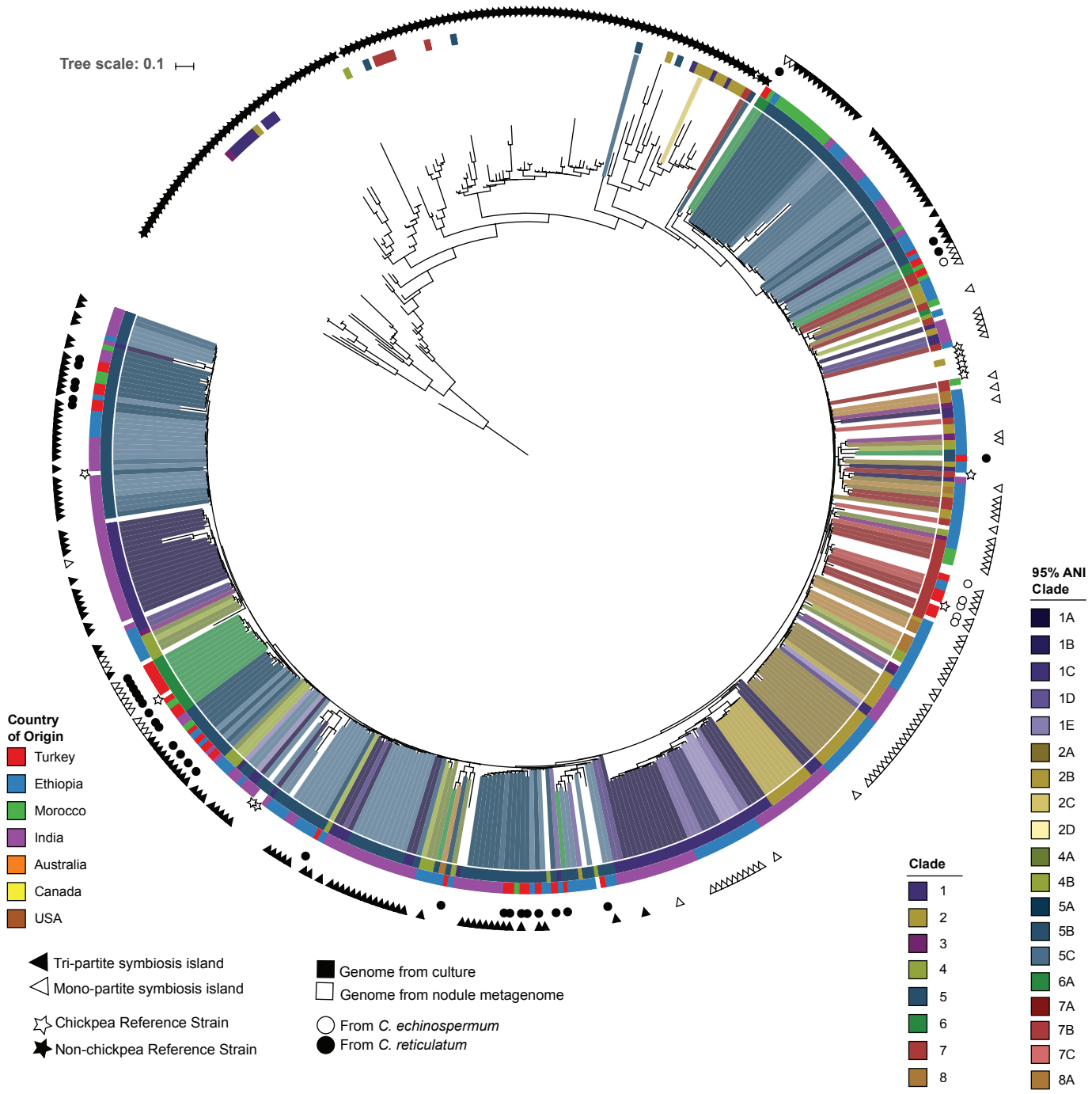


Fig. S13: Maximum-likelihood phylogeny inferred from concatenated nucleotide alignments of 16 symbiotic-nitrogen-fixation genes determined to be core to all nodule-forming Mesorhizobium genomes (both from bacterial cultures and DNA extract from chickpea nodules, as well as from the NCBI assembly database). Annotation order and color as per fig. 1A. This figure shows that phylogenetically the core symbiosis genes from chickpea isolates are monophyletic and closely-related relative to those from strains that nodulate other legume hosts.

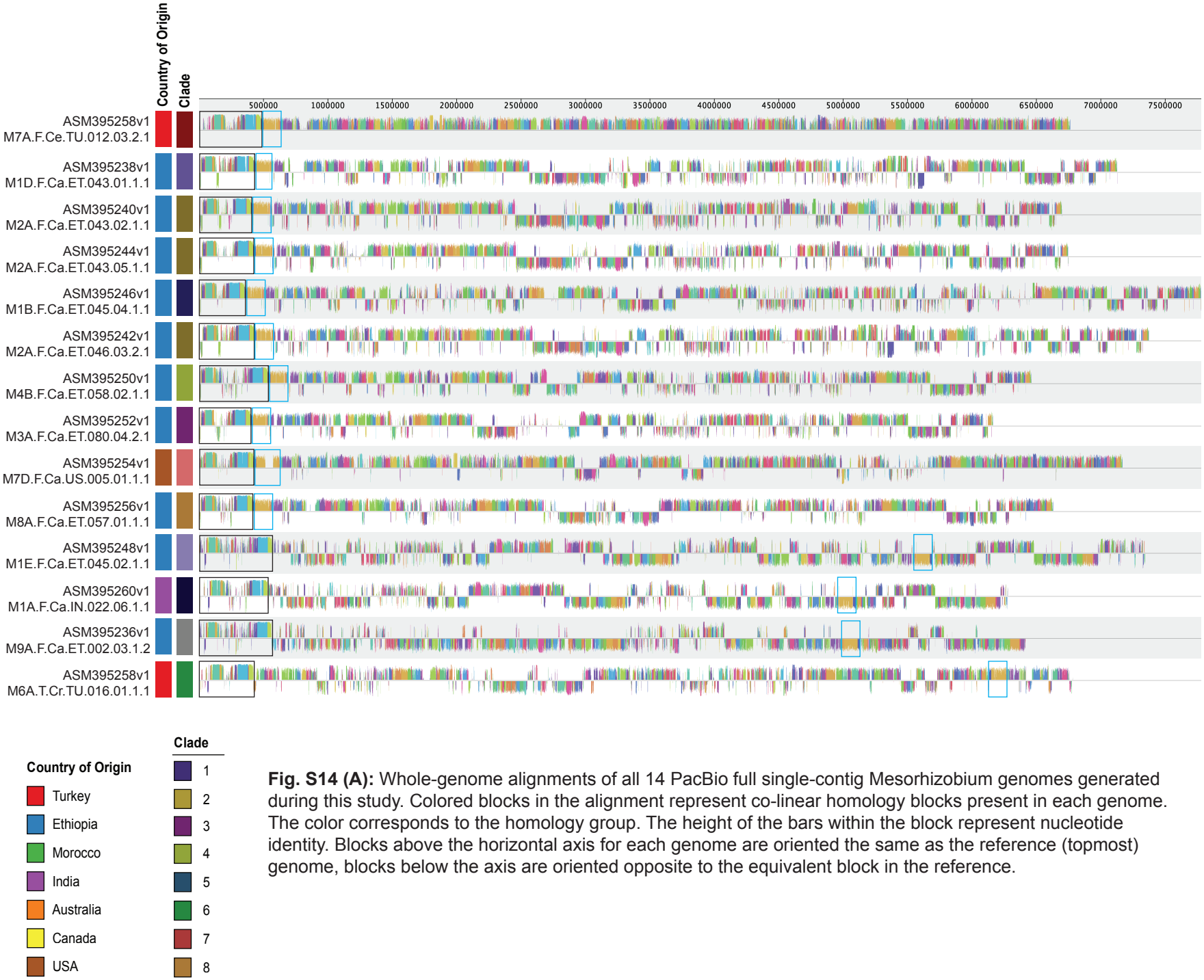


Fig. S14 (A): Whole-genome alignments of all 14 PacBio full single-contig Mesorhizobium genomes generated during this study. Colored blocks in the alignment represent co-linear homology blocks present in each genome. The color corresponds to the homology group. The height of the bars within the block represent nucleotide identity. Blocks above the horizontal axis for each genome are oriented the same as the reference (topmost) genome, blocks below the axis are oriented opposite to the equivalent block in the reference.

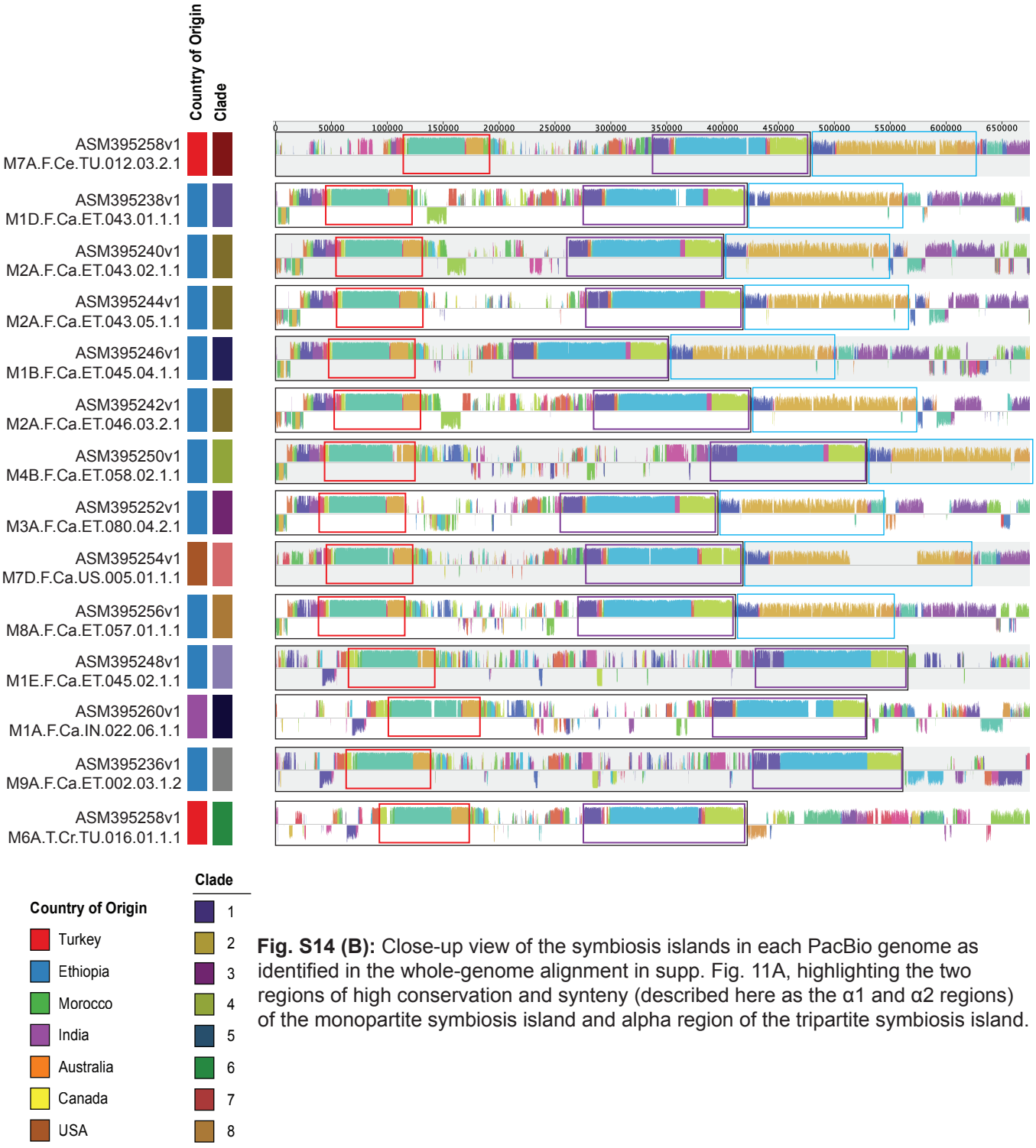


Fig. S14 (B): Close-up view of the symbiosis islands in each PacBio genome as identified in the whole-genome alignment in supp. Fig. 11A, highlighting the two regions of high conservation and synteny (described here as the α1 and α2 regions) of the monopartite symbiosis island and alpha region of the tripartite symbiosis island.

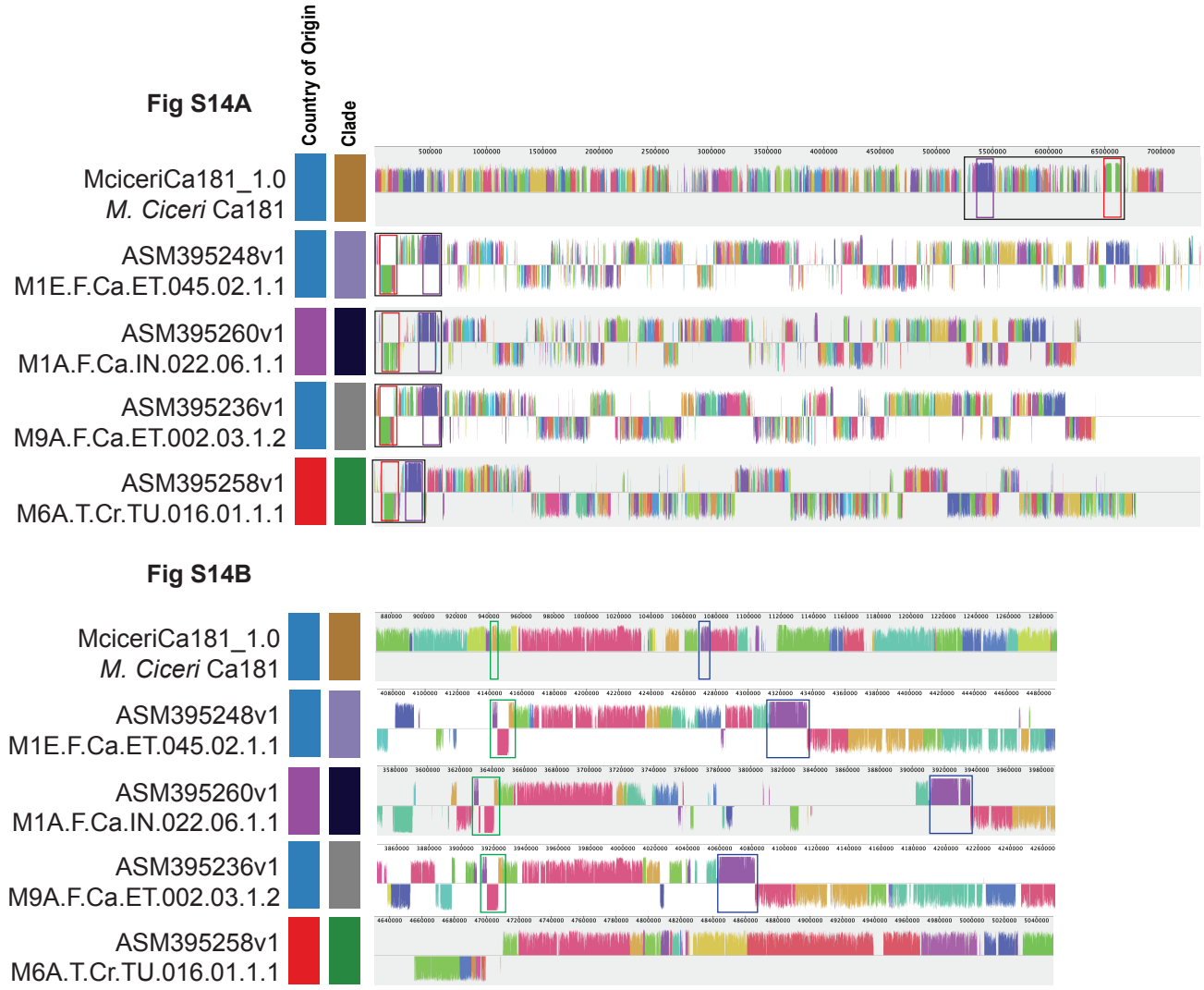


Fig S15: A: whole-genome alignment of pacbio genomes where primary symbiosis region has not inserted in tRNA- ser(cga) with the alpha region of the symbiosis highlighted against the published genome for strain ca181. Note that in the alpha region in ca181, the conserved flank containing the nod/nif genes is inverted relative to the other sequenced chickpea alpha symbiosis island, and is located a farther distance on the chromosome, suggesting either further chromosomal re-arrangement or assembly error. **B:** whole-genome alignment of pacbio genomes where primary symbiosis region has not inserted in tRNA- ser(cga) zoomed in on the beta and gamma regions of the tri-partite symbiosis island identified by Haskett et al (67) in strain ca181. These regions are present and contain the IntS2, IntG and IntM genes identified by Haskett et al., however have a different (and more conserved) structure in three of the four the genomes sequenced for this study. The fourth genome does not contain these regions.

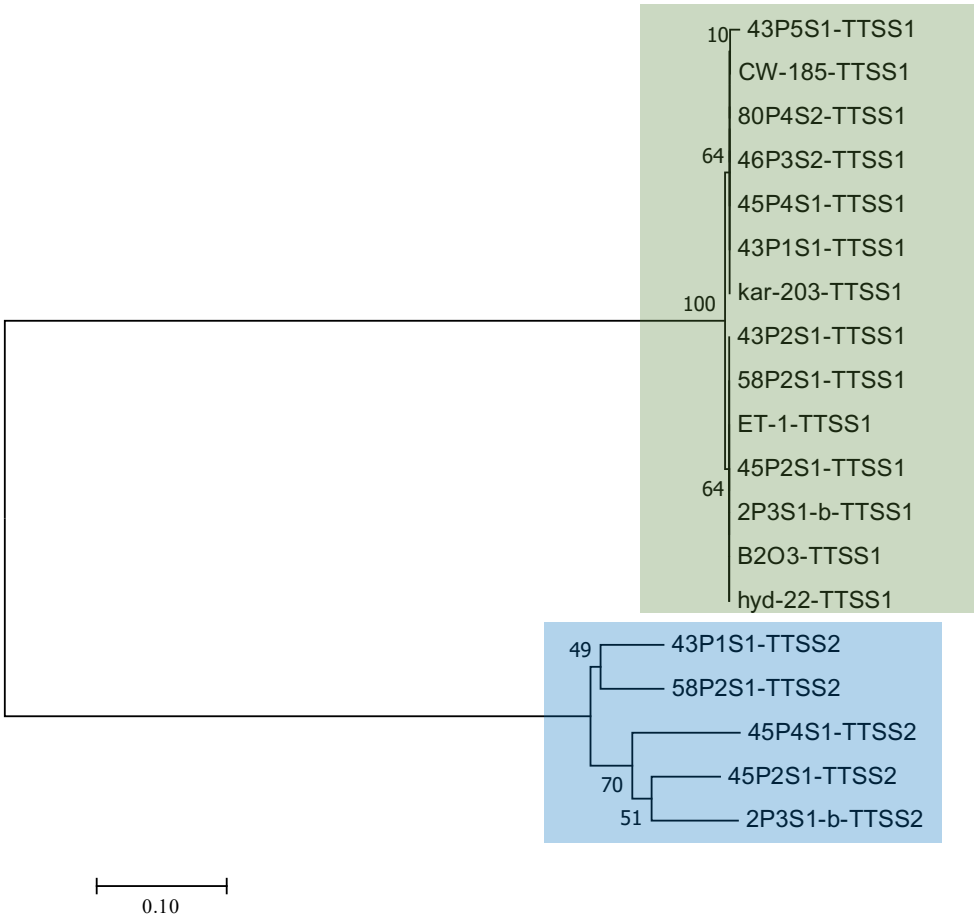


Fig. S16: Phylogeny of HrcC gene of TTSS from symbiosis island (orange box) and outsideof symbiosis island (blue box) in pacbio genomes.

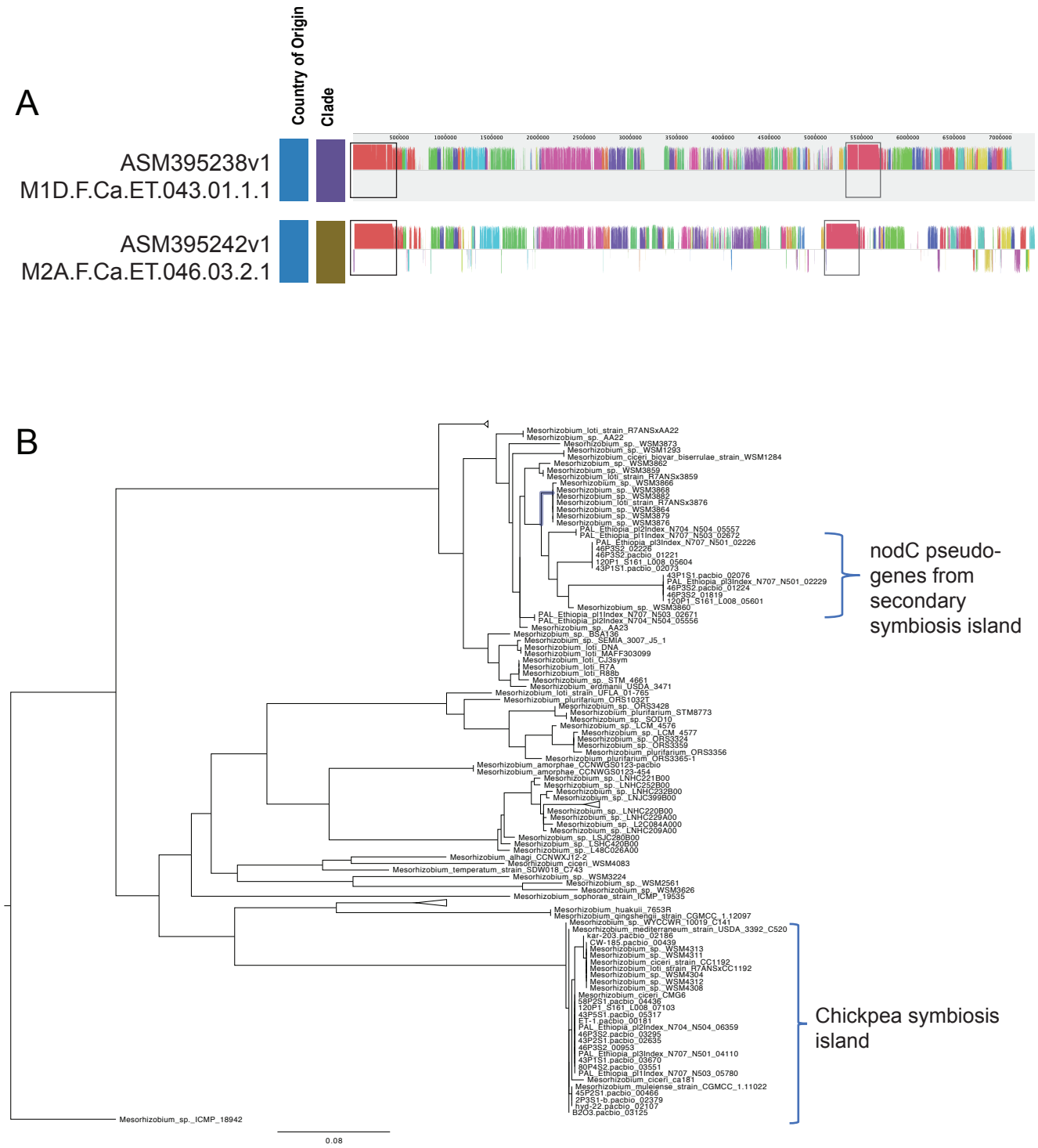


Fig. S17 (A): Whole-genome alignment of two PacBio genomes showing more-recently transferred symbiosis island (in magenta) as well as secondary non-functional symbiosis island (blue). **(B):** Phylogenetic tree depicting the relationship between the nod-factor synthesis gene nodC identified in the primary and secondary symbiosis islands in the genomes in supp. Fig. 14A.

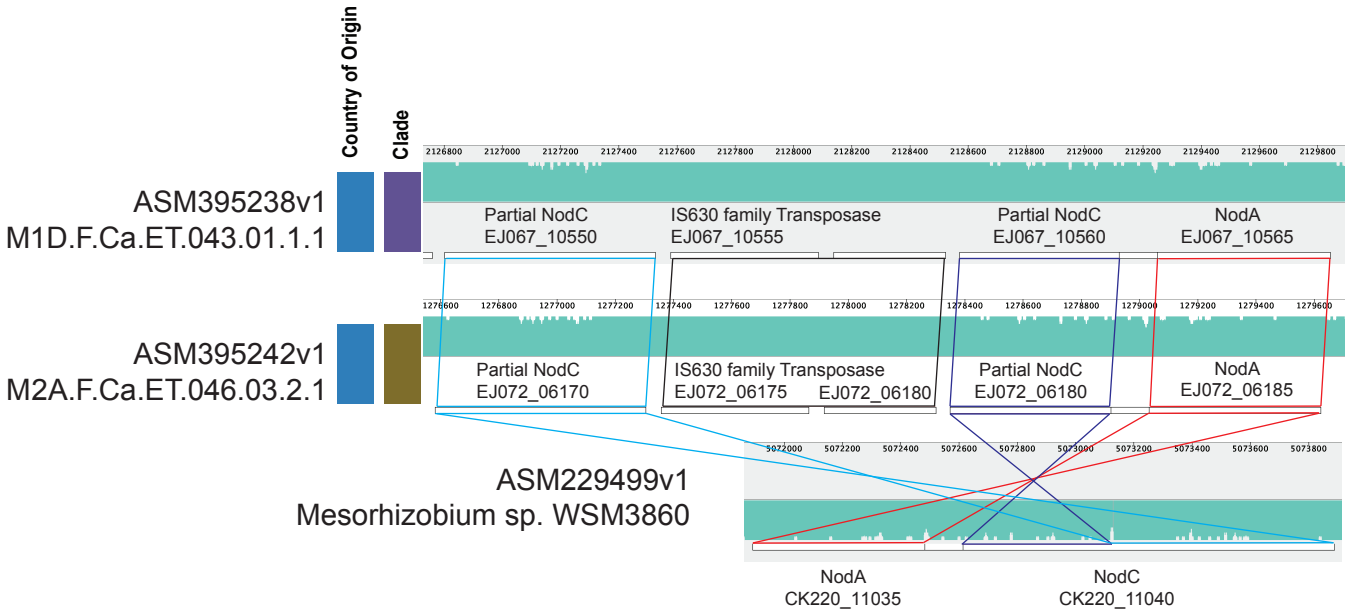


Fig. S17 (C): Secondary NodABC operon in two pacbio assemblies containing putative second symbiosis island, aligned to NodABC operon from reference genome of *Mesorhizobium* sp. WSM3860, the strain with the most closely-related nodC gene to the nodC pseudognes in the secondary symbiosis islands we describe. Alignment shows insertion of IS630 family transposase in nodC gene of the secondary symbiosis island of the chick-pea-nodulating strains.

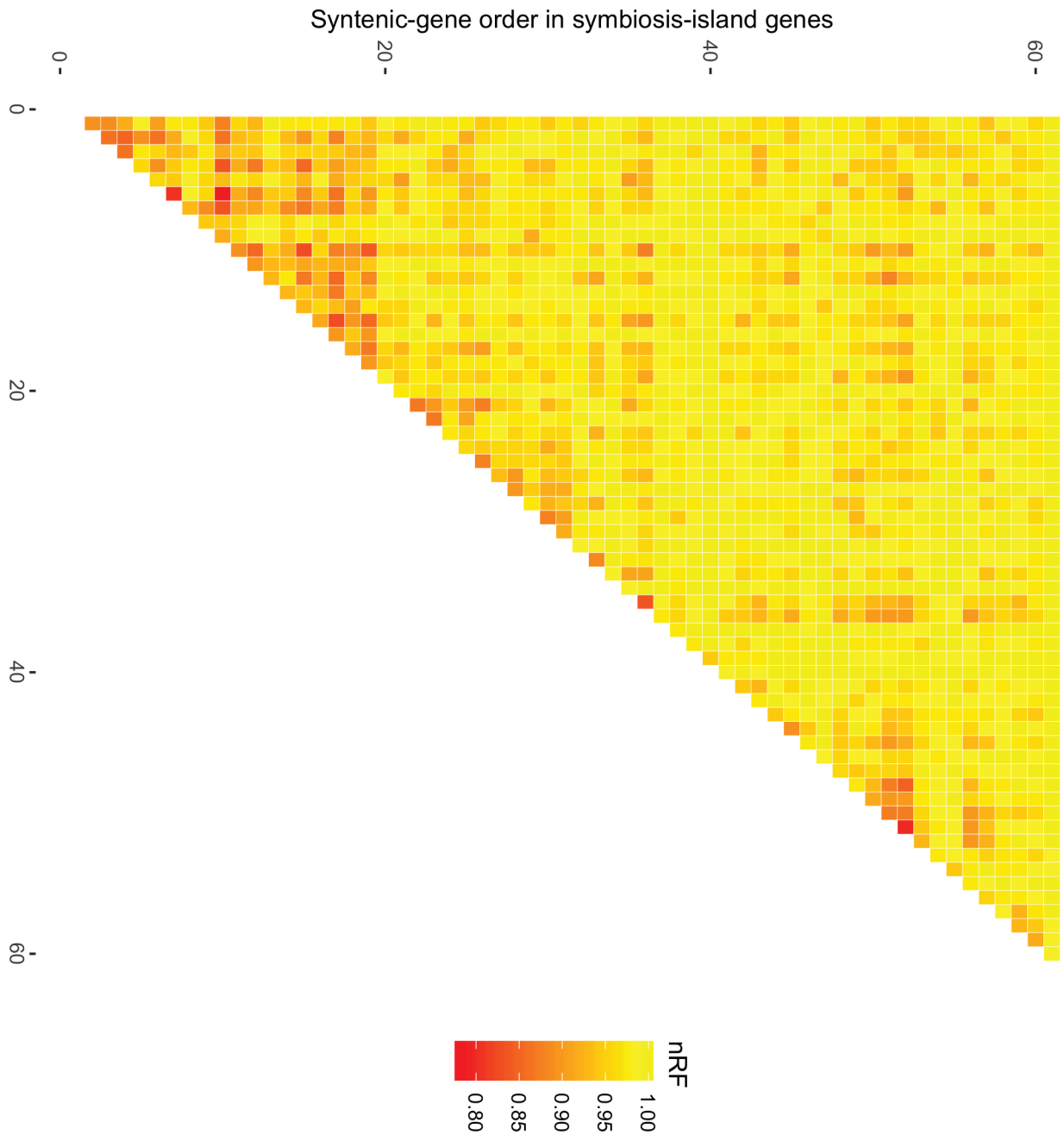


Fig. S18: Heatmap showing all pairwise comparisons of phylogenetic trees inferred from nucleotide sequences of symbiosis genes present in all 14 PacBio assemblies as well as 62 genomes assembled from root-nodules collected from wild-*Cicer* plants in southeastern Turkey, representing ANI95 groups 5A, 6A, and 7A.

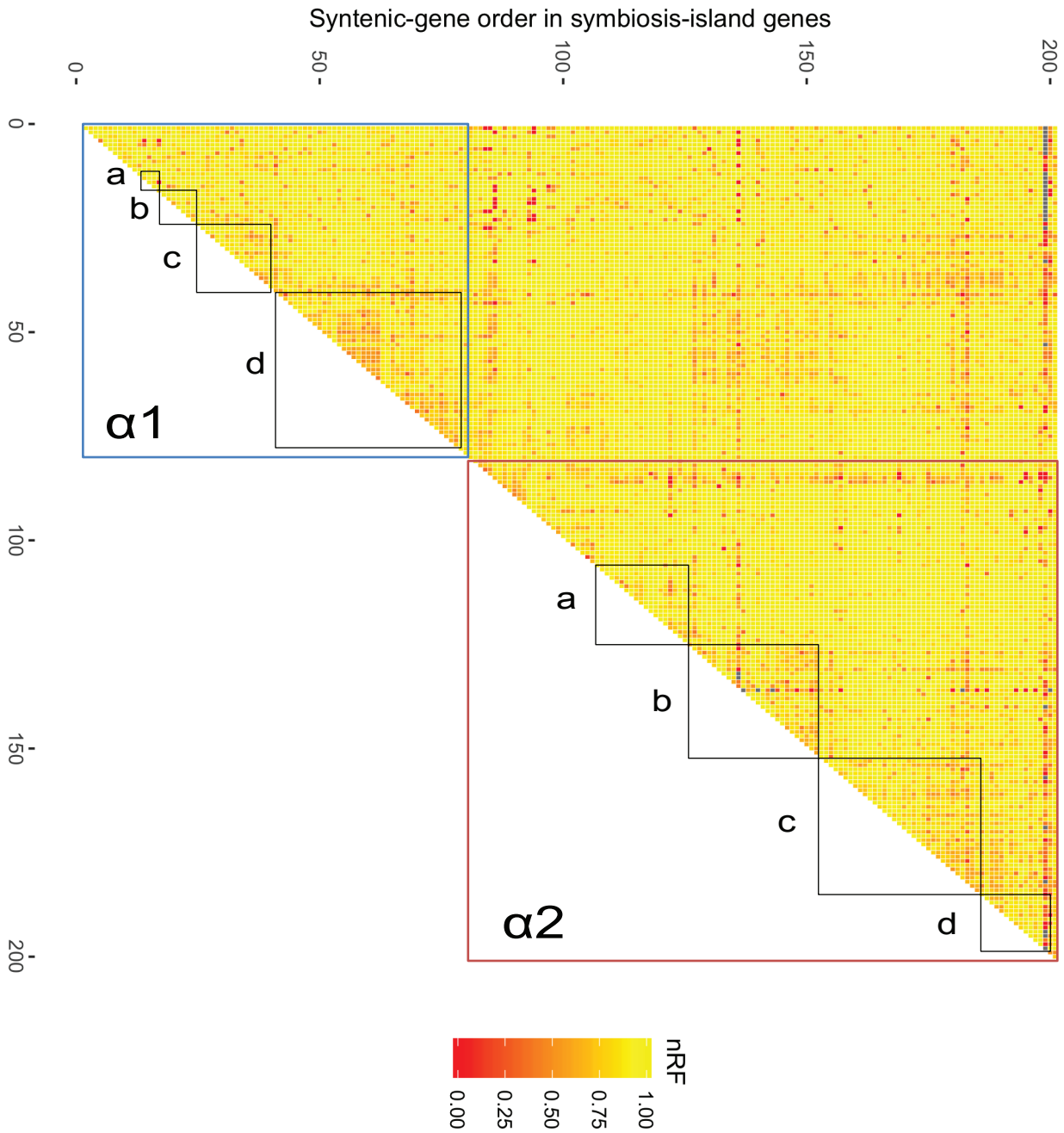


Fig. S19: Heatmap showing all pairwise comparisons of maximum-likelihood phylogenetic trees constructed from 200 individual genes inferred to be syntenic in all 14 PacBio symbiosis-islands, and present in 8 or greater genomes (whereas main-text **Fig. 4B** pairwise comparisons are of trees inferred from concatenated alignments of 10-gene slidinging windows). This figure shows that many of the same recombination patterns inferred in the sliding-window analysis are present albeit more weekly by comparisons of individual genes, and that certain individual genes have phylogenetic congruence with genes throughout the symbiosis island. Cells in the heat-map are colored by normalized Robinson-Foulds distance. Boxes are drawn around the two conserved regions of the symbiosis island, α1 and α2 also highlighted in **Fig S13B**. Within these regions, boxes are drawn around regulons of genes with related functions (as per Fig. 4B). These are: α1a: double-stranded DNA break repair; α 1b: hypothetical proteins; α1c: genes involved in nod-factor synthesis; α2d: genes involved in nitrogen fixation; α 2a: type-III secretion system and putative effectors; α2b: biofilm formation (including O-antigen, exopolysaccaride production, quorum-sensing genes, and the type-II secretion system); α2c: conjugation (type-IV secretion system, plasmid-transfer genes); α2d: cytochrome oxidases.