Local branch recrafting. Edmonds' algorithm attempts to minimize the sum of the edge lengths in the tree. However, the resulting dMST does not necessarily represent true phylogenetic relationships between strains because allelic distances do not always correlate with divergence time. We therefore implemented a subsequent branch optimization step that accounts for these discrepancies. Algorithm 1 gives an overview over the local branch recrafting (see also Supplemental Fig. S5), starting from the already computed dMST(V,E), where E is a distance matrix sorted in ascending order of allelic distances, and a forest F where each $u \in V$ is a single tree $t(u) \in F$.

Algorithm 1 Local branch recrafting

Input: Initial edge $(u \to v)$, Tree $t(u) \in F$, harmonic tiebreaker ht

Output: New edge $(u' \to v')$

1: Initialize $u' = u$

2: **for each** node $w \in TargetNodes\big(t(u')\big)$ **do**

3: $\quad P(M_A), P(M_B) = ModelSelection\big(d(u' \to w), d(w \to u'), d(u' \to v), d(w \to v)\big)$

4: $\quad$ **if** $P(M_A) \geq P(M_B) \wedge ht(u') > ht(w)$ **then**

5: $\quad\quad u' = w$

6: $\quad$ **else if** $P(M_A) < P(M_B) \wedge d(u' \to v) > d(w \to v)$ **then**

7: $\quad\quad u' = w$

8: Repeat $(1 - 7)$ on v to obtain v'

9: Return $(u' \to v')$

Optimizations are applied to both ends of each edge iteratively as shown in Supplemental Fig. S5D. The $TargetNodes()$ function picks a subset of the nodes in tree $t(u)$ which are the centroids and the nodes that are directly connected to u (Supplemental Fig. S5D). The $ModelSelection()$ function compares the maximum likelihoods of two models M_A and M_B (Supplemental Fig. S5B,C). Here we describe only the model selection process for u. Given $d(u \to w)$, $d(w \to u)$, $d(u \to v)$ and $d(w \to v)$, when assuming $d(u \to v) \geq d(w \to v)$, the proportions of invariable sites in branches l_A, k_A, l_B and k_B satisfy:

$$\underset{0 \leq l_A \leq 1, 0 \leq k_A \leq 1}{\operatorname{argmax}} \quad \log P(M_A | l_A, k_A)$$

$$= \underset{0 \leq l_A \leq 1, 0 \leq k_A \leq 1}{\operatorname{argmax}} \quad \log P(u \to w | l_A) P(u \to v | l_A, k_A) P(w \to v | l_A, k_A)$$

$$= \underset{0 \leq l_A \leq 1, 0 \leq k_A \leq 1}{\operatorname{argmax}} \quad |L| d(u \to w) \log(1 - l_A^2) + |L|\big(1 - d(u \to w)\big) \log(l_A^2)$$

$$+|L|d(u \to v)\log(1 - l_A k_A) + |L|\big(1 - d(u \to v)\big)\log(l_A k_A)$$

$$+|L|d(w \to v)\log(1 - l_A k_A) + |L|\big(1 - d(w \to v)\big)\log(l_A k_A) \tag{1}$$

$$\operatorname*{argmax}_{0 \le l_B \le 1, 0 \le k_B \le 1} \log P(M_B | l_B, k_B)$$

$$= \operatorname*{argmax}_{0 \le l_B \le 1, 0 \le k_B \le 1} \log P(w \to u | l_B) P(u \to v | l_B, k_B) P(w \to v | l_B, k_B)$$

$$= \operatorname*{argmax}_{0 \le l_B \le 1, 0 \le k_B \le 1} |L|d(w \to u)\log(1 - l_B) + |L|\big(1 - d(w \to u)\big)\log(l_B)$$

$$+|L|d(u \to v)\log(1 - l_B k_B) + |L|\big(1 - d(u \to v)\big)\log(l_B k_B)$$

$$+|L|d(w \to v)\log(1 - k_B) + |L|\big(1 - d(w \to v)\big)\log(k_B) \tag{2}$$

where *L* is a set of loci in an MLST profile. Note that the direction of the distances between *u* and *w* are different in the two equations. Model *A* assumes u as the centroid node and adopts $d(u \to w)$ in equation (1) whereas model *B* treats w as the centroid and thus uses $d(w \to u)$. We further denote

$$x = 1 - \frac{\big(1 - d(w \to u)\big)\big(1 - d(w \to v)\big) + \big(1 - d(u \to v)\big)}{2}$$

Then the parameters in equations (1) and (2) are calculated as:

$$l_A = \sqrt{1 - d(u \to w)}$$

$$k_A = \frac{1 - \frac{1}{2}\big(d(u \to v) + d(w \to v)\big)}{l_A}$$

$$l_B = 1 + \frac{x d(w \to u)}{d(u \to v) - 2x}$$

$$k_B = 1 + \frac{x d(w \to v)}{d(u \to v) - 2x}$$

These parameters can then be used to calculate $P(M_A)$ and $P(M_B)$ using equations (1) and (2).