Supporting Information

Pardi and Gascuel 10.1073/pnas.1118368109

Appendix 1

Here, we present an interpretation of the $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae as averages over a large number of simple branch length formulae, which allows us to prove Theorem 1: Because these simple formulae are correct—that is, they correctly provide the branch lengths of a tree whenever the input distances are additive with respect to that tree—it follows that also the $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae are correct. The detailed arguments follow below.

Let $\hat{\ell}_e(\boldsymbol{\delta})$ denote the length that is assigned to branch e by an adopted length estimation method. Let T^* be a tree where e has length ℓ_e. The adopted method is correct if, for any such tree T^*, $\hat{\ell}_e(\boldsymbol{d}^{T^*}) = \ell_e$.

Suppose e is an external branch, and define A, B, i as in Fig. 1*A* in the main text. Choose a taxon a from A and a taxon b from B. Then calculate the length of e with:

$$\hat{\ell}_e^{ab}(\boldsymbol{\delta}) = \frac{1}{2}(\delta_{ia} + \delta_{ib} - \delta_{ab}).$$

If instead e is an internal branch, let A_1, A_2, B_1, and B_2 be the four clades surrounding it, as in Fig. 1*B* in the main text. Choose taxa a_1, a_2, b_1, b_2 from A_1, A_2, B_1, B_2, respectively. It is clear that any drawing of the tree on the plane either places A_1 to the side of B_1 and therefore A_2 to the side of B_2 (as in Fig. S1*B*, *Top*) or alternatively A_1 to the side of B_2 and A_2 to the side of B_1 (Fig. S1*B*, *Bottom*). We associate the former drawing with the following formula for the length of e:

$$\hat{\ell}_e^{a_1b_1a_2b_2}(\boldsymbol{\delta}) = \frac{1}{2}(\delta_{a_1b_1} + \delta_{a_2b_2} - \delta_{a_1a_2} - \delta_{b_1b_2}).$$

The alternative drawing is associated to the formula $\hat{\ell}_e^{a_1b_2a_2b_1}(\boldsymbol{\delta})$. Note that $\hat{\ell}_e^{ab}(\boldsymbol{\delta})$, $\hat{\ell}_e^{a_1b_1a_2b_2}(\boldsymbol{\delta})$ and $\hat{\ell}_e^{a_1b_2a_2b_1}(\boldsymbol{\delta})$ are all trivially correct.

The $\boldsymbol{\gamma}^T$ and $\boldsymbol{\lambda}^T$ parameters introduced in the main text can be interpreted as controlling a probability distribution over all possible such formulae for calculating the length of a given branch e in T. First, as illustrated in Fig. S1*A*, the $\boldsymbol{\gamma}^T$ parameters determine the probability of choosing a given taxon out of any given clade (a from A, b from B in the case of external branches, a_1 from A_1, a_2 from A_2, b_1 from B_1, b_2 from B_2 in the case of internal branches). Second, as illustrated in Fig. S1*B*, the $\boldsymbol{\lambda}^T$ parameters determine the probability of choosing either of the two possible drawings for the clades around an internal branch e, and therefore either of $\hat{\ell}_e^{a_1b_1a_2b_2}(\boldsymbol{\delta})$ or $\hat{\ell}_e^{a_1b_2a_2b_1}(\boldsymbol{\delta})$ for the length of e: $\lambda_{A_1B_1} = \lambda_{A_2B_2}$ is the probability of drawing A_1 to the side of B_1 and A_2 to the side of B_2, whereas its complement $\lambda_{A_1B_2} = \lambda_{A_2B_1} = 1 - \lambda_{A_1B_1}$ is the probability of drawing A_1 to the side of B_2 and A_2 to the side of B_1.

Given this probability distribution, let us take the resulting expected value of the length assigned to e. In the (harder) case of an internal branch, this is given by

$$\begin{aligned}
&\sum_{\substack{a_1\in A_1, b_1\in B_1\\ a_2\in A_2, b_2\in B_2}} p_{a_1|A_1}p_{a_2|A_2}p_{b_1|B_1}p_{b_2|B_2}[\lambda_{A_1B_1}\hat{\ell}_e^{a_1b_1a_2b_2}(\boldsymbol{\delta})\\
&\quad + \lambda_{A_1B_2}\hat{\ell}_e^{a_1b_2a_2b_1}(\boldsymbol{\delta})]\\
&= \frac{1}{2}\sum_{\substack{a_1\in A_1, b_1\in B_1\\ a_2\in A_2, b_2\in B_2}} p_{a_1|A_1}p_{a_2|A_2}p_{b_1|B_1}p_{b_2|B_2}[\lambda_{A_1B_1}(\delta_{a_1b_1} + \delta_{a_2b_2})\\
&\quad + (1 - \lambda_{A_1B_1})(\delta_{a_1b_2} + \delta_{a_2b_1}) - \delta_{a_1a_2} - \delta_{b_1b_2}]\\
&= \frac{1}{2}[\lambda_{A_1B_1}(\delta_{A_1B_1} + \delta_{A_2B_2}) + (1 - \lambda_{A_1B_1})(\delta_{A_1B_2} + \delta_{A_2B_1})\\
&\quad - \delta_{A_1A_2} - \delta_{B_1B_2}].
\end{aligned}$$

Thus what we obtain (also in the easier case of an external branch; not shown) are precisely the $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae. In other words, these formulae can be seen as providing the expected length of a branch when this is assigned following the random procedure described above. Given this observation, the correctness of the $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae follows trivially from the correctness of the base formulae $\hat{\ell}_e^{ab}(\boldsymbol{\delta})$ and $\hat{\ell}_e^{a_1b_1a_2b_2}(\boldsymbol{\delta})$. Theorem 1 is therefore proved.

We note that the approach of expressing a length estimator as the combination of several simple formulae has already been considered by Willson (1). His base formulae, however, express the length of a path in the tree (as a function of the distances between three taxa) rather than a single branch (which we express as a function of the distances between three or four taxa, for exterior and interior branches, respectively). Moreover, the combination of his base formulae provide an estimate of the total length of the tree (in the ME_{+1} sense).

Appendix 2

Here, we prove the relationship between the M&P formulae (2) and our $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae, as stated in Theorem 2. We start by formally defining the M&P formulae (A2.1); then we introduce a few additional formalisms and a useful observation (A2.2) and then prove separately the two parts of Theorem 2 (*i* in A2.3 and *ii* in A2.4). This requires showing how to derive the parameters of each class of formulae from the parameters of the other class (i.e., $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$ from $\boldsymbol{w}$, and vice versa).

A2.1. The M&P Formulae. We assume that the weights $\boldsymbol{w} = (w_{ij})$ are multiplicative w.r.t. a binary topology T. Then, for any two clades A and B of T, define

$$Z_{AB} = \sum_{\substack{i\in A\\ j\in B}} w_{ij} \quad \text{and} \quad \delta^{\boldsymbol{w}}_{AB} = \frac{1}{Z_{AB}}\sum_{\substack{i\in A\\ j\in B}} w_{ij}\delta_{ij}.$$

Mihaescu and Pachter (2) have shown that the optimal branch lengths of T with respect to the WLS criterion **1** are then given by the following formula, applicable to any branch e in T:

$$\hat{\ell}_e(\boldsymbol{\delta}) = \begin{cases} \frac{1}{2}(\delta^{\boldsymbol{w}}_{iA} + \delta^{\boldsymbol{w}}_{iB} - \delta^{\boldsymbol{w}}_{AB}) & \text{if } e \text{ is external,}\\ \frac{1}{2}\left[\frac{Z_{A_1B_2}+Z_{A_2B_1}}{Z_{AB}}(\delta^{\boldsymbol{w}}_{A_1B_1} + \delta^{\boldsymbol{w}}_{A_2B_2}) + \frac{Z_{A_1B_1}+Z_{A_2B_2}}{Z_{AB}}(\delta^{\boldsymbol{w}}_{A_1B_2} + \delta^{\boldsymbol{w}}_{A_2B_1}) - \delta^{\boldsymbol{w}}_{A_1A_2} - \delta^{\boldsymbol{w}}_{B_1B_2}\right] & \text{if } e \text{ is internal,}\end{cases}$$

where, if e is external, we define A, B, i as in Fig. 1A in the main text and, if e is internal, we define A_1, A_2, $A = A_1 \cup A_2$, B_1, B_2, $B = B_1 \cup B_2$ as in Fig. 1B in the main text.

A2.2. Decomposition of Z_{XY}. Extend the w_{ij} notation to any pair of nodes x and y (possibly internal) in T:

$$w_{xy} = \prod_{e \in P_{xy}(T)} w_e, \qquad \textbf{[S1]}$$

where we recall that the w_e are the branch-associated weights that compose the pairwise weights w_{ij}, and $P_{xy}(T)$ is the set of branches on the path between x and y in T. Then define the *multiplicative weight of* X (a clade with root x) as:

$$Z_X = \sum_{i \in X} w_{ix}. \qquad \textbf{[S2]}$$

We assume $w_{xx} = 1$ for any node x, which implies $Z_{\{i\}} = 1$, for any one-taxon clade $\{i\}$. It is then easy to check that, if X and Y are any two disjoint clades in T, with roots x and y , respectively, then

$$Z_{XY} = w_{xy} Z_X Z_Y. \qquad \textbf{[S3]}$$

A2.3. The M&P Formulae are Also $(\gamma^{\text{‹}}\,\lambda)$-Formulae.

Lemma 1. *Given weights* $\boldsymbol{w} = (w_{ij})$ *multiplicative w.r.t. a binary topology* T*, define, for each pair of adjacent branches* e *and* f*:*

$$\gamma_{ef} = \frac{w_f Z_{A_1}}{Z_A}, \qquad \textbf{[S4]}$$

where A_1 *and* $A = A_1 \cup A_2$ *are the clades having* f *and* e *as root branches, respectively, as in Fig. 1B. Then,*

i. The resulting average distances between clades are such that $\delta_{XY} = \delta^{\boldsymbol{w}}_{XY}$, for any two disjoint clades X and Y in T.
ii. For any internal branch e, let $A = A_1 \cup A_2$ and $B = B_1 \cup B_2$ be the clades in the configuration of Fig. 1B; then $(Z_{A_1B_2} + Z_{A_2B_1})/Z_{AB} = \gamma_{ef} + \gamma_{eh} - 2\gamma_{ef}\gamma_{eh}$.

Proof:

i. Eq. **S3** allows us to express $\delta^{\boldsymbol{w}}_{XY}$ in a very similar form to that of δ_{XY}:

$$\delta^{\boldsymbol{w}}_{XY} = \sum_{\substack{i \in X \\ j \in Y}} \frac{w_{ij}}{Z_{XY}} \delta_{ij} = \sum_{\substack{i \in X \\ j \in Y}} \frac{w_{ix}}{Z_X} \frac{w_{jy}}{Z_Y} \delta_{ij},$$

   where x and y are the root nodes of X and Y, respectively. In order to have $\delta_{XY} = \delta^{\boldsymbol{w}}_{XY}$, it is then sufficient to prove that, for any clade X with root x, and any taxon $i \in X$,

$$p_{i|X} = \frac{w_{ix}}{Z_X}.$$

   Let the path from x to i traverse branches $e_1, e_2, \ldots, e_k$ (in this order), and let X_j be the subclade of X having e_j as root branch ($0 \le j \le k$, with $X_0 = X$ and e_0 being the root branch of X); then, $X_k = \{i\}$ and

$$p_{i|X} = \gamma_{e_0e_1} \cdot \gamma_{e_1e_2} \cdot \ldots \cdot \gamma_{e_{k-1}e_k} = \frac{w_{e_1} Z_{X_1}}{Z_X} \cdot \frac{w_{e_2} Z_{X_2}}{Z_{X_1}} \cdot \ldots$$
$$\cdot \frac{w_{e_k} Z_{X_k}}{Z_{X_{k-1}}} = \frac{w_{e_1} \cdot w_{e_2} \cdot \ldots \cdot w_{e_k}}{Z_X} Z_{\{i\}} = \frac{w_{ix}}{Z_X}.$$

ii. Using again Eq. **S3**,

$$\frac{Z_{A_1B_2} + Z_{A_2B_1}}{Z_{AB}} = \frac{w_f w_l Z_{A_1} Z_{B_2} + w_g w_h Z_{A_2} Z_{B_1}}{Z_A Z_B}$$
$$= \gamma_{ef}\gamma_{el} + \gamma_{eg}\gamma_{eh} = \gamma_{ef} + \gamma_{eh} - 2\gamma_{ef}\gamma_{eh}.$$

Points i and ii are thus both verified and Lemma 1 is proved.

Setting γ_{ef} as in Eq. **S4** has a simple intuitive meaning: If we call g the root branch of A_2 (as in Fig. 1B), then $Z_A = w_f Z_{A_1} + w_g Z_{A_2}$. The γ_{ef} above can then be seen as the relative multiplicative weight of the subtree corresponding to clade A_1 in the subtree corresponding to clade A. The random walk defined by these parameters is then "attracted" by the heavier subtrees, in a way that is directly proportional to the weights of the subtrees.

Proof of Theorem 2, part i: Given $\boldsymbol{w}$ multiplicative w.r.t. T, define the $\boldsymbol{\gamma}^T$ parameters as in Eq. **S4**. Note that because $w_e > 0$ for any branch e, then $Z_X > 0$ for every clade X. Moreover, for any three adjacent branches e, f , and g in the configuration of Fig. 1B, we have $w_f Z_{A_1} + w_g Z_{A_2} = Z_A$ and $w_f Z_{A_1}, w_g Z_{A_2} > 0$, which imply $\gamma_{ef} + \gamma_{eg} = 1$ and $0 < \gamma_{ef}, \gamma_{eg} < 1$. Therefore the definition of the $\boldsymbol{\gamma}^T$ parameters is admissible and implies that $\delta_{XY} = \delta^{\boldsymbol{w}}_{XY}$ (Lemma 1, part i).

As for the $\boldsymbol{\lambda}^T$ parameters, set $\lambda_{A_1B_1} = (Z_{A_1B_2} + Z_{A_2B_1})/Z_{AB}$, for every pair of clades A_1 and B_1 separated by 3 branches, and being in the configuration of Fig. 1B with $A = A_1 \cup A_2$ and $B = B_1 \cup B_2$. It is easy to check that this implies $\lambda_{A_1B_1} = \lambda_{A_2B_2} > 0$, $\lambda_{A_1B_2} = \lambda_{A_2B_1} > 0$ and $\lambda_{A_1B_1} + \lambda_{A_1B_2} = 1$, and therefore the definition of the $\boldsymbol{\lambda}^T$ parameters is also admissible.

It is now easy to verify that the resulting $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae coincide with the M&P formulae corresponding to $\boldsymbol{w}$: for the external branches this is an immediate consequence of $\delta_{XY} = \delta^{\boldsymbol{w}}_{XY}$, while for the internal branches we also use the above definition of $\lambda_{A_1B_1}$ and the fact that $1 - \lambda_{A_1B_1} = (Z_{A_1B_1} + Z_{A_2B_2})/Z_{AB}$. Finally, the $\boldsymbol{\gamma}^T$ and $\boldsymbol{\lambda}^T$ defined above satisfy properties P1 and P2: the first can be verified by using Eq. **S4** in P1, and the second is a direct consequence of Lemma 1, part ii.

A2.4. Characterization of the (γ,λ)-Formulae That are Also M&P Formulae.

Proof of Theorem 2, part ii: The proof has the following structure: As an intermediate step, we introduce—for every three clades A_1, A_2, and B whose respective root branches f, g, and e are incident to the same internal node (as in Fig. 1B)—three values φ_{A_1}, φ_{A_2} and φ_B such that $\varphi_{A_1} + \varphi_{A_2} + \varphi_B = 1$, $0 < \varphi_{A_1}, \varphi_{A_2}, \varphi_B < 1$ and

$$\gamma_{ef} = \frac{\varphi_{A_1}}{\varphi_{A_1} + \varphi_{A_2}}, \qquad \gamma_{fg} = \frac{\varphi_{A_2}}{\varphi_{A_2} + \varphi_B}, \qquad \gamma_{ge} = \frac{\varphi_B}{\varphi_B + \varphi_{A_1}}. \qquad \textbf{[S5]}$$

The existence of such values is guaranteed by property P1. In intuitive terms, we are requiring that each clade has somewhat a "preference value," such that the probabilities γ_{ef} and γ_{eg} are proportional to the preference values of the clades that f and g lead to. On the basis of these values, we then define a set of branch-associated weights w_e specifying a multiplicative model such that the preference values can be obtained as

$$\varphi_B = \frac{w_e Z_B}{w_e Z_B + w_f Z_{A_1} + w_g Z_{A_2}}. \qquad \textbf{[S6]}$$

(With A_1, A_2, B, f, g, e as above.) It is easy to see (shown below) that this implies that Eq. **S4** of Lemma 1, and therefore its conclusions, hold. This, together with the fact that the $\boldsymbol{\lambda}^T$ parameters satisfy P2, implies that the M&P formulae for $\boldsymbol{w}$ coincide with the given $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae.

Let us now look in detail at each step. First of all, property P1 implies that the determinant of the coefficient matrix for the system of linear equations in φ_{A_1}, φ_{A_2} , and φ_B corresponding to Eqs. **S5** is equal to 0; this system is then solved by the following subspace of solutions:

$$(\varphi_B, \varphi_{A_1}, \varphi_{A_2}) \in \left\{ \left(x, \frac{1-\gamma_{ge}}{\gamma_{ge}} x, \frac{1-\gamma_{ef}}{\gamma_{ef}} \frac{1-\gamma_{ge}}{\gamma_{ge}} x \right) | x \in \mathbb{R} \right\}.$$

If furthermore we impose $\varphi_{A_1} + \varphi_{A_2} + \varphi_B = 1$, it is easy to see that a unique solution is determined, such that $\varphi_{A_1}, \varphi_{A_2}, \varphi_B > 0$ (and therefore also <1).

Given the φ_X parameters, we now show how to define the branch-associated weights w_e. For any internal branch e separating clades A and $\overline{A}$, let

$$w_e = \sqrt{\frac{\varphi_A}{1-\varphi_A} \frac{\varphi_{\overline{A}}}{1-\varphi_{\overline{A}}}}. \qquad \textbf{[S7]}$$

As for the external branches, which for simplicity we call with the same names as the taxa they are incident to (e.g., branch i being the one incident to taxon i), we assign their weight in the following way: Choose arbitrarily $w_1 > 0$ and then, for any other external branch $i \in \{2, 3, \ldots, n\}$ define

$$w_i = w_1 \sqrt{\frac{\gamma^*_{1i}}{\gamma^*_{i1}} \frac{1-\varphi_1}{\varphi_1} \frac{\varphi_i}{1-\varphi_i}}, \qquad \textbf{[S8]}$$

where $\gamma^*_{ef} = \gamma_{ee_1} \cdot \gamma_{e_1 e_2} \cdot \ldots \cdot \gamma_{e_k f}$, for any pair of branches e, f linked by a path composed of the ordered sequence of branches $(e_1, e_2, \ldots, e_k)$, and for simplicity we write φ_1 and φ_i instead of $\varphi_{\{1\}}$ and $\varphi_{\{i\}}$.

The weights thus defined determine a multiplicative weighting $\boldsymbol{w} = (w_{ij})$ that satisfies Eq. **S6**. In order to show this, we first show that, for every clade A (whose root branch we call e) such that $|A| < n-1$ (i.e., the endpoint of e on the other side of A is not a leaf),

$$w_e Z_A = w_1 \sqrt{\frac{\gamma^*_{1e}}{\gamma^*_{e1}} \frac{1-\varphi_1}{\varphi_1} \frac{\varphi_A}{1-\varphi_A}}. \qquad \textbf{[S9]}$$

We prove this by induction on the size of A. If A consists of a single taxon, then either this is taxon 1, in which case both sides of the equation reduce to w_1, or this is another taxon i, in which case Eq. **S9** coincides with Eq. **S8**. In both cases Eq. **S9** trivially holds. If $|A| > 1$, let A_1, A_2, B, f and g be as in Fig. 1*B*. Then, by inductive hypothesis, Eq. **S9** holds for $w_f Z_{A_1}$ and $w_g Z_{A_2}$ and we have:

$$\begin{aligned} Z_A &= w_f Z_{A_1} + w_g Z_{A_2} \\ &= w_1 \sqrt{\frac{1-\varphi_1}{\varphi_1}} \left(\sqrt{\frac{\gamma^*_{1f}}{\gamma^*_{f1}} \frac{\varphi_{A_1}}{1-\varphi_{A_1}}} + \sqrt{\frac{\gamma^*_{1g}}{\gamma^*_{g1}} \frac{\varphi_{A_2}}{1-\varphi_{A_2}}} \right) \end{aligned}$$

Now note that

$$\frac{\gamma^*_{1f}}{\gamma^*_{f1}} = \frac{\gamma^*_{1e}}{\gamma^*_{e1}} \frac{\varphi_{A_1}(1-\varphi_{A_1})}{\varphi_B(1-\varphi_B)} \quad \text{and} \quad \frac{\gamma^*_{1g}}{\gamma^*_{g1}} = \frac{\gamma^*_{1e}}{\gamma^*_{e1}} \frac{\varphi_{A_2}(1-\varphi_{A_2})}{\varphi_B(1-\varphi_B)}, \qquad \textbf{[S10]}$$

which can be verified by noting that, depending on the position of taxon 1 (in A_1, A_2 or B), $\gamma^*_{1f}/\gamma^*_{f1}$ can either be equal to $(\gamma^*_{1e}/\gamma^*_{e1})(\gamma_{ef}/\gamma_{fe})$ or $(\gamma^*_{1e}/\gamma^*_{e1})(\gamma_{gf}\gamma_{eg})/(\gamma_{ge}\gamma_{fg})$, and $\gamma^*_{1g}/\gamma^*_{g1}$ can either be equal to $(\gamma^*_{1e}/\gamma^*_{e1})(\gamma_{eg}/\gamma_{ge})$ or $(\gamma^*_{1e}/\gamma^*_{e1})(\gamma_{fg}\gamma_{ef})/(\gamma_{fe}\gamma_{gf})$. The equations in **S10** can then be obtained from these expressions by making the substitutions $\gamma_{xy} = \varphi_Y/(1-\varphi_X)$ (equivalent to Eqs. **S5**) for $x, y \in \{e, f, g\}$, where X, Y are the clades separated by and having x, y as root branches, respectively. If now we use Eq. **S10** in the expression above for Z_A, we obtain after obvious simplifications

$$Z_A = w_1 \sqrt{\frac{\gamma^*_{1e}}{\gamma^*_{e1}} \frac{1-\varphi_1}{\varphi_1} \frac{1-\varphi_B}{\varphi_B}}. \qquad \textbf{[S11]}$$

If now we use Eqs. **S7** and **S11** to express w_e and Z_A in $w_e Z_A$, what we obtain is precisely Eq. **S9**, which therefore is proven.

We are now ready to prove Eq. **S6**. Note that Eq. **S11** holds for any "composite" clade A (i.e., one that can be decomposed into two other clades A_1 and A_2). Then,

$$\begin{aligned} \frac{w_e Z_B}{w_e Z_B + w_f Z_{A_1} + w_g Z_{A_2}} &= \frac{w_e Z_B}{w_e Z_B + Z_A} = \frac{\sqrt{\frac{\varphi_B}{1-\varphi_B}}}{\sqrt{\frac{\varphi_B}{1-\varphi_B}} + \sqrt{\frac{1-\varphi_B}{\varphi_B}}} \\ &= \varphi_B, \end{aligned}$$

where for the second equality we have used both Eqs. **S11** and **S9**.

But this implies that, for every composite clade $A = A_1 \cup A_2$ in the configuration of Fig. 1*B*,

$$\frac{w_f Z_{A_1}}{Z_A} = \frac{w_f Z_{A_1}}{w_f Z_{A_1} + w_g Z_{A_2}} = \frac{\varphi_{A_1}}{\varphi_{A_1} + \varphi_{A_2}} = \gamma_{ef}. \qquad \textbf{[S12]}$$

Eq. **S4** of Lemma 1 is therefore verified. But this ensures that $\delta_{XY} = \delta^{\boldsymbol{w}}_{XY}$, for any two disjoint clades X and Y (Lemma 1, part i), while the fact that the $\boldsymbol{\lambda}^T$ parameters satisfy P2 implies (Lemma 1, part ii) that $\lambda_{A_1 B_1} = (Z_{A_1 B_2} + Z_{A_2 B_1})/Z_{AB}$ for every pair of 3-separated clades A_1, B_1 in the configuration of Fig. 1*B*. That is, the M&P formulae for $\boldsymbol{w}$ coincide with the given $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae, which concludes the proof of Theorem 2, part ii.

Appendix 3

Here, we prove that the main criteria to score trees, LS and ME (in all their common variants), are statistically consistent when used in combination with our branch length formulae, as stated in Theorem 3. We start by showing that the consistency of any distance-based principle is essentially determined by its behavior on perfect data (A3.1). Next, we move on to proving the consistency of LS (A3.2) and then that of ME: For the latter, first we show a useful dependency property between different variants of ME (A3.3), and then we prove the consistency on perfect data of the classic version of ME (A3.4), which is the key nontrivial result of this appendix and allows us to conclude the proof of Theorem 3 (A.3.5).

We recall that a *branch length estimation scheme* is a method that, for any binary topology over the set of taxa under consideration $\{1, 2, \ldots, n\}$ $(n \geq 3)$, determines how to fit the length of its branches on the basis of an $n \times n$ distance matrix $\boldsymbol{\delta}$. We say that a branch length estimation scheme is *continuous* [*linear*] if, for any branch e in any binary topology, the function $\hat{\ell}_e(\boldsymbol{\delta})$ giving its fitted length is continuous [linear] in $\boldsymbol{\delta}$. A branch length estimation scheme is *correct* if, for any branch e in any binary tree with branch lengths, $\hat{\ell}_e(\boldsymbol{\delta})$ returns the length of e whenever $\boldsymbol{\delta}$ is additive with respect to that tree (as in Theorem 1). The branch length estimation schemes that we consider here are those *based on* (γ, λ)*-formulae*, whereby a collection of γ^T and λ^T parameters is chosen for each binary topology T, thus determining a set of (γ^T, λ^T)-formulae for T's branch lengths (with no assumed relation between the values of these parameters across different topologies). It is clear that the resulting branch length estimation schemes are linear (thus continuous) and correct (Theorem 1).

Any branch length estimation scheme can be combined to a number of principles identifying an optimal tree among all the topologically distinct fitted trees. The optimization principles we consider here are defined by a *tree score function*, which can depend on the topology of the tree, the assigned branch lengths and (in the case of LS but not ME) the input distances. An optimization principle $\mathscr{M}$ then consists of seeking the fitted tree(s) that minimize this function, and we denote this tree (or set of trees) with $\mathscr{M}(\boldsymbol{\delta})$. We say that $\mathscr{M}$ is statistically consistent if $\mathscr{M}(\boldsymbol{\delta})$ converges (in probability) to the correct tree.

The following assumption (the consistency of the distance estimates and the positive additivity of the correct evolutionary distances) applies to all the propositions that follow, and we state it here so that we do not have to repeat it in every statement.

Assumption 2. *Let the correct phylogenetic tree for the taxa under consideration,* T^*, *be a binary tree with positive branch lengths. Assume that the input distances* $\boldsymbol{\delta}$ *converge (in probability) to* $\boldsymbol{d}^{T^*}$.

A3.1. Consistency for Perfect Data Implies Statistical Consistency. The following is a well-known sufficient condition for consistency, which has been proven for ME with the same continuity arguments (e.g., ref. 3). It can be applied to most tree optimization principles (LS, ME, possibly combinations of the two or even totally different criteria).

Proposition 3. *Adopt a continuous branch length estimation scheme. Let* $\mathscr{M}$ *be an optimization principle based on a tree score function that is continuous in all its continuous parameters (i.e., all but the topology). If* $\mathscr{M}(\boldsymbol{d}^{T^*})$ *is unique and coincides with* T^*, *then* $\mathscr{M}$ *is statistically consistent.*

Proof: To any binary topology, $\mathscr{M}$ assigns a score by first assigning branch lengths to it and then applying the adopted tree score function. Note that because both the branch length estimation scheme and the tree score function are continuous, then also the score associated to any particular topology is continuous in $\boldsymbol{\delta}$. Because $\mathscr{M}(\boldsymbol{d}^{T^*}) = T^*$ is unique, when $\boldsymbol{\delta} = \boldsymbol{d}^{T^*}$ the score of the topology of T^* must be strictly smaller than the score of all other binary topologies. But then, because the scores of topologies are continuous in $\boldsymbol{\delta}$ and finite in number, this must still hold for every $\boldsymbol{\delta}$ in a neighborhood of $\boldsymbol{d}^{T^*}$; that is, for every $\boldsymbol{\delta}$ in this neighborhood, $\mathscr{M}(\boldsymbol{\delta})$ is unique and has the same topology as T^*. But because $\boldsymbol{\delta} \xrightarrow{p} \boldsymbol{d}^{T^*}$, the probability that $\boldsymbol{\delta}$ belongs to this neighborhood, and consequently $\mathscr{M}(\boldsymbol{\delta})$ has the same topology as T^*, converges to 1. Finally, when $\mathscr{M}(\boldsymbol{\delta})$ has the same topology as T^*, the continuity of the branch length estimation scheme implies that the branch lengths in $\mathscr{M}(\boldsymbol{\delta})$ converge in probability to those in $\mathscr{M}(\boldsymbol{d}^{T^*}) = T^*$. We can then conclude that both the topology and branch lengths of $\mathscr{M}(\boldsymbol{\delta})$ consistently converge to T^*.

A3.2. Consistency of Least Squares. We have briefly defined LS methods in the Introduction. Here we assume the most general form for LS and prove its consistency under very general conditions (Proposition 4 below). We define LS methods as those that use a tree score function with the following form:

$$Q(T, \boldsymbol{\delta}) = (\boldsymbol{\delta} - \boldsymbol{d}^T)^t W_{T,\boldsymbol{\delta}}(\boldsymbol{\delta} - \boldsymbol{d}^T), \qquad \textbf{[S13]}$$

where T is the tree fitted using the assumed branch length estimation scheme, $\boldsymbol{\delta}$ is a column vector with the $\binom{n}{2}$ input distances and $W_{T,\boldsymbol{\delta}}$ is a $\binom{n}{2} \times \binom{n}{2}$ matrix which may depend on the topology of T and, continuously, on $\boldsymbol{\delta}$ and the branch lengths of T. Additionally, we assume that, for any T and $\boldsymbol{\delta}$, the matrix $W_{T,\boldsymbol{\delta}}$ is positive-definite. ($W_{T,\boldsymbol{\delta}}$ should be interpreted as the inverse of the assumed covariance matrix for $\boldsymbol{\delta}$.)

Note that whereas the dependence of $W_{T,\boldsymbol{\delta}}$ on $\boldsymbol{\delta}$ is common (e.g., the version of WLS by Fitch and Margoliash (4) uses a diagonal matrix with $W_{ij,ij} = \delta_{ij}^{-2}$), the dependence on T is nonstandard, and we have included it here for completeness. (But for example the balanced version of WLS (5) at the basis of the balanced branch lengths (6) does assume a variance model that depends on tree topology.) Also recall that the criterion $Q(T, \boldsymbol{\delta})$ above is used to score trees with already-fitted branch lengths, so the dependence on the branch lengths does not cause any computational problem.

Proposition 4. *Adopt any correct and continuous branch length estimation scheme. Then, LS is consistent.*

Proof: We prove that for LS, the hypotheses of Proposition 3 are satisfied, and therefore LS is consistent. First, the branch length estimation scheme is continuous (by hypothesis) and the score function in Eq. **S13** is a continuous function of both $\boldsymbol{\delta}$ and of the branch lengths assigned to T (note that $\boldsymbol{d}^T$ is linear, and thus continuous, in the branch lengths). It remains then to show that, for $\boldsymbol{\delta} = \boldsymbol{d}^{T^*}$, LS uniquely identifies T^* as optimal. Because $W_{T,\boldsymbol{\delta}}$ is positive definite, $Q(T, \boldsymbol{d}^{T^*}) = 0$ if and only if $\boldsymbol{d}^{T^*} - \boldsymbol{d}^T = \boldsymbol{0}$, that is, if and only if $T = T^*$, whereas for all other trees $W \neq T^*$, $Q(W, \boldsymbol{d}^{T^*}) > 0$. Moreover, because the branch length estimation scheme is correct, T^* is precisely what is obtained when fitting its branch lengths. Therefore T^* uniquely minimizes the score function Q and is returned by LS.

Corollary 5. *Adopt a branch length estimation scheme based on* (γ, λ)*-formulae. Then, LS is consistent.*

A3.3. Dependency Between the Consistencies of Different Variants of Minimum Evolution. Recall that ME_i reconstructs the fitted tree that minimizes the following tree score function, where $\hat{\ell} = (\hat{\ell}_e)$ denotes the branch lengths in the fitted tree:

$$L_i(\hat{\ell}) = \sum_{e\,:\,\hat{\ell}_e > 0} \hat{\ell}_e + \sum_{e\,:\,\hat{\ell}_e < 0} i \cdot \hat{\ell}_e.\Lambda \qquad \textbf{[S14]}$$

In the main text, we assume $i \in \{-1, 0, +1\}$, but here we consider, more generally, ME_x with x being any real number. We do this not only for the sake of mathematical completeness but also to include variants of ME that may be considered in the future (e.g., $\mathrm{ME}_{-\infty}$, which corresponds to avoiding at all costs trees which are assigned negative branch lengths). The following proposition shows that if we can prove the consistency of ME_y using Proposition 3, then the same can be done for any ME_x with $x < y$.

Lemma 6. *Adopt any correct branch length estimation scheme and let* $x < y$. *If* $\mathrm{ME}_y(\boldsymbol{d}^{T^*}) = T^*$, *then also* $\mathrm{ME}_x(\boldsymbol{d}^{T^*}) = T^*$.

Proof: We adapt a line of reasoning that has appeared elsewhere (3, 7). Because of the correctness of the branch length estimation scheme, when the branches of the topology of T^* are fitted using $\boldsymbol{d}^{T^*}$, their lengths are set to their correct values. Because these are all positive, the scores assigned by ME_x and ME_y to T^* equal precisely the sum L^* of all branch lengths in T^*. Now let $\hat{\ell}$ denote the branch lengths assigned to an incorrect topology using $\boldsymbol{d}^{T^*}$; $x < y$ implies that $L_x(\hat{\ell}) \geq L_y(\hat{\ell})$, and because $\mathrm{ME}_y(\boldsymbol{d}^{T^*})$ is unique and equal to T^*, we also have $L_y(\hat{\ell}) > L^*$. But then $L_x(\hat{\ell}) > L^*$ for any incorrect topology. Because L^* coincides with the score assigned to T^* by ME_x, we can then conclude that $\mathrm{ME}_x(\boldsymbol{d}^{T^*})$ is unique and coincides with T^*.

A3.4. Consistency of the Classic Version of Minimum Evolution. We now concentrate on the consistency of ME_{+1}, which because of Lemma 6 implies that of ME_x for any $x < +1$. We use a standard framework for investigating the consistency of ME_{+1} (e.g., ref. 1), which consists in verifying a property ("Willson's condition") of branch length estimation in the presence of a special kind of binary distance matrix. In the following, we introduce and state Willson's condition (A3.4.1), then prove some properties of the $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae that are useful to verify it (A3.4.2) and finally prove the consistency of ME_{+1} via Willson's condition (A3.4.3). For simplicity, we write $\hat{L}^T(\boldsymbol{\delta})$ as a shorthand for $L_{+1}(\hat{\ell}_T(\boldsymbol{\delta}))$, that is, the tree length (sensu ME_{+1}) resulting from fitting the branch lengths of T using $\boldsymbol{\delta}$.

A3.4.1. Willson's condition. We denote by $\boldsymbol{d}^{S|\bar{S}}$ (where $S \subseteq \{1, 2, \ldots, n\}$ and $\bar{S} = \{1, 2, \ldots, n\} \setminus S$) the following collection of $\binom{n}{2}$ distances, indexed by $i \neq j$:

$$d_{ij}^{S|\bar{S}} = \begin{cases} 1 & \text{if } |S \cap \{i, j\}| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Willson's condition for ME consistency can be stated as follows: (1)

Lemma 7. *Adopt any linear branch length estimation scheme such that, for any binary topology T and any proper and nonempty subset S of taxa from T,*

$$\hat{L}^T(\boldsymbol{d}^{S|\bar{S}}) \begin{cases} =1 & \text{if } S \text{ is a clade in } T, \\ >1 & \text{otherwise.} \end{cases}$$

Then $\mathrm{ME}_{+1}(\boldsymbol{d}^{T^*}) = T^*$.

Informally, this holds because $\boldsymbol{d}^{T^*}$ is a weighed sum of all $\boldsymbol{d}^{S|\bar{S}}$ corresponding to the clades in T^*, the coefficients of this weighted sum being the branch lengths of T^*. The linearity of $\hat{L}^T$ then implies that, in turn, $\hat{L}^T(\boldsymbol{d}^{T^*})$ is a weighted sum of the branch lengths of T^*, where the coefficients are now either 1 or strictly greater than 1, depending on whether or not the corresponding S is a clade in T. Clearly this weighted sum is minimized when all the clades in T^* are also clades in T, that is for $T = T^*$.

A3.4.2. Tools to verify Willson's condition. Given any subset of taxa $S \subseteq \{1, 2, \ldots, n\}$ and any clade X in a binary topology T for which a set of $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae is defined, define $p_{S|X}$ as the probability of picking an element of S from the random distribution over X defined by the $\boldsymbol{\gamma}^T$ parameters. That is,

$$p_{S|X} = \sum_{i \in X \cap S} p_{i|X}.$$

Moreover, a clade X is monochromatic (w.r.t. S) if either $X \subseteq S$ or $X \subseteq \bar{S} = \{1, 2, \ldots, n\} \setminus S$. In this case it is clear that $p_{S|X} \in \{0, 1\}$.

Lemma 8. *Assume that the branch lengths of T are assigned with $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae using the input distances in $\boldsymbol{d}^{S|\bar{S}}$, for some $S \subseteq \{1, 2, \ldots, n\}$. Then,*

i. *For any two clades X and Y in T, the average distance between them $d_{XY}^{S|\bar{S}} = p_{S|X}p_{\bar{S}|Y} + p_{\bar{S}|X}p_{S|Y}$.*
ii. *If branch e belongs to a monochromatic clade, it is assigned length $\hat{\ell}_e = 0$.*
iii. *If adjacent branches f and g separate (and are the only ones to separate) two monochromatic clades A_1 and A_2, with $A_1 \subseteq S$ and $A_2 \subseteq \bar{S}$, then $\hat{\ell}_f + \hat{\ell}_g = 1$.*

Proof:

i. $d_{XY}^{S|\bar{S}} = \sum_{\substack{i \in X \\ j \in Y}} p_{i|X}p_{j|Y}d_{ij}^{S|\bar{S}} = \sum_{\substack{i \in X \cap S \\ j \in Y \cap \bar{S}}} p_{i|X}p_{j|Y} + \sum_{\substack{i \in X \cap \bar{S} \\ j \in Y \cap S}} p_{i|X}p_{j|Y} = p_{S|X}p_{\bar{S}|Y} + p_{\bar{S}|X}p_{S|Y}$.
ii. Branch e is either internal or external. We consider here only the case where it is internal, as the external case is analogous (and simpler). We assume clades A_1, A_2, B_1, and B_2 are defined as in Fig. 1*B*. Because e belongs to a monochromatic clade, at least three clades out of A_1, A_2, B_1, and B_2 are all subsets of S or all subsets of $\bar{S}$. Without loss of generality, we assume that A_1, A_2, B_1 are all subsets of S. Applying part i, it is easy to see that this implies $d_{A_1A_2}^{S|\bar{S}} = d_{A_1B_1}^{S|\bar{S}} = d_{A_2B_1}^{S|\bar{S}} = 0$ and $d_{B_1B_2}^{S|\bar{S}} = d_{A_2B_2}^{S|\bar{S}} = d_{A_1B_2}^{S|\bar{S}} = p_{\bar{S}|B_2}$. But then,

$$\begin{aligned}\hat{\ell}_e &= \frac{1}{2}[\lambda_{A_1B_1}(d_{A_1B_1}^{S|\bar{S}} + d_{A_2B_2}^{S|\bar{S}}) + (1 - \lambda_{A_1B_1})(d_{A_1B_2}^{S|\bar{S}} + d_{A_2B_1}^{S|\bar{S}}) \\ &\quad - d_{A_1A_2}^{S|\bar{S}} - d_{B_1B_2}^{S|\bar{S}}] \\ &= \frac{1}{2}[\lambda_{A_1B_1}p_{\bar{S}|B_2} + (1 - \lambda_{A_1B_1})p_{\bar{S}|B_2} - p_{\bar{S}|B_2}] = 0.\end{aligned}$$

iii. Let $B = \{1, 2, \ldots, n\} \setminus (A_1 \cup A_2)$, as in Fig. 1*B*. In a way analogous to part ii, it is easy to check that, independently of f being internal or external, $\hat{\ell}_f = 1/2(1 + p_{\bar{S}|B} - p_{S|B})$ and, similarly, $\hat{\ell}_g = 1/2(1 + p_{S|B} - p_{\bar{S}|B})$. Therefore, $\hat{\ell}_f + \hat{\ell}_g = 1$.

Points i, ii, and iii are thus all proved.

A3.4.3. Consistency of minimum evolution with perfect data.

Proposition 9. *Adopt a branch length estimation scheme based on $(\boldsymbol{\gamma}, \boldsymbol{\lambda})$-formulae. Then,* $\mathrm{ME}_{+1}(\boldsymbol{d}^{T^*}) = T^*$.

Proof: We show that any branch length estimation scheme based on $(\boldsymbol{\gamma}, \boldsymbol{\lambda})$-formulae satisfies Willson's condition (i.e. the hypotheses of Lemma 7), and therefore we must have $\mathrm{ME}_{+1}(\boldsymbol{d}^{T^*}) = T^*$.

First, it is trivial to see that any such branch length estimation scheme is linear. Second, it is correct (Theorem 1), which implies $\hat{L}^T(\boldsymbol{d}^{S|\bar{S}}) = 1$ whenever S is a clade of T: In this case, in fact, $\boldsymbol{d}^{S|\bar{S}}$ is additive with respect to a tree with topology T and with all branches of length 0 except the root branch of S, which has length 1; because of their correctness, the $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae result in assigning T precisely these branch lengths and therefore a total length $\hat{L}^T(\boldsymbol{d}^{S|\bar{S}}) = 1$.

It remains to prove that $\hat{L}^T(\boldsymbol{d}^{S|\bar{S}}) > 1$ whenever S is not a clade of T—for any branch length estimation scheme based on $(\boldsymbol{\gamma}, \boldsymbol{\lambda})$-formulae. We do this by induction on the size of T.

For $n = 3$ taxa, this is trivially true, as all proper, nonempty sets of $\{1, 2, 3\}$ are clades of T.

For $n > 3$, if S is not a clade of T, it is always possible to find a pair of 2-separated clades A_1, A_2 such that $A_1 \subsetneq S$ and $A_2 \subsetneq \bar{S}$—i.e., such that A_1 and A_2 are monochromatic—but $A_1 \cup A_2$ and

$B = \{1, 2, \ldots, n\} \setminus (A_1 \cup A_2)$ are not monochromatic. To see this, consider the tree $T^{(S)}$ that is obtained by substituting every monochromatic clade in T with a taxon; because S is not a clade of T, then $T^{(S)}$ must have at least two cherries (i.e., pairs of 2-separated taxa); any of these corresponds to a pair of clades A_1, A_2 in the original tree T with the required properties. Let e, f, and g be the root branches of B, A_1 , and A_2, respectively, and a their common endpoint, as in Fig. 1*B*. Because A_1 and A_2 are monochromatic, it is clear that all branches belonging to these clades are assigned length 0 when the input distances are $\boldsymbol{d} = \boldsymbol{d}^{S|\bar{S}}$ (Lemma 8, part ii).

Now let T' be the topology that is obtained by deleting from T all branches belonging to $A = A_1 \cup A_2$, so that a is a leaf of T'. It is clear that there is one-to-one correspondence between the branches/clades of T' and a subset of the branches/clades of T. When calculating the branch lengths in T', we assume that the γ_{ef} and λ_{XY} parameters are the same as those for the corresponding branches/clades in T. Now define the following distances over $\{1, 2, \ldots, n\} \cup \{a\} \setminus A$, i.e., the taxa in T':

$$\boldsymbol{d}^{(1)} = \boldsymbol{d}^{(S \cup \{a\} \setminus A_1)|(\bar{S} \setminus A_2)},$$

$$\boldsymbol{d}^{(2)} = \boldsymbol{d}^{(S \setminus A_1)|(\bar{S} \cup \{a\} \setminus A_2)},$$

$$\boldsymbol{d}' = \gamma_{ef}\boldsymbol{d}^{(1)} + \gamma_{eg}\boldsymbol{d}^{(2)}.$$

Note that $\boldsymbol{d}'$ coincides with $\boldsymbol{d}$ except for the distances that involve a. For these, we have

$$d'_{aj} = \gamma_{ef} d^{(1)}_{aj} + \gamma_{eg} d^{(2)}_{aj},$$

for every $j \in \{1, 2, \ldots, n\} \setminus A$. Note also that $d^{(1)}_{aj} = d_{A_1 j}$ and $d^{(2)}_{aj} = d_{A_2 j}$, which imply $d'_{aj} = d_{Aj}$. This in turn implies that, for any disjoint clades X and Y in T',

$$d'_{XY} = \begin{cases} d_{XY} & \text{if } a \notin X \cup Y, \\ d_{X'Y} & \text{if } a \in X \text{ and where } X' = X \cup A \setminus \{a\}, \end{cases} \qquad \textbf{[S15]}$$

that is, the average distances between disjoint clades remain the same when going from $(T, \boldsymbol{d})$ to $(T', \boldsymbol{d}')$. To see this, note that the first case is a simple consequence of the fact that $\boldsymbol{d}'$ coincides with $\boldsymbol{d}$ for the distances that do not involve a. As for the second case, it is a consequence of combining $d'_{aj} = d_{Aj}$ (shown above) with the first case (and it can be easily proved by induction on the size of X).

The important consequence of Eq. **S15** is that the lengths of branches belonging to B—which only depend on average distances between disjoint clades in T'—remain constant when going from $(T, \boldsymbol{d})$ to $(T', \boldsymbol{d}')$. Therefore, the only difference between the two tree lengths will come from the lengths of branches e, f and g:

$$\hat{L}^{T}(\boldsymbol{d}) = \hat{L}^{T'}(\boldsymbol{d}') + \hat{\ell}_f + \hat{\ell}_g + \hat{\ell}_e - \hat{\ell}'_e, \qquad \textbf{[S16]}$$

where $\hat{\ell}_b$ and $\hat{\ell}'_b$ represent the lengths assigned to branch b for $(T, \boldsymbol{d})$ and $(T', \boldsymbol{d}')$, respectively.

Because for ME_{+1} the tree length is a linear function of the branch lengths and the branch lengths themselves are linear functions of the input distances, $\hat{L}^{T'}(\boldsymbol{d}')$ is then linear in $\boldsymbol{d}'$ and we can write

$$\hat{L}^{T'}(\boldsymbol{d}') = \gamma_{ef}\hat{L}^{T'}(\boldsymbol{d}^{(1)}) + \gamma_{eg}\hat{L}^{T'}(\boldsymbol{d}^{(2)}).$$

Because $\boldsymbol{d}^{(1)}$ and $\boldsymbol{d}^{(2)}$ are equal to some $\boldsymbol{d}^{S'|\bar{S}'}$ where S' is a proper and nonempty subset of the $n'(< n)$ taxa in T', we can apply the induction hypothesis and infer that $\hat{L}^{T'}(\boldsymbol{d}^{(1)}) \geq 1$ and $\hat{L}^{T'}(\boldsymbol{d}^{(2)}) \geq 1$ (where equality is achieved when S' is a clade of T'). Therefore $\hat{L}^{T'}(\boldsymbol{d}') \geq 1$.

In order to prove that the tree length in Eq. **S16** is strictly greater than 1, we then just need to prove that $\hat{\ell}_f + \hat{\ell}_g + \hat{\ell}_e - \hat{\ell}'_e > 0$.

First, because of Lemma 8, part iii, $\hat{\ell}_f + \hat{\ell}_g = 1$. In order to calculate $\hat{\ell}_e$ and $\hat{\ell}'_e$, it is useful to note that $d_{A_1A_2} = 1$ and, for any clade $B' \subseteq B$, $d_{A_1B'} = p_{\bar{S}|B'} = 1 - p_{S|B'}$ and $d_{A_2B'} = p_{S|B'}$ (Lemma 8, part i). Then,

$$\begin{aligned} \hat{\ell}_e &= \frac{1}{2}[\lambda_{A_1B_1}(1 - p_{S|B_1} + p_{S|B_2}) + (1 - \lambda_{A_1B_1})(1 - p_{S|B_2} + p_{S|B_1}) \\ &\quad - 1 - d_{B_1B_2}] \\ &= \frac{1}{2}[(1 - 2\lambda_{A_1B_1})(p_{S|B_1} - p_{S|B_2}) - d_{B_1B_2}]. \end{aligned}$$

Similarly,

$$\begin{aligned} \hat{\ell}'_e &= \frac{1}{2}[d'_{aB_1} + d'_{aB_2} - d'_{B_1B_2}] \\ &= \frac{1}{2}[d_{AB_1} + d_{AB_2} - d_{B_1B_2}] \\ &= \frac{1}{2}[\gamma_{ef}(d_{A_1B_1} + d_{A_1B_2}) + (1 - \gamma_{ef})(d_{A_2B_1} + d_{A_2B_2}) - d_{B_1B_2}] \\ &= \frac{1}{2}[\gamma_{ef}(2 - p_{S|B_1} - p_{S|B_2}) + (1 - \gamma_{ef})(p_{S|B_!} + p_{S|B_2}) - d_{B_1B_2}] \\ &= \gamma_{ef} + \frac{1}{2}[(1 - 2\gamma_{ef})(p_{S|B_!} + p_{S|B_2}) - d_{B_1B_2}]. \end{aligned}$$

Then,

$$\begin{aligned} \hat{\ell}_f + \hat{\ell}_g + \hat{\ell}_e - \hat{\ell}'_e = 1 - \gamma_{ef} &+ p_{S|B_1}(\gamma_{ef} - \lambda_{A_1B_1}) \\ &+ p_{S|B_2}(\gamma_{ef} - (1 - \lambda_{A_1B_1})). \end{aligned}$$

But this is a linear function of $(p_{S|B_1}, p_{S|B_2})$ in the square $[0, 1]^2$, and is thus minimized in one of its four vertices. In (0, 0), (0, 1), (1, 0), and (1, 1), the function has values $1 - \gamma_{ef}$, $\lambda_{A_1B_1}$, $1 - \lambda_{A_1B_1}$, γ_{ef}, respectively. Because these are all strictly greater than 0 by hypothesis, then so is $\hat{\ell}_f + \hat{\ell}_g + \hat{\ell}_e - \hat{\ell}'_e$ and therefore $\hat{L}^{T}(\boldsymbol{d}) > 1$, which completes the proof of Proposition 9.

A3.5. Wrapping It All Together. By applying Proposition 9, Lemma 6, and Proposition 3, we then conclude:

Corollary 10. *Adopt a branch length estimation scheme based on (γ, λ)-formulae. Then, for any $x \leq +1$, ME_x is consistent.*

Which, together with Corollary 5 and Assumption 2, completes our proof of Theorem 3.

Appendix 4

Here, we prove the efficiency of calculating branch lengths with our formulae in hill climbing heuristics, as stated in Theorem 4. We start by showing that efficient branch length calculations essentially depend on the availability of the average distances between (some) clades in the current tree and that these can be calculated in quadratic time, which allows us to prove Theorem 4, part i (A4.1). When performing an NNI, calculating the new branch lengths can be done efficiently by recalculating only some of these average distances, which leads us to prove Theorem 4, part ii (A4.2). Finally, we show that updating the average distances following an NNI can also be done efficiently (A4.3), which is not a claim of Theorem 4 but is nevertheless potentially useful. The results and proofs here are inspired by those of Desper and Gascuel (8). However, their results were specific to the balanced and OLS branch lengths in combination with the ME_{+1} optimization principle. A key property of these estimators is that, when performing an NNI, the sum of the branch lengths in each of the four corner subtrees around the location of the NNI remains constant. Thanks to this property, the difference between the ME_{+1} lengths of any two NNI neighbors T and T' can be efficiently calculated using simple formulae. This property does not hold in general for $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-estimators, and so we have to recalculate all branch lengths every time we perform an NNI. The good news is that (*i*) the complexity of each iteration in a hill climbing heuristics for ME (computing the length of all NNI neighbors of a given topology and updating the data structures for the new best topology), which for BME was quadratic in the worst case, remains quadratic in the size of the tree, and that (*ii*) recalculating all branch lengths makes it possible to use optimization principles such as ME_0, ME_{-1} , and $\mathrm{ME}_{-\infty}$.

In the following, T, T' , and T_i always denote binary topologies, and $\boldsymbol{\gamma}^T, \boldsymbol{\gamma}^{T'}$, and $\boldsymbol{\gamma}^{T_i}$, collections of γ_{ef} parameters defined for them, in the way described in the main text.

A4.1. Computing the Branch Lengths of Fixed Topology.

Lemma 11. *Adopt a set of $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae for the branch lengths of T. Given $\boldsymbol{\delta}$ and δ_{XY} for every pair of 3-separated clades X, Y in T, the length of any branch in T can be calculated in $O(1)$ time.*

Proof: The $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae are simple linear combinations of average distances δ_{XY} between 2- and 3-separated clades X and Y and can be computed in $O(1)$ once these average distances are available. Because we assume that the average distances between 3-separated clades are given, it remains to show that δ_{XY} between any pair X, Y of 2-separated clades can be obtained in $O(1)$. But this is trivial: Either both X and Y consist of one taxon only (i.e., $X = \{i\}$ and $Y = \{j\}$) in which case $\delta_{XY} = \delta_{ij}$, or at least one of the two clades, say X, is such that $X = X_1 \cup X_2$, where both X_1 and X_2 are clades, in which case $\delta_{XY} = \gamma_{ef}\delta_{X_1Y} + \gamma_{eg}\delta_{X_2Y}$, where e, f, and g are the root branches of X, X_1, and X_2, respectively, and both δ_{X_1Y} and δ_{X_2Y} are known, as X_i and Y are 3-separated (for $i \in 1, 2\}$). Lemma 11 is therefore proved.

Although the one above is a straightforward observation, it determines the minimum amount of information necessary to determine any branch length in T in constant time. Motivated by it, we define $\Delta^T(\boldsymbol{\gamma}^T)$ as a data structure holding all the average distances δ_{XY} between pairs of disjoint clades in T and make explicit its dependence on $\boldsymbol{\gamma}^T$. Note that $\Delta^T(\boldsymbol{\gamma}^T)$ specifies the average distances between 3-separated clades as a particular case.

Lemma 12. *Given $\boldsymbol{\delta}$, T and $\boldsymbol{\gamma}^T$, the calculation of $\Delta^T(\boldsymbol{\gamma}^T)$ requires $O(n^2)$ time.*

Proof: Consider any total ordering $A_1, A_2, \ldots, A_{2(2n-3)}$ of the clades in T, such that if $A_k = A_i \cup A_j$ then $i < k$ and $j < k$. Finding one such ordering is trivial and can be done in a number of ways, for example by sorting the clades in ascending order of depth (**2**), or by rooting the tree in one of its leaves and then performing a postorder traversal, listing the clades oriented away from the root, followed by a preorder traversal, listing the clades oriented toward the root. The following procedure then calculates δ_{XY} for all pairs of clades (including non-disjoint ones):

For $i = 2, \ldots, 2(2n-3)$,
for $j = 1, \ldots, i-1$,

$$\delta_{A_iA_j} = \begin{cases} \delta_{xy} & \\ \qquad \text{if } A_i = \{x\} \quad \text{and} \quad A_j = \{y\}, & \\ \gamma_{ef}\delta_{A_{i_1}A_j} + \gamma_{eg}\delta_{A_{i_2}A_j} & \\ \qquad \text{if } A_i = A_{i_1} \cup A_{i_2} \text{ for some clades } A_{i_1}, A_{i_2}, & \\ \gamma_{ef}\delta_{A_iA_{j_1}} + \gamma_{eg}\delta_{A_iA_{j_2}} & \\ \qquad \text{if } A_j = A_{j_1} \cup A_{j_2} \text{ for some clades } A_{j_1}, A_{j_2}. & \end{cases} \quad \textbf{[S17]}$$

In the second case of Eq. **S17** we assume that e, f, g are the root branches of A_i, A_{i_1}, A_{i_2}, respectively, whereas in the third case they are the root branches of A_j, A_{j_1}, A_{j_2}, respectively. Note that these two cases are not mutually exclusive, and the result is the same independently of which case is applied. Moreover, because of the way the ordering is defined, we must have $i_1, i_2 < i$, in the second case, or $j_1, j_2 < j$, in the third case, which means that $\delta_{A_{i_1}A_j}$ and $\delta_{A_{i_2}A_j}$ (second case), or $\delta_{A_iA_{j_1}}$ and $\delta_{A_iA_{j_2}}$ (third case) have already been calculated and are available when we calculate $\delta_{A_iA_j}$. Because each $\delta_{A_iA_j}$ can be calculated in constant time, the whole calculation requires $O(n^2)$ time. Lemma 12 is thus proved.

The complexity we obtain in Lemma 12 is optimal. Even if we restrict the calculation to 3-separated clades, we still cannot do better than $O(n^2)$, as the average distances between such pairs of clades still depend on $O(n^2)$ input distances.

Proof of Theorem 4, part i: Combining Lemma 12 with Lemma 11 yields that the branch lengths determined by a set of $(\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)$-formulae for a binary topology T can be calculated in $O(n^2)$ time.

A4.2. Computing the Branch Lengths of the NNI Neighbors of a Given Topology.

Lemma 13. *Let T and T' be NNI neighbors and let $\boldsymbol{\gamma}^T$ and $\boldsymbol{\gamma}^{T'}$ be almost identical. Let δ_{XY} and δ'_{XY} denote the average clade distances in $\Delta^T(\boldsymbol{\gamma}^T)$ and $\Delta^{T'}(\boldsymbol{\gamma}^{T'})$, respectively. Then, given $\Delta^T(\boldsymbol{\gamma}^T)$, the calculation of δ'_{XY} for every pair of 3-separated clades X, Y in T', requires $O(n)$ time.*

Proof: We assume that T is as in Fig. 1*B* and T' as in Fig. S2. Let the elements of $\boldsymbol{\gamma}^T$ and $\boldsymbol{\gamma}^{T'}$ be denoted by $\gamma_{e_1e_2}$ and $\gamma'_{e_1e_2}$, respectively.

First, we show that δ'_{XY} is straightforward to obtain in the case of pairs of 3-separated clades in T' such that none or 1 of the 3 branches separating X and Y belongs to one of the corner clades A_1, A_2, B_1, B_2. Let $A_1 = A'_1 \cup A''_1$, $A_2 = A'_2 \cup A''_2$, $B_1 = B'_1 \cup B''_1$ and $B_2 = B'_2 \cup B''_2$, where all the sets involved are also clades of T' (and therefore T). It is trivial to verify that

$$\begin{array}{lll} \delta'_{A_1A_2} = \delta_{A_1A_2}, & \delta'_{B_1B_2} = \delta_{B_1B_2}, & \delta'_{A_1B_1} = \delta_{A_1B_1}, \\ \delta'_{A_2B_2} = \delta_{A_2B_2}, & \delta'_{A'_1B_2} = \delta_{A'_1B_2}, & \delta'_{A''_1B_2} = \delta_{A''_1B_2}, \\ \delta'_{A_1B'_2} = \delta_{A_1B'_2}, & \delta'_{A_1B''_2} = \delta_{A'_1B''_2}, & \delta'_{A'_2B_1} = \delta_{A'_2B_1}, \\ \delta'_{A''_2B_1} = \delta_{A''_2B_1}, & \delta'_{A_2B'_1} = \delta_{A_2B'_1}, & \delta'_{A_2B''_1} = \delta_{A_2B''_1}, \end{array} \quad \textbf{[S18]}$$

as $\boldsymbol{\gamma}^T$ and $\boldsymbol{\gamma}^{T'}$ are the same within all the clades in the subscripts above. Now observe that $A_1 \cup B_2$ and $A_2 \cup B_1$ are clades in T'

but not in T. Their average distances with other 3-separated clades must then be obtained with expressions such as

$$\delta'_{A'_1 A_2 \cup B_1} = \gamma'_{e'g}\delta_{A'_1 A_2} + \gamma'_{e'h}\delta_{A'_1 B_1}. \qquad \textbf{[S19]}$$

(Similar formulae are easy to obtain for $\delta'_{A''_1 A_2 \cup B_1}$, $\delta'_{B'_2 A_2 \cup B_1}$, $\delta'_{B''_2 A_2 \cup B_1}$ and $\delta'_{A'_2 A_1 \cup B_2}$, $\delta'_{A''_2 A_1 \cup B_2}$, $\delta'_{B'_1 A_1 \cup B_2}$, $\delta'_{B''_1 A_1 \cup B_2}$.) We have therefore proved that these δ'_{XY} can be obtained from one or two corresponding entries in $\Delta^T(\boldsymbol{\gamma}^T)$ in $O(1)$ time.

We still have to show how to derive δ'_{XY} when two or all three of the branches separating X and Y belong to a corner clade. Without loss of generality, we assume this clade to be A_1. If both $X, Y \subset A_1$, then we trivially have $\delta'_{XY} = \delta_{XY}$. Assume then $Y \supseteq B_1 \cup B_2 \cup A_2$. Let clades $Y_1, Y_2, \ldots, Y_k$ be defined as in Fig. S2 (and note that if $Y = B_1 \cup B_2 \cup A_2$, no such Y_i clade is defined). Also, if clade Y contains clade Y' in T, define $p_{Y'|Y}$ as the probability that the random walk defined by the $\boldsymbol{\gamma}^T$ parameters reaches Y', assuming that it enters Y from its root branch: $p_{Y'|Y} = \gamma_{e_0 e_1} \cdot \gamma_{e_1 e_2} \cdot \ldots \cdot \gamma_{e_{k-1} e_k}$, where e_0 is the root branch of Y and $e_1, e_2, \ldots, e_k$ are the branches on the path between the roots of Y and Y'. Define $p'_{Y'|Y}$ similarly for T' and $\boldsymbol{\gamma}^{T'}$. Then,

$$\begin{aligned}
\delta'_{XY} &= p'_{Y_1|Y}\delta'_{XY_1} + \ldots + p'_{Y_k|Y}\delta'_{XY_k} + p'_{B_1|Y}\delta'_{XB_1} \\
&\quad + p'_{B_2|Y}\delta'_{XB_2} + p'_{A_2|Y}\delta'_{XA_2} \\
&= p'_{Y_1|Y}\delta'_{XY_1} + \ldots + p'_{Y_k|Y}\delta'_{XY_k} \\
&\quad + p'_{B_1 \cup B_2 \cup A_2|Y}(\gamma'_{fe'}\gamma'_{e'h}\delta'_{XB_1} + \gamma'_{fl}\delta'_{XB_2} + \gamma'_{fe'}\gamma'_{e'g}\delta'_{XA_2}) \\
&= p_{Y_1|Y}\delta_{XY_1} + \ldots + p_{Y_k|Y}\delta_{XY_k} \\
&\quad + p_{B_1 \cup B_2 \cup A_2|Y}(\gamma'_{fe'}\gamma'_{e'h}\delta_{XB_1} + \gamma'_{fl}\delta_{XB_2} + \gamma'_{fe'}\gamma'_{e'g}\delta_{XA_2}),
\end{aligned}$$

where the last equality uses the almost identity of $\boldsymbol{\gamma}^T$ and $\boldsymbol{\gamma}^{T'}$. Similarly,

$$\begin{aligned}
\delta_{XY} &= p_{Y_1|Y}\delta_{XY_1} + \ldots + p_{Y_k|Y}\delta_{XY_k} \\
&\quad + p_{B_1 \cup B_2 \cup A_2|Y}(\gamma_{fe}\gamma_{eh}\delta_{XB_1} + \gamma_{fe}\gamma_{el}\delta_{XB_2} + \gamma_{fg}\delta_{XA_2}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\delta'_{XY} - \delta_{XY} &= p_{B_1 \cup B_2 \cup A_2|Y}[(\gamma'_{fe'}\gamma'_{e'h} - \gamma_{fe}\gamma_{eh})\delta_{XB_1} \\
&\quad + (\gamma'_{fl} - \gamma_{fe}\gamma_{el})\delta_{XB_2} + (\gamma'_{fe'}\gamma'_{e'g} - \gamma_{fg})\delta_{XA_2}]. \qquad \textbf{[S20]}
\end{aligned}$$

It is easy to derive similar equations for the cases where (a) $X \subset A_2$, $Y \supseteq B_1 \cup B_2 \cup A_1$, (b) $X \subset B_1$, $Y \supseteq A_1 \cup A_2 \cup B_2$, (c) $X \subset B_2$, $Y \supseteq A_1 \cup A_2 \cup B_1$, which allow us to derive δ'_{XY} in $O(1)$ time from four entries in $\Delta^T(\boldsymbol{\gamma}^T)$ (including δ_{XY}) and $p_{B_1 \cup B_2 \cup A_1|Y}, p_{A_1 \cup A_2 \cup B_2|Y}, p_{A_1 \cup A_2 \cup B_1|Y}$ for cases (a), (b), (c), respectively. Now consider the following procedure:

1. For every clade $Y \supseteq B_1 \cup B_2 \cup A_2$, calculate $p_{B_1 \cup B_2 \cup A_2|Y}$.
2. Do the same as above, for every clade $Y \supseteq B_1 \cup B_2 \cup A_1$, for every $Y \supseteq A_1 \cup A_2 \cup B_2$ and for every $Y \supseteq A_1 \cup A_2 \cup B_1$.
3. Use Eq. **S20**, or similar equation, to derive δ'_{XY} for all 3-separated clades X, Y in T' such that two or all three of the branches separating X and Y belong to a corner clade A_1, A_2, B_1, B_2.
4. Use the simple equations in **S18** and **S19** to calculate δ'_{XY} for the remaining 3-separated clades.

Step 1 can be done in $O(n)$ time, by starting with the smallest clades including $B_1 \cup B_2 \cup A_2$ and using the derived values to calculate those for the larger clades. The same holds for step 2. Then, each δ'_{XY} can be calculated in $O(1)$ time. Because there are $O(n)$ 3-separated pairs of clades, the entire algorithm runs in $O(n)$ time and thus Lemma 13 is proved.

We are now ready to complete the proof of Theorem 4.

Proof of Theorem 4, part ii: Recall that all the branch lengths in T_0 and its NNI neighbors $T_1, T_2, \ldots, T_{2(n-3)}$ are defined by $(\boldsymbol{\gamma}^{T_i}, \boldsymbol{\lambda}^{T_i})$-formulae, with the constraint that $\boldsymbol{\gamma}^{T_i}$ and $\boldsymbol{\gamma}^{T_0}$ are almost identical. We wish to prove that the branch lengths of $T_1, T_2, \ldots, T_{2(n-3)}$ can be calculated in $O(n^2)$ time. Let $\delta^{(i)}_{XY}$ denote the average clade distances in $\Delta^{T_i}(\boldsymbol{\gamma}^{T_i})$. Because of Lemma 12, $\Delta^{T_0}(\boldsymbol{\gamma}^{T_0})$ can be calculated in $O(n^2)$ time. From this, the calculation of $\delta^{(i)}_{XY}$ for every pair of 3-separated clades in T_i, requires $O(n)$ time (Lemma 13). Combining this to Lemma 11 yields that all $O(n)$ branch lengths in T_i can be calculated in $O(n)$ time. Because there are $O(n)$ neighbors of T and each is treated in $O(n)$ time, the whole calculation requires $O(n^2)$ time. The proof of Theorem 4 is thus complete.

A4.3. Updating the Accessory Information When Performing an NNI. The proof of Lemma 13 above suggests a related result that may also be useful for hill climbing, when the best NNI neighbor T' of T has been identified and we need to calculate $\Delta^{T'}(\boldsymbol{\gamma}^{T'})$ in order to explore efficiently the NNI neighborhood of T'. Define $\mathrm{diam}(T)$, the diameter of T, as the maximum number of branches separating any two leaves of T.

Proposition 14. *Let T and T' be NNI neighbors and let $\boldsymbol{\gamma}^T$ and $\boldsymbol{\gamma}^{T'}$ be almost identical. Given $\Delta^T(\boldsymbol{\gamma}^T)$, its update into $\Delta^{T'}(\boldsymbol{\gamma}^{T'})$ requires $O(n \cdot \mathrm{diam}(T))$ time.*

Proof: Let T be as in Fig. 1*B* and T' as in Fig. S2. Let δ_{XY} and δ'_{XY} denote average clade distances from $\Delta^T(\boldsymbol{\gamma}^T)$ and $\Delta^{T'}(\boldsymbol{\gamma}^{T'})$, respectively. In order to obtain $\Delta^{T'}(\boldsymbol{\gamma}^{T'})$ from $\Delta^T(\boldsymbol{\gamma}^T)$, one needs to calculate the entries of $\Delta^{T'}(\boldsymbol{\gamma}^{T'})$ that have no corresponding entry in $\Delta^T(\boldsymbol{\gamma}^T)$ or those that have changed. These are the δ'_{XY} for all pairs of clades X, Y in T' such that some of the branches f, g, h, l belong to X or Y. The only case where both X and Y have at least one of f, g, h, l belonging to them is that where $X = A_1 \cup B_2$ and $Y = A_2 \cup B_1$. In this case, δ'_{XY} can be obtained from $\Delta^T(\boldsymbol{\gamma}^T)$ with $\delta'_{XY} = \gamma'_{e'f}\gamma'_{e'g}\delta_{A_1 A_2} + \gamma'_{e'f}\gamma'_{e'h}\delta_{A_1 B_1} + \gamma'_{e'g}\gamma'_{e'l}\delta_{A_2 B_2} + \gamma'_{e'h}\gamma'_{e'l}\delta_{B_1 B_2}$.

All the other cases correspond to a pair of clades X, Y such that one of them, say X, is included in one of the four corner clades A_1, A_2, B_1, B_2 and the other, Y, includes two or three of the other clades (see, e.g., Fig. S2, where $X \subset A_1$ and $Y \supseteq B_1 \cup B_2 \cup A_2$). It is clear that for any such X, the number of possible choices for Y equals the number of branches in the path starting with e' and ending in the root of X. In other words, there are $O(n)$ possible choices for X, each of which corresponds to at most $\mathrm{diam}(T)$ choices for Y. Therefore we need to consider $O(n\mathrm{diam}(T))$ pairs of clades. For each of these pairs, we now prove that δ'_{XY} can be calculated in $O(1)$ time from $\Delta^T(\boldsymbol{\gamma}^T)$, once steps 1 and 2 from the proof of Proposition 13 have been executed (in $O(n)$ time): If $Y = A_2 \cup B_1$ or $Y = A_1 \cup B_2$, then it is straightforward to obtain δ'_{XY} as $\gamma'_{e'g}\delta_{XA_2} + \gamma'_{e'h}\delta_{XB_1}$ or as $\gamma'_{e'f}\delta_{XA_1} + \gamma'_{e'l}\delta_{XB_2}$, respectively; otherwise, if Y includes three of A_1, A_2, B_1, B_2, it is easy to see that Eq. **S20**, and similar equations for $X \subset A_2, B_1, B_2$, still hold (without the assumption, made in the proof of Lemma 13, that X and Y are 3-separated). It is then possible to calculate each δ'_{XY} in constant time.

1. Willson S (2005) Consistent formulas for estimating the total lengths of trees. *Discrete Appl Math* 148:214–239.
2. Mihaescu R, Pachter L (2008) Combinatorics of least squares trees. *Proc Natl Acad Sci USA* 105:13206–13211.
3. Denis F, Gascuel O (2003) On the consistency of the minimum evolution principle of phylogenetic inference. *Discrete Appl Math* 127:63–77.
4. Fitch W, Margoliash E (1967) Construction of phylogenetic trees. *Science* 155:279–284.
5. Desper R, Gascuel O (2004) Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol Biol Evol* 21:587–598.
6. Pauplin Y (2000) Direct calculation of a tree length using a distance matrix. *J Mol Evol* 51:41–47.
7. Gascuel O, Bryant D, Denis F (2001) Strengths and limitations of the minimum evolution principle. *Syst Biol* 50:621–627.
8. Desper R, Gascuel O (2002) Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comp Biol* 9:687–705.

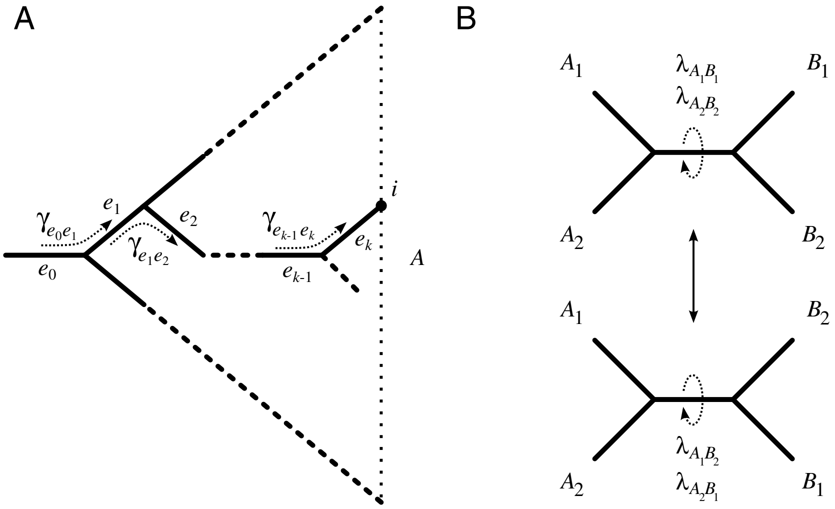

Fig. S1. (*A*) $p_{i|A} = \gamma_{e_0e_1} \cdot \gamma_{e_1e_2} \cdot \ldots \cdot \gamma_{e_{k-1}e_k}$ is the probability of ending up in *i* when entering clade *A* from its root and following the random walk rules described in the main text; (*B*) $\lambda_{A_1B_1} = \lambda_{A_2B_2}$ can be seen as the probability of drawing the tree in the top configuration, while $\lambda_{A_1B_2} = \lambda_{A_2B_1} = 1 - \lambda_{A_1B_1}$ can be seen as the probability of drawing the tree in the bottom configuration.

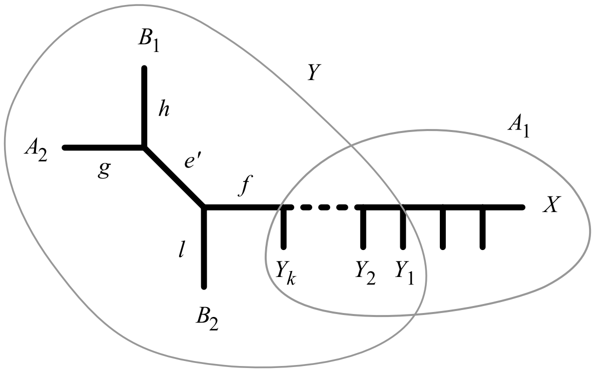

Fig. S2. Illustration for the proof of Lemma 13.