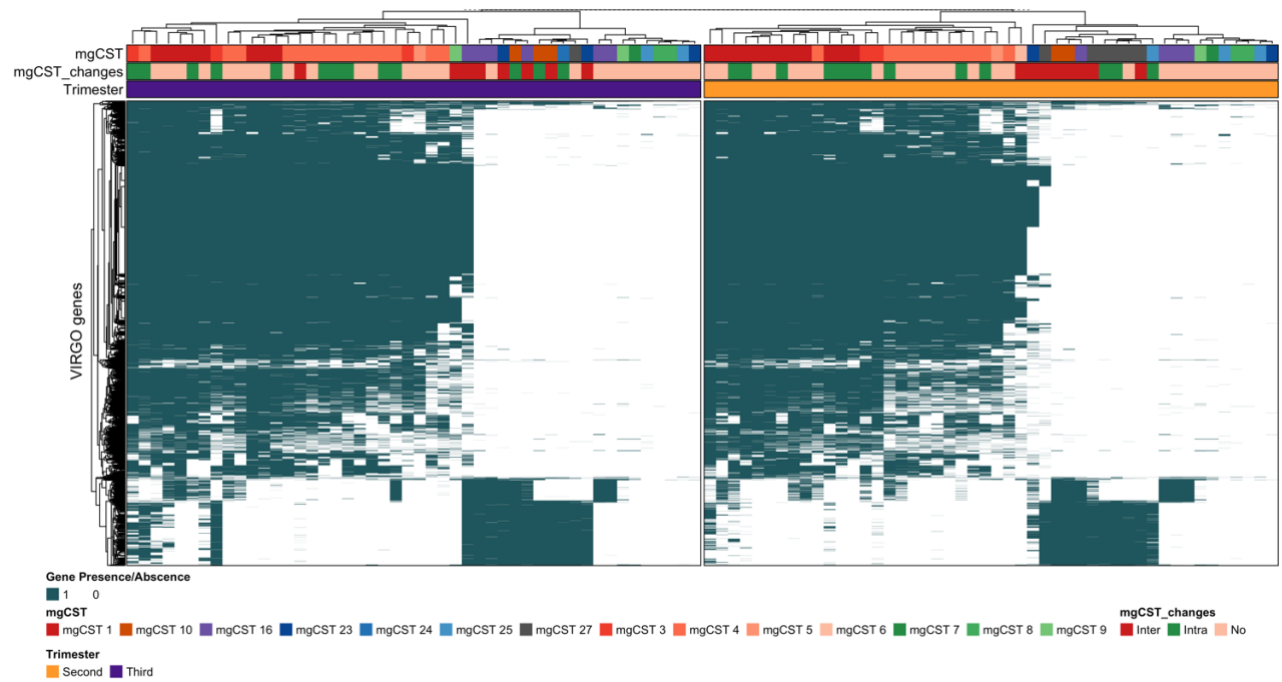
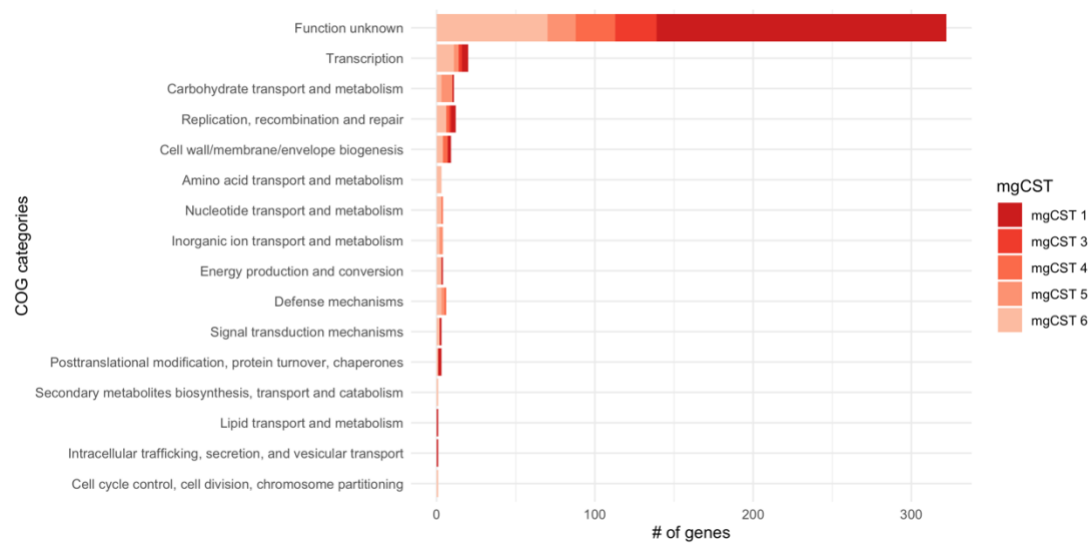


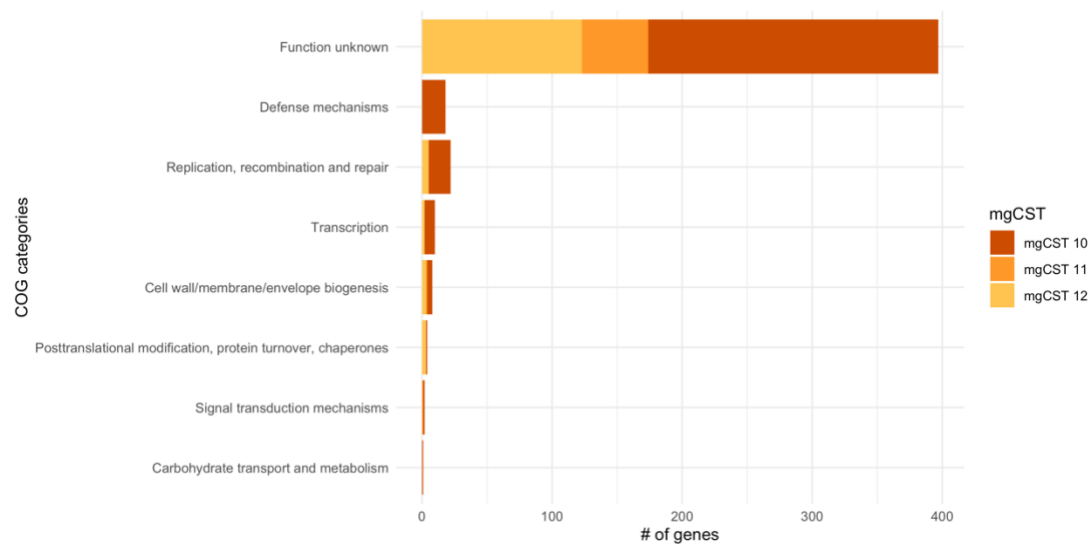


Supplementary Figure 1. Heatmap showing the pan-mgCST gene content between second and third trimesters in paired samples (n=48). Heatmap is clustered based on trimester and indicated by color strip on the top. For each non-redundant gene, a dark green cell indicates its presence (1), and a white cell (0) indicates its absence in the corresponding sample.

a)

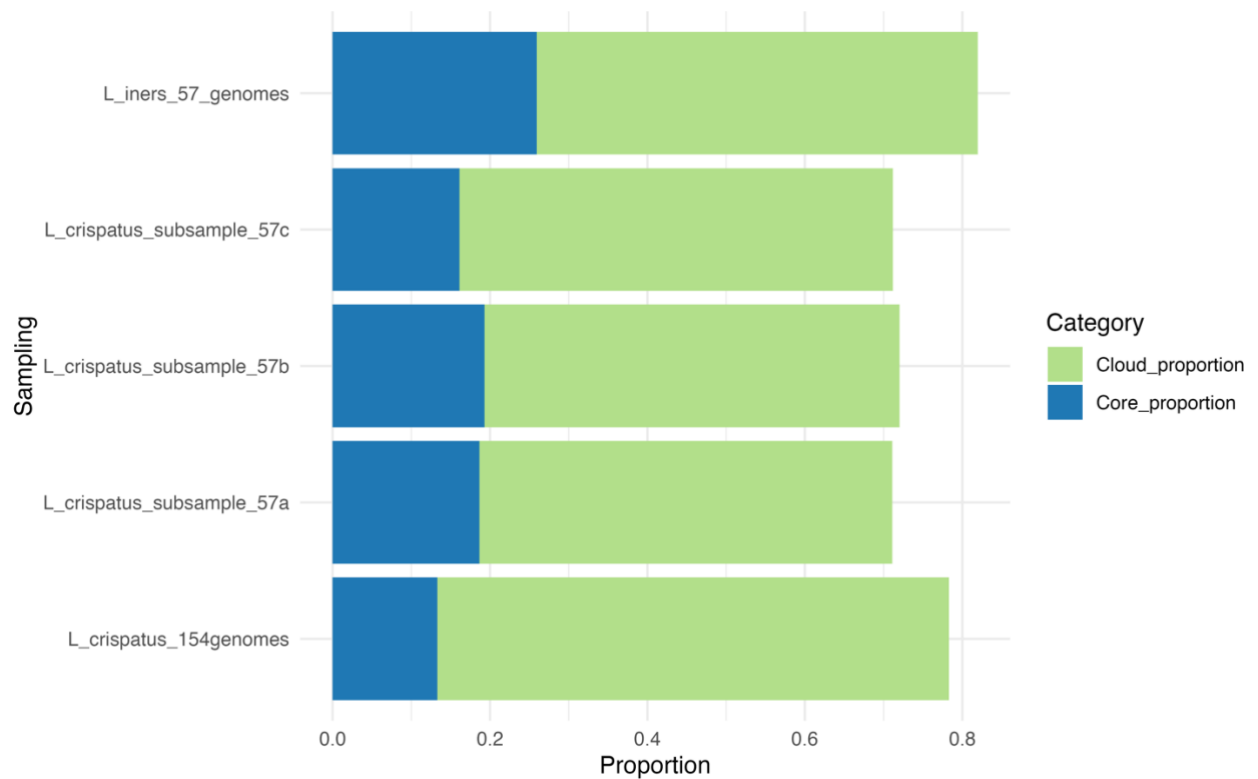


b)

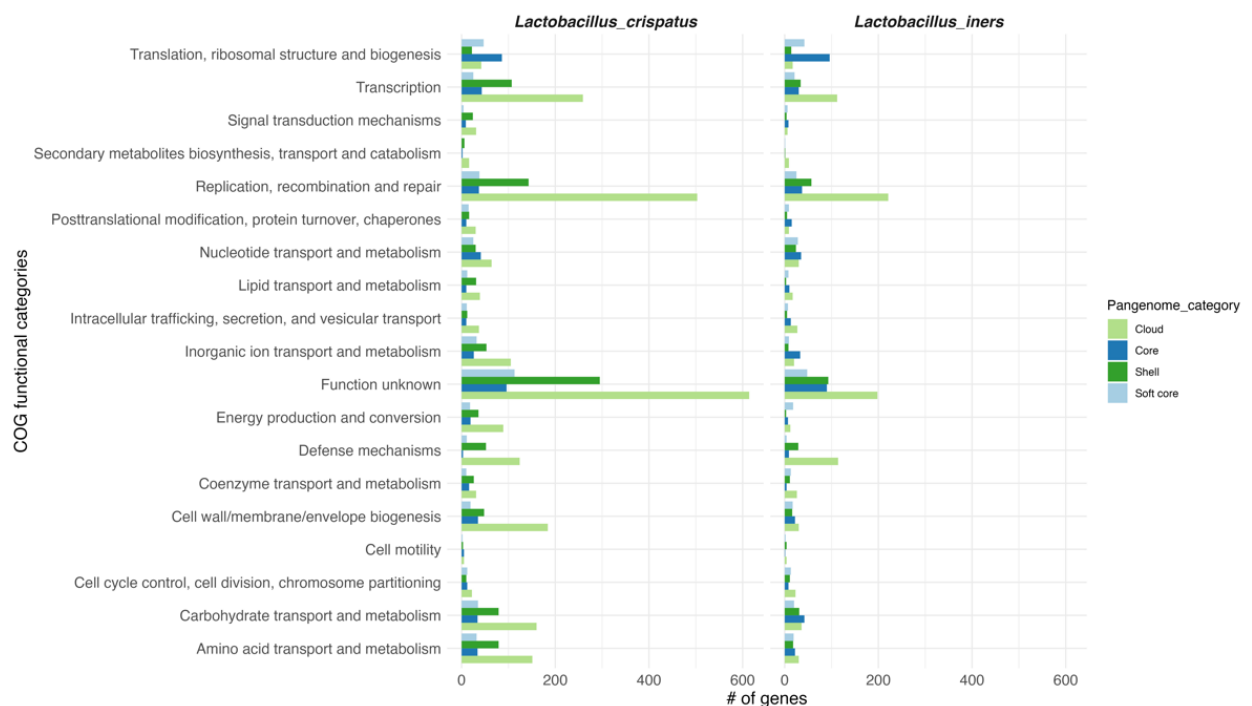


Supplementary Figure 2. a, b. Barplots showing the distribution of mgCST-specific genes across various functional categories in *L. crispatus* (a) and *L. iners* (b).

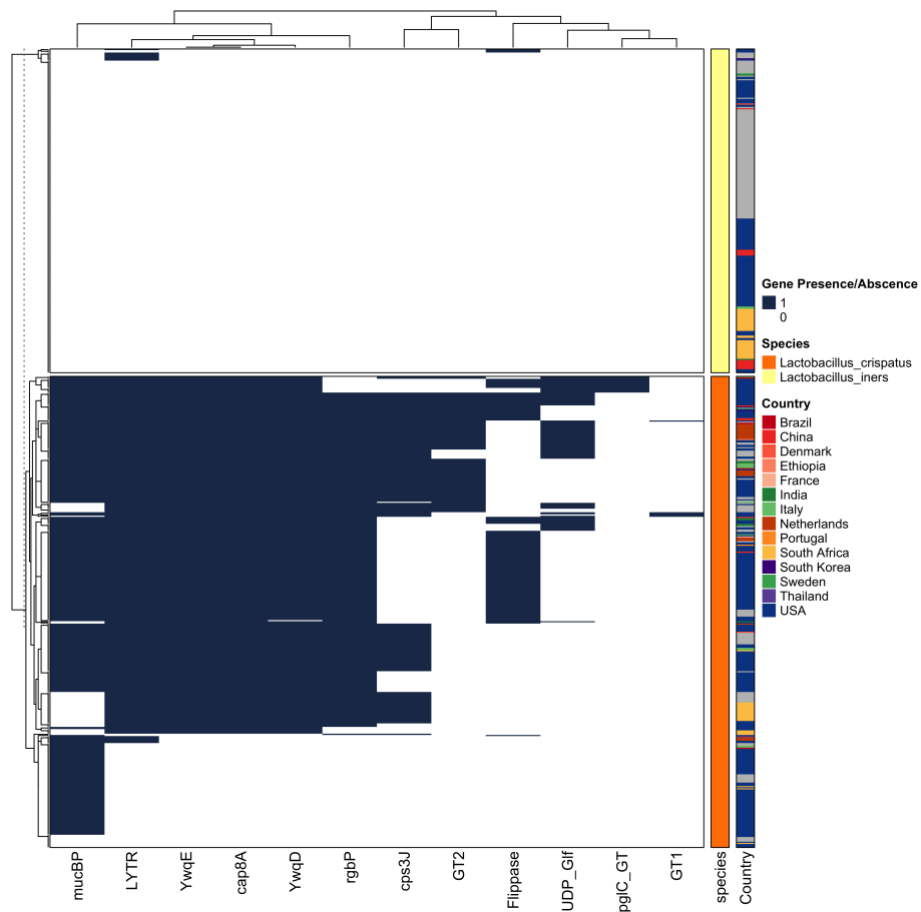
a)



b)



Supplementary figure 3. a. Stacked barplots showing the comparison of proportion of the cloud and core gene content between *L. iners* and three random samplings of *L. crispatus*. b. Barplots of *L. crispatus* ($n=170$) and *L. iners* ($n=76$) showing the distribution of the genes of each pangenome category across various functional categories.



Supplementary Figure 4. Heatmap showing the presence and absence of genes from cell surface glycan gene cluster in publicly available *L. crispatus* and *L. iners* genomes. The x-axis represents the genes from the gene cluster. For each gene, a blue cell indicates its presence, and a white cell indicates its absence in the corresponding genome. The heatmap was clustered based on the species indicated by the color strip on the side and other color strip represent the isolation country of the corresponding genome.