Article - Discoveries

Pervasive selection biases inferences of the species tree

Rui Borges[1], Bastien Boussau[2] Gergely J. Szöllősi[3,4,5] and Carolin Kosiol[1,6,*]

1. Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Wien, Austria
2. Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5558, LBBE, F-69100, Villeurbanne, France
3. Department of Biological Physics, Eötvös University, Pázmány P. stny. 1A, H-1117 Budapest, Hungary
4. MTA-ELTE "Lendület" Evolutionary Genomics Research Group, Pázmány P. stny. 1A, H-1117 Budapest, Hungary
5. Evolutionary Systems Research Group, Centre for Ecological Research, Hungarian Academy of Sciences, 8237 Tihany, Hungary
6. Centre for Biological Diversity, University of St Andrews, St Andrews, Fife KY16 9TH, UK

*** Corresponding author**: `ck202@st-andrews.ac.uk`

Abstract

Despite the importance of natural selection in species' evolutionary history, phylogenetic methods that take into account population-level processes typically ignore selection. The assumption of neutrality is often based on the idea that selection occurs at a minority of loci in the genome and is unlikely to compromise phylogenetic inferences significantly. However, selection might behave more pervasively, as in the case of nearly neutral evolving mutations. Genome-wide processes like GC-bias and some variation segregating at the coding regions are known to evolve in the nearly neutral range. As we are now using genome-wide data to estimate species trees, it is natural to ask whether weak but pervasive selection is likely to blur species tree inferences. We developed a polymorphism-aware phylogenetic model tailored for measuring signatures of nucleotide usage biases to test the impact of selection in the species tree. Our analyses indicate that while the inferred relationships among species are not significantly compromised, the genetic distances are systematically underestimated in a node-height dependent manner: i.e., the deeper nodes tend to be more underestimated than the shallow ones. Such biases have implications for molecular dating. We dated the evolutionary history of 30 worldwide fruit fly populations, and we found signatures of GC-bias considerably affecting the estimated divergence times (up to 23%) in the neutral model. Our findings call for the need to account for selection when quantifying divergence or dating species evolution.

Keywords: species tree, selection, nearly neutral evolution, GC-bias, molecular dating

Significance statement

Although little is known about the impact of natural selection on species tree estimation, expectations are that it occurs at a minority of loci in the genome and is thus unlikely to affect the divergence process. However, growing evidence suggests that a large amount of the genomic variation evolves under weak but pervasive selection (e.g., GC-bias gene conversion). We tested the impact of unaccounted-for nearly neutral selection on species tree estimation and found that the estimated branch lengths are systematically biased. Our results highlight the need for selection-aware models in species tree estimation and molecular dating.

Introduction

Species trees provide a framework to understand the divergence process and speciation and are nowadays used routinely in integrative research to address many biological questions. Despite their generalized use, modeling species evolution using DNA sequences poses significant challenges. The main one being that different genes or genomic regions narrate different evolutionary histories, leading to discordant gene and species topologies (Maddison and Knowles, 2006). Incomplete lineage sorting (ILS), i.e., the maintenance of ancestral polymorphisms due to random genetic drift, is a primary cause of such discordance (Pollard et al., 2006). However, other processes were also described, such as gene gain or loss, horizontal gene transfer across the species boundaries, and gene flow among diverging populations (reviewed in Szöllősi et al. (2015)). Apart from the gene/species tree discordance, difficulties arise when quantifying divergence. Xu and Yang (2016) reported that calendar times and evolutionary rates can be particularly challenging to untie for deeper evolutionary scales where one violates the molecular clock hypothesis.

Despite the challenges posed by modeling species evolution, the last two decades have seen an explosion of sophisticated statistical methods for inferring species trees. The multispecies coalescent model (MSC) has arisen as a leading framework for inferring species phylogenies while accounting for ILS and gene tree-species tree conflict (Rannala and Yang, 2003). The coalescent traces the genealogical history of a sample of sequences from a population backward in time, describing the stochastic process of lineage joining (Kingman, 1982a,b).

Nevertheless, alternative approaches to the MSC have been proposed, as the polymorphism-aware phylogenetic models (PoMos; fig. 1) (De Maio et al., 2013, 2015). PoMos model the allele content of a set of populations over time at a particular locus, thus naturally integrating over all the possible allelic histories to directly estimate the species tree. (To be technically accurate, PoMos estimate population trees. However, because of PoMo ability to model closely related populations as well as diverged populations it most prominent application are populations sampled to estimate species trees. Therefore, we will adhere to this terminology in this article.) PoMos naturally account for ILS while avoiding using genealogy samplers that are computationally costly. Different from the MSC, PoMos assume independence between sites, which allows easily gathering information from multiple individuals and several populations to testing hypotheses genome-wide (Schrempf et al., 2016, 2019). More importantly, PoMos are versatile, for they can be straightforwardly rebuilt to account for other population forces (e.g., allelic selection (Borges et al., 2019)). MSC-based methods with selection are notoriously difficult.

Despite the remarkable effort made to account for forces that can derail the species evolutionary history, models of species evolution generally ignore natural selection. While the literature acknowledges the need for selection-aware models of species evolution, the expec-

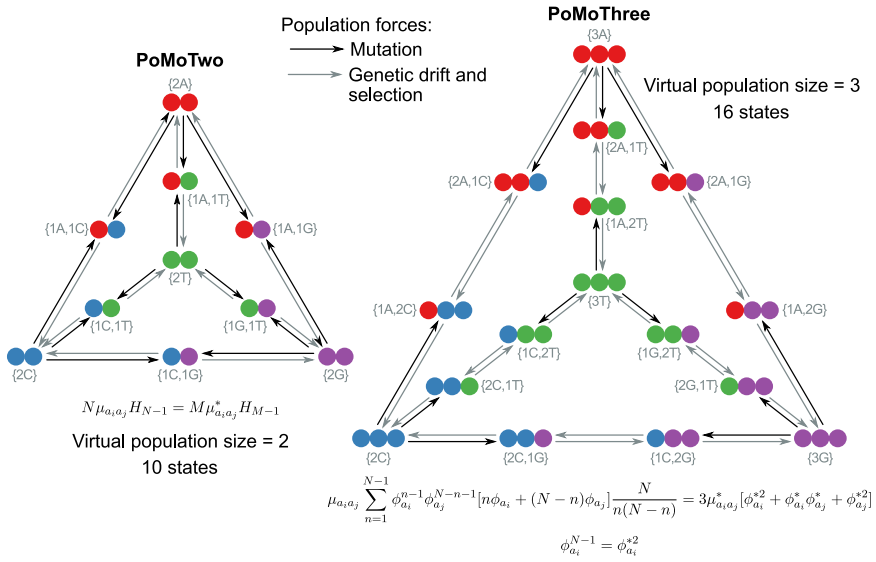

Figure 1: **State-space and parameter scaling of the virtual PoMos.** The tetrahedrons represent the state-space of the virtual PoMos: the fixed sites $\{Ma_i\}$ are placed in the vertices, while the edges represent the polymorphic states $\{ma_i, (M-m)a_j\}$. Black and grey arrows distinguish respectively mutation from genetic drift plus selection events. PoMo is a particular case of the four-variate Moran model with boundary mutations and selection, in which the alleles are the four nucleotide bases. The formulas represent the parameter scaling between the effective and the virtual populations (supplementary text S1). $\mu_{a_i a_j}$ is the mutation rate from allele a_i to a_j; and σ_{a_i} is the selection coefficient of allele a_i on the effective dynamic of population size N; $\mu^*_{a_i a_j}$ and $\sigma^*_{a_i}$ are the corresponding parameters in the virtual dynamic.

tation is that most forms of natural selection are not greatly compromising phylogenetic analysis (Edwards, 2009). Two main arguments justify this expectation. The first argument is that selection often occurs at a minority of loci in the genome, in which case any spurious signals would likely be swamped out by the many neutral loci that are sampled (Edwards et al., 2016). For example, stabilizing selection (i.e., the most common type of selection acting on protein-coding regions), directional selection, and selective sweeps are presumed to mildly lower the overall rate of evolution or moderately violate the molecular clock (Kimura, 1983). The second argument is that the forms of selection that most likely affect the species tree are rare. This is the case of selection-driven convergent evolution and balancing selection, which tend to preserve beneficial alleles at a gene for long periods of time (Castoe et al., 2009; Edwards, 2009). As corroboration for these expectations, a recent study found that species-specific positive selection has mild effects on phylogeny estimation across an extensive range of conditions encountered in empirical data sets (Adams et al., 2018).

However, selection may act in a more pervasive, though weaker, way. Theoretical expectations predict the existence of an intermediate category between the neutral and the selected modes of evolution, known as nearly neutral. The study of variants with small selection coefficients, particularly slightly deleterious mutations, has been the focus of the so-called nearly neutral theory (Ohta, 1973). This theory provided criteria to define the fate of nearly neutral variants,

which depends on both selection and random genetic drift. Ohta (2002) has suggested that if the relative advantage or disadvantage of a particular allele σ is less than twice the reciprocal of the effective population size N (i.e., the scaled selection coefficient $N|\sigma| < 2$) the allele's trajectory is nearly neutral. However, more lenient thresholds were also reasoned (e.g., Nei (2005)).

Empirical evidence for nearly neutral evolution has become more substantial within the last few years due to the possibility of sequencing many genomes and multiple individuals. One such example is GC-biased gene conversion (gBGC), a bias favoring G and C alleles over A and T alleles at mismatch positions during recombination (Galtier et al., 2001). gBGC affects the fixation probability of GC alleles and is best modeled as selection (Nagylaki, 1983). Integrative analysis considering the recombination landscape and nucleotide substitution patterns along the genome provided evidence that gBGC acts in both eukaryotes and bacteria (Galtier et al., 2009; Lassalle et al., 2015). In humans, gBGC was estimated one order of magnitude lower than the reciprocal of the effective population size ($\sim 1.17 \times 10^{-5}$) (Glémin, 2010), which suggests that gBGC operates in the nearly neutral range. The same was observed for other apes where most of their exons evolve under weak gBGC ($N\gamma < 1$; where γ is the GC-bias rate) (Lartillot, 2013). In fruit flies, most amino acid replacements have weak signatures of positive selection. However, most of the selective effects are nearly neutral, with around 46% of amino acid replacements exhibiting scaled selection coefficients lower than two and 84% lower than four (Sawyer et al., 2007). Also, many non-synonymous polymorphisms in functionally important sites of human and bacterial populations were shown to segregate at frequencies of around 1–10%; i.e., much more frequent than variants associated with classic Mendelian diseases (Hughes et al., 2003; Hughes, 2005). Such an observation suggests that ongoing purifying selection on protein-coding regions is weak.

As a substantial part of the genome evolves nearly neutrally, this form of weak but pervasive selection has the potential to impact species' evolutionary history. Here, we present novel PoMos specially tailored to measure nucleotide usage bias throughout the genome and use them to assess the impact of unaccounted-for weak selection (or selection-like processes) on species tree estimation. We show that near neutrality significantly blurs inferences of the species tree by biasing estimations of the species divergence and discuss the implications of these biases on molecular dating.

Results

Modelling species evolution with selection

PoMos offer a versatile approach to describe species evolution (Leaché and Oaks, 2017). They add a new layer of complexity to the standard models of sequence evolution by accounting

for population-level processes (such as mutations, genetic drift, and selection) to describe sequence evolution. To do that, PoMos expand the 4×4 state-space to model a population of individuals in which changes in allele content and frequency are both possible. The PoMo state-space includes fixed (or boundary) states $\{Na_i\}$, in which all N individuals have the same allele $a_i \in \{A, C, G, T\}$, but also polymorphic states $\{na_i, (N-n)a_j\}$, if two alleles a_i and a_j are present in the population with absolute frequencies n and $N-n$ (fig. 1). PoMos do not consider triallelic sites, a simplification that is generally acceptable for eukaryotes, for which mutations occur at a much lower rate than fixation by genetic drift (Lynch et al., 2016). Furthermore, like many phylogenetic methods, PoMos assume that sites are independent (i.e., linkage equilibrium).

One of the main challenges of PoMos is the size of their state-space grows with the effective population size: i.e., $4 + 6(N-1)$ states. This prohibits inferences with realistic population sizes as they can easily have $10^4 - 10^6$ individuals (Lynch et al., 2016), PoMos would typically employ 10 individuals because it results in manageable state-space and robust estimates of population properties such as the genetic diversity θ (De Maio et al., 2015; Schrempf et al., 2016). Recently, we developed theory that permitted us to distinguish between the effective and virtual population size (Borges et al., 2019). Here, we use those results to develop the virtual PoMos. The idea is to mimic a population dynamic that unfolds on the effective population N, using a virtual population of smaller size M. It can be shown that by matching the expected diversity (i.e., the proportion of fixed and polymorphic sites) in both the effective and the virtual populations (Borges et al., 2019), one can obtain scaling laws for the mutation rates and selection coefficients (supplementary Text S1):

$$\phi_{a_i}^{N-1} = \phi_{a_i}^{*M-1} \quad \wedge \quad \mu_{a_i a_j} = \mu^*_{a_i a_j} \frac{\sum_{m=1}^{M-1} \phi_{a_i}^{*m-1} \phi_{a_j}^{*M-m-1} [m\phi^*_{a_i} + (M-m)\phi^*_{a_j}] \frac{M}{m(M-m)}}{\sum_{n=1}^{N-1} \phi_{a_i}^{n-1} \phi_{a_j}^{N-n-1} [n\phi_{a_i} + (N-n)\phi_{a_j}] \frac{N}{n(N-n)}} \quad , \tag{1}$$

where $\mu_{a_i a_j}$ is the mutation rate from allele a_i to a_j and ϕ_{a_i} the fitness coefficient of allele a_i on the effective dynamic, while $\mu^*_{a_i a_j}$ and $\phi^*_{a_i}$ are the corresponding parameters in the virtual one.

Here, we defined the virtual PoMos Two and Three, by setting the virtual population size M to two and three individuals, respectively. PoMoTwo has only one polymorphic state per allelic pair, i.e., $\{1a_i, 1a_j\}$, and is used to describe species evolution under neutrality (fig. 1) while still accounting for mutation bias. For the neutral case, the mutation rates are scaled by the ratio of harmonic numbers (denoted here by $H_.$) of the virtual and effective population sizes:

$$\mu_{a_i a_j} = \mu^*_{a_i a_j} \frac{M H_{M-1}}{N H_{N-1}} \quad . \tag{2}$$

PoMoThree includes two polymorphic states per allelic pair, i..e, $\{1a_i, 2a_j\}$ and $\{2a_i, 1a_j\}$, and additionally allows one to infer allelic selection (at least two polymorphic states are

Scenario	θ	$N\gamma$	τ
1	0.02	0.5	N
2	0.02	0.5	$10N$
3	0.02	1.5	N
4	0.02	1.5	$10N$
5	0.002	0.5	N
6	0.002	0.5	$10N$
7	0.002	1.5	N
8	0.002	1.5	$10N$

Table 1: **Description of the simulated scenarios.** θ represents the genetic diversity, γ the GC-bias coefficient (which relates to the GC-bias fitness by $\phi_S = 1 + \gamma$) and τ the time of the most recent common ancestor in the phylogeny (defined in units of generations).

necessary to make the selection coefficients identifiable; fig. 1). The scalings in equation (1) simplify to:

$$\phi_{a_i}^{N-1} = \phi_{a_i}^{*2} \quad \wedge \quad \mu_{a_i a_j} = \mu_{a_i a_j}^{*} \frac{3}{N} \frac{\phi_{a_i}^{*2} + \phi_{a_i}^{*}\phi_{a_j}^{*} + \phi_{a_j}^{*2}}{\phi_{a_i}^{N-1} + \sum_{n=1}^{N-2} \phi_{a_i}^{n} \phi_{a_j}^{N-n-1} [\frac{1}{N-n} + \frac{1}{n+1}] + \phi_{a_j}^{N-1}} \quad . \tag{3}$$

The virtual PoMoThree, in particular, allows quantifying signatures of selection and its impact on evolution.

Virtual population dynamics provide a good fit for their effective counterpart

To validate the concept behind the virtual PoMos, we assessed their fit and computational performance using simulated data. We simulated phylogenetic data sets with signatures of weak but pervasive nucleotide usage biases as observed in populations of great apes and fruit flies (Borges et al., 2019; Prado-Martinez et al., 2013; Begun et al., 2007). In particular, we simulated the typically observed mutation bias from G or C (also known as strong alleles) to A and T (also known as weak alleles) opposing gBGC (Glémin et al., 2015).

Mutation rates were assumed reversible, where the mutation rate is defined by the product of an exchangeability term and an allele frequency: e.g., the mutation rate from a strong (S) to a weak (W) allele is defined as $\mu_{SW} = \rho_{SW}\pi_W$. The four possible mutation rates are thus fully determined by three parameters: $\pi_W = 1 - \pi_S$, $\rho_{SW} = \rho_{WS}$ and $\rho_{SS} = \rho_{WW}$. The GC-bias preference was modeled by assigning a fitness of 1 to the weak alleles and a fitness $\phi_S = 1+\gamma$ to the strong alleles. A four-allelic Moran model was employed to simulate the evolutionary history of four populations across eight evolutionary scenarios by combining dynamics with high and low genetic diversity, weak and strong regimes of GC-bias, and shallow and deep divergence times. The specific parameters used in each simulated scenario are listed in table 1.

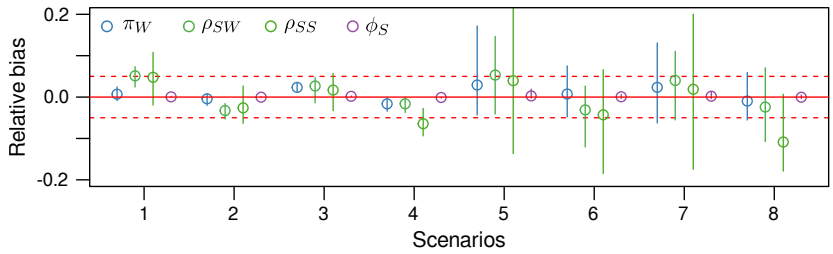

Figure 2: **Estimating mutation rates and selection coefficients with a virtual dynamic.** The simulated data from a four-allelic Moran dynamic with mutation bias and selection was fed to the virtual PoMoThree and the population parameters estimated using the Bayesian phylogenetic tool RevBayes (Höhna et al., 2016). As the population parameters were estimated in a population of three virtual individuals, we recalibrated them back to the effective population via the parameter scalings in equation (1). The estimated parameters are four: π_W, ρ_{SW} and ρ_{SS} are terms of the mutation rates that were assumed reversible (i.e., $\mu_{SW} = \rho_{SW}\pi_W$); ϕ_S is the fitness coefficient of the strong alleles (i.e., the GC-bias fitness). The average bias of the estimates was obtained based on alignments of 10^5 sites and 25 replicates per scenario. The specific parameters used in each simulated scenario are listed in table 1. The red dotted line represents a 5% relative error threshold.

Phylogenetic data from the eight simulated scenarios and the virtual PoMoThree model were employed to infer the mutation rates and the GC-bias fitness on the virtual population. These estimates were recalibrated back to the effective dynamic using equation (1), and the relative bias was calculated to evaluate their accuracy. The GC-bias fitness shows the highest accuracy, being estimated with a consistent relative error smaller than 1% across the eight simulated scenarios. This result demonstrates that PoMoThree is particularly suited to quantify selection bias on genomic sequences. For most of the simulated scenarios, the mutation rate parameters are reasonably well estimated, with an average relative error smaller than 5% (fig. 2). These analyses indicate that the virtual PoMos provide a good approximation for an effective dynamic with mutation and selection biases. We have also estimated the mutation rate parameters using the virtual PoMoTwo and observed that the equilibrium base composition of strong alleles is systematically overestimated (i.e., $\pi_S = 1 - \pi_W$), particularly in scenarios where GC-bias is stronger (supplementary fig. S1). This shows that the mutation rates try to accommodate the effect of selection towards the strong alleles by increasing the rate of weak to strong mutations (notice that $\mu_{WS} = \rho_{SW}\pi_S$).

We further compared the computational efficiency of the virtual PoMos, by comparing it to the standard general time-reversible nucleotide substitution model (GTR; Tavaré (1986)) in multiple sequence alignments, including 10 to 100 species and 100 to 1 million sites. For the three methods, similar settings for the phylogenetic inferences were employed in RevBayes (Höhna et al., 2016), including a standard MCMC chain with 50 000 generations and no parallelization. The computational performance of these three methods was assessed using three criteria: memory usage, CPU time, and average effective sample size of the sampled posterior parameters (ESS; i.e., the number of effectively independent draws from the posterior distribution that the Markov chain is equivalent to).

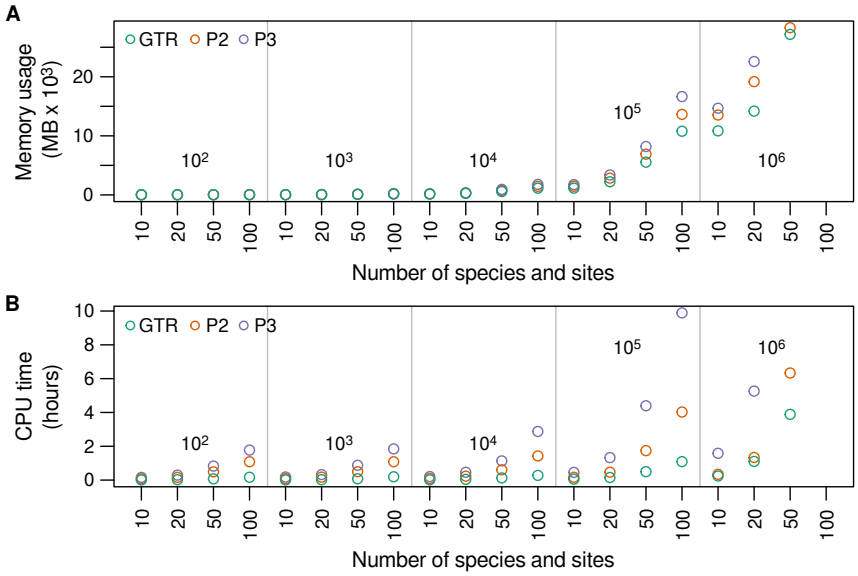

Figure 3: **Computational performance of the virtual PoMos.** The computational performance of the virtual PoMos Two and Three was compared with the general time-reversible model (GTR) based on **(A)** memory usage and **(B)** CPU time. The Bayesian inferences consisted of a standard MCMC chain with 50 000 generations and on simulated alignments with 10 to 100 species and 100 to 1 million sites, with levels of diversity similar to those of great apes (using the parameters defined for the simulated scenario 5). These analyses ran on an iMac desktop with an Intel Core i5 processor of 3.4 GHz of speed and 32GB of memory, without any kind of parallelization. Missing dots represent situations where the computer's maximum memory was reached, and RevBayes was not able to start the analyses.

Despite PoMoThree and Two having a larger state-space (16 and 10 states respectively) and therefore more possibilities for different site patterns at the tips, the memory usage is relatively similar among the methods (fig. 3A). The memory usage of PoMoThree is on average 16% higher than that of PoMoTwo and 39% higher than that of GTR. However, we observed that in some cases, RevBayes could not start the analyses because the memory reached the computers' maximum. Thus, while memory usage does not differ greatly among the methods, it certainly might be a limitation for genome-wide applications. Different from memory usage, the CPU time varies considerably among methods. We observed that the time consumed by each method increases exponentially with the alignment size (fig. 3B). PoMoThree requires, on average, 10.5 times more CPU time than the GTR and 2.3 times than PoMoTwo. This is because the PoMoThree has a larger state-space and consequently more site patterns to evaluate the likelihood. Regarding the average ESS, PoMoThree is the slowest mixing method with an average ESS of 2000 (supplementary table S1), 21.9% and 17.6% smaller than the average ESS of GTR and PoMoTwo, respectively.

Ignored selection affects species tree topology mildly

The impact of unaccounted-for selection on the reconstructed species tree can only be truly known by assessing the correctness of the estimation of both the topology and the branch lengths. Such aspects can be easily tested using simulated data on a known species tree. We used a species tree with four evolving populations, including two closely related populations and a third one recently diverging from the other two: `(A,(B,(C,D)))`. We set the evolutionary distances matching so that the first, second, and third divergence times are equal to τ, $\tau/2$ and $\tau/4$ respectively, where τ is the time of the most recent common ancestor. We thus expect higher levels of ILS when τ is equal to N generations than when it is equal to $10N$ (table 1).

To investigate the impact of selection on the species tree topology, we compared the trees inferred by the virtual PoMos (i.e., neutral PoMoTwo and the selection-aware PoMoThree) and the standard GTR model. If not stated otherwise, fixed and polymorphic counts from a sampled 10 individuals per population was used to perform the phylogenetic inferences with the virtual PoMos. As the GTR model only includes fixed states, a single random individual was used. The posterior probability of the correct topology was used to evaluate the accuracy of each method. Our results show that the virtual PoMos are more accurate than the GTR in all the tested scenarios, especially for alignments with fewer sites (fig. 4 and supplementary fig. S2). Because there are 15 different topologies, the expected posterior probability of the true tree if there is no signal in the data is around 6.7%, which is close to the performance of our methods in some conditions when there is little data. An interesting feature is that the topologies estimated by the GTR model tend to pick the first diverging branch incorrectly with greater probability than PoMos, especially for polymorphism-poor data sets (scenarios 1 and 5; fig. 4). This suggests that apart from being able to correct for ILS among closely related taxa, the PoMos are also superior in establishing deeper relationships. Despite that, we observed that the GTR's accuracy increases with the alignment size, showing that the GTR topologies should converge to the correct species tree topology with genome-scale data.

PoMoThree is only slightly more accurate than PoMoTwo, which indicates that neutral models may already provide fairly good estimates of the species tree topology (fig. 4). This is still valid when we increase the intensity of GC-selection (scenario 5 vs. 7; fig. 4)), suggesting that weak selection regimes should not greatly impact inferences of the relationships among species. Overall, our results confirm that the three methods perform similarly in estimating the species tree topology if enough genomic data is available. The specific number of sites necessary to return the correct topology will depend on several aspects, as the number of taxa and the complexity of the phylogenetic relationships among them. In any case, our results suggest that regardless of the employed method, better estimates of the species tree topology are expected for well-diverged taxa and polymorphism-rich data sets (scenario 1 vs. 2 and 7; fig. 4).

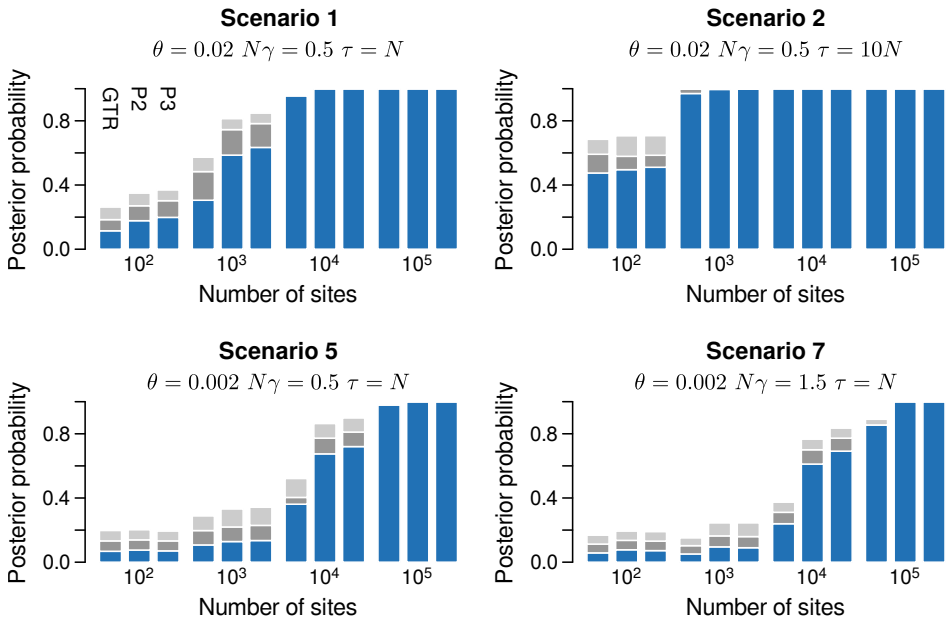

Figure 4: **Estimating the species tree topology with selection.** Data sets of different sizes (x-axis) and simulated under different scenarios were used to perform species tree inferences with three methods: GTR, PoMoTwo (P2 in the figure), and PoMoThree (P3 in the figure). We investigated the accuracy of the topology by assessing the posterior probability of the true topology (`(A,(B,(C,D)))`; blue sub-bars) under these three methods. The remaining probability corresponds to wrongly estimated topologies. The posterior probabilities of wrong topologies that are still able to correctly identify the first taxon to diverge are represented in shades of gray: `(A,(C,(B,D)))` in dark gray and `(A,(D,(C,B)))` in light gray. PoMoTwo and Three trees were inferred based on a sample of 10 individuals per taxa (totaling 40 sequences). The results for the eight scenarios can be found in supplementary fig. S2.

Similar conclusions can be drawn from when a sample of two individuals per population is taken instead of 10. Though the overall accuracy of PoMos decreases (supplementary fig. S3), we again observed that the posterior probability of the correct topology increases with the alignment size. This result is not surprising if we consider the results we have already discussed for the GTR model. Choosing a single virtual sequence to represent the whole population, as is often done when employing the standard nucleotide models with multi-individual data, corresponds to sample a single individual for the population (i.e., the situation where no polymorphisms are observed). Nevertheless, despite the dramatic reduction of the observed diversity, we observed that the GTR model can estimate the correct topology with enough genomic data.

Neglected selection distorts the measured divergence

We investigated the impact of selection on the branch lengths by comparing the genetic distances estimated by the virtual PoMos and the standard GTR model. To assess the accuracy of the branch length estimates, we evaluated the absolute error between the estimated

and expected pairwise distances among the four taxa. This strategy has the advantage of permitting comparing methods even if the estimated topologies are not exactly matching. Normalized distances were used for this comparison, as the GTR and PoMos operate in different evolutionary units: number of substitutions and Moran events (i.e., mutations plus frequency shifts), respectively.

Our results show that the error in the estimated branch lengths decreases as the size of the alignments increases (fig. 5 and supplementary fig. S4). Again, we observed that the GTR is the model of evolution producing the least accurate estimates, with an average error 2.5 times higher than PoMoThree, indicating that polymorphisms have a considerable impact on branch length estimation. PoMoThree outperforms PoMoTwo for all the observed scenarios, with PoMoTwo having an increased error of 61% on average. In some scenarios, PoMoThree has a negligible error (e.g., scenarios 2 and 8; Fig. 5 and supplementary figs. S4 and S5), demonstrating that the virtual PoMos provide a good fit for the divergence process operating in an effective dynamic with mutation and selection biases. If we only consider the scenarios where selection is more intense (i.e., $N\gamma = 1.5$), the error increases to 75%, indicating that ignored selection significantly contributes to bias the estimated divergence (e.g., scenarios 6 vs. 8; Fig. 5 and supplementary figs. S4 and S5).

We have also observed that the error on the estimated branch lengths does not decay to zero as fast as for the topology case, persisting in some cases even when the topologies are perfectly estimated (inferences with 10^5 sites in scenarios 1 and 7; fig. 5). Because of the sampling of individuals, polymorphisms in the original population may appear as fixed differences and provide misleading phylogenetic signal that could contribute to this slow decay. We observed that decreasing the sample size from 10 to two individuals increases the error of the estimated branch lengths by 75% on average in the virtual PoMos (supplementary fig. S5). As the patterns of diversity on the original and sampled populations are different, the models of evolution are unable to determine the exact contribution of mutations, fixations, and frequency shifts for the divergence process. The impact of such sampling bias should be particularly problematic among recently diverged populations, where several populations likely share segregating polymorphisms. Indeed, our simulations corroborate this expectation: scenarios 1 vs. 2 (fig. 5).

To establish that the superior performance of PoMoThree was not due to a larger state space with more polymorphic states, we have calculated the accuracy of the estimated topologies and branch lengths by employing the PoMoThree model with selection coefficients fixed to 0.0. We observed that this neutral PoMoThree has an intermediate accuracy between PoMoTwo and PoMoThree with selection (supplementary figs. S6 and S7), suggesting that the selection coefficients increase the fit of the model beyond that one would expect by using a finer space-space.

To further investigate the source of bias on the branch lengths, we tested whether the place-

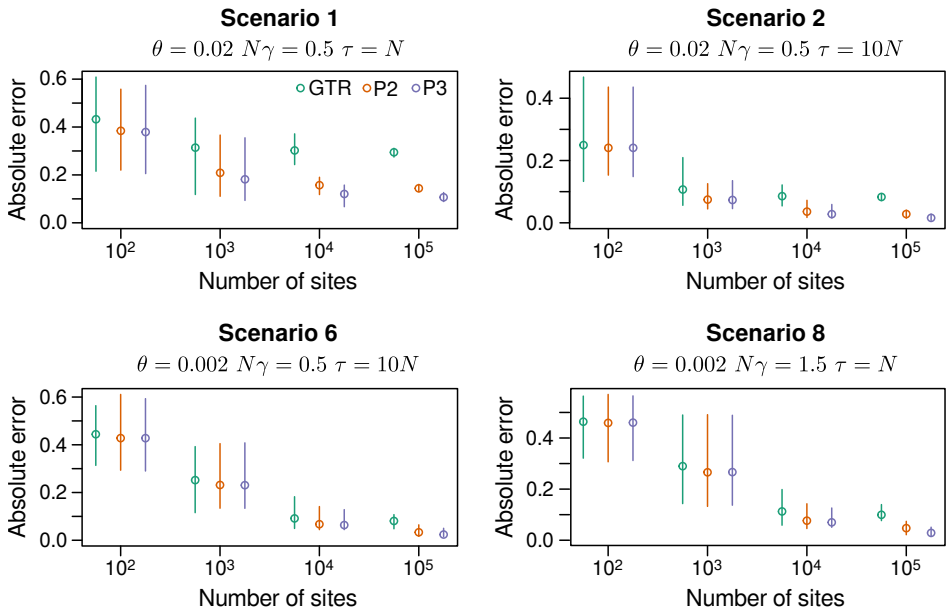

Figure 5: **Estimating the species tree branch lengths with selection.** Data sets of different sizes (x-axis) and simulated under different scenarios were used to perform species tree inferences with three methods: GTR, PoMoTwo (P2 in the figure), and PoMoThree (P3 in the figure). We investigated the accuracy of the branch lengths by calculating the absolute error of all six pairwise comparisons of genetic distances among the four taxa. As the three methods employ different evolutionary units, the branch lengths were normalized so that the three models could be compared. PoMoTwo and Three trees were inferred based on a sample of 10 individuals per taxa (totaling 40 sequences).

ment of the estimated divergence varies across evolutionary scales. To do that, we compare the height of each node on the true tree as estimated by the neutral PoMoTwo and the selection-aware PoMoThree models. For this comparison, we have only used node heights estimated under the 100 000 sites alignments, where both methods estimate the true topology correctly. We observed that PoMoTwo produces smaller node heights on average than PoMoThree (clouds always below the identity line in fig. 6A). This result suggests that weak selection signatures can mislead neutral models to estimate species histories with more recent divergence events. However, and more importantly, this bias is not uniform across nodes. The shallow nodes are less underestimated than the deeper nodes, with the ratio of node heights moving away from 1.0 from the shallow to the deeper nodes in all the tested scenarios (fig. 6A and supplementary figs. S8 and S9). The rate of evolution is known to be challenging to untie for deeper evolutionary scales (Xu and Yang, 2016); here, we show that ignored pervasive weak selection may contribute to it.

While our simulations intend to represent realistic scenarios, one might want to know what is the overall impact of selection (not necessarily nearly neutral) on the branch lengths. An advantage of PoMos is that some quantities can be formally obtained, such as the expected divergence per generation in diverse mutation-selection regimes (fig. 6B and supplementary text S2).

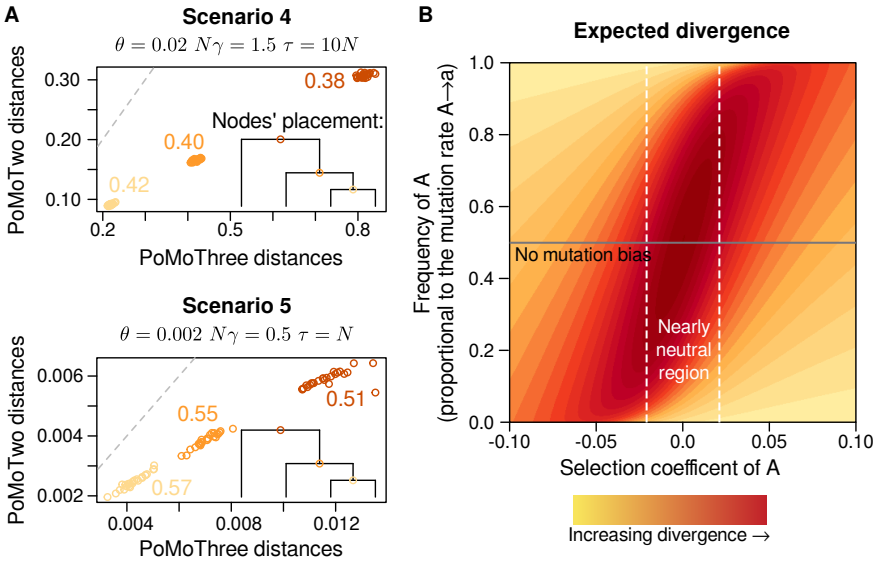

Figure 6: **Estimating divergence in the presence of selection. (A)** We investigated the association between the node heights as estimated by the virtual PoMos Two and Three. The phylogenetic tree depicts the three nodes' placement in the true phylogeny: the node heights double as we approach the root. Values accompanying each cloud denote the average ratio of the node heights in PoMoTwo to those in PoMoThree; these ratios are always smaller than 1.0 for all the other simulated scenarios (supplementary figs. S8 and S9). Each cloud represents 25 replicated alignments. The depicted node heights were all estimated using the 100 000 alignments, where both methods return the correct topology with maximum accuracy. **(B)** The expected divergence was calculated for different regimes of selection and mutation. For the biallelic case, the base composition π is proportional to the mutation bias: $\frac{\mu_{Aa}}{\mu_{aA}} = \frac{\pi_a}{\pi_A}$, where μ is the mutation rate. If the base composition is equal (gray line), there is no mutation bias (supplementary text S2).

We observed that the expected divergence is generally higher for the neutral case and tends to decrease as selection becomes more intense (fig. 6B; such trend is not qualitatively affected by the effective population size as shown in supplementary fig. S10). Directional selection fades alleles that could potentially be drifting in the population to fixation, overall reducing the expected divergence: i.e., for the same amount of time, we expect more homogeneous sequences if directional selection is acting. Consequently, to match the same patterns of diversity, selection-aware models tend to return larger branch lengths than the neutral models, corroborating the results obtained in fig. 6A. Furthermore, the underlying population dynamic strongly determines the magnitude of biases on the estimated branch lengths. For instance, the expected divergence is generally higher when selection counteracts mutation (i.e., rarer alleles are favored and *vice versa*, as is the case with gBGC) than when selection and the mutational bias act concordantly (i.e., rarer alleles are disfavoured and *vice versa*; fig. 6B).

Molecular dating with fruit fly populations

The geographic origins of the globally distributed fruit fly (*Drosophila melanogaster*) are still not fully understood. This is certainly not due to the lack of genomic data, as hundreds of sequenced genomes are available for the fruit flies (e.g., Hervas et al. (2017)), but primarily due to the methods of species tree estimation not scaling with multiple populations and genome-wide data. We employed the virtual PoMos to estimate and date the evolutionary history of 30 fruit fly populations. We used 966 sequences (supplementary table S2) from 1 million distantly located genomic sites, acknowledging the sites' independence assumption. We also assumed a global clock for molecular dating analyses, which is generally valid for short time scales. We set a uniform prior of 60 000 $\pm$ 15 000 years dating the African population expansion, as is suggested in the field's literature (e.g., Li and Stephan (2006); Stephan and Li (2007); Laurent et al. (2011)). The phylogenies estimated by the virtual PoMos agreed for most of the major clades (fig. 7), overall supporting a similar phylogeographic history for the fruit fly populations. Our analyses showed that the fruit fly populations have flourished in the southeast of the African continent. It is generally accepted that fruit flies originated in equatorial Africa (Lachaise et al., 1988; Stephan and Li, 2007). Our analyses suggested further that the species spread to the north, establishing in western equatorial Africa. From there, fruit flies colonized the western equatorial and the northern regions of Africa, with the northern wave leading to the spread of fruit flies worldwide. This phylogeography is not captured by the GTR model (supplementary fig. S11), suggesting high levels of ILS among fruit flies.

The time inferences are less consistent between the virtual PoMos, with PoMoThree providing, as expected by our formal results and simulations, more ancient node ages (around 23%; figs. 7 and supplementary figs. S12 and S13). Our divergence times are also comparatively more ancient than other studies. For example, some studies suggested a split of the African lineage at about 16 000 years ago (Stephan and Li, 2007; Laurent et al., 2011), while our analyses estimated it around 38 000 years ago. While unaccounted signatures of selection can underestimate the divergence times, we also notice that the non-monophyly of the out-of-Africa fruit fly populations (i.e., the Egyptian population (EG in fig. 7) clades together with these populations). This can cause such volatility in the estimated divergence times among studies.

An additional advantage of using selection-aware methods is that we can quantify selection coefficients. We measured the fruit flies' GC-bias rate at around 6.17×10^{-7} per site per generation, one order of magnitude lower than the reciprocal of their effective population size (approximately 1 million individuals according to Lynch et al. (2016)) and matching the expectations of gBGC. It remains unclear whether fruit flies have gBGC (Robinson et al., 2014). While it is difficult to think of any other process than gBGC favoring G and C alleles genome-wide, further studies associating these signatures with the recombination landscape

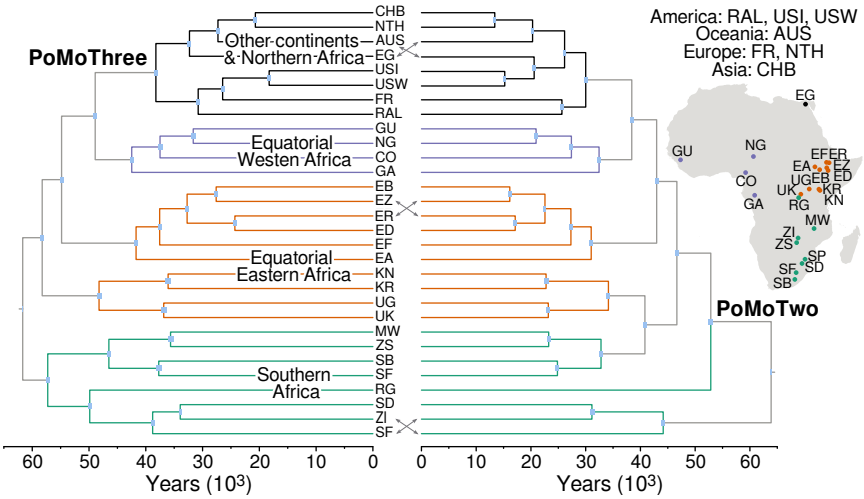

Figure 7: **Dating the evolutionary history of 30 worldwide fruit fly populations.** We dated the evolutionary history of 30 populations with 4 to 205 individuals (966 sequences in total, supplementary table S1) of *Drosophila melanogaster* distributed across the globe (data taken from Hervas et al. (2017)). The phylogenetic analyses were employed in RevBayes (Höhna et al., 2016), using alignments of 1 million distantly located genomic sites and the PoMoTwo and PoMoThree models of evolution, the latter being selection-aware. The gray lines between the trees indicate disagreement between the two models. The maximum *a posteriori* tree is depicted in both cases. The blue rectangles represent the 95% credibility intervals of the age of each node.

are still needed.

Discussion

Methods of species tree estimation have become an essential tool in studies of molecular evolution. Despite their widespread use, most of these methods assume neutrality. Here, we showed that unaccounted-for pervasive selection and selection-like processes significantly impact the inferences of the species tree. Our extensive analyses in both simulated and real data have two main implications for phylogenomic studies.

The first one is that assumed neutrality is likely to bias genetic distances in typical population genetic data sets. Modeling natural selection within the species evolution framework is not theoretically or computationally trivial, which has undoubtedly contributed to most tree inference methods making the neutral assumption. The virtual PoMos are full-likelihood methods that include selection and scale with genome-wide data from multiple populations and individuals. Other species tree inference methods, like the MSC, experience a data bottleneck for a few sequences and a handful of populations. Constraints between the species tree and the genealogical histories make it difficult to perform their joint inference. Furthermore, the space of unknown genealogies is large, and the genealogy samplers computationally intensive (Flouri et al., 2020; Rannala et al., 2020). In contrast, PoMos directly estimate the

species tree and naturally integrate over all possible genealogies. Hence, we expect that our methods will permit us to understand better the practical consequences of assuming neutrality in real contexts. Here, we have focused on the case of weak pervasive selection for populations with patterns of diversity similar to fruit flies and great apes. However, our results are likely to extend to several other biological groups, as GC-biased gene conversion is found all across the major groups of eukaryotes (Pessia et al., 2012).

The second one is that the bias created by ignoring selection is not uniform across nodes; it is more significant in the deeper nodes than in the shallow ones. Previous reports reasoned that directional selection should only mildly affect the rate of evolution (Edwards, 2009). By demonstrating that this effect is node-height dependent, we show that merely recalibrating the overall expected divergence is insufficient to correct divergence time inferences. Therefore, selection-aware models are particularly crucial for molecular dating or other phylogenetic analyses relying on the estimated branch lengths. By showing that one cannot ignore the net effect of selection in the evolutionary process (even in its weakest forms as tested here), we question the neutral assumption's suitability in species tree inference.

As the number of genomes sequenced increases rapidly, the development of methods describing the evolutionary process in all its complexity emerges as of fundamental importance. The methods presented here represent an important extension of existing approaches by allowing selection and permitting inference with large-scale data. In the future, we envision that our methods will be particularly useful for molecular dating using genome-wide data sets.

Materials and Methods

Simulations

As our methods aim at estimating species trees with genomic sequences, our simulations were designed to imitate general patterns of molecular evolution. In particular, we set a molecular dynamic with GC-bias favoring the fixation of G and C alleles (or strong alleles) and mutational bias to A and T alleles (or weak alleles) (Prado-Martinez et al., 2013; Begun et al., 2007). Mutation rates were assumed reversible and defined by the product of an exchangeability term and an allele frequency: e.g., the mutation rate from a strong to a weak allele is $\mu_{SW} = \rho_{SW}\pi_W$. Exchangeabilities between weak and strong alleles (i.e., $\rho_{SW} = \rho_{WS}$) were set three times more likely than between similar alleles (i.e., $\rho_{SS} = \rho_{WW}$). We have also set the frequency of the weak alleles to be higher than the strong alleles (i.e., $\pi_W = 0.7$ and $\pi_S = 0.3$). By combining the two exchangeabilities and allele frequencies, four mutation rates can be computed: $\mu_{SW} = \rho_{SW}\pi_W$, $\mu_{WS} = \rho_{SW}\pi_S$, $\mu_{SS} = \rho_{SS}\pi_S$ and $\mu_{WW} = \rho_{SS}\pi_W$. A fitness of $\phi_S = 1 + \gamma$ was set to the strong bases, where γ is the GC-bias rate, while the weak bases were assumed neutral with fitness 1.

We simulated the evolution of a population of four populations of 100 individuals using a four-variate Moran model with boundary mutations and allelic selection, according to the previously defined molecular dynamic. The mutation rates and the selection coefficient were rescaled accordingly. To make our simulation more realistic, we defined different evolutionary scenarios. Following the patterns of diversity observed in populations of fruit flies and great apes (Prado-Martinez et al., 2013; Begun et al., 2007)), we varied the levels of genetic diversity by setting $\theta = 0.02$ or 0.002. We added two regimes of GC-bias with scaled selection coefficient $N\gamma = 0.5$ or 1.5, both under the nearly neutral range (i.e., $N\gamma < 2$). We have also considered different divergence times by setting the time of the most recent common ancestor to $\tau = N$ or $10N$ generations. The evolution of four populations was simulated according to the following phylogenetic tree `(((pop4:`$\tau/4$`, pop3:`$\tau/4$`):`$\tau/4$`, pop2:`$\tau/2$`):`$\tau/2$`, pop1:`τ`);`. As some populations are closely related, we expect considerable levels of ILS for shorter divergence times. The simulated scenarios and respective parameters are listed in table 1. For each scenario, we produced alignments of 100 to 10^5 sites and 25 repetitions. We further generated two additional data sets by sampling 10 and 2 individuals per population. To assess the computational performance of the virtual PoMos, we simulated phylogenetic data for 10 to 100 populations and alignments of 100 to 1 million sites based on the parameters of scenario 5.

Because the simulations were performed in terms of the effective dynamic, the simulated data was converted to match the GTR, PoMoTwo, and PoMoThree state spaces. The fixed states in the original population $\{Na_i\}$ were directly converted to the fixed states $\{3a_i\}$ and $\{2a_i\}$ in the virtual PoMoThree and Two respectively. The polymorphic state $\{na_i, (N-n)a_j\}$ in the original population was directly converted to the state $\{1a_i, 1a_j\}$ PoMoTwo state, as this model has only one polymorphic state per polymorphic type. Differently, PoMoThree has two polymorphic states per polymorphic type. In this case, we used the allele frequency $f_{a_i} = n/N$ in the original state and attributed the state $\{2a_i, 1a_j\}$ with probability f_{a_i} and $\{1a_i, 2a_j\}$ with probability $1 - f_{a_i}$. As the GTR state-space only includes the four alleles, fixed states were set to the fixed allele $\{a_i\}$ while polymorphic ones to one of the alleles, $\{a_i\}$ or $\{a_j\}$, by randomly drawing according to the allele frequency. The state conversion for the GTR model corresponds to sample a single individual from the population.

Bayesian phylogenetic inferences

Phylogenetic trees were estimated using Bayesian inference in the RevBayes software (Höhna et al., 2016). The virtual PoMos are implemented in RevBayes and can be employed using the functions `fnReversiblePoMoTwo4N` and `fnReversiblePoMoThree4N`, which can be found in the RevBayes GitHub project (`https://github.com/revbayes`). Alternatively, the general function `fnReversiblePoMo4N` can also be used by setting the population size to two or three. In total, four free parameters were estimated during the Bayesian inferences: π_W,

ρ_{WS}, $\rho_{SS} = \rho_{WW}$ and ϕ_S. We set a Beta prior on π_W and an exponential prior on the exchangeabilities and the GC-bias fitness ϕ_S. We assumed a strict global clock rate for the molecular clock, drawn from an exponential prior with a root age fixed to 1.0. We fixed the root age to allow distance comparisons between the GTR and the virtual PoMos, as they all operate with different evolutionary units. A uniform time tree prior was set for the phylogeny. MCMC estimation was carried out with two chains with 50 000 generations for the GTR model and 75 000 for the virtual PoMos. The ESS was checked for convergence. The MCMC samples were accepted provided that the ESS of all model parameters was above 200. Similar priors were used for the GTR model, with the exception that the exchangeabilites were also set with a beta prior instead of an exponential one.

Application to fruit fly populations

Genome-wide data from 30 worldwide distributed populations of *Drosophila melanogaster* with 4 to 205 individuals (966 individuals) were extracted from the PopFly database (Hervas et al., 2017). We have only used genomic data from the 2R, 2L, 3R, and 3L autosomes. In total, one million genomic sites were randomly sampled for the phylogenetic inferences. The fruit flies allele counts were converted to the GTR, PoMoTwo, and Three state spaces following the same procedure as for the simulated data. The resulting multiple sequence alignments were then used to perform molecular dating analyses in RevBayes (Höhna et al., 2016). We employed a global clock model together with a uniform prior of 60 000 $\pm$ 15 000 years dating the African population expansion, following the results of Laurent et al. (2011). For the remaining parameters, we set the same priors as for the Bayesian phylogenetic inferences with the simulated data.

Data Availability Statement

The simulated data, RevBayes scripts, and fruit fly alignments generated and analyzed during the current study are available at Zenodo (`https://doi.org/10.5281/zenodo.5202827`).

Author Contributions

R.B., C.K., and G.S. designed research; R.B. and B.B. implemented the PoMo models in RevBayes; R.B. and C.K. analyzed data; R.B., C.K., G.S. and B.B. wrote and revised the paper.

Acknowledgements

The molecular dating analyses of the fruit fly data set have been achieved using the Vienna Scientific Cluster (VSC). This work was supported by the Vienna Science and Technology Fund (WWTF) [MA16-061]. GJS received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program under grant agreement no. 714774 and the grant GINOP-2.3.2.-15-2016-00057.

References

Adams, R. H., Schield, D. R., Card, D. C., and Castoe, T. A. (2018). Assessing the Impacts of Positive Selection on Coalescent-Based Species Tree Estimation and Species Delimitation. *Systematic Biology*, 67(6):1076–1090.

Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y.-P., Hahn, M. W., Nista, P. M., Jones, C. D., Kern, A. D., Dewey, C. N., Pachter, L., Myers, E., and Langley, C. H. (2007). Population Genomics: Whole-Genome Analysis of Polymorphism and Divergence in Drosophila simulans. *PLoS Biology*, 5(11):e310.

Borges, R., Szöllősi, G. J., and Kosiol, C. (2019). Quantifying GC-Biased Gene Conversion in Great Ape Genomes Using Polymorphism-Aware Models. *Genetics*, 212(4):1321–1336.

Castoe, T. A., de Koning, A. P. J., Kim, H.-M., Gu, W., Noonan, B. P., Naylor, G., Jiang, Z. J., Parkinson, C. L., and Pollock, D. D. (2009). Evidence for an ancient adaptive episode of convergent molecular evolution. *Proceedings of the National Academy of Sciences*, 106(22):8986–8991.

De Maio, N., Schlötterer, C., and Kosiol, C. (2013). Linking great apes genome evolution across time scales using polymorphism-aware phylogenetic models. *Molecular Biology and Evolution*, 30(10):2249–2262.

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.

Edwards, S. V. (2009). Natural selection and phylogenetic analysis. *Proceedings of the National Academy of Sciences*, 106(22):8799–8800.

Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., Zhong, B., Wu, S., Lemmon, E. M., Lemmon, A. R., Leaché, A. D., Liu, L., and Davis, C. C. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, 94:447–462.

Flouri, T., Jiao, X., Rannala, B., and Yang, Z. (2020). A bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution*, 37(4).

Galtier, N., Duret, L., Glémin, S., and Ranwez, V. (2009). GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates.

Galtier, N., Piganeau, G., Mouchiroud, D., and Duret, L. (2001). GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics*, 159(2):907–11.

Glémin, S. (2010). Surprising Fitness Consequences of GC-Biased Gene Conversion: I. Mutation Load and Inbreeding Depression. *Genetics*, 185(3):939–959.

Glémin, S., Arndt, P. F., Messer, P. W., Petrov, D., Galtier, N., and Duret, L. (2015). Quantification of GC-biased gene conversion in the human genome. *Genome Research*, 25(8):1215–1228.

Hervas, S., Sanz, E., Casillas, S., Pool, J. E., and Barbadilla, A. (2017). PopFly: the Drosophila population genomics browser. *Bioinformatics*, 33(17):2779–2780.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., and Ronquist, F. (2016). RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. *Systematic Biology*, 65(4):726–736.

Hughes, A. L. (2005). Evidence for Abundant Slightly Deleterious Polymorphisms in Bacterial Populations. *Genetics*, 169(2):533–538.

Hughes, A. L., Packer, B., Welch, R., Bergen, A. W., Chanock, S. J., and Yeager, M. (2003). Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proceedings of the National Academy of Sciences*, 100(26):15754–15757.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Kingman, J. (1982a). The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248.

Kingman, J. F. C. (1982b). On the genealogy of large populations. *Journal of Applied Probability*, 19(A):27–43.

Lachaise, D., Cariou, M.-L., David, J. R., Lemeunier, F., Tsacas, L., and Ashburner, M. (1988). Historical Biogeography of the Drosophila melanogaster Species Subgroup. In *Evolutionary Biology*, pages 159–225. Springer US, Boston, MA.

Lartillot, N. (2013). Phylogenetic Patterns of GC-Biased Gene Conversion in Placental Mammals and the Evolutionary Dynamics of Recombination Landscapes. *Molecular Biology and Evolution*, 30(3):489–502.

Lassalle, F., Périan, S., Bataillon, T., Nesme, X., Duret, L., and Daubin, V. (2015). GC-Content Evolution in Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. *PLOS Genetics*, 11(2):e1004941.

Laurent, S. J., Werzner, A., Excoffier, L., and Stephan, W. (2011). Approximate Bayesian Analysis of Drosophila melanogaster Polymorphism Data Reveals a Recent Colonization of Southeast Asia. *Molecular Biology and Evolution*, 28(7):2041–2051.

Leaché, A. D. and Oaks, J. R. (2017). The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 48(1):69–84.

Li, H. and Stephan, W. (2006). Inferring the demographic history and rate of adaptive substitution in Drosophila. *PLoS Genetics*.

Lynch, M., Ackerman, M. S., Gout, J.-F., Long, H., Sung, W., Thomas, W. K., and Foster, P. L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, 17(11):704–714.

Maddison, W. P. and Knowles, L. L. (2006). Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic Biology*, 55(1):21–30.

Nagylaki, T. (1983). Evolution of a Finite Population under Gene Conversion. *Proceedings of the National Academy of Sciences of the United States of America*, 80(20):6278–6281.

Nei, M. (2005). Selectionism and Neutralism in Molecular Evolution. *Molecular Biology and Evolution*, 22(12):2318–2342.

Ohta, T. (1973). Slightly Deleterious Mutant Substitutions in Evolution. *Nature*, 246(5428):96–98.

Ohta, T. (2002). Near-neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences*, 99(25):16134–16137.

Pessia, E., Popa, A., Mousset, S., Rezvoy, C., Duret, L., and Marais, G. A. B. (2012). Evidence for Widespread GC-biased Gene Conversion in Eukaryotes. *Genome Biology and Evolution*, 4(7):675–682.

Pollard, D. A., Iyer, V. N., Moses, A. M., and Eisen, M. B. (2006). Widespread Discordance of Gene Trees with Species Tree in Drosophila: Evidence for Incomplete Lineage Sorting. *PLoS Genetics*, 2(10):e173.

Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., Veeramah, K. R., Woerner, A. E., O'Connor, T. D., Santpere, G., Cagan, A., Theunert, C., Casals, F., Laayouni, H., Munch, K., Hobolth, A., Halager, A. E., Malig, M., Hernandez-Rodriguez, J., Hernando-Herraez, I., Prüfer, K., Pybus, M., Johnstone, L., Lachmann, M., Alkan, C., Twigg, D., Petit, N., Baker, C., Hormozdiari, F., Fernandez-Callejo, M., Dabad, M., Wilson, M. L., Stevison, L., Camprubí, C., Carvalho, T., Ruiz-Herrera, A., Vives, L., Mele, M., Abello, T., Kondova, I., Bontrop, R. E., Pusey, A., Lankester, F., Kiyang, J. A., Bergl, R. A., Lonsdorf, E., Myers, S., Ventura, M., Gagneux, P., Comas, D., Siegismund, H., Blanc, J., Agueda-Calpena, L., Gut, M., Fulton, L., Tishkoff, S. A., Mullikin, J. C., Wilson, R. K., Gut, I. G., Gonder, M. K., Ryder, O. A., Hahn, B. H., Navarro, A., Akey, J. M., Bertranpetit, J., Reich, D., Mailund, T., Schierup, M. H., Hvilsom, C., Andrés, A. M., Wall, J. D., Bustamante, C. D., Hammer, M. F., Eichler, E. E., and Marques-Bonet, T. (2013). Great ape genetic diversity and population history. *Nature*, 499(7459):471–475.

Rannala, B., Edwards, S. V. S. V., Leaché, A., and Yang, Z. (2020). The Multi-species Coalescent Model and Species Tree Inference. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, pages 3.3:1–3.3:21. No commercial publisher — Authors open access book.

Rannala, B. and Yang, Z. (2003). Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci. *Genetics*, 164(4):1645–1656.

Robinson, M. C., Stone, E. A., and Singh, N. D. (2014). Population Genomic Analysis Reveals No Evidence for GC-Biased Gene Conversion in Drosophila melanogaster. *Molecular Biology and Evolution*, 31(2):425–433.

Sawyer, S. A., Parsch, J., Zhang, Z., and Hartl, D. L. (2007). Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. *Proceedings of the National Academy of Sciences*, 104(16):6504–6510.

Schrempf, D., Minh, B. Q., De Maio, N., von Haeseler, A., and Kosiol, C. (2016). Reversible polymorphism-aware phylogenetic models and their application to tree inference. *Journal of Theoretical Biology*, 407:362–370.

Schrempf, D., Minh, B. Q., von Haeseler, A., and Kosiol, C. (2019). Polymorphism-Aware Species Trees with Advanced Mutation Models, Bootstrap, and Rate Heterogeneity. *Molecular Biology and Evolution*, 36(6):1294–1301.

Stephan, W. and Li, H. (2007). The recent demographic and adaptive history of Drosophila melanogaster. *Heredity*, 98(2):65–68.

Szöllősi, G. J., Tannier, E., Daubin, V., and Boussau, B. (2015). The Inference of Gene Trees with Species Trees. *Systematic Biology*, 64(1):e42–e62.

Tavaré, S. (1986). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lectures on Mathematics in the Life Sciences*, 17:57–86.

Xu, B. and Yang, Z. (2016). Challenges in Species Tree Estimation Under the Multispecies Coalescent Model. *Genetics*, 204(4):1353–1368.