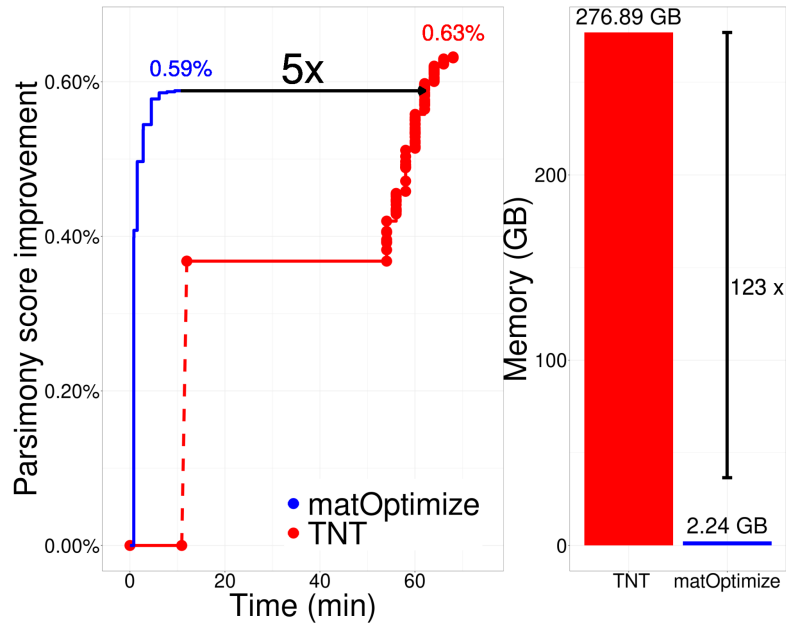
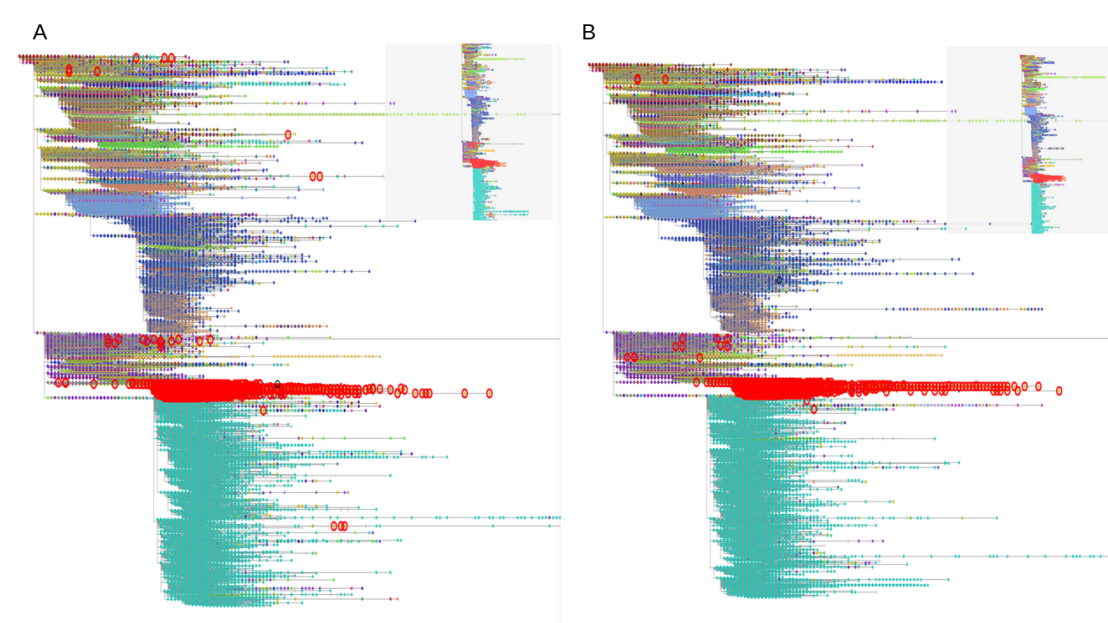


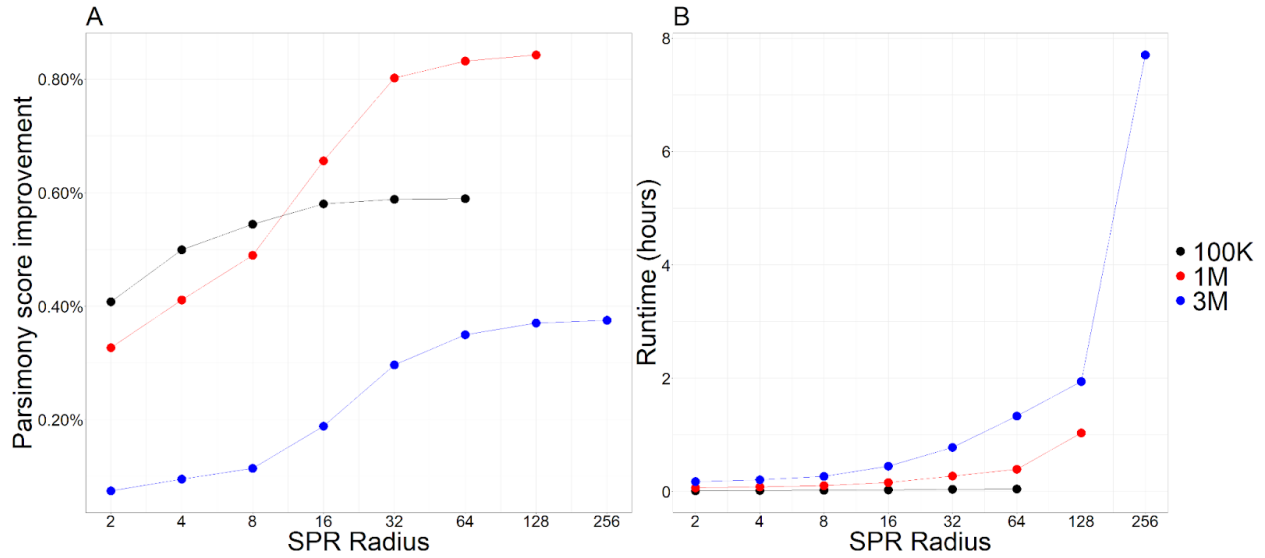
## Supplementary Figures And Tables



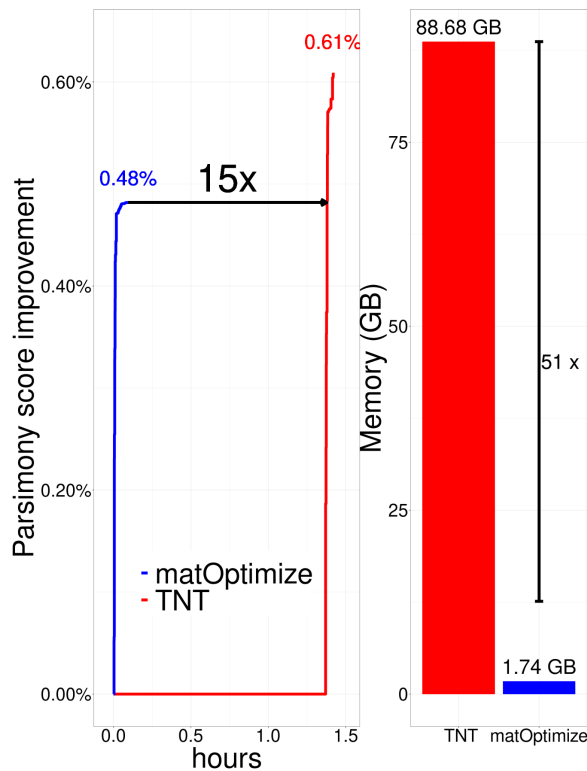
**Figure S1.** Comparison of parsimony score improvement and peak memory requirement of matOptimize and TNT, both using 40 threads on m1-ultramem-40 instance, starting from the SARS-CoV-2 based UShER-derived 100K-sample tree.



**Figure S2:** Taxonium (<https://taxonium.org/>) view of the 1M-sample tree (A) before and (B) after optimization. The tips of the trees are colored based on the lineage assignments derived from a trained PangoLEARN model (29), with P.1 labels highlighted using red circles.



**Figure S3: (A)** Parsimony score improvement and **(B)** the total runtime for different SPR radius achieved through the radius doubling mode in matOptimize.



**Figure S4.** Comparison of parsimony score improvement and peak memory requirement of matOptimize and TNT starting from the based UShER-derived 10K-sample tree of the *Mycobacterium tuberculosis* complex dataset. Benchmarking was done on iso-cost e2 instances of the Google Cloud Platform (GCP). TNT provided noticeably higher parsimony score improvement – 0.61% compared to 0.48% through matOptimize, but the log likelihood score of the two trees differed only by 0.007%.

Tree size (number of samples)	Parsimony Score Improvement	
	Radius Doubling	Fixed Radius
100K	0.182%	0.182%
1M	0.842%	0.798%
3M	0.375%	0.376%

**Table S1:** Parsimony score improvement with fixed radius and radius doubling optimization strategies in matOptimize.