

Supplementary information for the manuscript entitled “Biogeographic and genomic signatures of thermal adaptation in facultative symbionts”

This includes:

Supplementary Methods

Supplementary Figures 1-5

Growth assays

Microplate reader setting. At the later stage of incubation, uneven water evaporation may alter the liquid volume in individual microplate wells, which in turn introduces measurement errors in the optical density. To account for evaporation, the microplate reader was set to consider the path length correction set at a test wavelength of 977 nm, a reference wavelength of 900 nm, and the K-factor at 0.18. To further account for liquid volume change throughout incubation, in our plate design, we intentionally left 32 wells uninoculated as the cell-free control and blank.

Growth traits estimation. We smoothed and extracted three growth traits from the growth curves: lag time, maximum growth rate, and maximum yield. Briefly, we first subtracted the OD_{600} of cell cultures from the cell-free cultures as blank. To compute the maximum growth rate, we smoothed the $\log(OD_{600})$ of each replicate by fitting a generalized additive model using the *gam* function from the R package *mgcv* (Wood 2011). We excluded the first 1 hr and any OD_{600} below 0.015 to avoid artifacts introduced by measurement and fitting noises, respectively. We then took the maximum derivative as the maximum growth rate. Lastly, the yield is determined by the maximum OD reading. As a result, we obtained 12 traits per strain (three growth traits at four temperatures) from the growth curve experiments.

SUPPLEMENTARY FIGURES

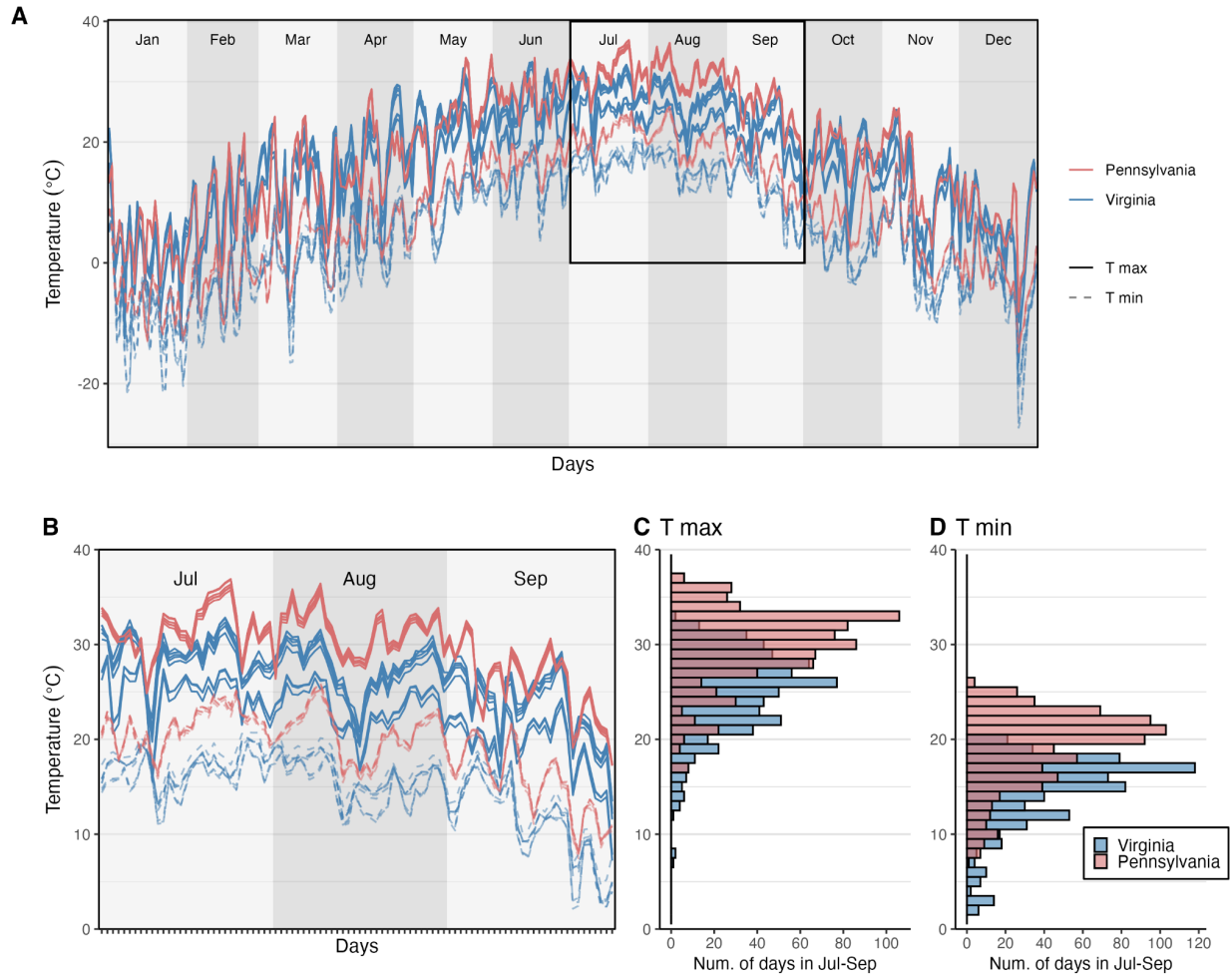


Figure S1. Daily temperature in the regions (Pennsylvania and Virginia) of sampling. (A) Daily temperature maximum ($T_{\max}$) and minimum ($T_{\min}$) in the year of 2022. (B) Daily temperature in the sampling season July to September. In panels A and B, each line represents a distinct sampling site, with line colors indicating regions. (C) Histogram of daily maximum by the number of days during July to September. (D) Daily minimum.

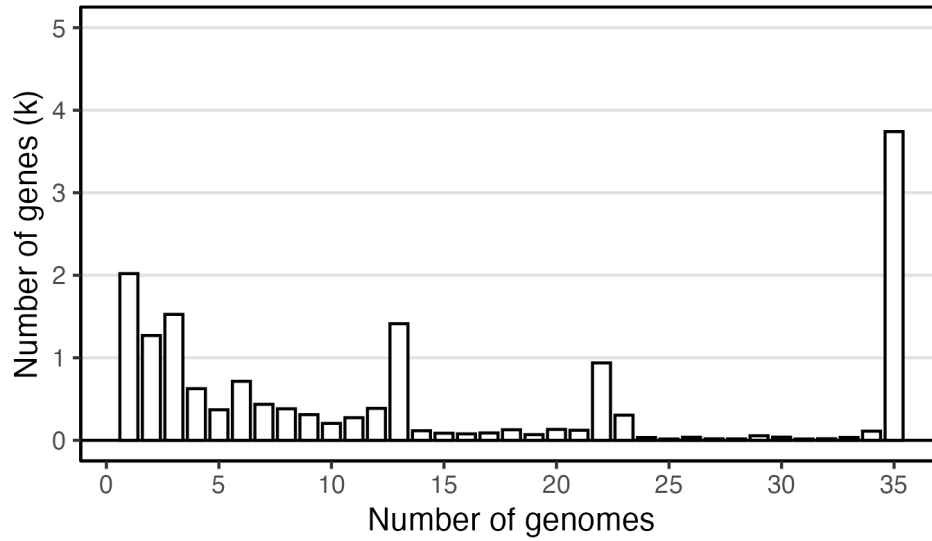
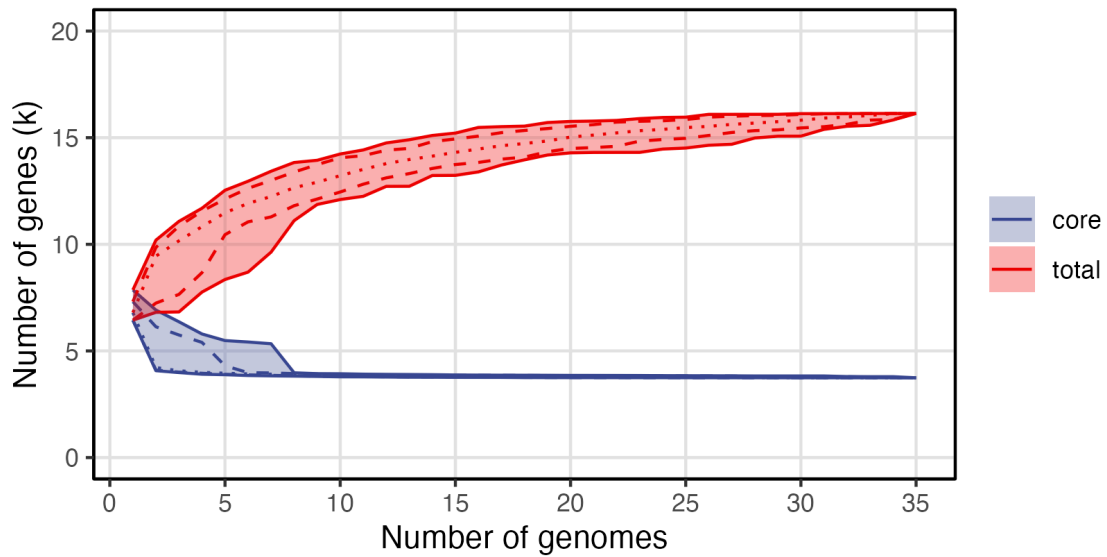
A**B**

Figure S2. Pangenome sampling of 35 *Sinorhizobium* genomes. (A) Gene frequency spectrum shows the 2,020 singletons (i.e., only present in one but not other genomes), 3,741 core genes (i.e., shared by all genomes; including 2,929 single-copy core genes), and a total of 16,135 genes in the pangenome of 35 *Sinorhizobium* genomes. (B) We estimated the sampling effort of genomes. At each number of genomes, we performed 100 bootstraps and computed the number of core genes and the number of total genes. The solid lines represent the maximum and minimum of the bootstrapped values. The dashed lines represent the 5% and 95 percentiles and the dotted line denotes the median.

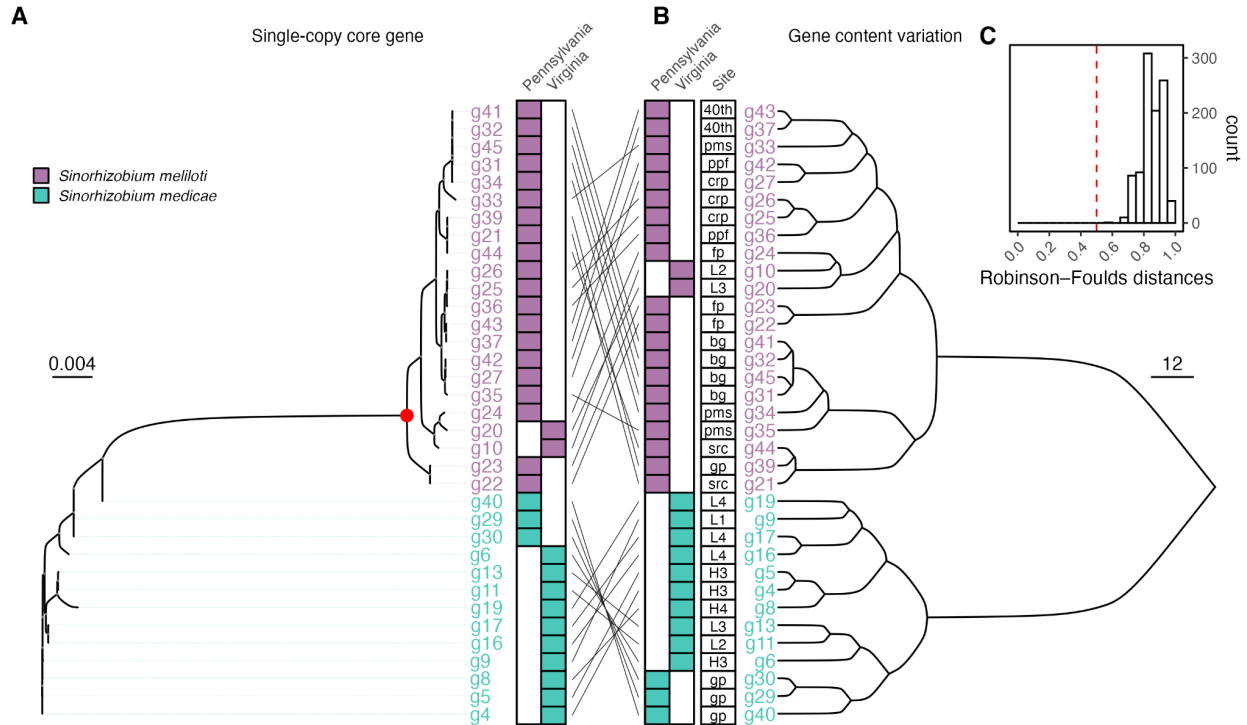


Figure S3. Congruence of core and accessory gene phylogeny. Panel (A) shows the phylogeny inferred from 2,929 single-copy core genes. The single-copy core-gene phylogeny represents the consensus tree of 1000 bootstrapped trees using the maximum likelihood approach. The branches with < 95% bootstrap support are collapsed. The internal node marked with a red point represents a 1/10 scaling to branch length for better presentation. (B) denotes the hierarchical clustering (ward.D2) of gene content variation inferred from the presence and absence of 16,135 genes. The heatmap represents the isolates' origin in the Philadelphia or Virginia region. (C) The core and accessory phylogeny shown in panels A and B are more congruent than random expectation. The red line denotes observation and the histogram represents 1000 randomizations. The distance between the two trees was calculated using the R package TreeDist::Robinson-Foulds .

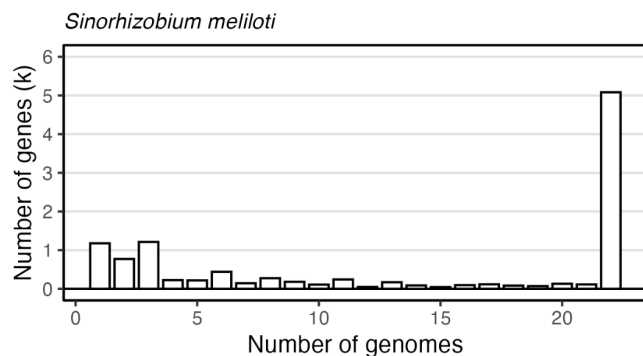
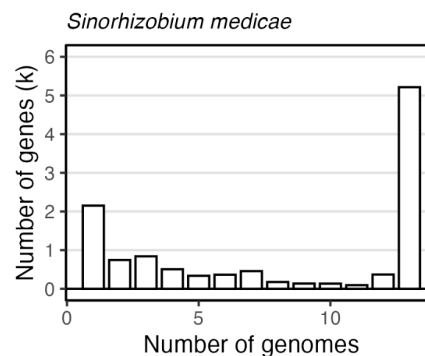
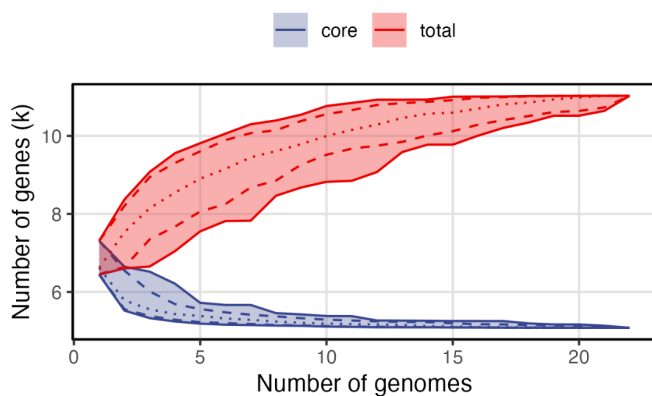
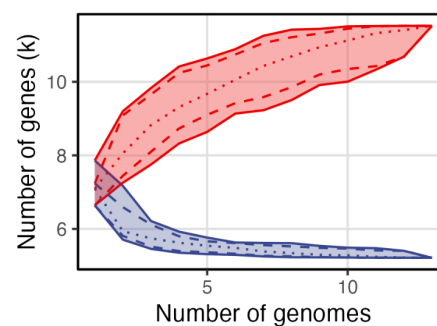
A**B****C****D**

Figure S4. Pangenome sampling within *Sinorhizobium* species. Panels A and C show the gene frequency spectrum and pangenome sampling in *S. meliloti*, respectively; Panels B and C show the corresponding figures in *S. medicae*. At each number of genomes, we performed 100 bootstraps and computed the number of core genes and the number of total genes. The solid lines represent the maximum and minimum of the bootstrapped values. The dashed lines represent the 5% and 95 percentiles and the dotted line denotes the median.

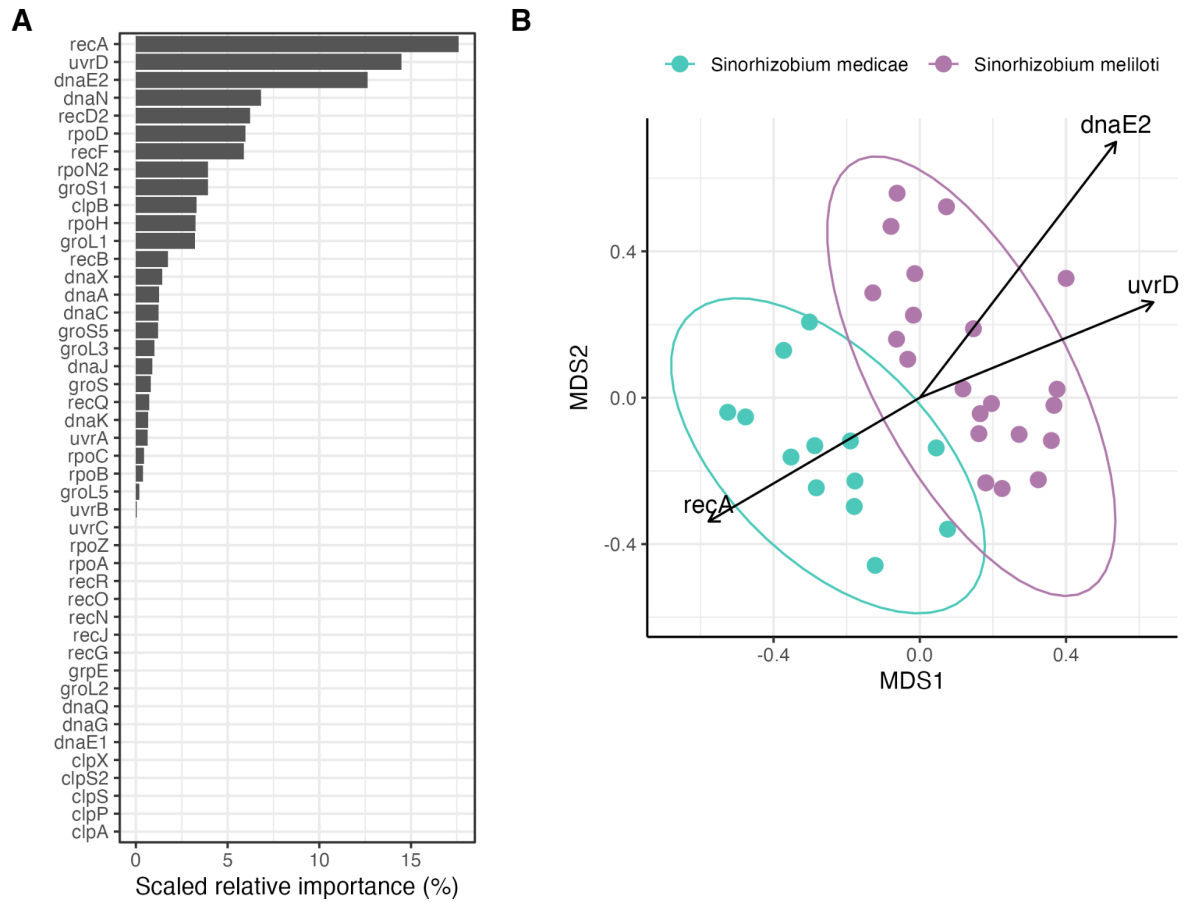


Figure S5. Random forest and ordination analyses of heat-relevant accessory genes differentiating *S. meliloti* and *S. medicae*. (A) Random forest variable importance scores (MeanDecreaseGini) were scaled to percentages to show the relative contribution of each gene to species classification. (B) Principal coordinates analysis (PCoA) based on Bray–Curtis dissimilarity of heat-relevant genes, with species shown by ellipses. Vectors from *envfit* indicate the direction and strength of key genes (*recA*, *uvrD*, *dnaE2*), showing that *uvrD* and *dnaE2* are enriched toward *S. meliloti*, whereas *recA* is more associated with *S. medicae*.