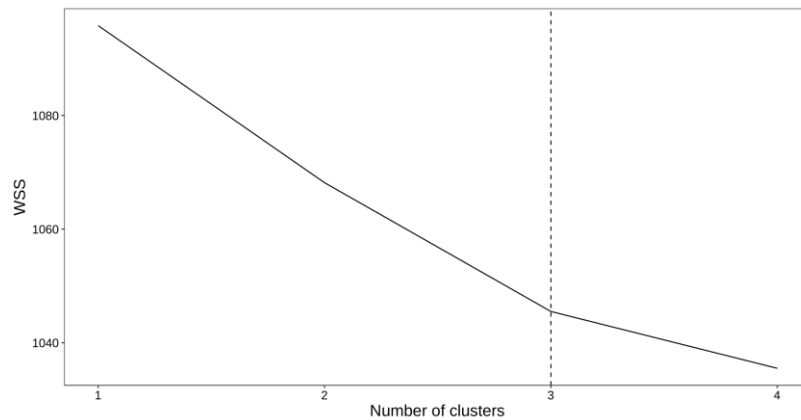
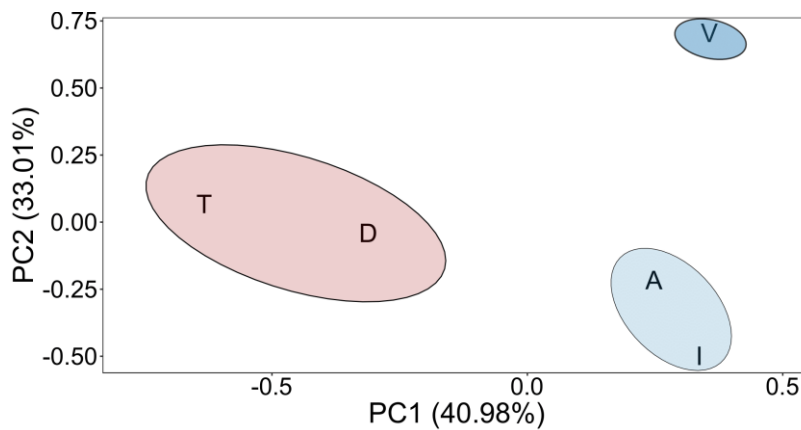


**Figure S1.** The optimal number of clusters for the k-means clusterization of the binarized 2-DIGE data using the elbow method. The dashed line represents the number selected (2) depicted in Figure 3a of the main text.

**a**



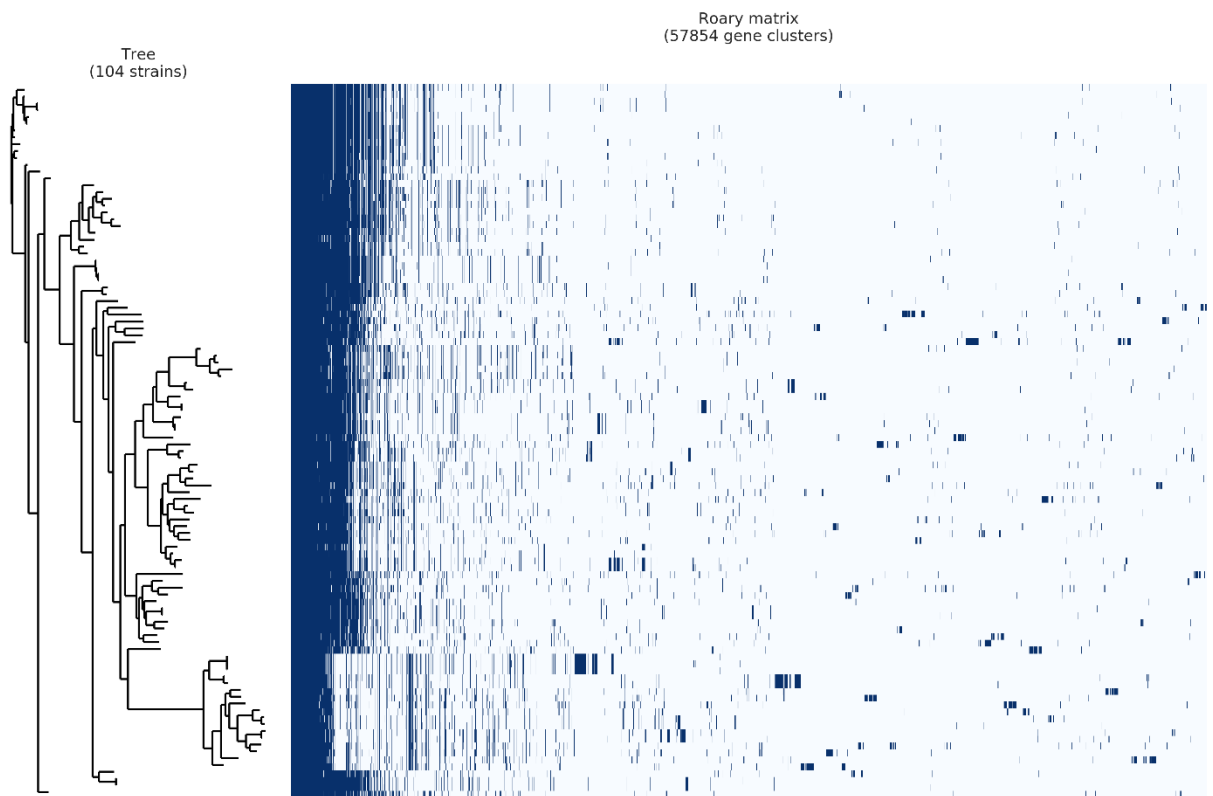
**b**



**Figure S2.** (a) The optimal number of clusters for the k-means clusterization of all DIGE spots using the elbow method. The dashed line represents the number selected. (b) The PCA and k-means clustering results obtained by utilizing all the protein spots found in the 2D-DIGE gel for strain 800/3 vegetative culture (V), strain 800/3 virulent sporulating (S) culture, and strain 800/3-15 avirulent sporulating (A).

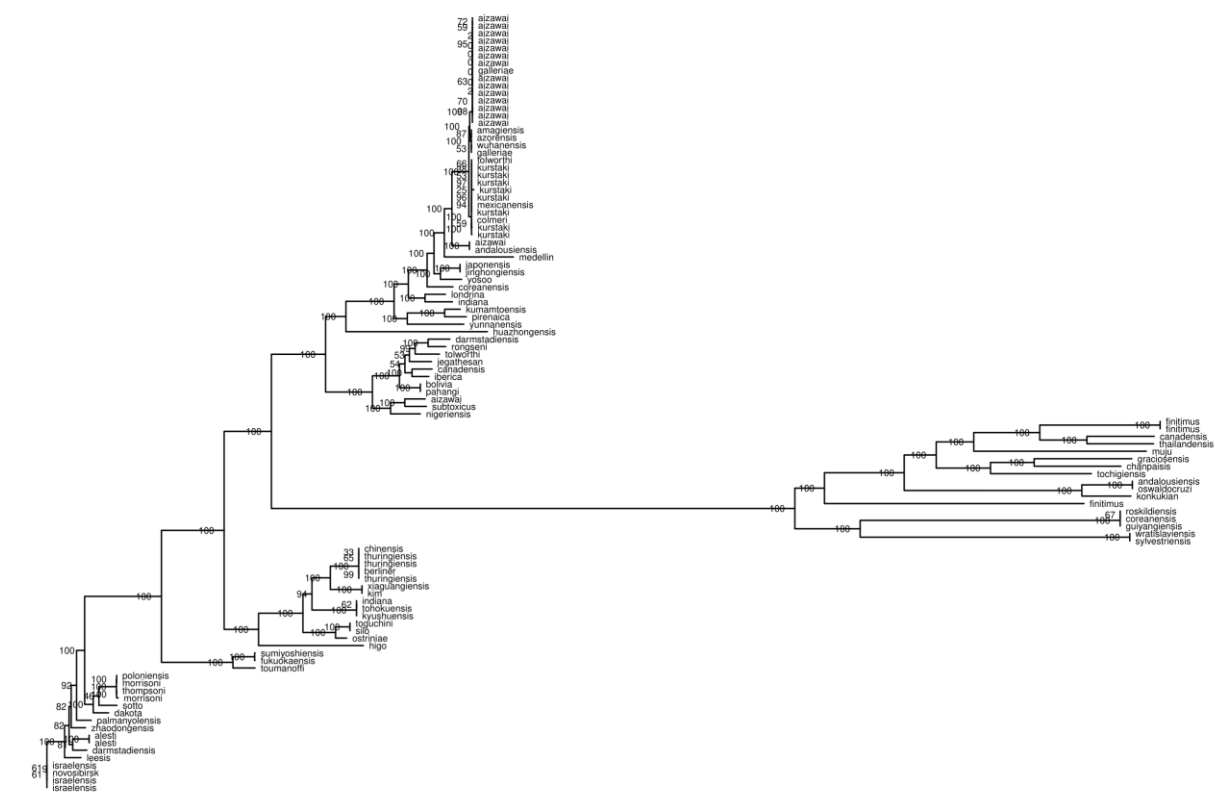


**c**



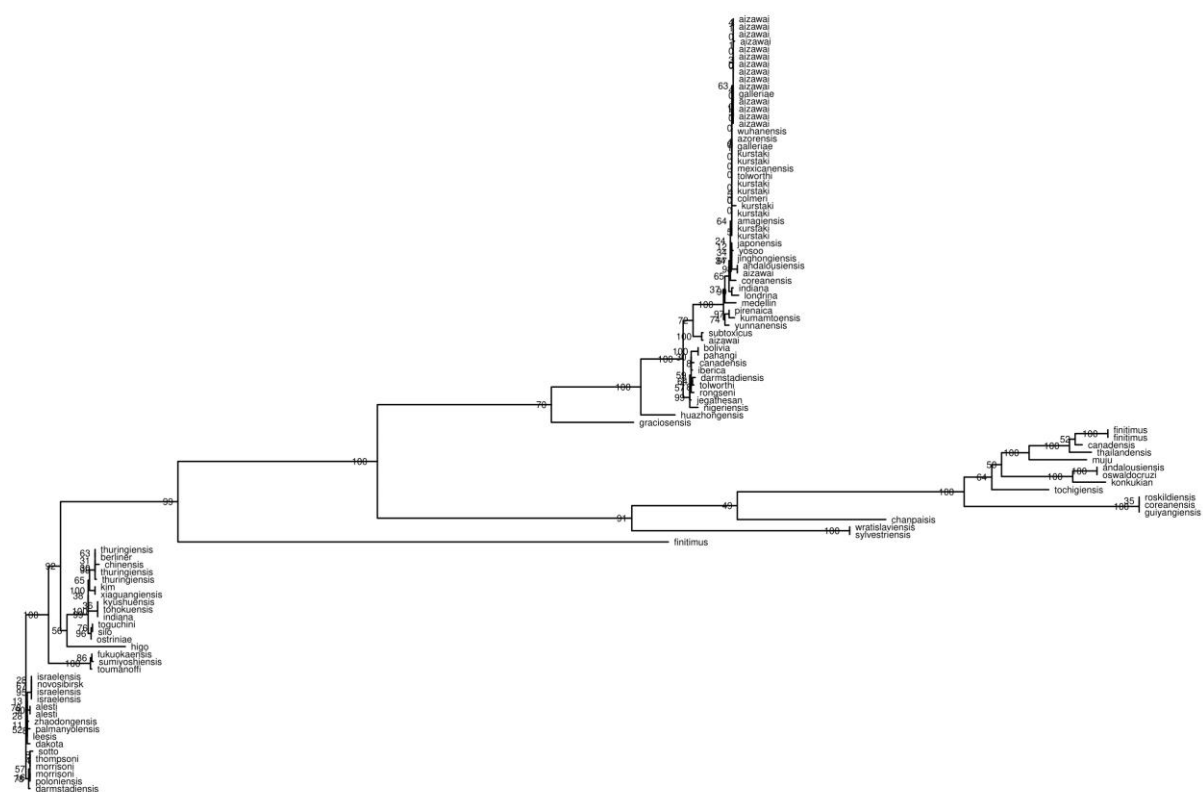
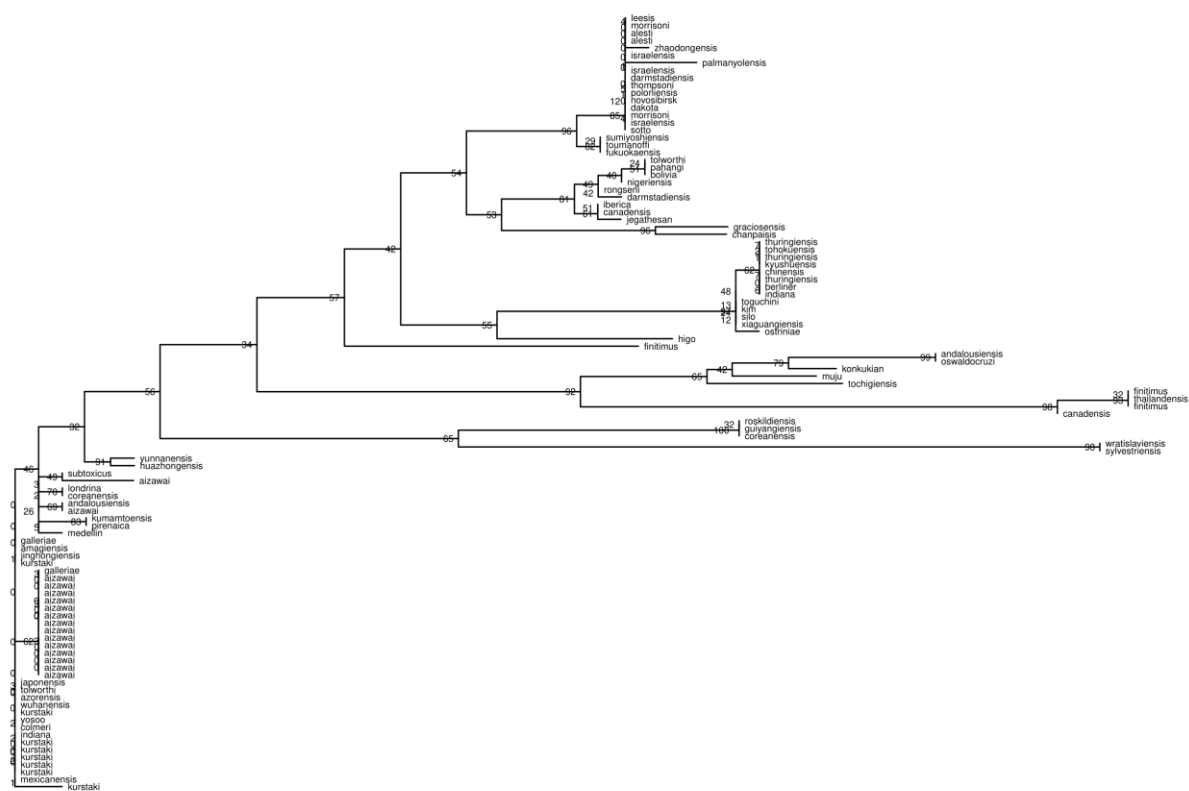
**Figure S4:** Visualization of Roary-obtained pangenome reconstructed on 104 pre-filtered Bt assemblies. (a) The link between the number of genomes and the number of new genes in the pangenome. (b) A pie chart visualizing the number of core and cloud genes in the pangenome. (c) The presence/absence matrix based on the accessory genes and the corresponding tree. The detailed description of the plots obtained is available at <http://sanger-pathogens.github.io/Roary/>.

**a**

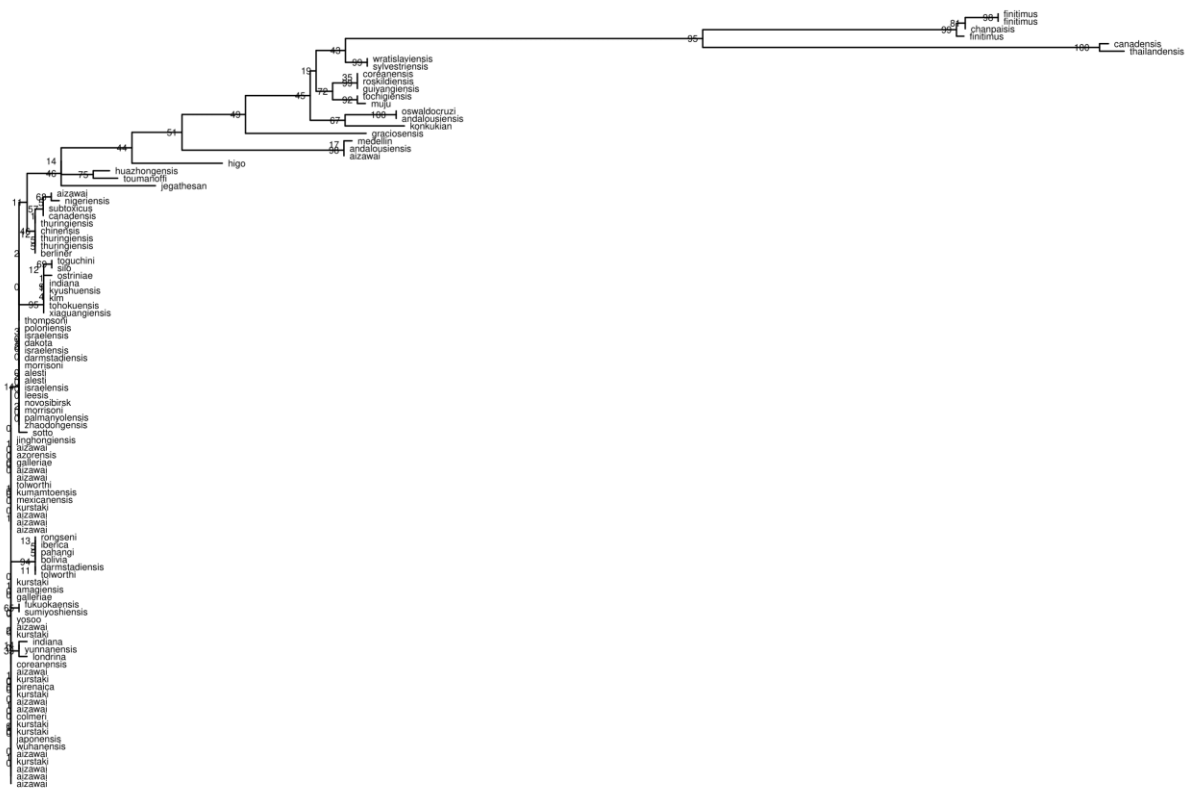
**b**

[illegible][illegible]

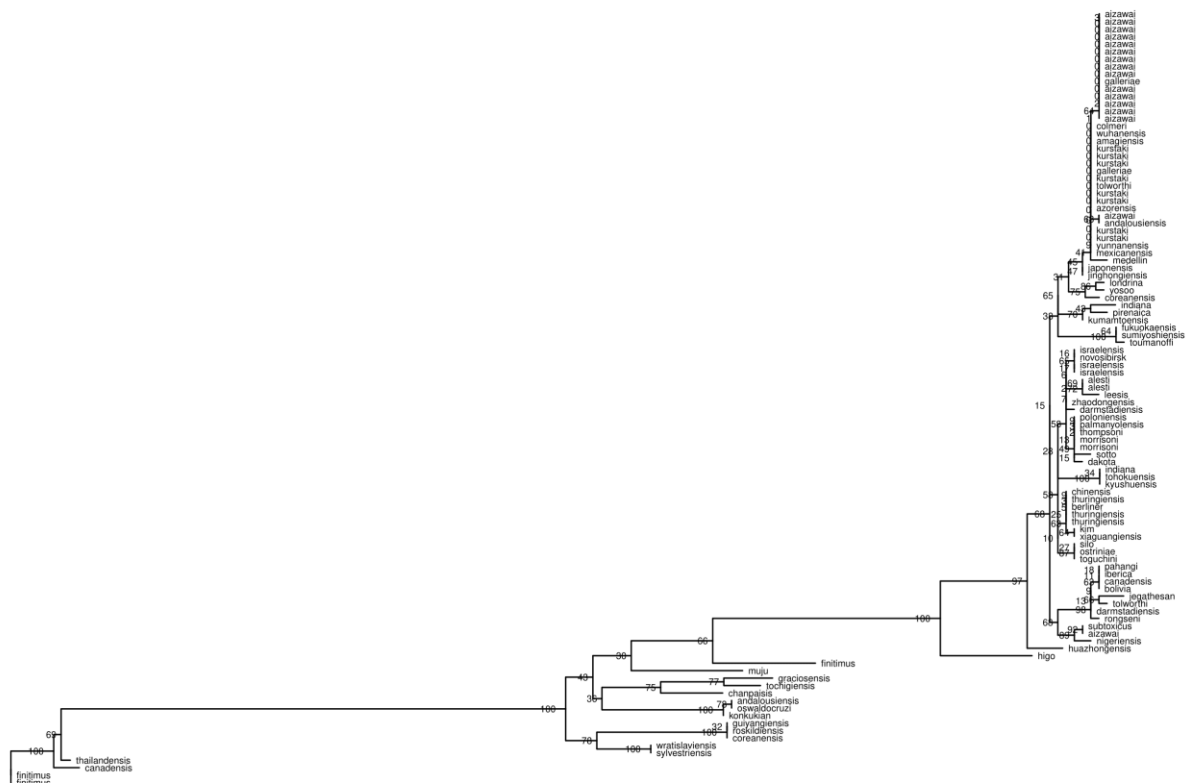
**e**

**f**

g



h

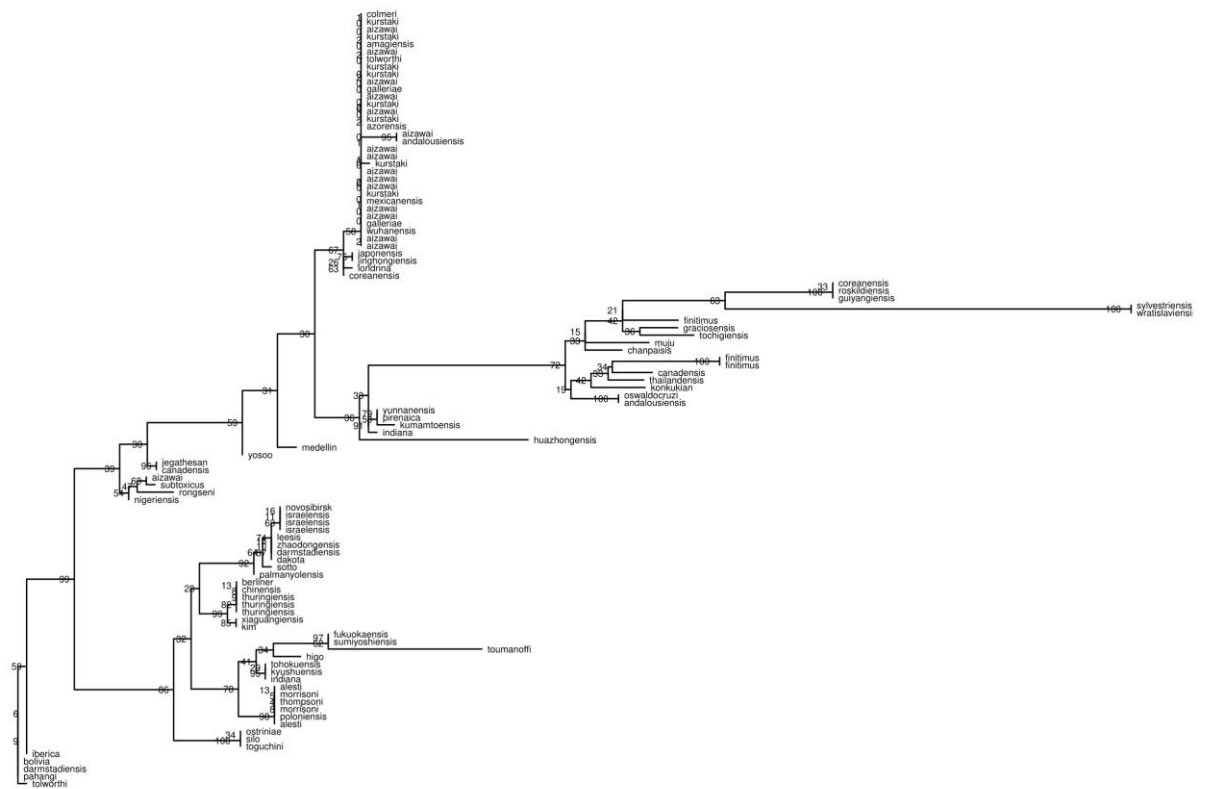


A detailed phylogenetic tree illustrating the evolutionary relationships among numerous species. The tree is rooted on the left and branches out to the right. Bootstrap values are provided at many of the internal nodes, indicating the confidence in the branching order. The species names are listed along the branches, often with a small number preceding them. The tree shows several major clades, including a large group of species starting from the bottom left (e.g., aizawai, wuhanensis, azorensis) and another large group starting from the top left (e.g., finitimus, canadensis, thailandensis). The tree also includes a small cluster of species on the far right (e.g., israelensis, alesti, mansongens).

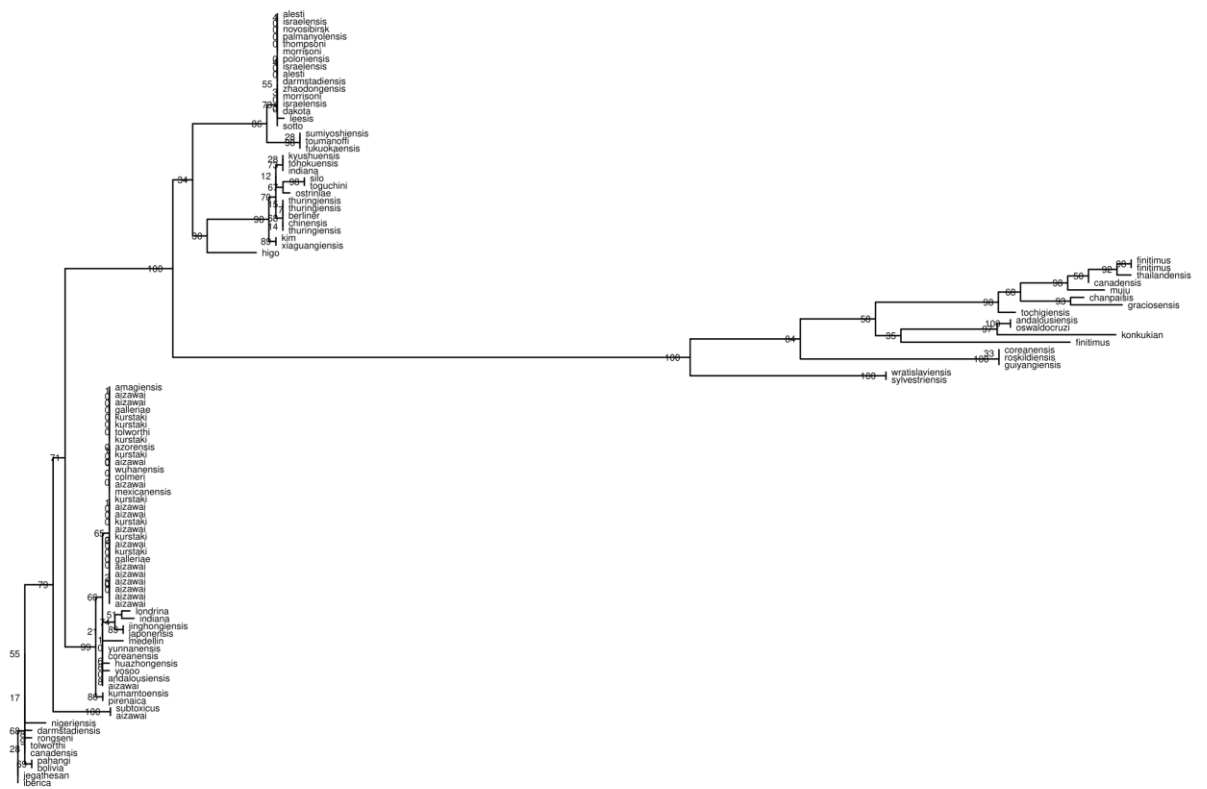
[illegible]



**k**



1

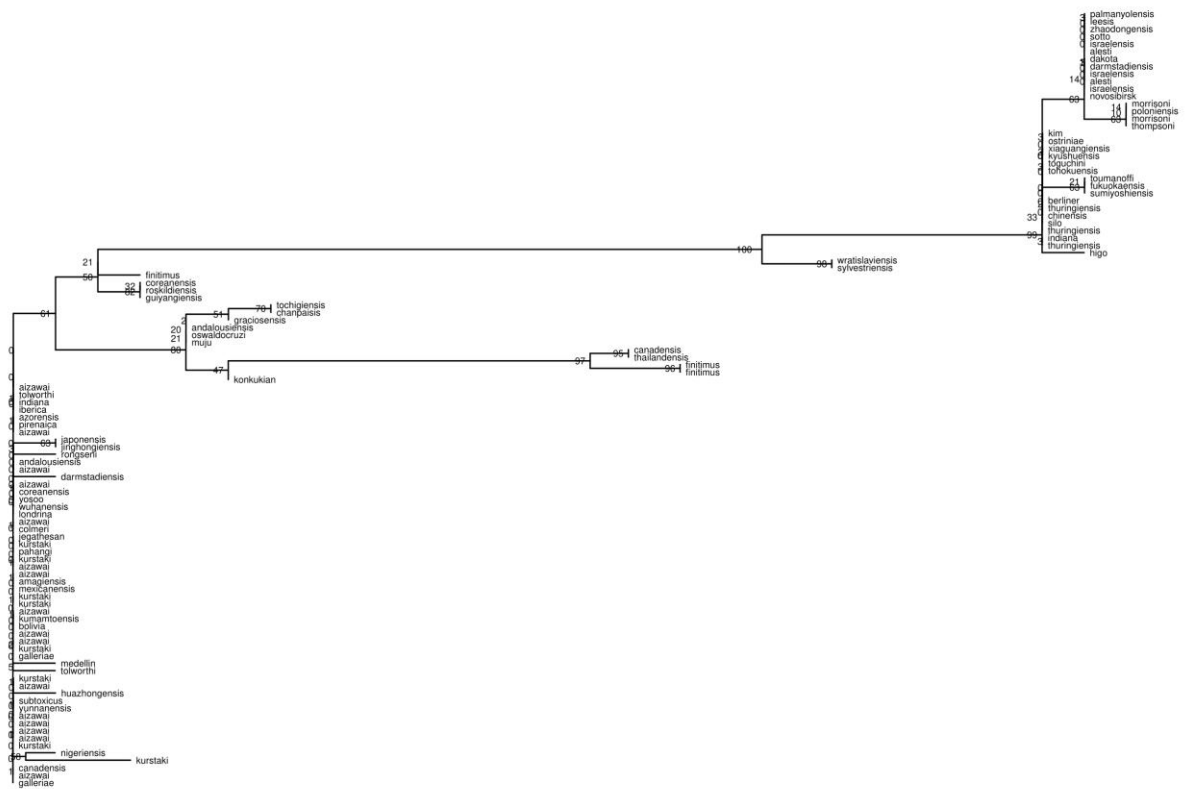




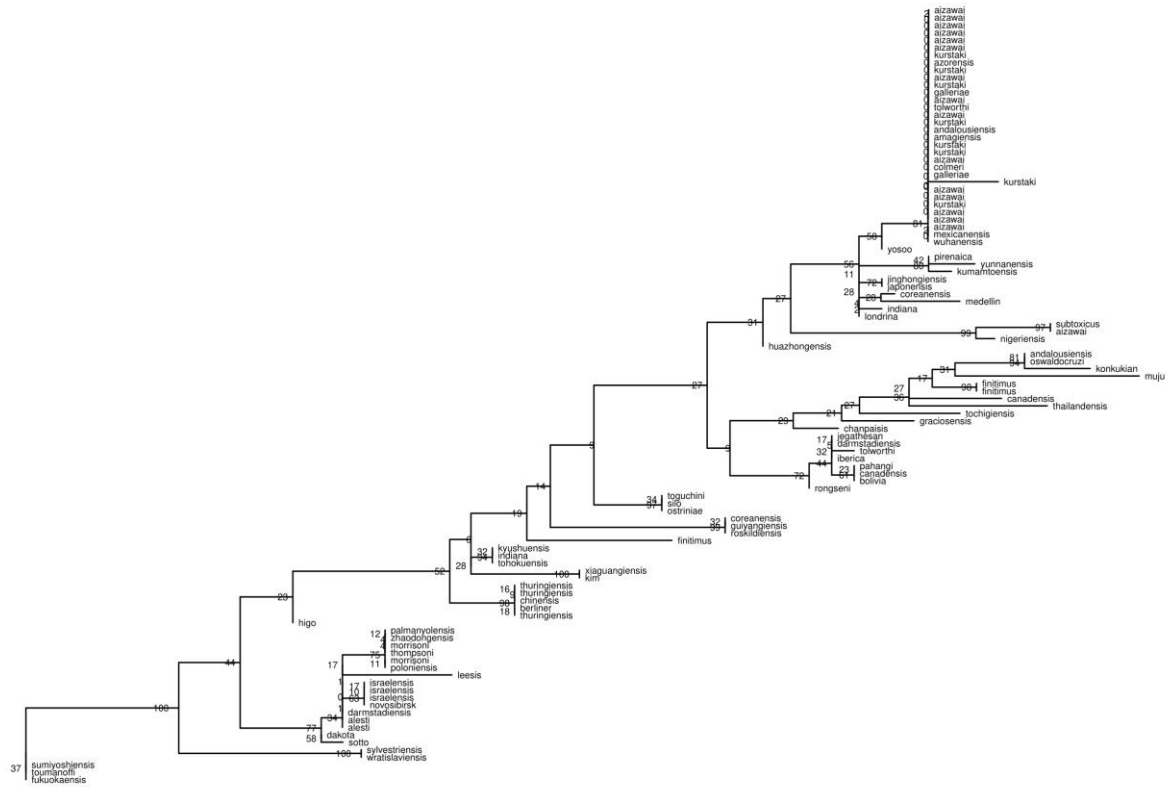
[illegible][illegible]

[illegible]

**t**

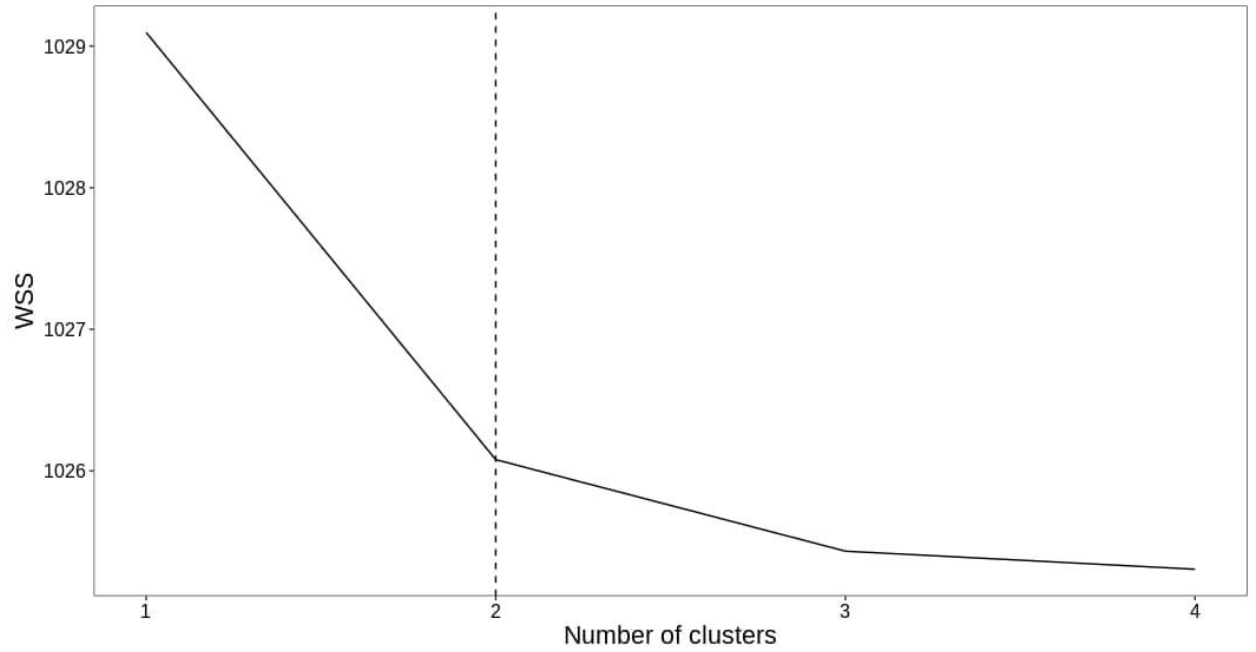


u

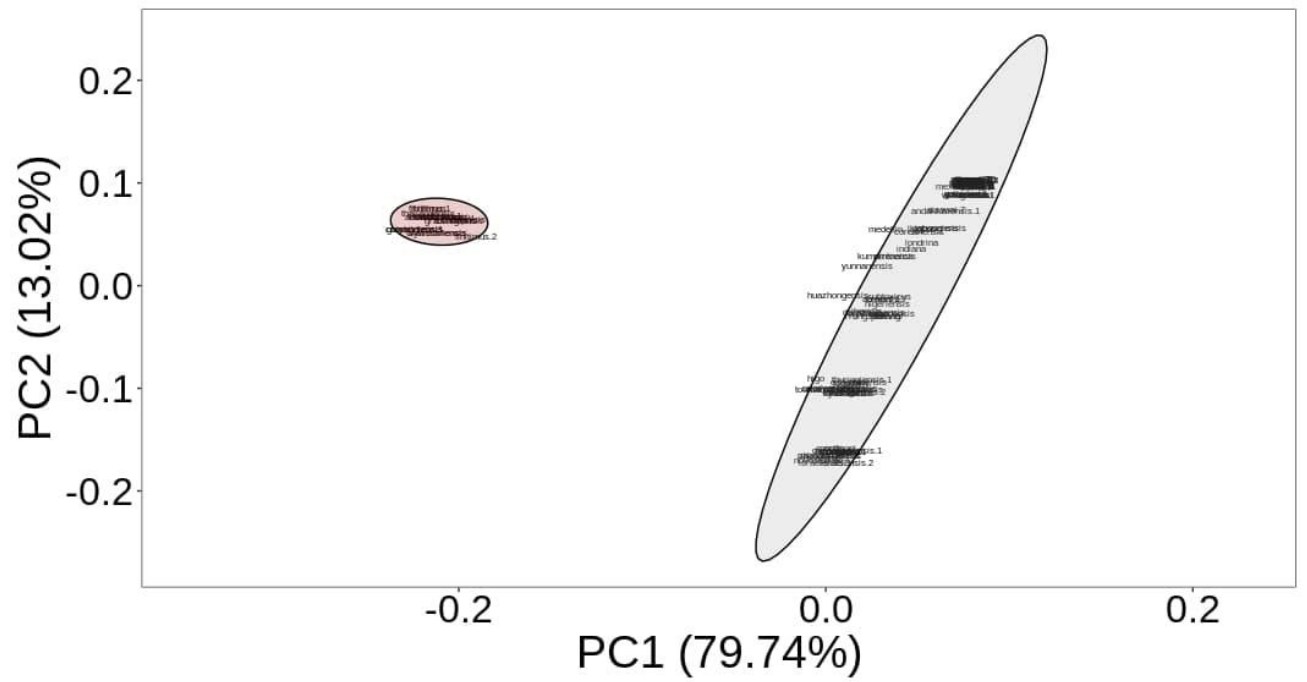


**Figure S5:** All phylogenomic and single-loci phylogenetic trees based on Bt-genomes. Trees are reconstructed using raxml-ng. Supporting values for 1000 bootstrap replicates are shown. Used evolutionary models and mean supporting values are presented in Table S8. **(a)** Presence/absence-based tree. **(b)** Core SNPs' tree. **(c)** Protein *gyrB*. **(d)** Concatenated protein *gyrB* and *gyrA* alignments-based tree. **(e)** Concatenated nucleotide *gyrB* and *gyrA* alignments-based tree **(f)** flagellin orthologs. **(g)** *calY*. **(h)** *dnaK*. **(i)** *fusA*. **(j)** *atpD*. **(k)** *groEL*. **(l)** *guaB*. **(m)** *inhA*. **(n)** *mmsA*. **(o)** *nprB*. **(p)** *phbB*. **(q)** *roCA*. **(r)** *rph*. **(s)** *sucC*. **(t)** *tuf*. **(u)** *yjLD*.

**a**

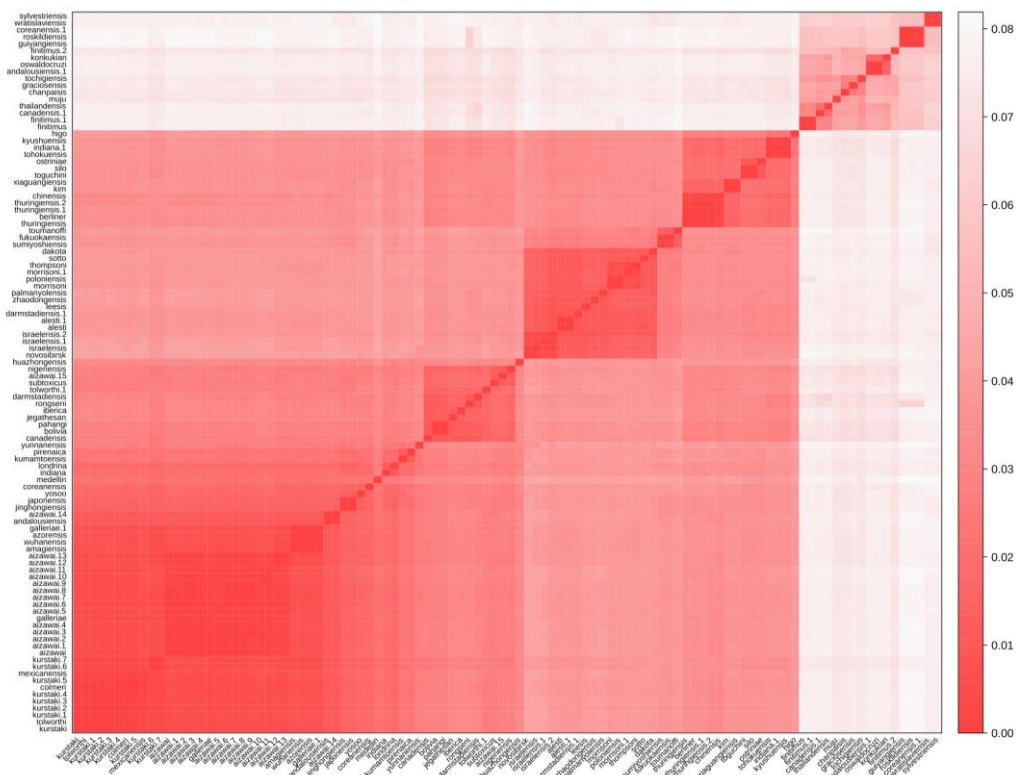


**b**

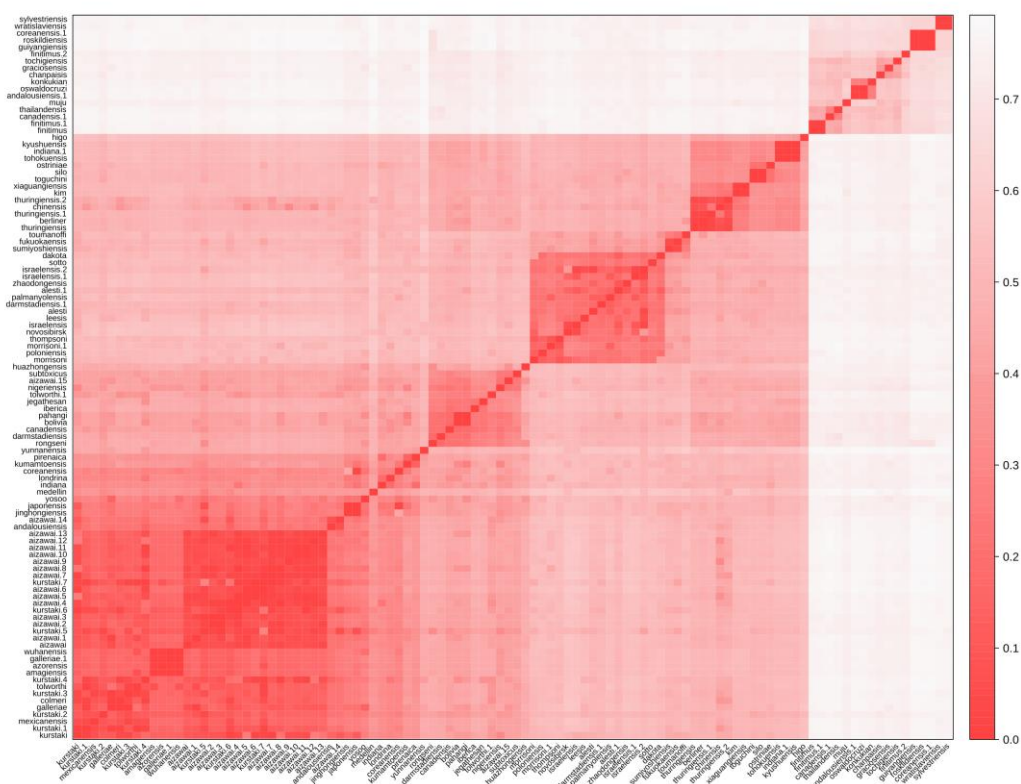


**Figure S6:** (a) The optimal number of clusters for the k-means clusterization of the data based on the ANI matrix. (b) The results of k-means clusterization of the serovars.

**a**

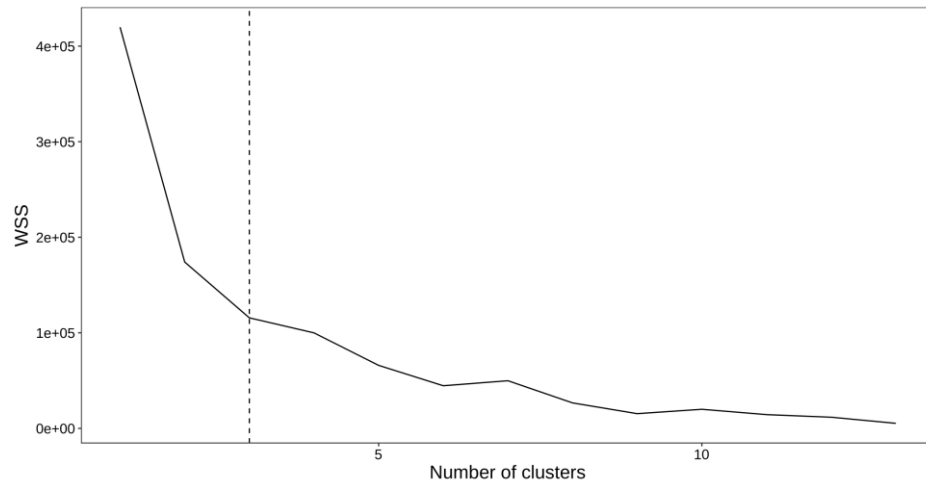


**b**



**Figure S7:** Heatmap-visualisation of clusters based on mash-distance and mean genome identity. **(a)** A mash distance-based heatmap. The intensity of the color is proportional to ANI (1-mash distance). **(b)** A minimap-2-based heatmap. The intensity of the color is proportional to the mean identity between genomes.





**Figure S8:** The optimal number of clusters for the k-means clusterization of the data based on subtrees' length containing all representatives of each serovar using the elbow method. The dashed line represents the number selected (3) depicted in Figure 5d of the main text.