

Accession	id
GCA_046118645.1	vp_0
GCA_028228685.1	vp_1
GCA_028551295.1	vp_2
GCA_049816165.1	vp_3
GCA_038070605.1	vp_4
GCA_030064615.2	vp_5
GCA_045278325.1	vp_6
GCA_048284885.2	vp_7
GCA_009734325.1	vp_8
GCA_018135645.1	vp_9
GCA_020041945.1	vp_10
GCA_015779285.1	vp_11
GCA_049816605.1	vp_12
GCA_048914355.1	vp_13
GCA_045287955.1	vp_14
GCA_046118655.1	vp_15
GCA_040254425.1	vp_16
GCA_039658515.1	vp_17
GCA_030064635.1	vp_18

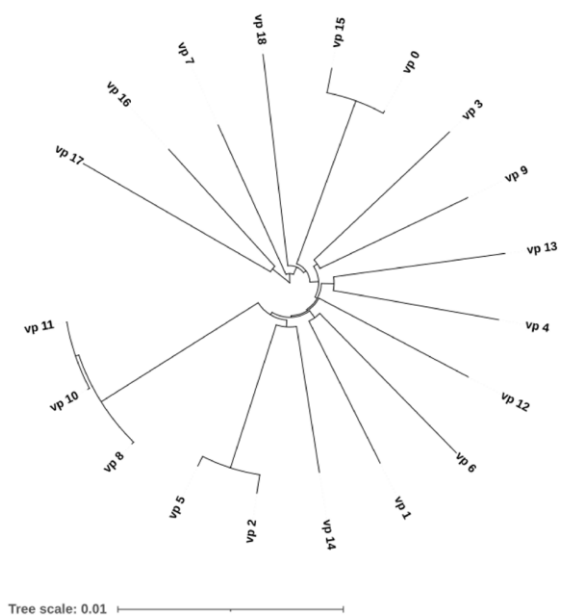


Figure S1: List of accessions and their Mash distance-based clustering, used for benchmark dataset simulation.

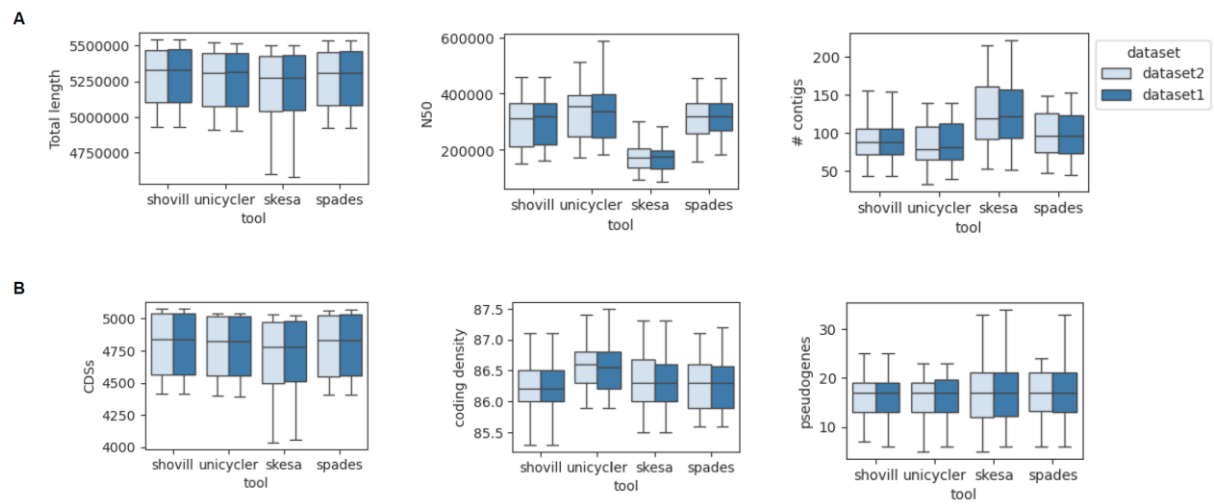


Figure S2: Reproducibility and comparison of assemblies and annotations across the tools used. A. Comparison of genome assemblies. B. Comparison of annotation results for the assemblies.

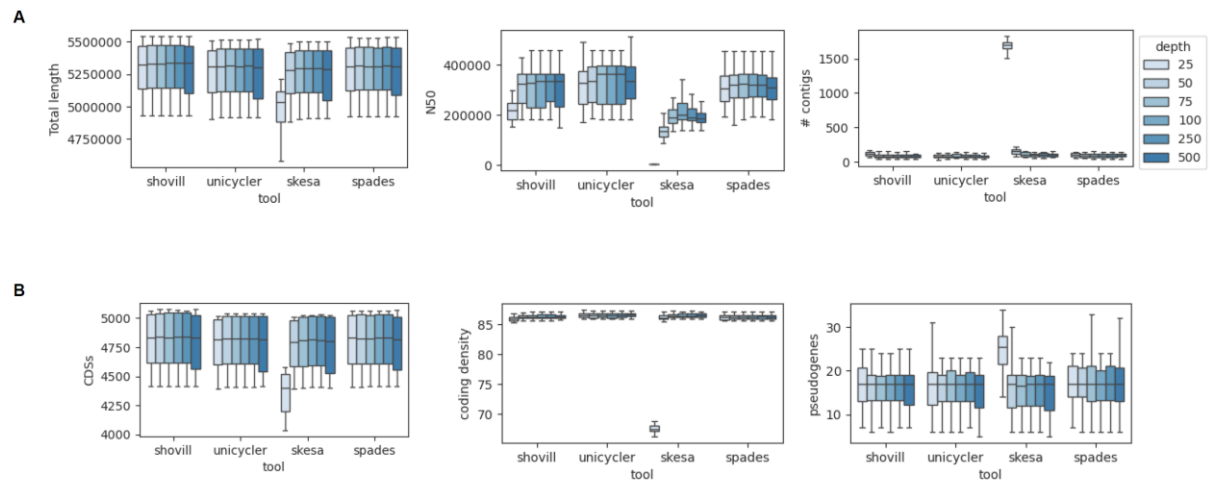


Figure S3: Reproducibility and comparison of assemblies and annotations across the tools used and sequencing simulated depths. A. Comparison of genome assemblies. B. Comparison of annotation results for the assemblies.

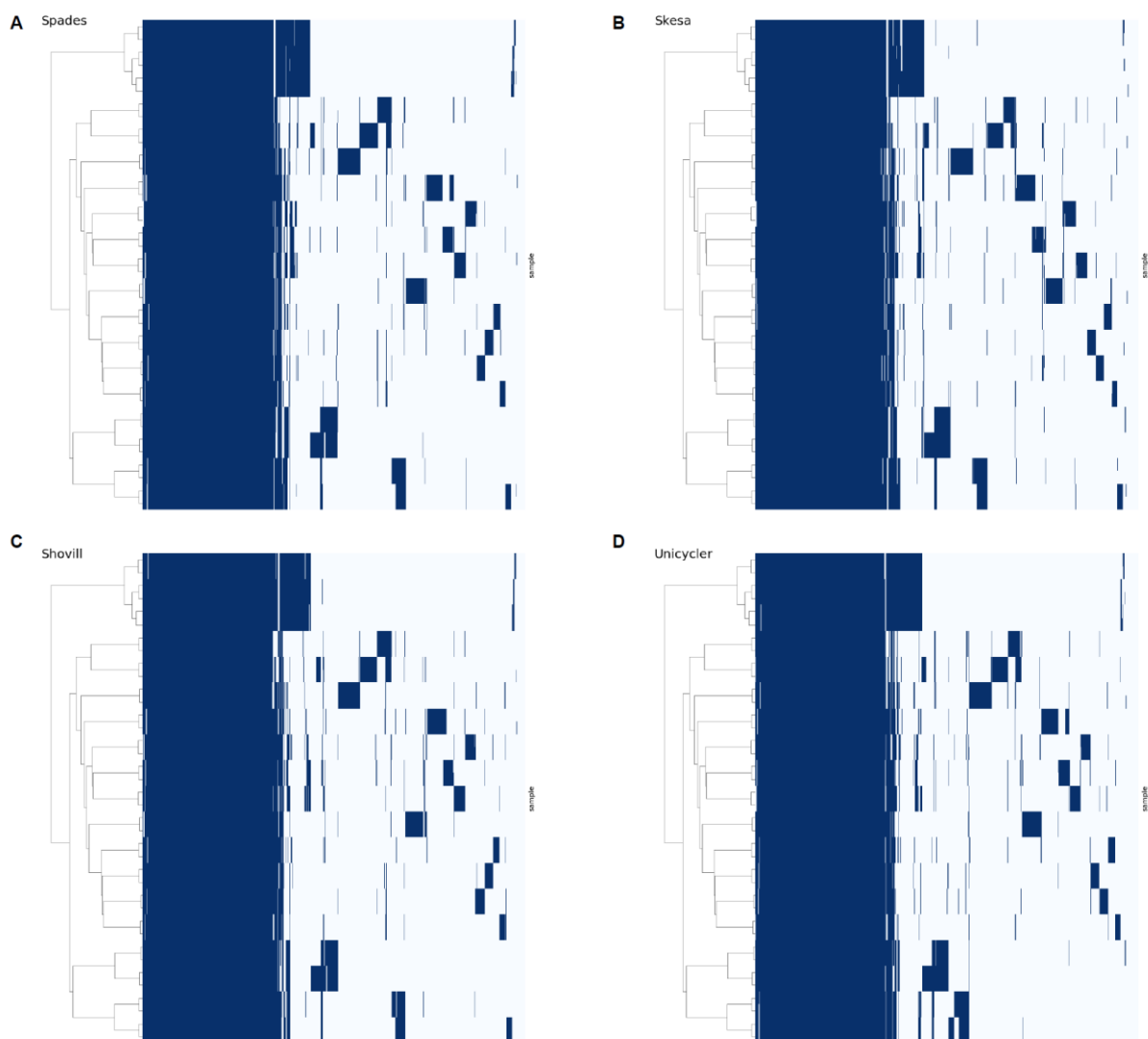


Figure S4: Pangenome clustering using benchmark 100X simulated datasets regarding assembly tools. A. SPADES. B. SKESA. C. SHOVILL. D. UNICYCLER.

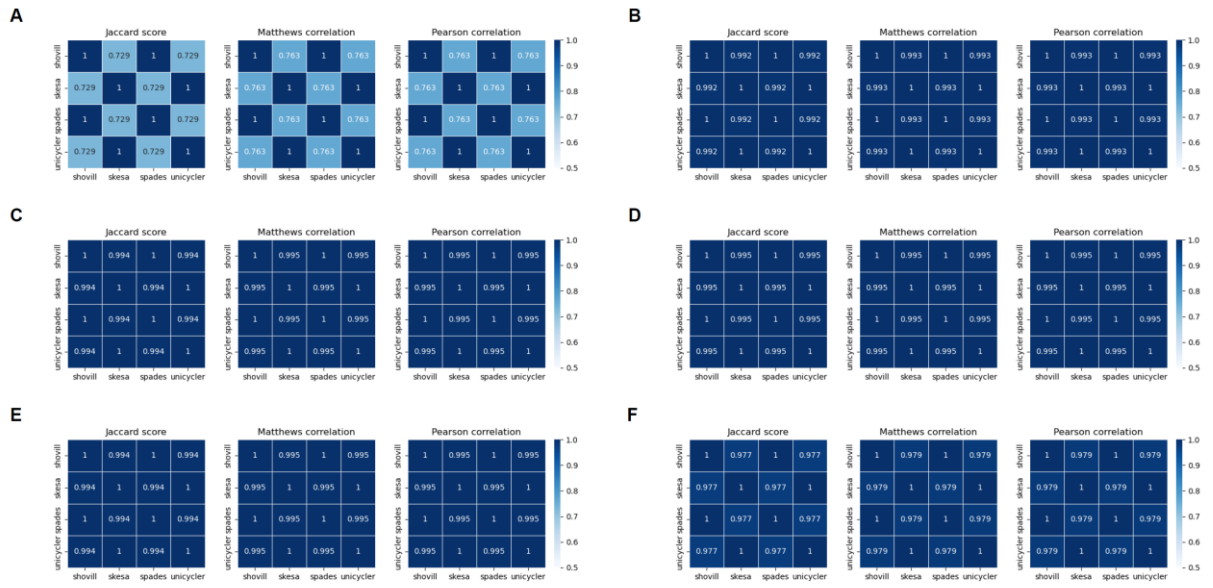


Figure S5: Comparison of pangenome computation regarding to the assembly tools used. Jaccard score, Pearson and Mathews correlation were computed to compare obtained pangenome matrices at different sequencing depth. A. 25X. B. 50X. C. 75X. D. 100X. E. 250X. F. 500X

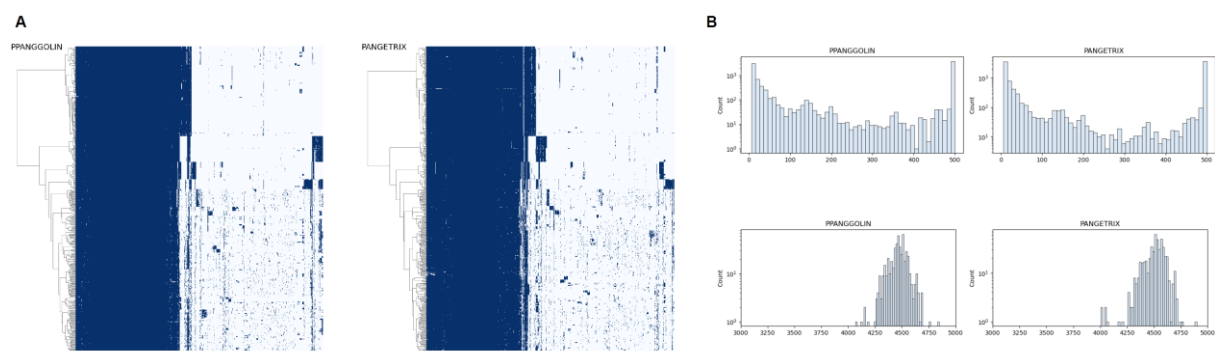


Figure S6. Comparison between pangenomes computation tools using 500 *V. parahaemolyticus* samples from dataset 1. A. Heatmap of pangenomes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenomes per samples.

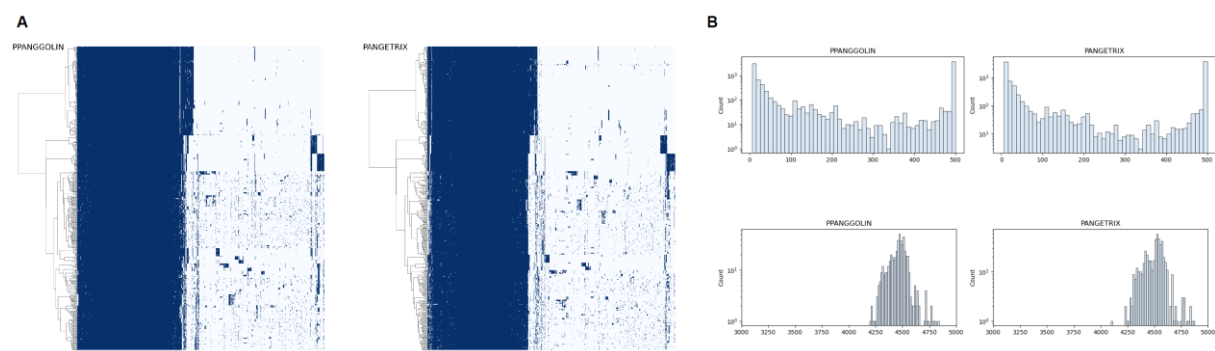
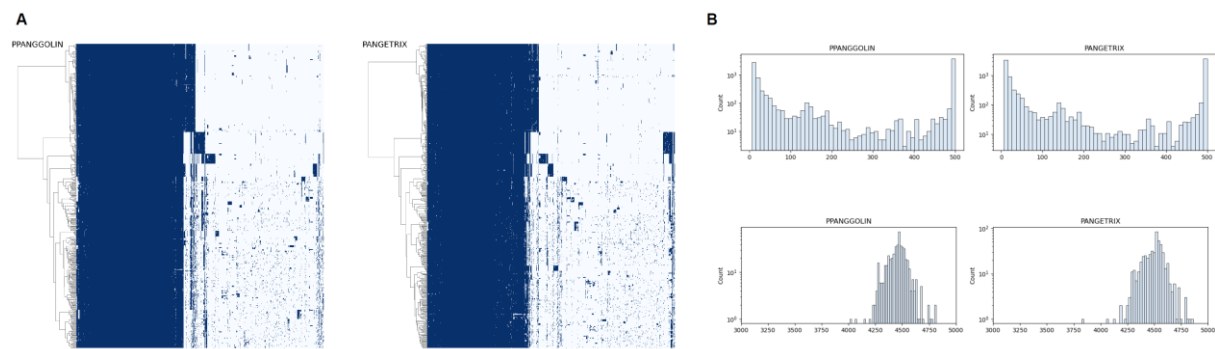


Figure S7. Comparison between pangenomes computation tools using 500 *V. parahaemolyticus* samples from dataset 2. A. Heatmap of pangenomes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenomes per samples.



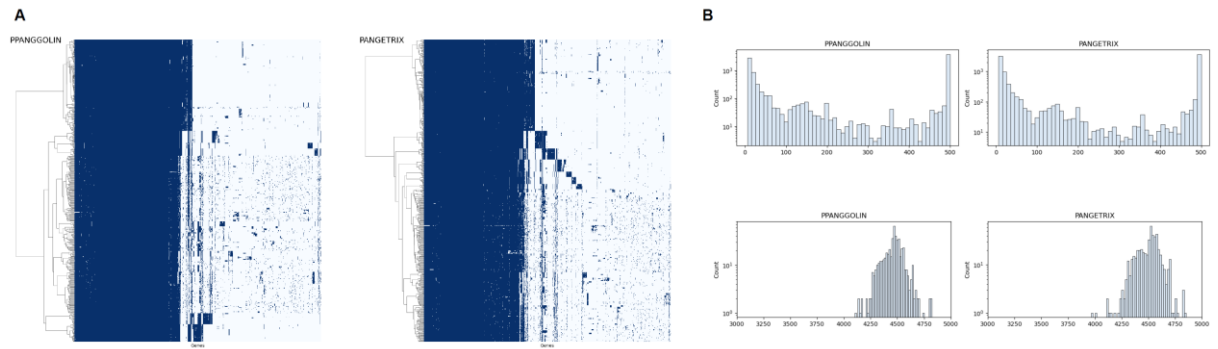


Figure S9. Comparison between pangenomes computation tools using 500 *V. parahaemolyticus* samples from dataset 4. A. Heatmap of pangenomes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenomes per samples.

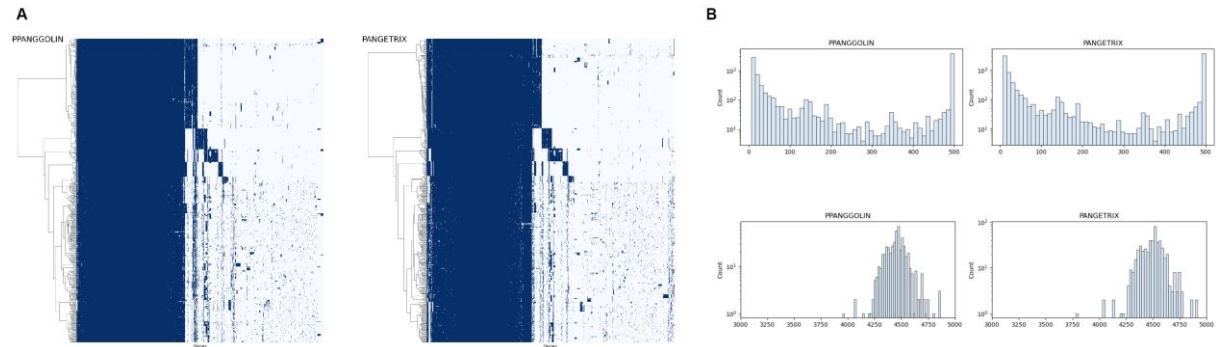


Figure S10. Comparison between pangenomes computation tools using 500 *V. parahaemolyticus* samples from dataset 5. A. Heatmap of pangenomes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenomes per samples.

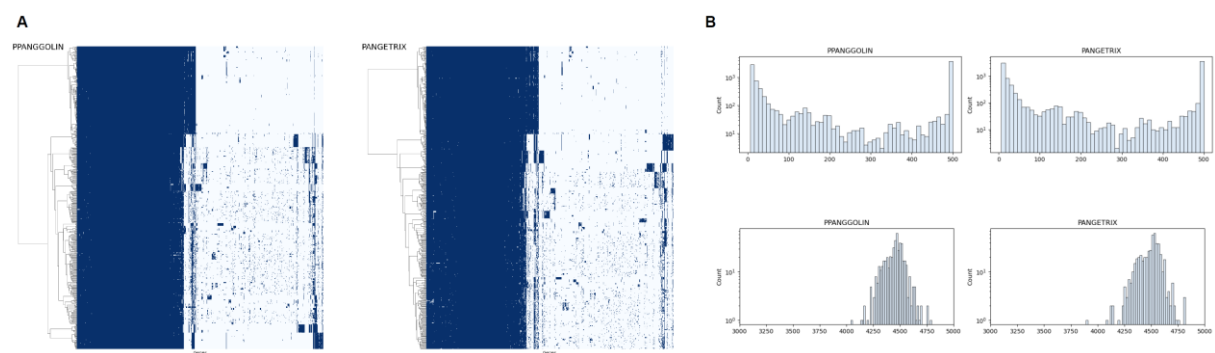


Figure S11. Comparison between pangenomes computation tools using 500 *V. parahaemolyticus* samples from dataset 6. A. Heatmap of pangenomes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenomes per samples.

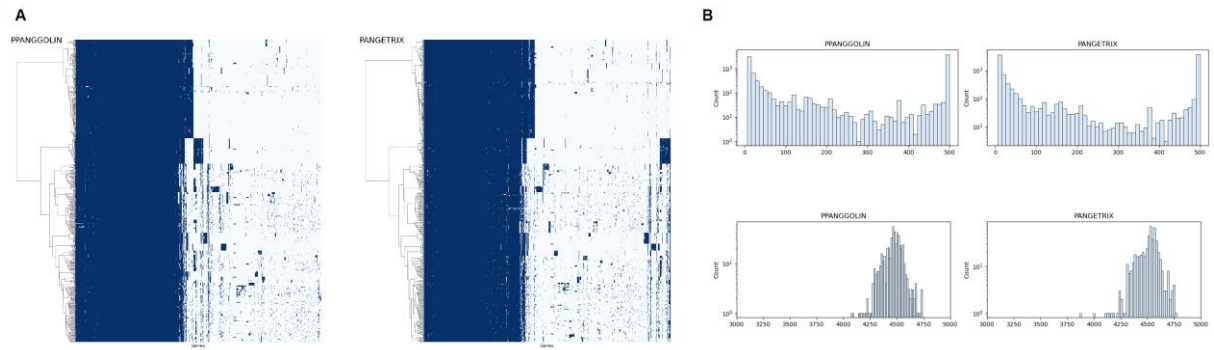


Figure S12. Comparison between pangenomes computation tools using 500 *V. parahaemolyticus* samples from dataset 7. A. Heatmap of pangenomes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenomes per samples.

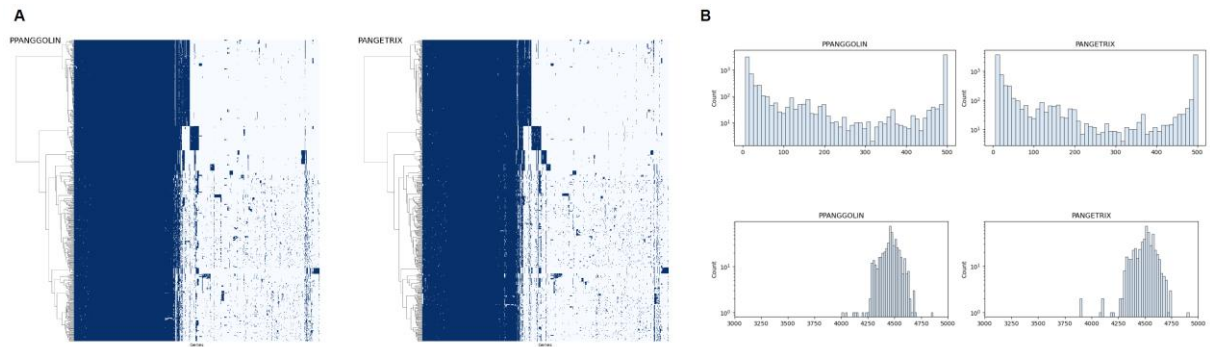


Figure S13. Comparison between pangenomes computation tools using 500 *V. parahaemolyticus* samples from dataset 8. A. Heatmap of pangenomes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenomes per samples.

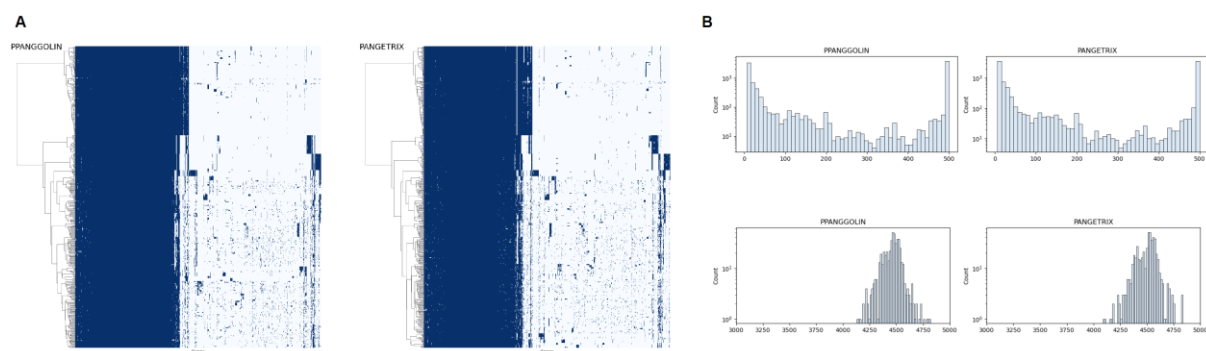


Figure S14. Comparison between pangomes computation tools using 500 *V. parahaemolyticus* samples from dataset 9. A. Heatmap of pangenes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenes per samples.

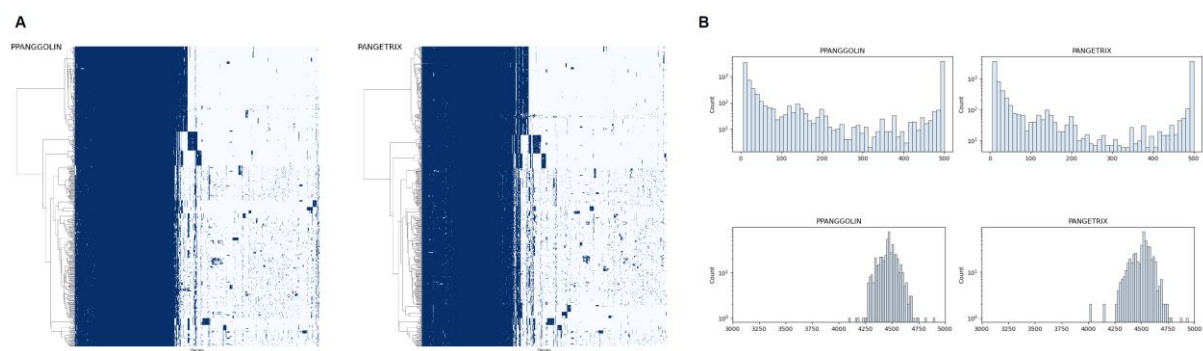


Figure S15. Comparison between pangenomes computation tools using 500 *V. parahaemolyticus* samples from dataset 10. A. Heatmap of pangenomes distribution across the 500 samples. B. Genes occurrence distribution. C. Distribution of the number of pangenomes per samples.