Reversible Polymorphism-Aware Phylogenetic Models and their Application to Tree Inference

—

Supplementary Material

Dominik Schrempf[a,b], Bui Quang Minh[c], Nicola De Maio[d,e], Arndt von Haeseler[c], Carolin Kosiol[a,*]

[a]*Institut für Populationsgenetik, Vetmeduni Vienna, Wien, Austria*
[b]*Vienna Graduate School of Population Genetics, Wien, Austria*
[c]*Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, Austria*
[d]*Nuffield Department of Medicine, University of Oxford, UK*
[e]*Oxford Martin School, University of Oxford, UK*

Contents

[*]Corresponding author; *carolin.kosiol@vetmeduni.ac.at*.

S1. Proof of Theorem 1 (Reversibility)

Here, we present a proof by construction of Theorem 1 (Section 2.5). $\boldsymbol{Q}_{revPoMo}$ with entries q^{ij}_{xy} is defined in Appendix A. For $0 \leq i \leq N$ note the tautologies

$$q^{ij}_{xy} = q^{(N-i)(N-j)}_{yx}. \tag{S1}$$

During the construction of $\boldsymbol{p}$, it is convenient to work with a stationary measure $\boldsymbol{\lambda} = (\lambda_s)_{s \in \mathscr{A}_{PoMo}}$ which fulfills stationarity but is not normalized; i.e., $\sum_{s \in \mathscr{A}_{PoMo}} \lambda_s \neq 1$ and $\boldsymbol{\lambda} = c\boldsymbol{p}$, where c is a normalization constant. Like before, λ_x and λ^i_{xy} are the elements of $\boldsymbol{\lambda}$ corresponding to boundary states and polymorphic states, respectively (Section 2.5). A sufficient condition of reversibility of an irreducible Markov process (i.e., if all states are connected) with a finite number of states is the existence of such a stationary measure $\boldsymbol{\lambda}$ that fulfills detailed balance $\lambda_r q_{rs} = \lambda_s q_{sr}$ ($r, s \in \mathscr{A}_{PoMo}$; e.g., Norris 1998, p. 125). $\boldsymbol{Q}_{revPoMo}$ is irreducible because we can reach every state irrespective of the starting state. The conditions of detailed balance lead to $\binom{4}{2}$ equations for the boundary states

$$\lambda_x \mu_{xy} = \lambda^{N-1}_{xy} q^{N-1}. \tag{S2}$$

For changes within polymorphic states we get $(N-2)\binom{4}{2}$ conditions

$$\lambda^i_{xy} q^i = \lambda^{i+1}_{xy} q^{i+1} \quad (1 \leq i \leq N-2), \tag{S3}$$

which can be rewritten in the recursive form

$$\lambda^{i+1}_{xy} = \lambda^i_{xy} \frac{q^i}{q^{i+1}} \quad (1 \leq i \leq N-2). \tag{S4}$$

Together with Eq. (S2) and realizing that $\lambda^1_{xy} = \lambda^{N-1}_{yx}$, as well as $q^1 = q^{N-1}$, this leads to

$$\lambda^i_{xy} = \frac{\lambda_y \mu_{yx}}{q^1} \frac{q^1 \cdots q^{i-1}}{q^2 \cdots q^i} = \lambda_y \pi_x m_{xy} \frac{1}{q^i} \quad (1 \leq i \leq N-1). \tag{S5}$$

We only need to determine λ_y so that it conforms to the boundary conditions because the π_x and m_{xy} are model parameters and $q^i = i(N-i)/N$. If $i = N$ we find with Eq. (S2) and (S5)

$$\frac{\lambda_y}{\lambda_x} = \frac{\pi_y}{\pi_x}. \tag{S6}$$

Note, that to arrive at this equation we required reversibility of the associated mutation model. A possible solution to this set of $\binom{4}{2}$ equations is $\lambda_y = \pi_y$, which is the one considered here. However, other solutions might be obtained by scaling by a common factor. In brief,

$$\lambda_x = \pi_x, \tag{S7}$$

$$\lambda_{xy}^i = \pi_x \pi_y m_{xy}/q^i. \tag{S8}$$

is a solution to the detailed balance conditions and revPoMo is reversible. □

The conditions of detailed balance hold for any reversible DNA mutation model that is nested within the GTR model (e.g., the HKY model of Hasegawa et al. 1985). It is important that the reversibility of revPoMo is not a mere consequence of the reversibility of general birth-death processes with a finite number of states (e.g., Norris, 1998). The situation in revPoMo is more complicated because there are $\binom{4}{2} = 6$ connections between the boundary states. If the total rate of traversal along an arbitrarily chosen circular path depends on its direction, the process ceases to be reversible. This is also known as the Kolmogorov criterion (Kelly, 1979, p. 21). We want to emphasize, that only the symmetry of the coefficients $q^i = q^{N-i}$ as well as $q^{i,i+1} = q^{i,i-1} = q^i$ was used and not their functional form per se. This prevents, at least in this setting, a treatment with unequal frequency bin sizes. For example, smaller bins close to the boundaries would be a very appealing idea (cf. one-step process, Malaspinas et al., 2012).

Finally, we calculate the normalization constant c. Let $\mathscr{F}$ and $\mathscr{P}$ be the sets of boundary and polymorphic states in $\mathscr{A}_{PoMo}$, respectively. Then, $c^{-1} = \sum_{s \in \mathscr{F}} \lambda_s + \sum_{s \in \mathscr{P}} \lambda_s$. For the boundary, we easily find

$$\sum_{s \in \mathscr{F}} \lambda_s = \sum_{x \in \mathscr{A}} \pi_x = 1. \tag{S9}$$

The sum over all polymorphic states is a little bit more tedious:

$$\begin{aligned}
\sum_{s \in \mathscr{P}} \lambda_s &= \frac{1}{2} \sum_{\substack{x,y \in \mathscr{A} \\ x \neq y}} \sum_{i=1}^{N-1} \lambda_{xy}^i \\
&= \frac{1}{2} \sum_{\substack{x,y \in \mathscr{A} \\ x \neq y}} \sum_{i=1}^{N-1} \pi_x \pi_y m_{xy}/q^i \\
&= \frac{1}{2} \sum_{\substack{x,y \in \mathscr{A} \\ x \neq y}} \pi_x \pi_y m_{xy} N \sum_{i=1}^{N-1} \frac{1}{i(N-i)} \\
&= a_N \sum_{\substack{x,y \in \mathscr{A} \\ x \neq y}} \pi_x \pi_y m_{xy},
\end{aligned} \tag{S10}$$

where we used

$$\sum_{i=1}^{N-1} \frac{1}{i(N-i)} = \frac{2}{N} a_N \text{ and } a_N = \sum_{j=1}^{N-1} \frac{1}{j} \tag{S11}$$

S2. Symmetry of Stationary Distribution

The elements of $\boldsymbol{p}$ corresponding to polymorphic states are symmetric with respect to a permutation of bases. That is, for $1 \leq i \leq N$, we have $p_{xy}^{i} = p_{xy}^{N-i}$. This conforms to the high symmetry of the model itself (Eq. S1). We expect that high mutational biases lead to a skewed stationary distribution not only at the boundaries but to a lower extent also for polymorphic states. This is reflected in the solution of the diffusion equation for unequal scaled mutation rates θ_1 and θ_2, $\Phi(\nu, \theta_1, \theta_2) \sim \mathrm{Beta}(\theta_1, \theta_2)$ which allows mutations also when the population is polymorphic. Mutational biases in population genetics theory are captured by a non-uniform $\boldsymbol{\pi}$ in phylogenetics. Also the reversible PoMo approach (revPoMo) does this at the boundary but not in between them because of the symmetry of the polymorphic elements of $\boldsymbol{p}$. However, we can convince ourselves that the boundary mutation model is correct within its assumptions, i.e., non-uniform π does not contradict a symmetric distribution of the polymorphic states for boundary mutation only.

Let the distribution of allele frequencies be very skewed such that $\pi_x = n\pi_y$ ($n \in \mathbb{N}, n \gg 1$; we do not assume that $\sum_{x\in\mathscr{A}_{PoMo}} \pi_x = 1$). In this case, the entries of the stationary distribution are:

$$p_y = c\pi_y, \tag{S12}$$

$$p_x = c\pi_x = cn\pi_y, \text{ and} \tag{S13}$$

$$p_{xy}^{i} = cn\pi_y\pi_y m_{xy}/q^{i}. \tag{S14}$$

The mutation coefficients are $\mu_{xy} = m_{xy}\pi_y$ and $\mu_{yx} = nm_{xy}\pi_y$. At equilibrium, the frequency of state $\{Nx\}$ is n times higher than the one of state $\{Ny\}$ but the mutation rate from x to y is also n times smaller. In total, the mutation rates from $\{Nx\}$ in direction of y and from $\{Ny\}$ in direction of x are $p_x\mu_{xy} = p_y\mu_{yx} = cn\pi_y\pi_y m_{xy}$. They level out and the effective frequency of polymorphic states at equilibrium is symmetric.

S3. Description of Simulations

S3.1. Small Trees

Here, we provide command lines for the simulation pipeline of the Incomplete Lineage Sorting (ILS) scenario that is also treated in the main manuscript and another 4-species scenario with Recent Radiation (RR) events. Two scenarios with 8 species (balanced, BAL and unbalanced UNB) have also been included. An exact description of the RR, BAL and UNB scenarios can be found in our previous publication (De Maio, Schrempf, and Kosiol, 2015).

1. Simulate gene trees for a fixed species tree with MSMS v.3.2 (Ewing and Hermisson, 2010); e.g., ILS scenario, tree height 1 N_e, 10 Samples (S) 3 Genes (G) and the first replicate (R).

```
java -Xmx1g -jar msms3.2rc-b163.jar 40 3 -t 0.01 \
     -I 4 10 10 10 10 -ej 0.5 4 3 -ej 0.6 3 2 \
     -ej 1 2 1 -oSeqOff -T \
     > 01a_msms_out/ILS_1Ne_10S_0003G_00R
```

The MSMS output has to be modified slightly before Seq-Gen can read it.

2. Creation of sequence data with Seq-Gen v.1.3 (Rambaut and Grass, 1997). The HKY model is used, the stationary frequencies for A, C, G and T are set to 0.3, 0.2, 0.2 and 0.3 respectively and the transition to transversion ratio is 3.0. Note that this means that the ratio of the transition rate to the transversion rate is 6.0. One gene has 1000 base pairs (bp). The height of the species tree is 1.0 which is scaled by a factor of 0.0025 by Seq-Gen. This means that the average number of substitutions from the tip to the root is 0.0025.

```
seq-gen -mHKY -f0.3,0.2,0.2,0.3 -t3.0 -l1000 -n1 -on \
        -s0.0025 < 02_seqgen_in/ILS_1Ne_10S_0003G_00R \
        > 03_seqgen_out/ILS_1Ne_10S_0003G_00R
```

3. The resulting sequences are either concatenated, for the use of concatenation methods or converted to Counts Files with *libPoMo* (De Maio, Schrempf, and Kosiol, 2015).

S3.2. Large Trees

Here, we describe the simulation pipeline for the large trees simulated under the Yule speciation model. The species and gene trees with 50 and 60 species were simulated with SimPhy v.1.0 (Mallo et al., 2015). It randomly simulates species trees and subsequently gene trees. The creation of sequences with Seq-Gen and the preparation of the input files for concatenation methods and PoMo is the same as described above. This is the command line for the trees with 60 species and a tree height of $3 N_e$.

```
simphy -sb f:0.0001226623470983912 -rs 10 -rl f:1 -rg 3 \
       -sp f:10000 -sg f:1 -sl f:60 -st f:30000 \
       -si f:10 -su f:0.00005 \
       -o data/01b_simphy_out/Y50_1Ne_10S_0003G
```

S4. Description of Analysis

Three different methods were used to analyze sequence data: an analysis of the concatenated input with standard DNA substitution models (concatenation), the older non-reversible version of PoMo in HyPhy (PoMo, De Maio, Schrempf, and Kosiol, 2015) and the new reversible version of PoMo in IQ-TREE (revPoMo, Nguyen et al., 2015). Source code and binaries are available online under

- *https://github.com/pomo-dev/PoMo/releases/tag/v1.1.0* and
- *https://github.com/Cibiv/IQ-TREE/tree/PoMo*, respectively.

S4.1. Concatenation

The concatenated input was analyzed with IQ-TREE and the HKY model.

```
iqtree -s ILS_1Ne_10S_0003G_00R -m HKY
```

S4.2. Non-Reversible PoMo

The non-reversible version of PoMo runs within a Python interpreter. The *.cf* file ending indicates that the file is a Counts File.

```
PoMo.py "HYPHYMP CPU=1"  -t 0.0025 ILS_1Ne_10S_0003G_00R.cf
```

S4.3. Reversible PoMo

The reversible version of PoMo has been implemented into IQ-TREE (Nguyen et al., 2015). It automatically recognizes Counts File data and uses PoMo with a virtual population size of 9 and the *weighted* sampling technique by default.

```
iqtree -s ILS_1Ne_10S_0003G_00R_CF09.cf -m HKY
```

The sampling technique and virtual population size can be adjusted with the `-st` flag. The following code snippet runs PoMo with a virtual population size of 7 and the *sampled* input method.

```
iqtree -s ILS_1Ne_10S_0003G_00R_CF09.cf -m HKY -st CR07
```

Options and detailed help are available with `iqtree -h` and online at *https://github.com/Cibiv/IQ-TREE/tree/PoMo*.

S5. Details on Implementation

S5.1. Numerical Stability

To increase numerical stability we fix the last element of $\boldsymbol{\pi}$ to 1.0 as opposed to fixing the sum of its elements to 1.0, which is done by standard DNA substitution models. We want to emphasize that the nature of the model may lead to numerical instabilities especially because the stationary frequency vector $\boldsymbol{p}$ may have very low entries for some polymorphic states compared to the entries of boundary states. This huge span may lead to problems upon eigendecomposition which may be avoided by using a Taylor expansion for the calculation of the matrix exponential (Pond et al., 2005). Subsequently, the partial likelihood vector of inner nodes has the same structure and the standard numerical precision might be insufficient. Most numerical problems can be avoided by using the *weighted* input method because then the likelihoods of many states at the leaves of the tree are nonzero.

The difficulties in implementing complex models and the corresponding numerical challenges are fundamental in maximum likelihood inference, especially once the amount of data or the model state space increases. The limit of numerical precision where the likelihood computation breaks down and the Felsenstein's algorithm does not work with extremely small numbers anymore might be reached. We have seen this happening for codon models, as well as for DNA models when we have more than 2000 tips in the tree.

S5.2. Weighted Input Method

If the observed samples are fixed for a specific nucleotide x at a specific site the weighted input method tends to underestimate the partial likelihood of the boundary state $\{Nx\}$ and overestimate the likelihoods for polymorphic PoMo states that include at least one x. We multiply the partial likelihood of the boundary state $\{Nx\}$ with a factor of $\binom{4}{2}$ because this is the number of nucleotide combinations. This improves estimates, especially when N is much larger than the actual number of samples per population in the data.

S5.3. Beta-Binomial with Pool-Seq Data Input

A beta-binomial distribution can be used to allow pool sequence (Futschik and Schlötterer, 2010) input data with sequencing errors to revPoMo. Then, even if the state at the tip of the tree is $\{NA\}$, the probability of this state given a sample $\{(M-1)A, 1C\}$ is not 0 because of a possible sequencing error. Similar to Eq. (13), we have

$$\mathbb{P}(\{jx,(M-j)y\}|\{ix,(N-i)y\}) = \mathrm{BetaBin}\left(j; M, \frac{i}{N}, \rho\right), \tag{S15}$$

where ρ is the dispersion factor. It models the probability of sequencing errors by defining the amount of extra variance with respect to the binomial distribution

$$\begin{aligned}\mathrm{Var}\left[\mathrm{BetaBin}(j; M, \frac{i}{N}, \rho)\right] &= M\frac{i(N-i)}{N^2}(1+(M-1)\rho)\\ &= \mathrm{Var}\left[\mathrm{Bin}(j; M, \frac{i}{N})\right](1+(M-1)\rho).\end{aligned} \tag{S16}$$

S6. Further Results — Runtimes

Additionally to the ILS scenario, we also analyzed the RR scenario (De Maio, Schrempf, and Kosiol, 2015). The tree height was either $1\,N_e$ or $10\,N_e$. The sample number was 2, 3, 10 and 20 and also the virtual population size of PoMo was varied. The next figures are a more complete subset of the results (runtimes and branch score distances, BSDs) that have been produced. In all figures we abbreviate the concatenation approach, the non-reversible PoMo approach, revPoMo with weighted input method and revPoMo with sampled input method with “IQ-TREE, HKY+Conc”, “HyPhy, HKY+PoMo”, “IQ-TREE, HKY+revPoMo+*Weighted*” and “IQ-TREE, HKY+revPoMo+*Sampled*”, respectively.

S6.1. Incomplete Lineage Sorting Scenario

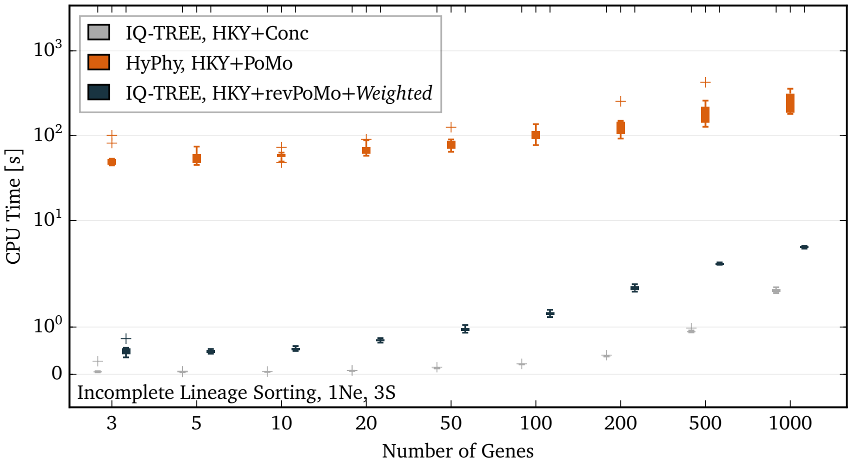

Figure S1: Runtimes of the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *weighted*) approaches for the ILS scenario with three samples and a tree height of $1\,N_e$ and different amounts of data on the x-axis. The HKY model was used for all methods. Each gene has a length 1000 bp.

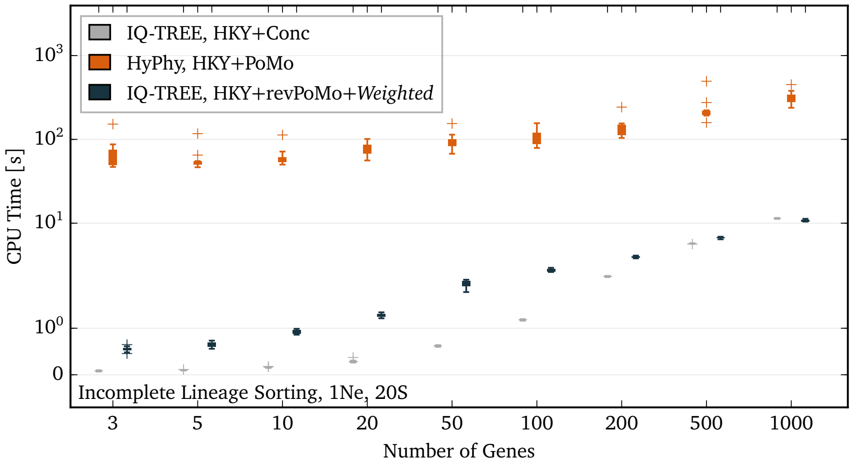

Figure S2: The runtimes of the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *weighted*) approaches for the ILS scenario with 20 samples and a tree height of $1\,N_e$ and different amounts of data on the x-axis. The HKY model was used for all methods. Each gene has a length 1000 bp.

S6.2. Recent Radiation Scenario

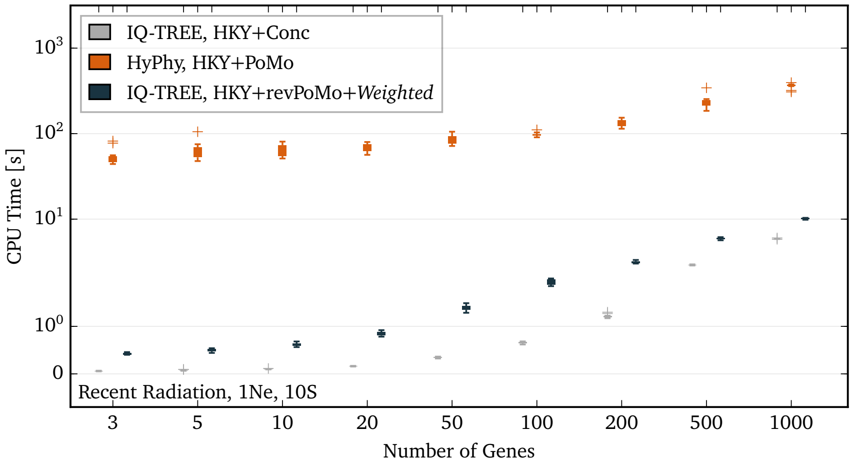

Figure S3: The runtimes of the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *weighted*) approaches for the recent radiation (RR) scenario with ten samples and a tree height of $1\,N_e$ and different amounts of data on the x-axis. The HKY model was used for all methods. Each gene has a length 1000 bp.

S7. Further Results — Tree Errors

S7.1. Incomplete Lineage Sorting Scenario

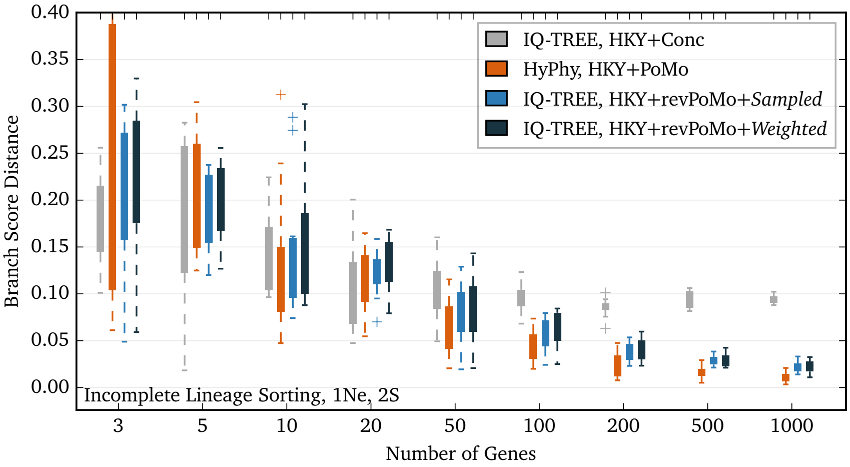

Figure S4: The tree error measured by the branch score distance for the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *sampled* and *weighted*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the ILS scenario with two samples and a tree height of $1\,N_e$; one gene has 1000 bp.

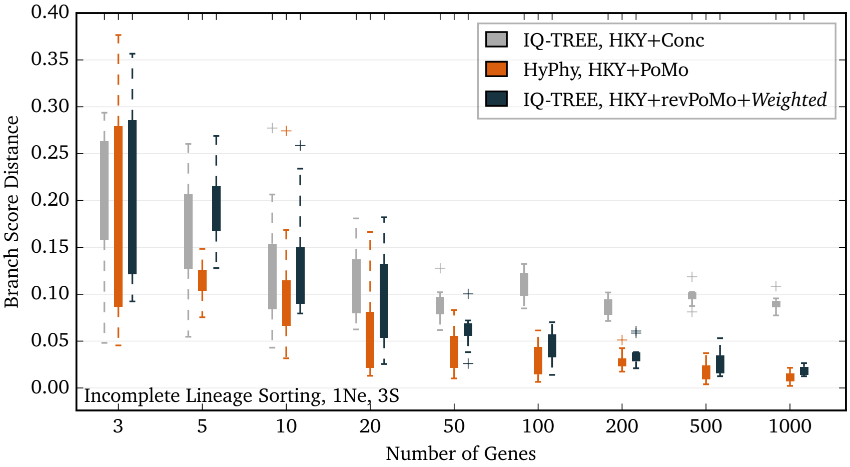

Figure S5: The tree error measured by the branch score distance for the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *weighted*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the ILS scenario with three samples and a tree height of $1\,N_e$; one gene has 1000 bp.

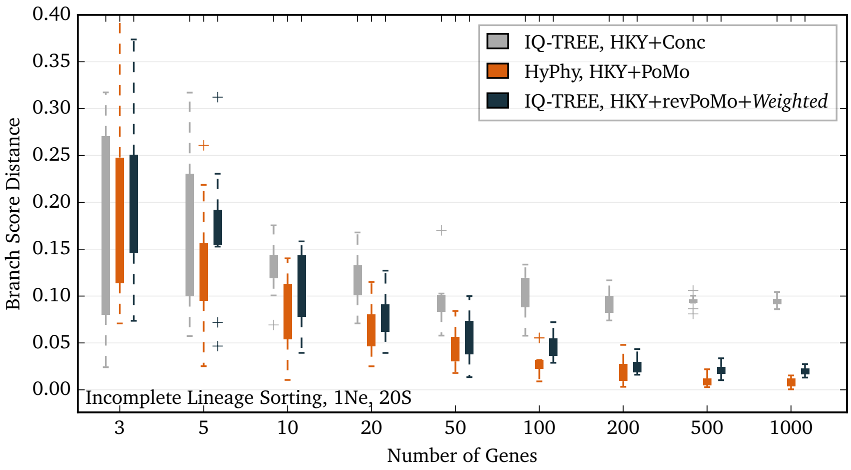

Figure S6: The tree error measured by the branch score distance for the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *weighted*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the ILS scenario with 20 samples and a tree height of $1\,N_e$; one gene has 1000 bp.

S7.2. *Recent Radiation Scenario*

The recent radiation (RR) scenario involves three species that are very closely related.

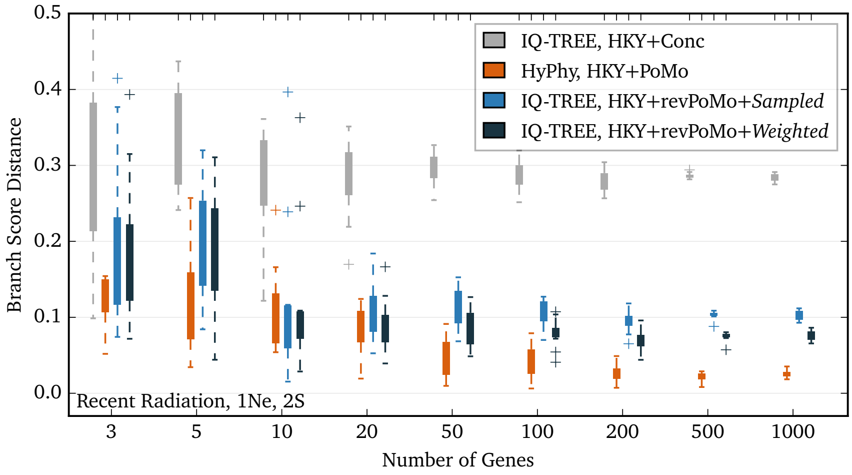

Figure S7: The tree error measured by the branch score distance for the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *sampled* and *weighted*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the RR scenario with two samples and a tree height of $1\,N_e$; one gene has 1000 bp. In this specific case, revPoMo has a higher tree error. The *weighted* input method is more accurate.

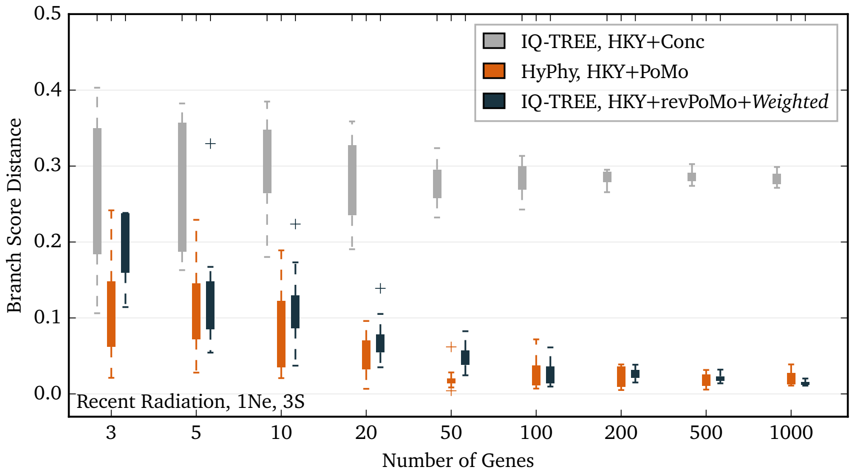

Figure S8: The tree error measured by the branch score distance for the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *weighted*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the RR scenario with three samples and a tree height of $1\,N_e$; one gene has 1000 bp. The reversible model performs best.

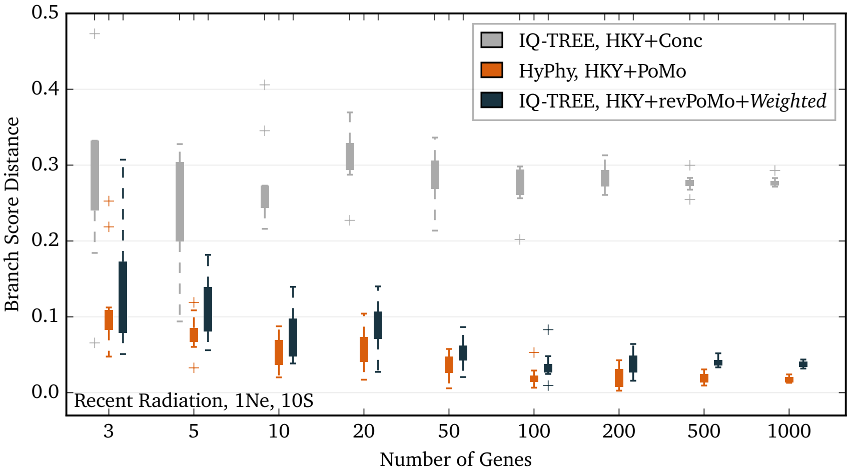

Figure S9: The tree error measured by the branch score distance for the concatenation, non-reversible PoMo and revPoMo ($N = 9$, *weighted*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the RR scenario with ten samples and a tree height of $1\,N_e$; one gene has 1000 bp.

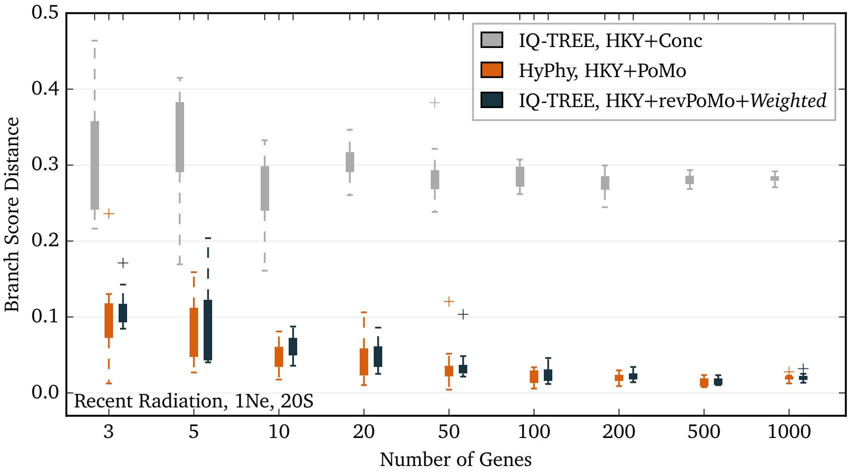

Figure S10: The tree error measured by the branch score distance for the concatenation, non-reversible PoMo and revPoMo ($N = 19$, *weighted*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the RR scenario with 20 samples and a tree height of $1\,N_e$; one gene has 1000 bp.

S7.3. Balanced Tree Scenario

The balanced tree (BAL) has eight species. Every split creates two clades with an equal number of species. I.e., the oldest split divides the tree into two clades with four species. Hardly any incomplete lineage sorting is expected in this scenario. Indeed, PoMo does not perform considerably better, even when a lot of data is available.

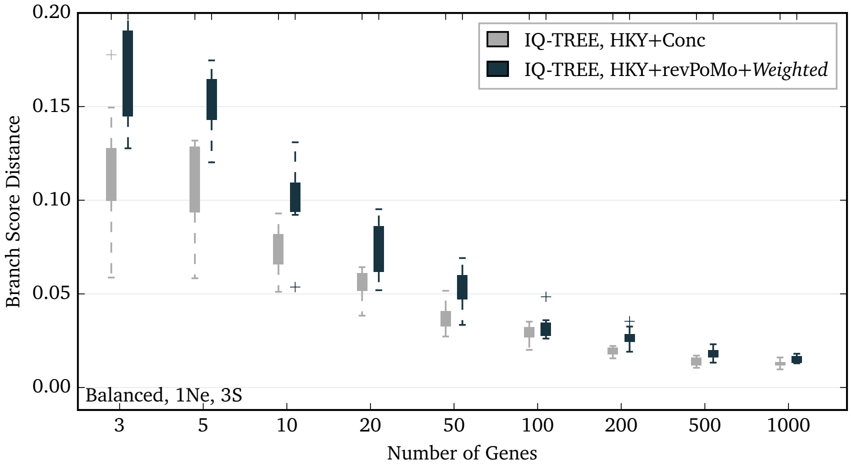

Figure S11: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *sampled*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the BAL scenario with three samples and a tree height of $1\,N_e$; one gene has 1000 bp. This is the only case where PoMo performs worse.

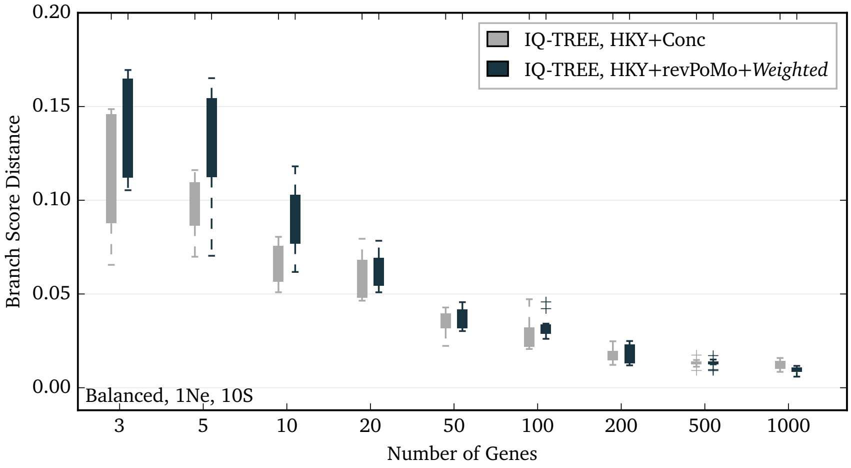

Figure S12: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *sampled*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the BAL scenario with ten samples and a tree height of $1\,N_e$; one gene has 1000 bp.

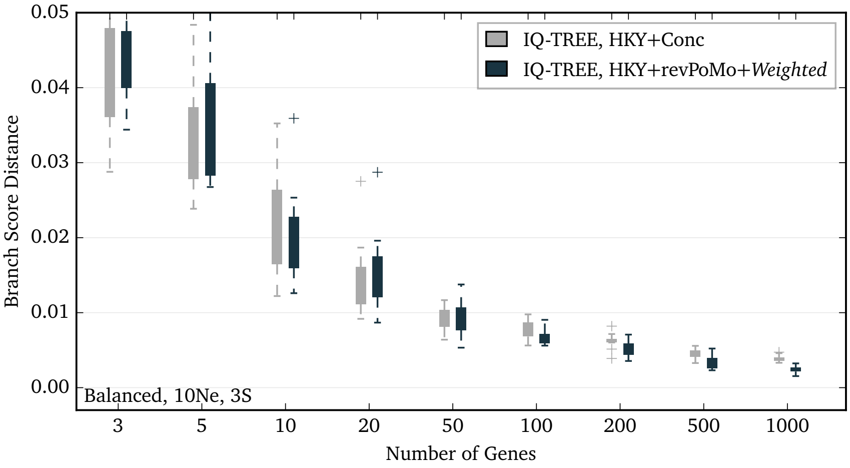

Figure S13: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *sampled*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the BAL scenario with three samples and a tree height of $10N_e$; one gene has 1000 bp.

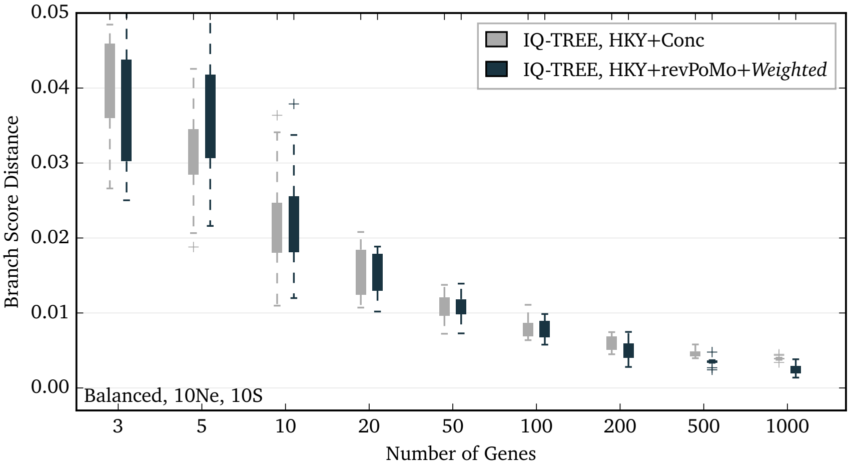

Figure S14: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *sampled*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the BAL scenario with ten samples and a tree height of $10N_e$; one gene has 1000 bp.

S7.4. Unbalanced Tree Scenario

The unbalanced tree (UNB) scenario has eight species. Every split takes away one species. I.e., the oldest split divides the tree into clades with one and seven species. Here, we expect a lot of incomplete lineage sorting, especially when the tree is short.

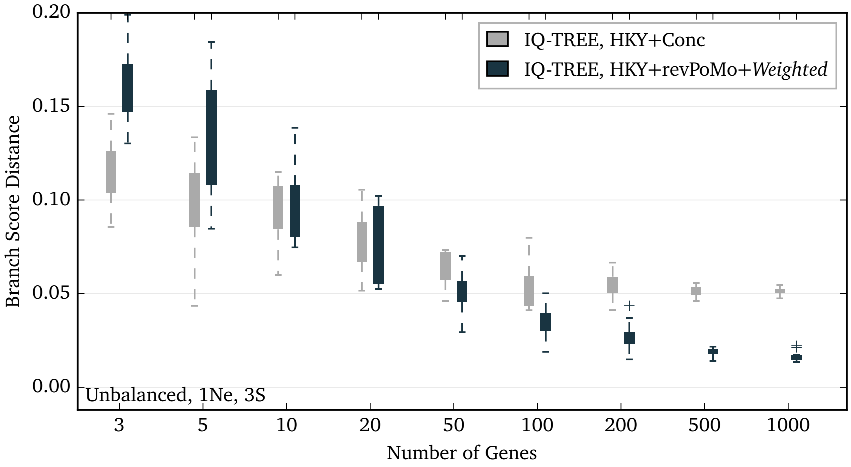

Figure S15: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *sampled*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the UNB scenario with three samples and a tree height of $1\,N_e$; one gene has 1000 bp. PoMo performs way better here because a high level of incompletely sorted lineages is expected for this scenario.

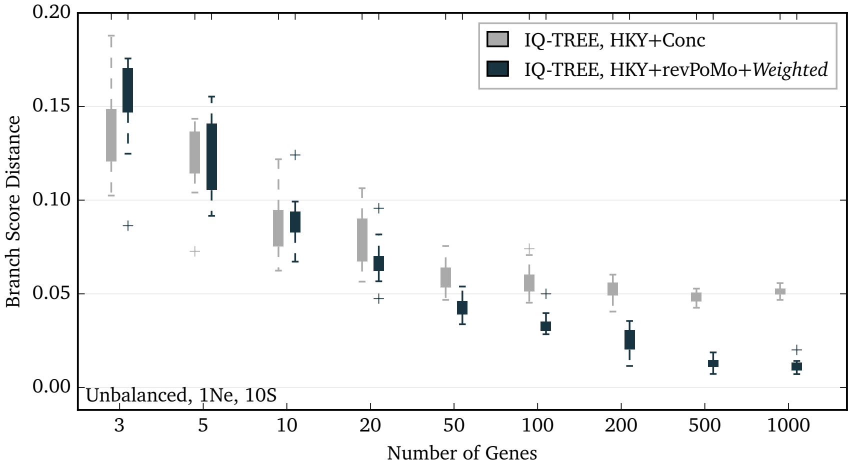

Figure S16: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *sampled*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the UNB scenario with ten samples and a tree height of $1\,N_e$; one gene has 1000 bp. PoMo performs way better here because a high level of incompletely sorted lineages is expected for this scenario.

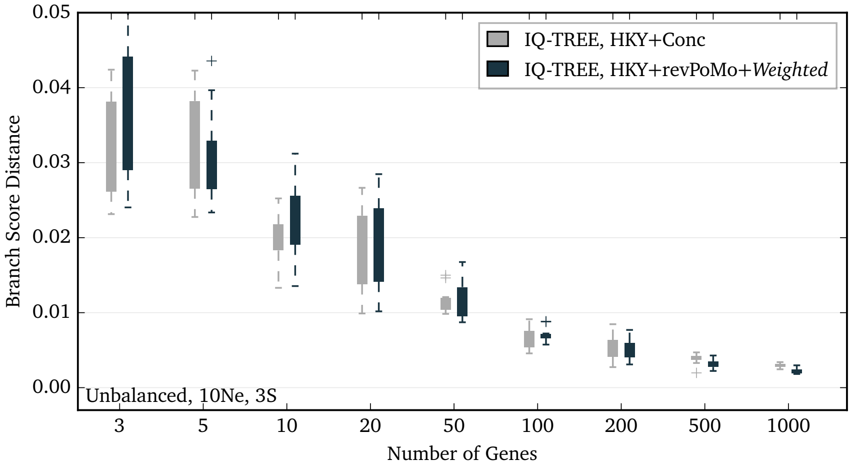

Figure S17: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *sampled*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the UNB scenario with three samples and a tree height of $10\,N_e$; one gene has 1000 bp.

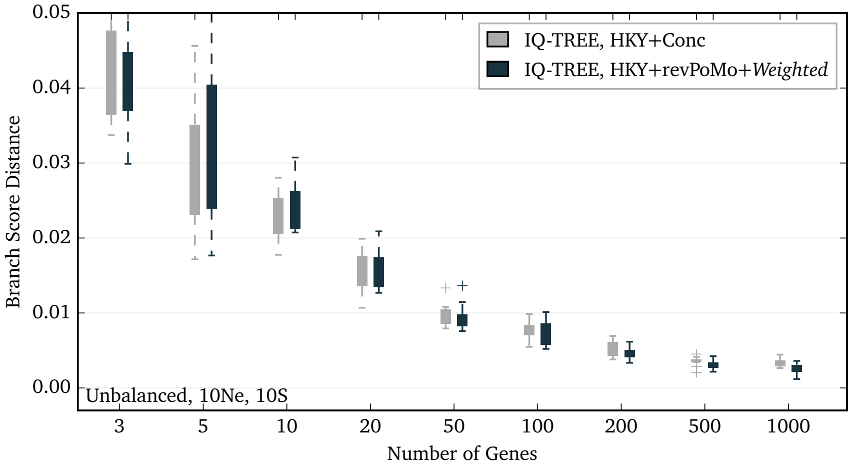

Figure S18: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *sampled*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under the UNB scenario with ten samples and a tree height of $10\,N_e$; one gene has 1000 bp.

S7.5. *Yule Tree with 50 Species*

We also simulated a Yule tree for 50 species with $1\,N_e$ tree height and a rate of three speciation per coalescent time unit.

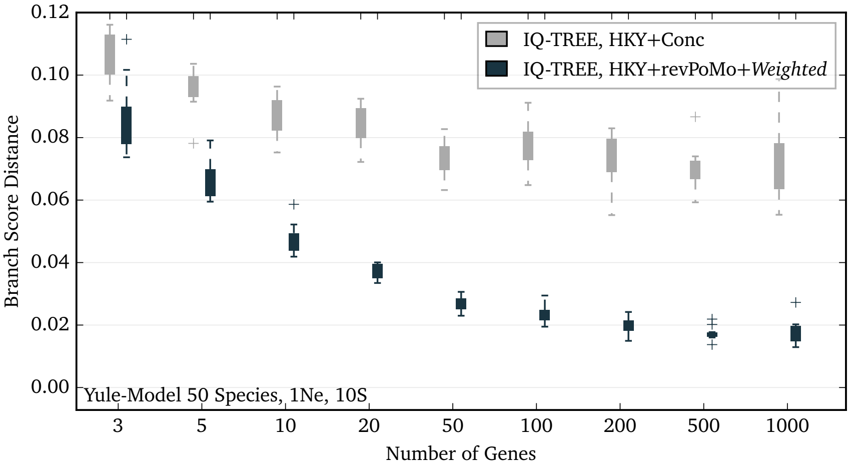

Figure S19: The tree error measured by the branch score distance for the concatenation and revPoMo ($N = 9$, *weighted*) approaches with the HKY model in dependence of the amount of data. The analyzed sequences were simulated under a Yule speciation model with ten samples per species and a tree height of $1\,N_e$; one gene has 1000 bp.

S8. Dependence on the Virtual Population Size

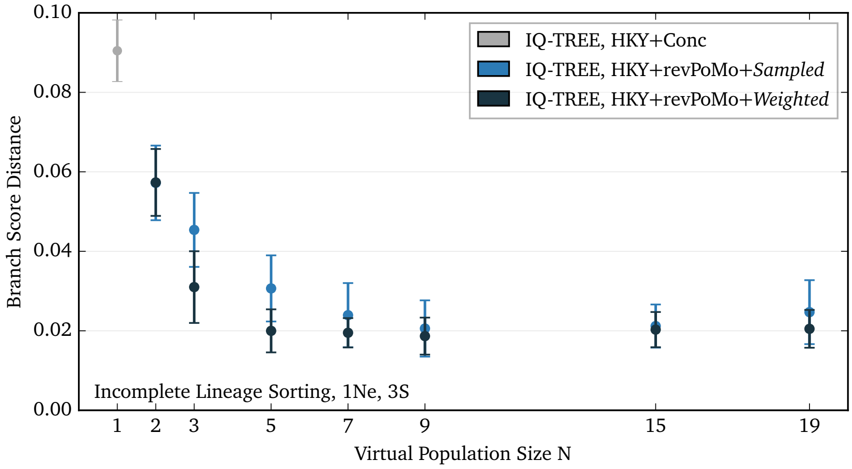

Figure S20: The branch score distance in dependence of the virtual population size N for both sampling techniques of revPoMo and the incomplete lineage sorting scenario with three samples and a tree height of $1\,N_e$. The error bars are standard deviations of ten runs. For $N = 1$, the estimate of the concatenation approach was used because PoMo requires $N \geq 2$. All models use the HKY model. The accuracy of revPoMo improves up to a virtual population size of $N = 9$. We believe that an underestimation of actual substitutions (Fig. 8) and maybe also the introduction of numerical errors due to an oversized state space leads to the deterioration of the branch score distance for large N.

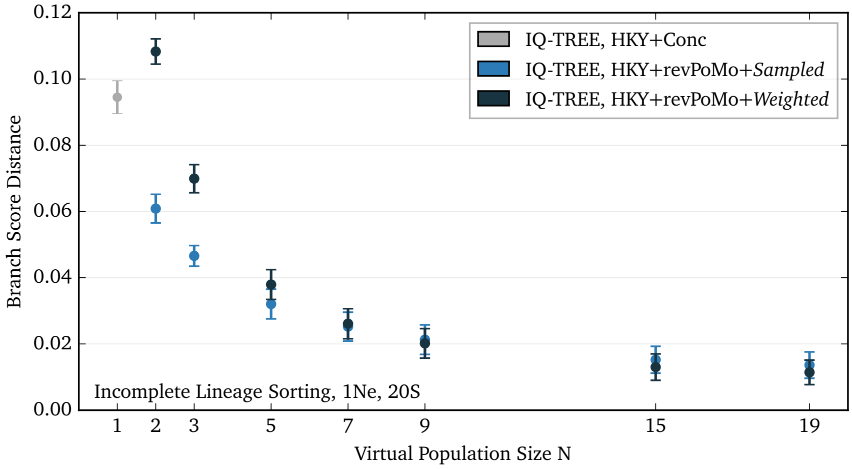

Figure S21: The branch score distance in dependence of the virtual population size N for both sampling techniques of revPoMo and the incomplete lineage sorting scenario with 20 samples and a tree height of $1\,N_e$. The error bars are standard deviations of ten runs. For $N = 1$, the estimate of the concatenation approach was used because PoMo requires $N \geq 2$. All models use the HKY model. Here, the accuracy of revPoMo slightly improves even for high virtual population sizes.

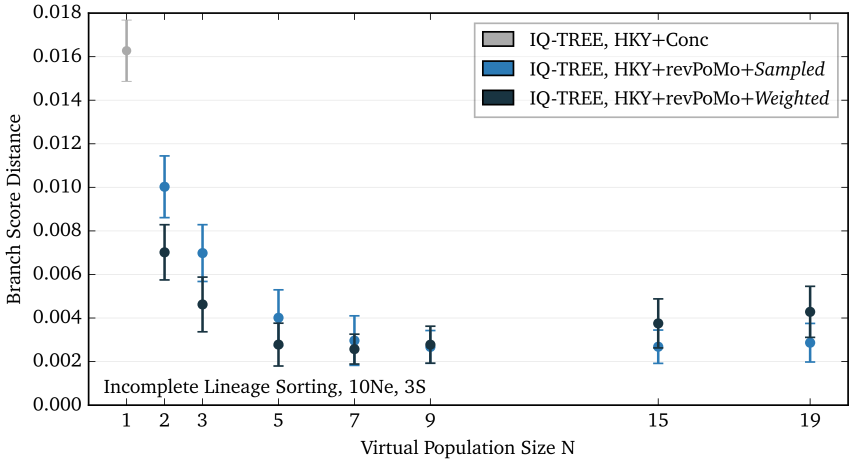

Figure S22: The branch score distance in dependence of the virtual population size N for both sampling techniques of revPoMo and the incomplete lineage sorting scenario with three samples and a tree height of $10N_e$. The error bars are standard deviations of ten runs. For $N = 1$, the estimate of the concatenation approach was used because PoMo requires $N \geq 2$. All models use the HKY model. The accuracy of the reversible PoMo improves up to a virtual population size of $N = 9$. We believe that an underestimation of actual substitutions (Fig. 8) and maybe also the introduction of numerical errors due to an oversized state space leads to the deterioration of the branch score distance for large N.

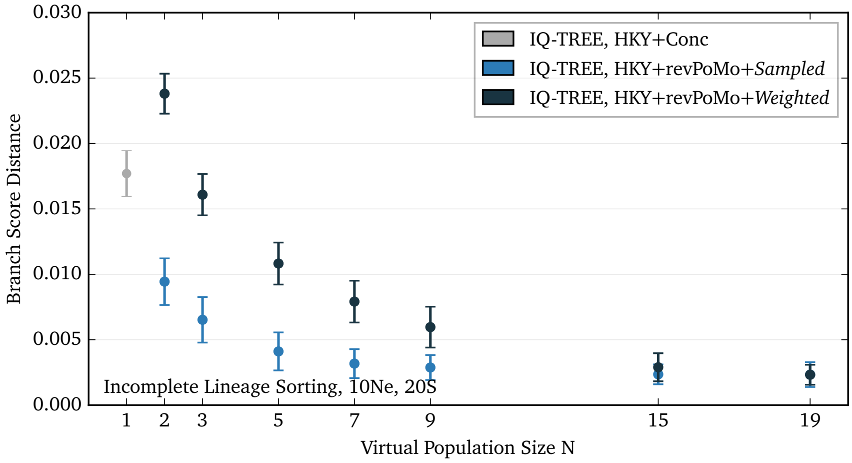

Figure S23: The branch score distance in dependence of the virtual population size N for both sampling techniques of revPoMo and the incomplete lineage sorting scenario with 20 samples and a tree height of $10N_e$. The error bars are standard deviations of ten runs. For $N = 1$, the estimate of the concatenation approach was used because PoMo requires $N \geq 2$. All models use the HKY model. The accuracy of the reversible PoMo minimally improves even for high virtual population sizes.

S9. Bibliography

De Maio N, Schrempf D, and Kosiol C, 2015. PoMo: An Allele Frequency-Based Approach for Species Tree Estimation. *Systematic Biology*, 64(6):1018–1031.

Ewing G and Hermisson J, 2010. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics*, 26(16):2064–2065.

Futschik A and Schlötterer C, 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, 186(1):207–18.

Hasegawa M, Kishino H, and Yano T a, 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2):160–174.

Kelly F P, 1979. *Reversibility and Stochastic Networks*. Wiley, Chichester.

Malaspinas A S, Malaspinas O, Evans S N, and Slatkin M, 2012. Estimating Allele Age and Selection Coefficient from Time-Serial Data. *Genetics*, 192(2):599–607.

Mallo D, de Oliveira Martins L, and Posada D, 2015. SimPhy: Phylogenomic Simulation of Gene, Locus and Species Trees. *Systematic biology*, page syv082.

Nguyen L T, Schmidt H A, von Haeseler A, and Minh B Q, 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1):268–274.

Norris J R, 1998. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Pond S L K, Frost S D W, and Muse S V, 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, 21(5):676–679.

Rambaut A and Grass N C, 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences : CABIOS*, 13(3):235–238.