



Figure S1. Analysis of the impact of early-life factors on the infant gut microbiota. Bar plot shows the proportion of gut microbiome variance (r^2 , envfit analysis) explained by various infant-associated variables across different age groups. Significant factors (FDR-adjusted p -value < 0.05) are marked with an asterisk. In the lower panel, the effect of the most influential factors within specific age groups, as determined by envfit analysis, were further evaluated on individual microbial species. Balloon sizes represent the average relative abundance of species, while green-shaded rectangles indicate statistically significant differences between the levels of the assessed variable (Mann-Whitney U Test p -value < 0.05). Since information on feeding type and delivery mode was not consistently available across all studies, and their impact on the microbiota was therefore assessed using the subset of samples for which these metadata were available (Neonatal, $n=868$; pre-weaning, $n=1,261$; post-weaning, $n=2,002$).

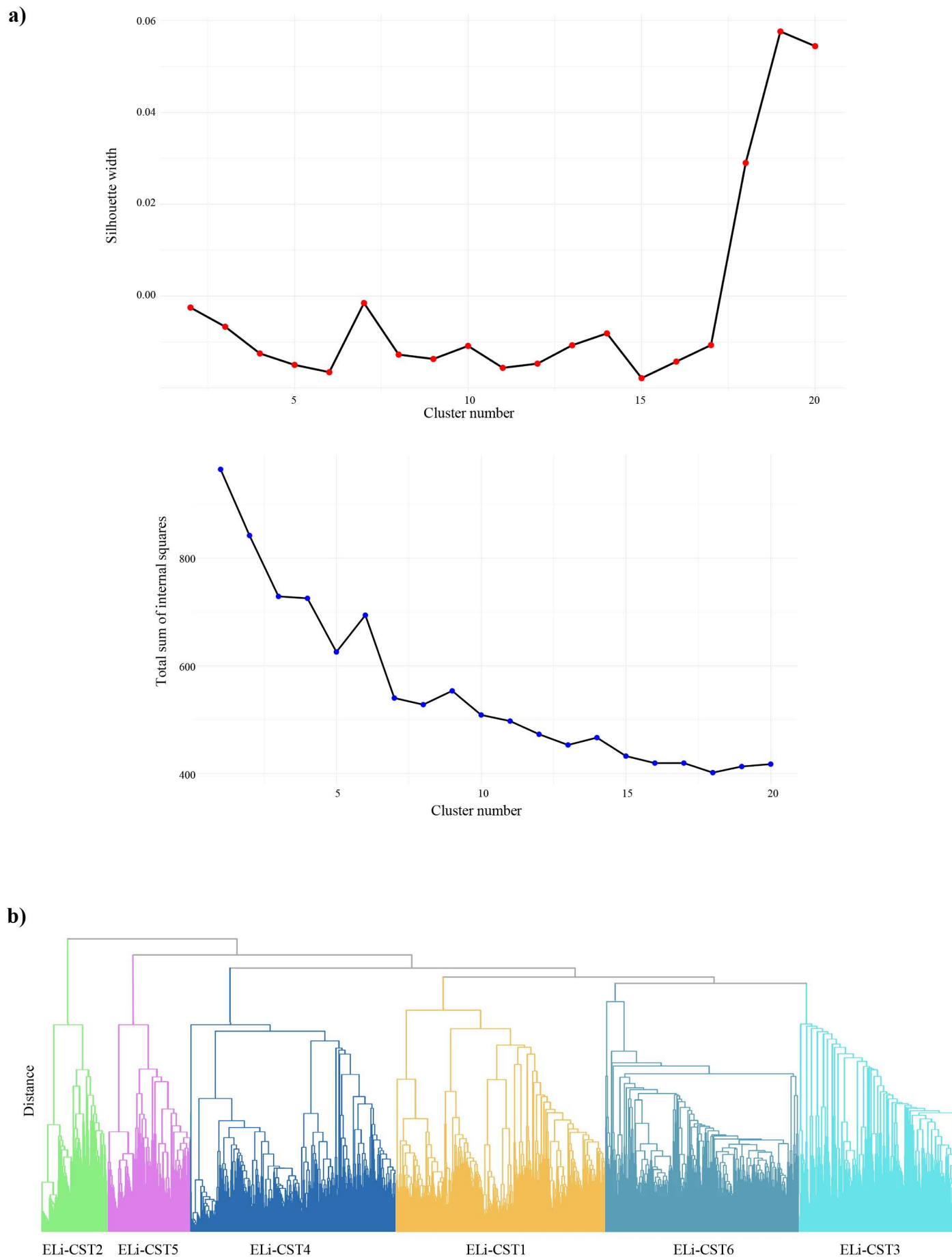


Figure S2. Clustering the infant microbiota dataset. Panel (a) shows average silhouette width and elbow curve computed on Bray-Curtis dissimilarities matrix. Panel (b) displays hierarchical clustering dendrogram with samples annotated by ELi-CST labels.

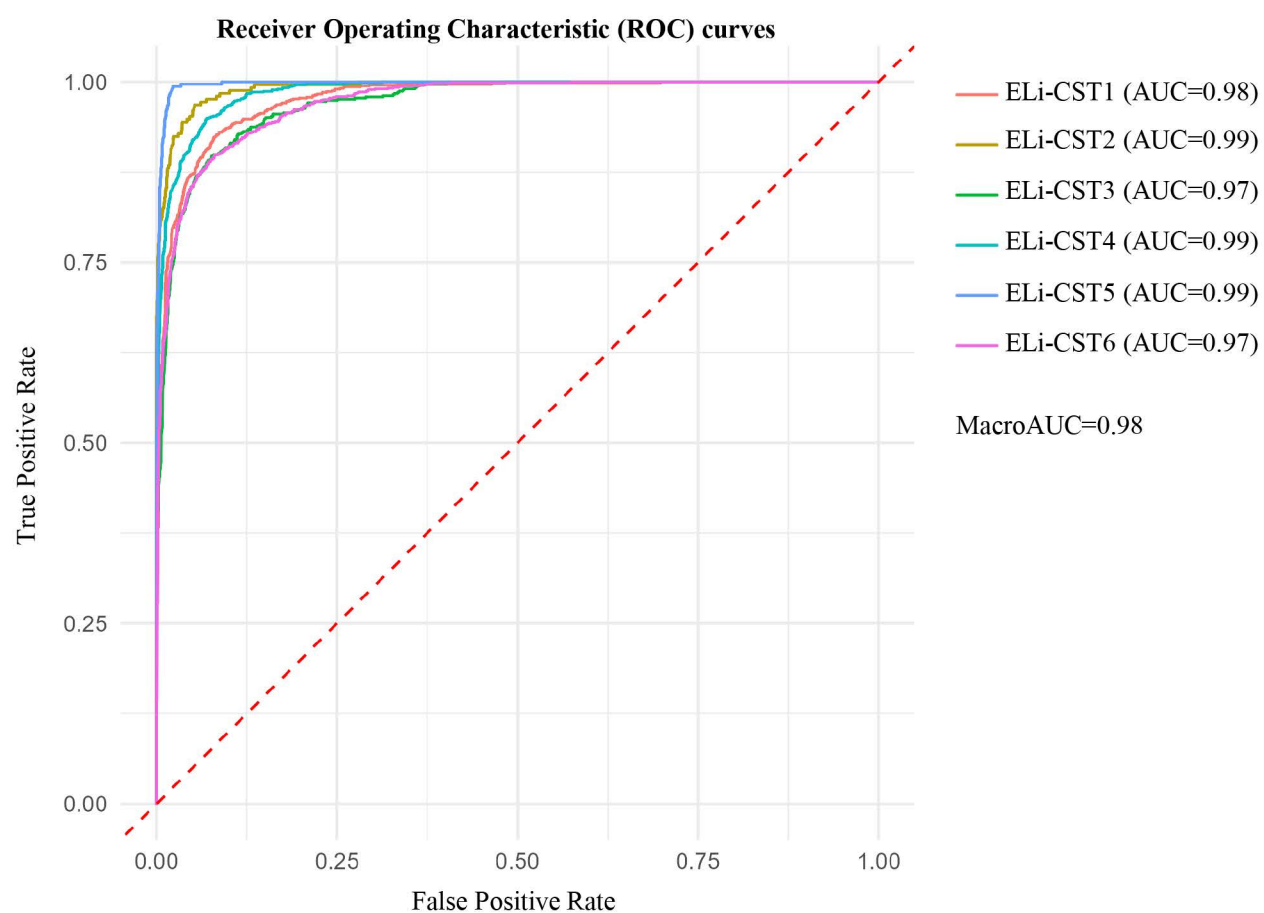


Figure S3. ROC Curves for Multiclass Classification (One-vs-Rest). The dashed red line denotes the no-discrimination reference (AUC = 0.5), and the legend reports per-class AUCs.

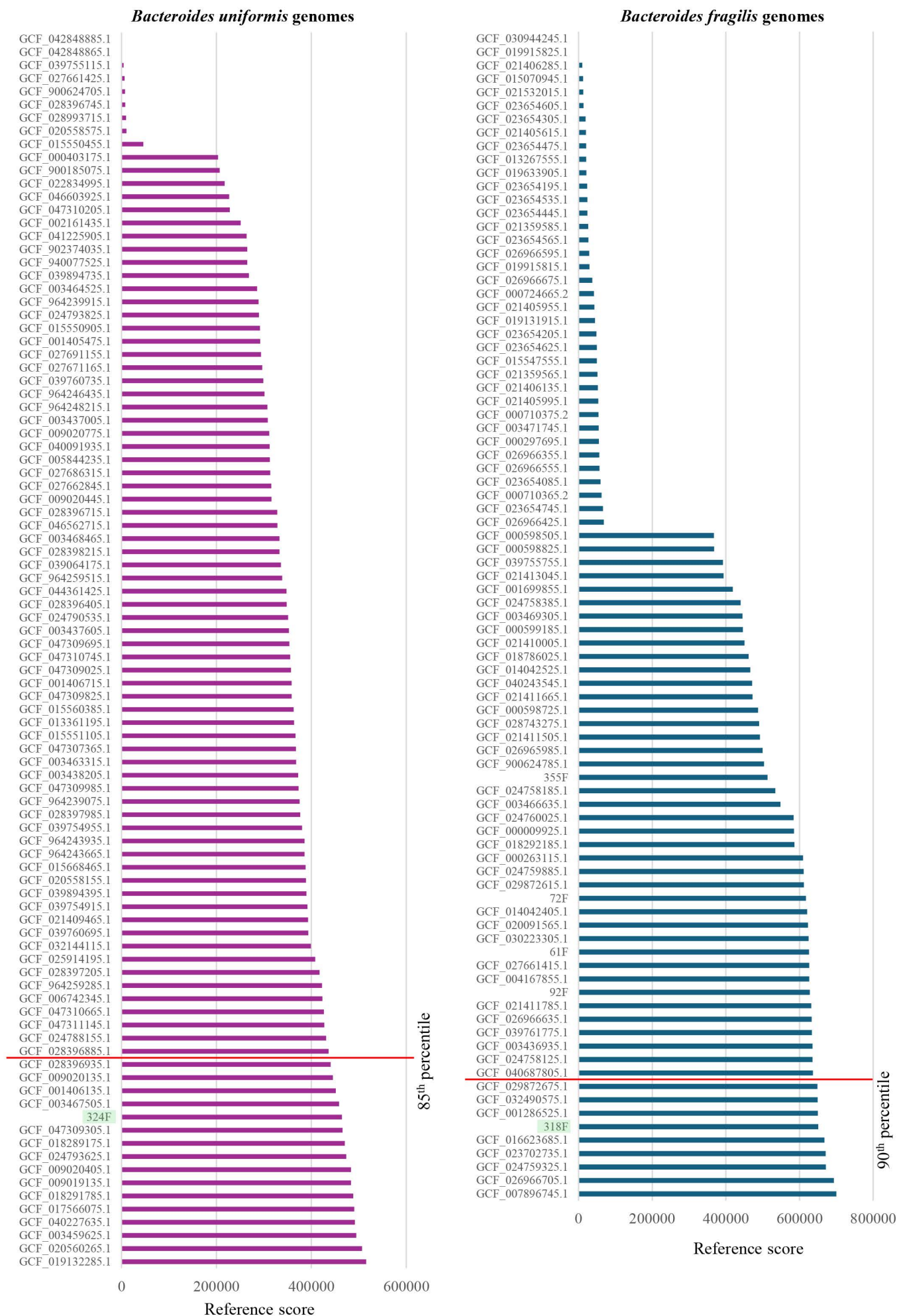
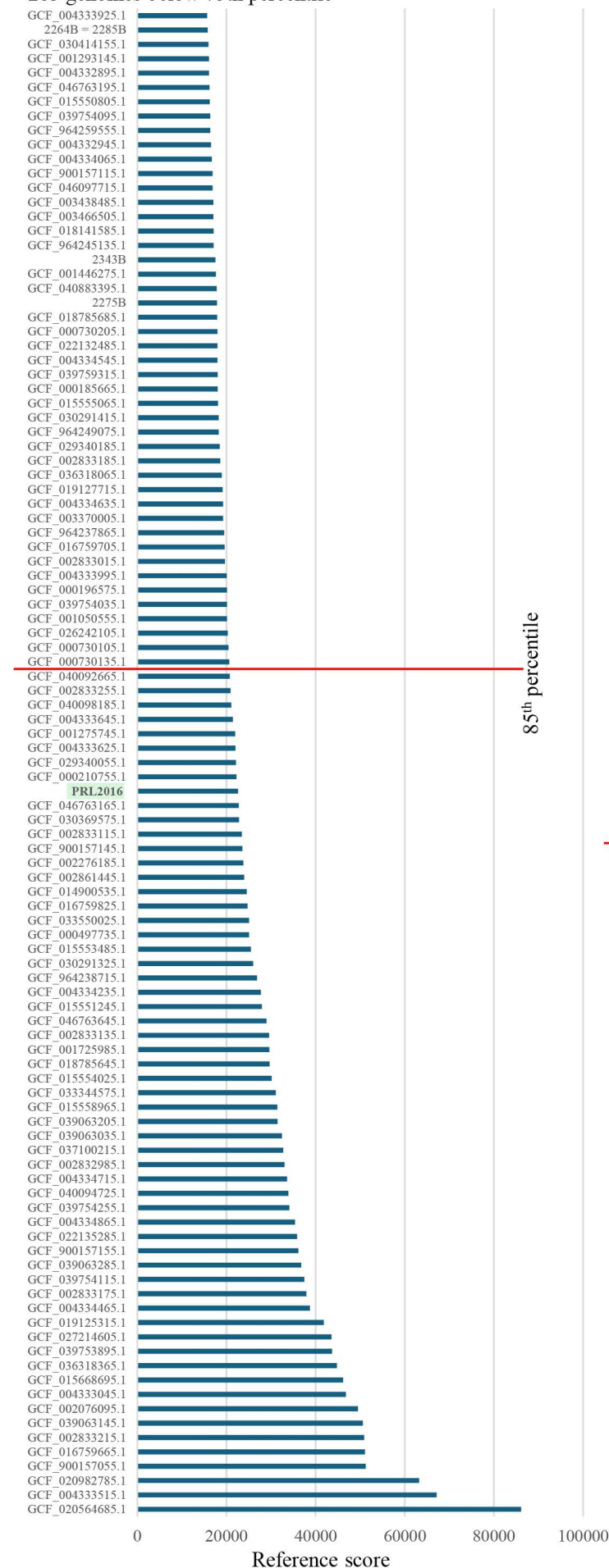


Figure S4. Optimal reference strain (ORS) analysis of *Bacteroides fragilis* and *Bacteroides uniformis* genomes. Bar graphs show the reference score obtained through the ORS selector pipeline for each assessed genome. The red line indicates the 85th-90th percentile threshold.

Bifidobacterium longum subsp. *longum* genomes

+ 283 genomes below 75th percentile



Bifidobacterium bifidum

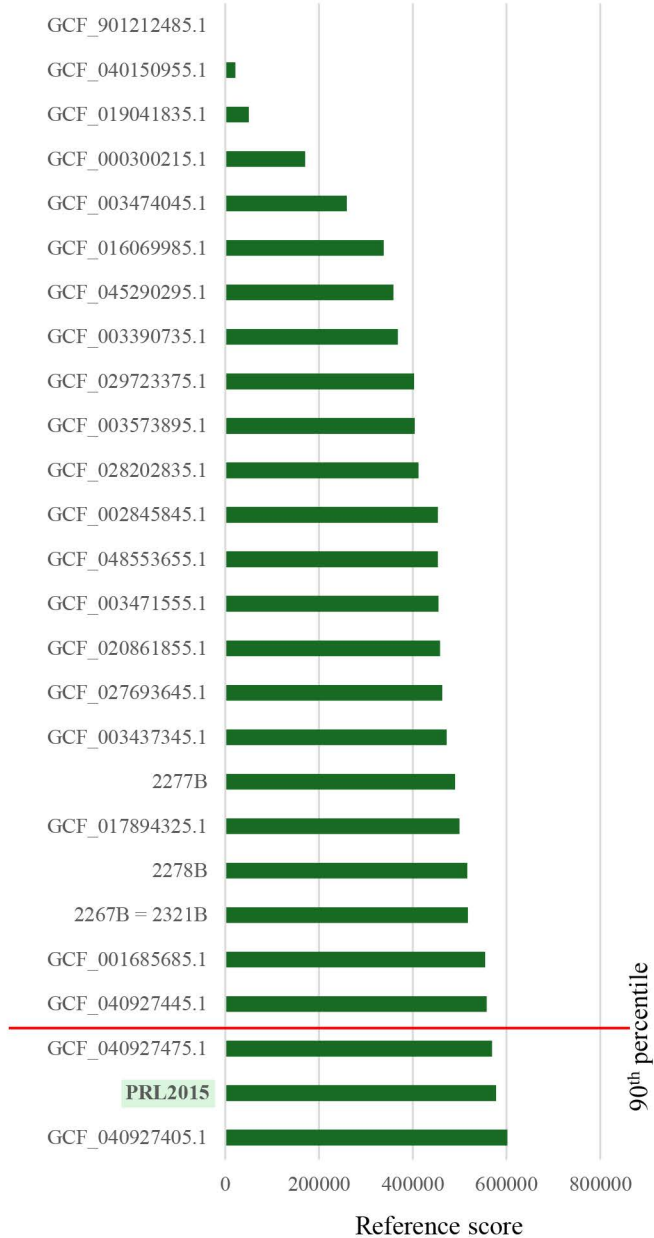


Figure S5. Optimal reference strain analysis of *Bifidobacterium bifidum* and *Bifidobacterium longum* subsp. *longum* genomes. Bar graphs show the reference score obtained through the ORS selector pipeline for each assessed genome. The red line indicates the 85th-90th percentile threshold.