

Supplementary Materials for Median Quartet
Tree Search Algorithms using Optimal Subtree
Prune and Regraft

Shayesteh Arasti

Siavash Mirarab

1 Supplementary Tables

Rooted Quartet		$CC \rightarrow C$		$GG \rightarrow G$
	1 2 3 4 5, 6	$n_0^{C_1} n_{-1}^{C_1} [n_{(\bullet\Box)}^{C_2} + n_{(\bullet\bullet)}^{C_2}]$ $+ \binom{n_{\bullet}^{C_1}}{2} n_{(0-1)}^{C_2}$ $+ n_0^{C_1} [n_{-1\uparrow(\bullet\Box)}^{C_2} + n_{-1\uparrow(\bullet\bullet)}^{C_2}]$ $+ n_{-1}^{C_1} [n_{0\uparrow(\bullet\Box)}^{C_2} + n_{0\uparrow(\bullet\bullet)}^{C_2}]$ $+ n_{\bullet}^{C_1} n_{\bullet\uparrow(0-1)}^{C_2}$	1 2	$n_{(0-1)}^{G_1} [n_{(\bullet\Box)}^{G_2} + n_{(\bullet\bullet)}^{G_2}]$ $+ n_{(0-1)}^{G_2} [n_{(\bullet\Box)}^{G_1} + n_{(\bullet\bullet)}^{G_1}]$
	1 2 3 4 5, 6	$+ n_0^{C_1} [n_{(-1(\bullet\Box))}^{C_2} + n_{(-1(\bullet\bullet))}^{C_2}]$ $+ [n_{[-1(\bullet\Box)]}^{C_1} + n_{[-1(\bullet\bullet)]}^{C_1}] n_0^{C_2}$ $+ n_{-1}^{C_1} [n_{(\bullet\Box)\uparrow 0}^{C_2} + n_{(\bullet\bullet)\uparrow 0}^{C_2}]$ $+ \binom{n_{\bullet}^{C_1}}{2} n_{-1\uparrow 0}^{C_2}$ $+ n_{\bullet}^{C_1} n_{\bullet\uparrow -1\uparrow 0}^{C_2}$	1 2	$+ n_0^{G_1} [n_{[-1(\bullet\Box)]}^{G_2} + n_{[-1(\bullet\bullet)]}^{G_2}]$ $+ n_0^{G_2} [n_{[-1(\bullet\Box)]}^{G_1} + n_{[-1(\bullet\bullet)]}^{G_1}]$
	1 2 3 4 5, 6	$+ n_{-1}^{C_1} [n_{(0(\bullet\Box))}^{C_2} + n_{(0(\bullet\bullet))}^{C_2}]$ $+ [n_{[0(\bullet\Box)]}^{C_1} + n_{[0(\bullet\bullet)]}^{C_1}] n_{-1}^{C_2}$ $+ n_0^{C_1} [n_{(\bullet\Box)\uparrow -1}^{C_2} + n_{(\bullet\bullet)\uparrow -1}^{C_2}]$ $+ \binom{n_{\bullet}^{C_1}}{2} n_{0\uparrow -1}^{C_2}$ $+ n_{\bullet}^{C_1} n_{\bullet\uparrow 0\uparrow -1}^{C_2}$	1 2	$+ n_{-1}^{G_1} [n_{[0(\bullet\Box)]}^{G_2} + n_{[0(\bullet\bullet)]}^{G_2}]$ $+ n_{-1}^{G_2} [n_{[0(\bullet\Box)]}^{G_1} + n_{[0(\bullet\bullet)]}^{G_1}]$
	1 2 3 4 5 6	$+ n_{\bullet}^{C_1} n_{(\bullet(0-1))}^{C_2}$ $+ n_{[\bullet(0-1)]}^{C_1} n_{\bullet}^{C_2}$ $+ n_{\bullet}^{C_1} n_{(0-1)\uparrow\bullet}^{C_2}$ $+ n_0^{C_1} n_{-1}^{C_1} [n_{\bullet\uparrow\Box}^{C_2} + n_{\bullet\uparrow\bullet}^{C_2}]$ $+ n_0^{C_1} [n_{-1\uparrow\bullet\uparrow\Box}^{C_2} + n_{-1\uparrow\bullet\uparrow\bullet}^{C_2}]$ $+ n_{-1}^{C_1} [n_{0\uparrow\bullet\uparrow\Box}^{C_2} + n_{0\uparrow\bullet\uparrow\bullet}^{C_2}]$	1 2	$+ n_{\bullet}^{G_1} n_{[\bullet(0-1)]}^{G_2}$ $+ n_{\bullet}^{G_2} n_{[\bullet(0-1)]}^{G_1}$

Table S1: **Equations for computing the HDT counter ρ_{comb}^X for component types $CC \rightarrow C$ and $GG \rightarrow G$.** Each row shows one rooted topology of the unrooted quartet $((i, j), (-1, 0))$. ρ_{comb}^X for each component type is the sum over the equations of all the rooted topologies.

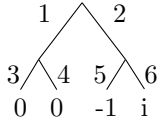
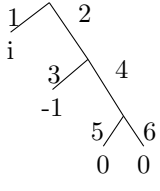
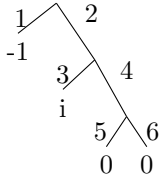
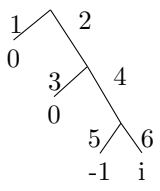
Rooted Quartet		$CC \rightarrow C$		$GG \rightarrow G$
	1 2 3, 4 5 6	$+\binom{C_1}{2}n_{(-1\bullet)}^{C_2}$ $+n_{-1}^{C_1}n_{(00)}^{C_1}n_{(00)}^{C_2}$ $+n_0^{C_1}n_{0\uparrow(-1\bullet)}^{C_2}$ $+n_{-1}^{C_1}n_{\bullet\uparrow(00)}^{C_2}$ $+n_{\bullet}^{C_1}n_{-1\uparrow(00)}^{C_2}$	1 2	$+n_{(00)}^{G_1}n_{(-1\bullet)}^{G_2}$ $+n_{(00)}^{G_2}n_{(-1\bullet)}^{G_1}$
	1 2 3 4 5, 6	$+n_{\bullet}^{C_1}n_{(-1(00))}^{C_2}$ $+n_{[-1(00)]}^{C_1}n_{\bullet}^{C_2}$ $+n_{-1}^{C_1}n_{(00)\uparrow\bullet}^{C_2}$ $+\binom{C_1}{2}n_{-1\uparrow\bullet}^{C_2}$ $+n_0^{C_1}n_{0\uparrow-1\uparrow\bullet}^{C_2}$	1 2	$+n_{\bullet}^{G_1}n_{[-1(00)]}^{G_2}$ $+n_{\bullet}^{G_2}n_{[-1(00)]}^{G_1}$
	1 2 3 4 5, 6	$+n_{-1}^{C_1}n_{(\bullet(00))}^{C_2}$ $+n_{[\bullet(00)]}^{C_1}n_{-1}^{C_2}$ $+n_{\bullet}^{C_1}n_{(00)\uparrow-1}^{C_2}$ $+\binom{C_1}{2}n_{\bullet\uparrow-1}^{C_2}$ $+n_0^{C_1}n_{0\uparrow\bullet\uparrow-1}^{C_2}$	1 2	$+n_{-1}^{G_1}n_{[\bullet(00)]}^{G_2}$ $+n_{-1}^{G_2}n_{[\bullet(00)]}^{G_1}$
	1 2 3 4 5 6	$+n_0^{C_1}n_{(0(-1\bullet))}^{C_2}$ $+n_{[0(-1\bullet)]}^{C_1}n_0^{C_2}$ $+n_0^{C_1}n_{(\bullet-1)\uparrow 0}^{C_2}$ $+n_{\bullet}^{C_1}n_{-1}^{C_1}n_{0\uparrow 0}^{C_2}$ $+n_{-1}^{C_1}n_{\bullet\uparrow 0\uparrow 0}^{C_2}$ $+n_{\bullet}^{C_1}n_{-1\uparrow 0\uparrow 0}^{C_2}$	1 2	$+n_0^{G_1}n_{[0(-1\bullet)]}^{G_2}$ $+n_0^{G_2}n_{[0(-1\bullet)]}^{G_1}$

Table S2: **Equations for computing the HDT counter ρ_{comb}^X for component types $CC \rightarrow C$ and $GG \rightarrow G$ (Cont.)** Each row shows one rooted topology of the unrooted quartet $((0, 0), (-1, i))$. ρ_{comb}^X for each component type is the sum over the equations of all the rooted topologies in Table S1 and S2.

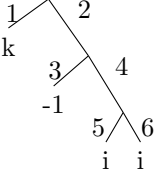
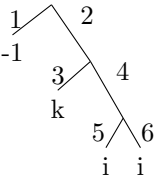
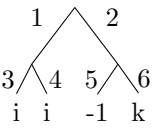
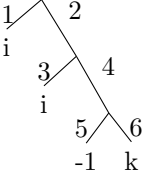
Rooted Quartet		$CC \rightarrow C$		$GG \rightarrow G$
	1 2 3 4 5, 6	$\begin{aligned} & \sum_{i \neq j} n_i^{C_1} (n_{(-1)(\bullet\bullet)}^{C_2} - n_{(-1)(ii)}^{C_2} - n_{(-1)(jj)}^{C_2}) \\ & + \sum_{i \neq j} n_{[-1(ii)]}^{C_1} (n_{\bullet\bullet}^{C_2} - n_i^{C_2} - n_j^{C_2}) \\ & + n_{-1}^{C_1} (n_{(\bullet\bullet)\uparrow\Box}^{C_2} - n_{(jj)\uparrow\bullet}^{C_2} - n_{(\bullet\bullet)\uparrow j}^{C_2}) \\ & + \sum_{i \neq j} \binom{n_i^{C_1}}{2} (n_{-1\uparrow\bullet}^{C_2} - n_{-1\uparrow i}^{C_2} - n_{-1\uparrow j}^{C_2}) \\ & + \sum_{i \neq j} n_i^{C_1} (n_{i\uparrow-1\uparrow\bullet}^{C_2} - n_{i\uparrow-1\uparrow j}^{C_2}) \end{aligned}$	1 2	$\begin{aligned} & \sum_{i \neq j} n_{[-1(ii)]}^{G_2} (n_{\bullet\bullet}^{G_1} - n_i^{G_1} - n_j^{G_1}) \\ & + \sum_{i \neq j} n_{[-1(ii)]}^{G_1} (n_{\bullet\bullet}^{G_2} - n_i^{G_2} - n_j^{G_2}) \end{aligned}$
	1 2 3 4 5, 6	$\begin{aligned} & + n_{-1}^{C_1} (n_{(\bullet\bullet)(\Box\Box)}^{C_2} - n_{(j)(\bullet\bullet)}^{C_2} - n_{(\bullet\bullet)(jj)}^{C_2}) \\ & + (n_{[\bullet\bullet](\Box\Box)}^{C_1} - n_{[j](\bullet\bullet)}^{C_1} - n_{[\bullet\bullet](jj)}^{C_1}) n_{-1}^{C_2} \\ & + \sum_{i \neq j} n_i^{C_1} (n_{(\bullet\bullet)\uparrow-1}^{C_2} - n_{(ii)\uparrow-1}^{C_2} - n_{(jj)\uparrow-1}^{C_2}) \\ & + \sum_{i \neq j} \binom{n_i^{C_1}}{2} (n_{\bullet\bullet\uparrow-1}^{C_2} - n_{i\uparrow-1}^{C_2} - n_{j\uparrow-1}^{C_2}) \\ & + \sum_{i \neq j} n_i^{C_1} (n_{i\uparrow\bullet\uparrow-1}^{C_2} - n_{i\uparrow j\uparrow-1}^{C_2}) \end{aligned}$	1 2	$\begin{aligned} & + n_{-1}^{G_1} (n_{[\bullet\bullet](\Box\Box)}^{G_2} - n_{[j](\bullet\bullet)}^{G_2} - n_{[\bullet\bullet](jj)}^{G_2}) \\ & + n_{-1}^{G_2} (n_{[\bullet\bullet](\Box\Box)}^{G_1} - n_{[j](\bullet\bullet)}^{G_1} - n_{[\bullet\bullet](jj)}^{G_1}) \end{aligned}$
	1 2 3, 4 5 6	$\begin{aligned} & + \sum_{i \neq j} \binom{n_i^{C_1}}{2} (n_{(-1)\bullet}^{C_2} - n_{(-1)i}^{C_2} - n_{(-1)j}^{C_2}) \\ & + \sum_{i \neq j} n_i^{C_1} n_{-1}^{C_1} (n_{(\bullet\bullet)}^{C_2} - n_{(ii)}^{C_2} - n_{(jj)}^{C_2}) \\ & + \sum_{i \neq j} n_i^{C_1} (n_{(\bullet\bullet)\uparrow(-1)\bullet}^{C_2} - n_{(ii)\uparrow(-1)j}^{C_2}) \\ & + n_{-1}^{C_1} (n_{\bullet\bullet\uparrow(\Box\Box)}^{C_2} - n_{j\uparrow(\bullet\bullet)}^{C_2} - n_{\bullet\bullet\uparrow(jj)}^{C_2}) \\ & + \sum_{i \neq j} n_i^{C_1} (n_{-1\uparrow(\bullet\bullet)}^{C_2} - n_{-1\uparrow(ii)}^{C_2} - n_{-1\uparrow(jj)}^{C_2}) \end{aligned}$	1 2	$\begin{aligned} & + \sum_{i \neq j} n_{(ii)}^{G_1} [n_{(-1)\bullet}^{G_2} - n_{(-1)i}^{G_2} - n_{(-1)j}^{G_2}] \\ & + \sum_{i \neq j} n_{(ii)}^{G_2} [n_{(-1)\bullet}^{G_1} - n_{(-1)i}^{G_1} - n_{(-1)j}^{G_1}] \end{aligned}$
	1 2 3 4 5 6	$\begin{aligned} & + \sum_{i \neq j} n_i^{C_1} (n_{(i(-1)\bullet)}^{C_2} - n_{(i(-1)j)}^{C_2}) \\ & + \sum_{i \neq j} (n_{[i(-1)\bullet]}^{C_1} - n_{[i(-1)j]}^{C_1}) n_i^{C_2} \\ & + \sum_{i \neq j} n_i^{C_1} (n_{(\bullet(-1)\uparrow i)}^{C_2} - n_{(j(-1)\uparrow i)}^{C_2}) \\ & + \sum_{i \neq j} n_i^{C_1} n_{-1}^{C_1} (n_{\bullet\bullet}^{C_2} - n_{j\uparrow j}^{C_2} - n_{i\uparrow i}^{C_2}) \\ & + n_{-1}^{C_1} (n_{\bullet\bullet\uparrow\Box\Box}^{C_2} - n_{j\uparrow\bullet}^{C_2} - n_{\bullet\bullet\uparrow j}^{C_2}) \\ & + \sum_{i \neq j} n_i^{C_1} (n_{-1\uparrow\bullet\uparrow}^{C_2} - n_{-1\uparrow i\uparrow}^{C_2} - n_{-1\uparrow j\uparrow}^{C_2}) \end{aligned}$	1 2	$\begin{aligned} & + \sum_{i \neq j} n_i^{G_1} (n_{[i(-1)\bullet]}^{G_2} - n_{[i(-1)j]}^{G_2}) \\ & + \sum_{i \neq j} n_i^{G_2} (n_{[i(-1)\bullet]}^{G_1} - n_{[i(-1)j]}^{G_1}) \end{aligned}$

Table S3: **Equations for computing the HDT counter $\pi_{comb_j}^X$ for component types $CC \rightarrow C$ and $GG \rightarrow G$** Each row shows one rooted topology of the unrooted quartet $((i, i), (-1, k))$. $\pi_{comb_j}^X$ for each component type is the sum over the equations of all the rooted topologies.

Table S4: **E2 Results.** Comparing the running time of Q-SPR and ASTRAL-III in minutes. For the Q-SPR search method, the running time of building the starting tree is also considered. Both the heuristic and random versions of Q-SPR search are included.

Genes	ILS-Level	Runtime (minutes)			
		ASTRAL-III	Q-SPR (random)	Q-SPR (heuristic)	starting tree
$k = 50$	High	1.28	59.68	50.23	5.22
	Med	0.23	41.00	29.97	2.68
	Low	0.17	25.75	25.9	2.52
$k = 200$	High	4.37	179.72	146.52	11.12
	Med	0.48	151.77	107.97	9.98
	Low	0.27	93.82	98.60	10.02

2 Supplementary Figures

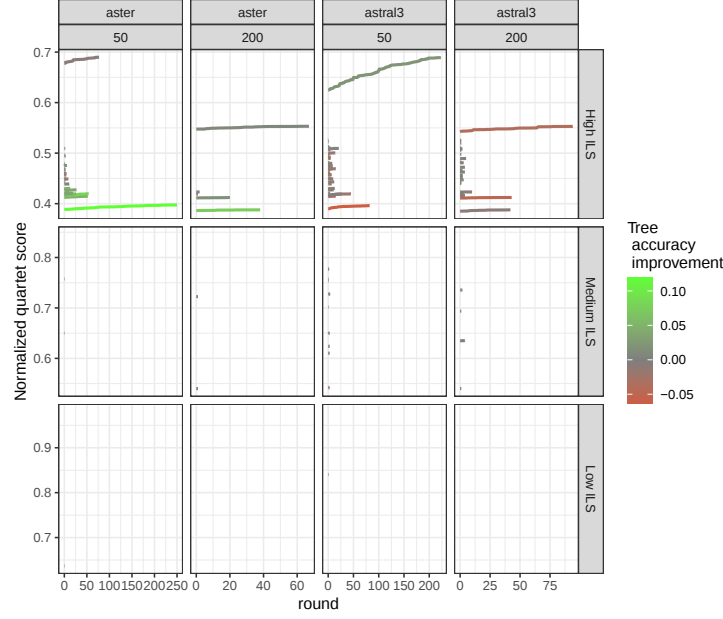


Figure S2: **E4 Results: increase in quartet score across runs.** The normalized quartet score between the Q-SPR tree at the end of each SPR round and the gene trees for all model conditions. The final improvement of the Q-SPR tree with respect to the true species tree is shown in colors.

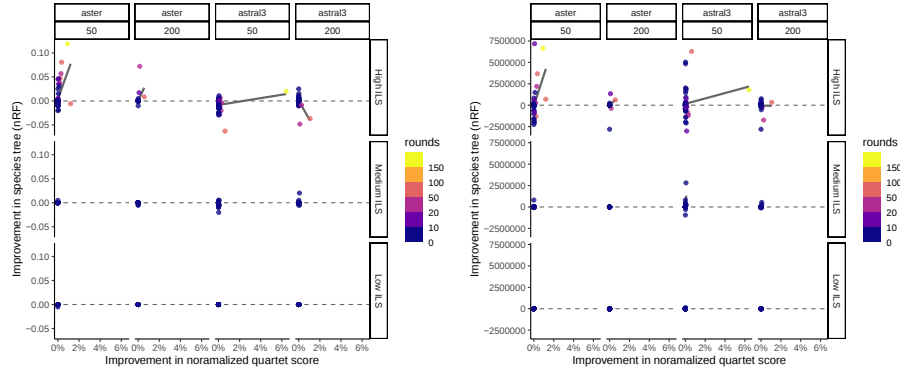


Figure S3: **E4 Results: optimization score versus accuracy.** Improvement in the quartet score of the Q-SPR algorithm above the ASTER and ASTRAL-III tree with respect to the gene trees versus the improvement in the normalized RF or the quartet distance between the ASTER tree and the true species tree. The number of SPR rounds performed for each replicate is shown in colors.

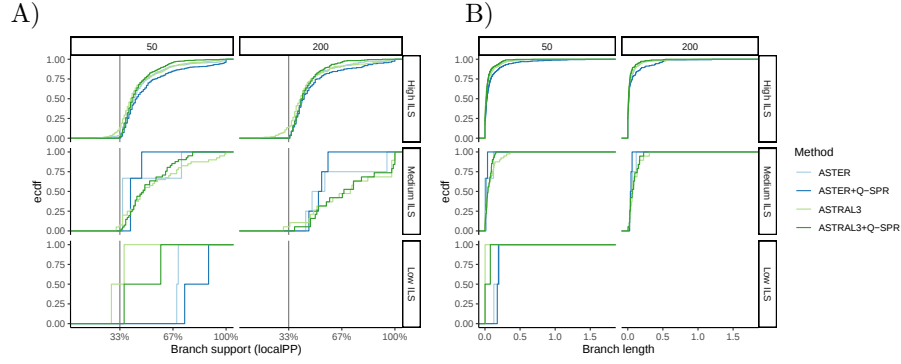


Figure S4: **E4 Results: improvements in branch support and branch length.** The results of the analysis on the non-matching branch lengths and branch supports of the improved tree using Q-SPR and the starting ASTER/ASTRAL-III tree. A) The branch supports for the improved tree. The non-matching branches refer to the branches that are either removed from the starting tree or added to the improved tree. The branch support threshold $1/3$ is shown by the dotted horizontal line. B) The empirical cumulative density function (ECDF) of the branch lengths of the improved and the starting tree for the non-matching branches.