

Additional file 1: Efficient inference of large prokaryotic pangenomes with PanTA

Method	Sp600			Pa800			Kp1500		
	Core genes	Accessory genes	Total genes	Core genes	Accessory genes	Total genes	Core genes	Accessory genes	Total genes
Roary	1539	3577	5116	4798	34369	39167	4181	27566	31747
PIRATE	1569	1742	3311	4712	12447	17159	4039	14163	18202
PPanGGOLiN	1565	2856	4421	5152	18578	23730	4197	19490	23687
Panaroo	1616	2211	3827	5194	17740	22934	4258	15636	19894
PanTA	1578	2145	3723	4935	15397	20332	4106	16564	20670

Table S1: The numbers of core-, accessory- and total genes in the pangenomes constructed by the completing methods. The gene counts were from the pangenomes of 600 genomes for all the datasets so that PIRATE and PPanGGOLiN could complete the pangenome construction.

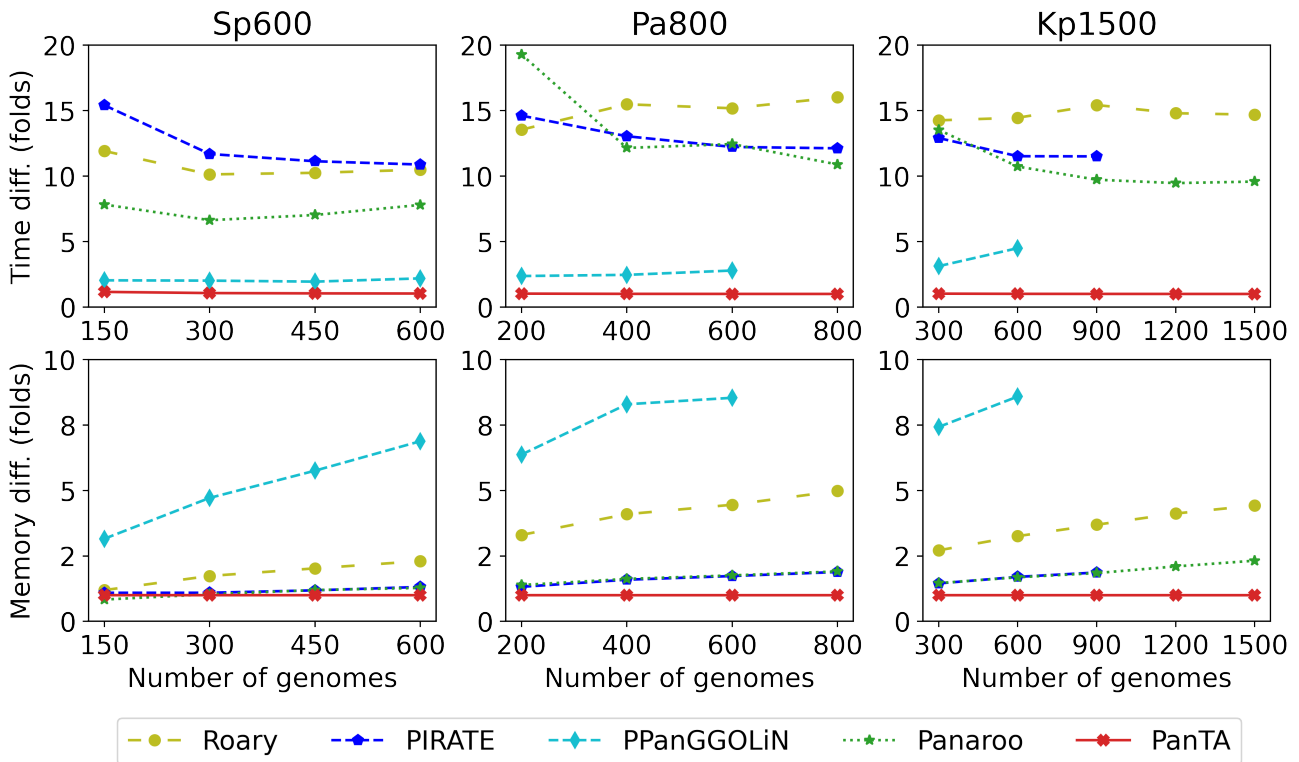


Fig. S1: Fold difference between computational time (top panel) and memory (bottom panel) required by existing methods against PanTA.

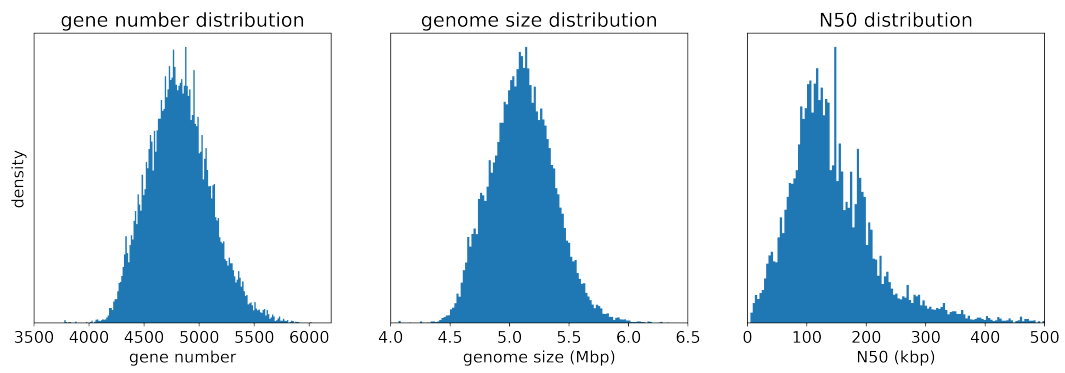


Fig. S2: Distributions of the gene number, genome size and N50 of the *Escherichia coli* genomes collected from RefSeq database.