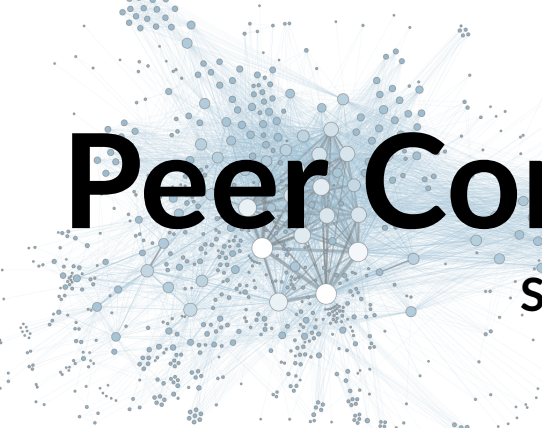




Peer Community Journal

Section: Mathematical & Computational Biology

Research article

Published
2026-09-24

Cite as

Cécile Ané and Paul Bastide
(2026) *Trait evolution with
incomplete lineage sorting and
gene flow: the Gaussian
Coalescent model*, Peer
Community Journal, 6: e96.

Correspondence

cecile.ane@wisc.edu
paul.bastide@u-pariscite.fr

Peer-review

Peer reviewed and
recommended by
PCI Mathematical &
Computational Biology,
<https://doi.org/10.24072/pci.mcb.100819>



This article is licensed
under the Creative Commons
Attribution 4.0 License.

Trait evolution with incomplete lineage sorting and gene flow: the Gaussian Coalescent model

Cécile Ané^{, #,1} and Paul Bastide^{, #,2}

Volume 6 (2026), article e96

<https://doi.org/10.24072/pcjournal.789>

Abstract

Most phylogenetic comparative methods use a species-level phylogeny, ignoring the effect of incomplete lineage sorting (ILS) and hemiplasy on the traits of interest. We consider here a trait controlled additively by one or more unknown loci. Their gene trees may differ from the species phylogeny due to ILS. The species phylogeny can be any phylogenetic tree or network or admixture graph on any number of taxa or populations. Our process of trait evolution accounts for introgression represented in the phylogeny by hybrid (or admixture) edges, which capture gene flow, admixture or hybridization events. The process models the trait in each individual, for any number of individuals from each taxon. Our model allows for polymorphism in the ancestral population at the root of the species phylogeny, and we can derive the heritable within-population variation expected from our model. Even if each locus evolves according to a Brownian motion, the joint distribution of the trait across all measured individuals is not generally Gaussian due to ILS. We provide a Gaussian approximation, named the Gaussian Coalescent (GC), and show how to compute its variance matrix efficiently using a single traversal of the species phylogeny. In simulations, this model is much more accurate than the model ignoring ILS. In simulations and on a data set of tomato floral traits, it is favored over the standard Brownian motion model with extra within-population variance. The GC model opens new avenues for various phylogenetic methods, accounting for hemiplasy and gene flow simultaneously on any general network or admixture graph. It is implemented in phylolm v2.7.0 and in PhyloTraits v1.2.0.

¹Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, United States,

²Université Paris Cité, CNRS, MAP5, F-75006 Paris, France, #Equal contribution



Contents

1	Introduction	2
2	Continuous trait evolution model with ILS.....	4
2.1	Polygenic evolution model on a phylogeny	4
2.2	Covariances on a species network for a single locus effect.....	6
2.3	Additivity and branch length transformation*	9
2.4	Phylogenetic model for the mean trait in populations*	11
2.5	The Gaussian-Coalescent model	13
2.6	Comparison with alternate approaches accounting for ILS*	15
3	Impact on comparative methods	19
3.1	Phylogenetic regression and implementation	19
3.2	Estimation under ILS and gene flow	20
3.3	Floral traits in the wild tomato clade.....	22
4	Discussion.....	24
	Acknowledgements.....	27
	Funding	27
	Conflict of interest disclosure	27
	Scripts and code availability	27
A	Graphical model and proof of covariance formulas	28
A.1	Graphical model.....	28
A.2	Proofs of main results	30
A.3	Mean population trait evolution on a network	33
B	Non-Gaussian joint distribution.....	34
C	Limits in small or large populations	35
D	Supplementary figures	36
D.1	Numerical experiments with <i>seastaR</i>	36
D.2	Simulations on a network	39
D.3	Simulations on the tomato tree	42
	References	44

1. Introduction

Phylogenetic comparative methods (PCMs) leverage the historical relationships between taxa of interest to discover patterns of trait evolution. They have a vast array of applications in evolutionary biology (Felsenstein, 2004; Harmon, 2019; Revell and Harmon, 2022). Traditionally, the evolution of a trait is modeled along the species phylogeny. Only recently was it recognized that spurious conclusions may arise in the presence of gene-tree discordance, when individual genes have genealogies that differ from the species phylogeny. For binary traits, hemiplasy (Avice and Robinson, 2008) is a source of trait incongruence relative to the species phylogeny, when a single transition on a discordant gene tree can explain the trait pattern whereas homoplasy would instead require several convergent transitions along the species phylogeny, either a tree (Guerrero and Hahn, 2018) or a network with reticulations (Hibbins et al., 2020; Wang et al., 2021). Incomplete lineage sorting (ILS) is one of the main identified biological causes for gene-tree discordance (Degnan and Rosenberg, 2009; Maddison, 1997; Schrempf and Szöllösi, 2020). Within a population, it can be modeled by the coalescent process (Kingman, 1982). When applied to each population of a phylogenetic tree or network, this process is typically called the multispecies coalescent (MSC, Degnan, 2018; Rannala and Yang, 2003). On a network, parental inheritance can be assumed independent, correlated or even shared across lineages in admixed populations (Fogg et al., 2023).

Continuous polygenic traits are controlled by a large number of genes, each of which being possibly affected by ILS. Classical models use a single stochastic process such as the Brownian motion (BM) on the species tree (Felsenstein, 1973) or network (Bastide et al., 2018). These models induce phylogenetic correlation, which can then be accounted for in statistical analyses such as linear regression or ANOVA, to avoid the detection of spurious effects if ignored (Felsenstein, 1985). However, traditional models do not account for the diverse histories of the gene trees underlying the trait. Mendes et al. (2018) pioneered the use of the MSC in this context. They assume that the measured trait X is the sum of the effects $Y^{(l)}$ of L genes, each following its own stochastic process on its own gene tree. Under the MSC, they derived theoretical formulas for the covariance matrix of such a trait measured at the tips of a phylogenetic tree, and provided a closed-form expression for triplets. Hibbins et al. (2023) built on this approach and implemented the computation of a triplet-based approximation for trees with branches scaled in coalescent units in the R package *seastaR*, and demonstrated its use on a wild tomato flower trait dataset. Their work underlined the importance of taking ILS into account in PCMs. Their method has some limitations: it requires one trait value per population, does not consider within-population variation, and cannot be used on a phylogenetic network. More importantly, it is based on conditioning the evolution of each effect $Y^{(l)}$ to its value at the root of its own gene tree, which can be arbitrarily far in the past. As we show below, this induces some unexpected behavior of the model, that, for a phylogeny with a given root population, can predict a different covariance between the trait of two populations A and B whether a third population C is included in the study or not.

In this work, we propose a general model for the evolution of a polygenic trait on a species network with ILS. We assume that the gene trees are distributed in the phylogeny according to the network multispecies coalescent process with independent inheritance of parental lineages at reticulations (Degnan, 2018; Fogg et al., 2023). As in Mendes et al. (2018), we assume that the continuous trait is the sum of the effects of many genes, each evolving on its own gene tree, but we use a different conditioning at a fixed point in time, at or before the root of the phylogeny, to ensure a consistent model across genes. We derive the covariance matrix of the univariate trait, and show that the distribution of the vector of measured traits converges under mild conditions to a multivariate Gaussian. We coin this distribution the “Gaussian-Coalescent” (GC). The computation of the GC covariance matrix involves one traversal of the phylogeny, and has an explicit closed form in the tree case. It depends on two parameters: the rate of evolution per coalescent unit σ_L^2 , and the trait variance in the ancestral population v_0 . While classical BM models only describe the evolution of the population mean (Felsenstein, 2004), we derive formulas for how both the mean and the variance of the trait within each population propagate along the phylogeny. This implies that the GC can directly give an expectation of the within-population variance on the phylogeny, that can be compared to the classical BM model with additional within-population variation (Teo et al., 2023). Further, our model recovers several classical frameworks: the standard BM when ILS disappears, a single population when ILS dominates, and the evolution of an allele frequency when considering a single locus and no mutation.

In the remainder, we most often refer to taxa as “populations”, but our model and theory equally applies to species-level phylogenies. We keep referring to “species tree” and “species networks” because these terms are frequently used in studies of the coalescent. We also refer to the genealogy of any locus as a “gene tree”, even for a single non-coding position in the genome, also to follow conventional terminology. The paper is organized as follows. In Section 2, we formally define our model of polygenic trait evolution under ILS (Sec. 2.1), give exact recursive formulas for computing its moments on a phylogenetic network (Sec. 2.2), show how the model can be re-written as a sum of independent Brownian motions on re-scaled phylogenies (Sec. 2.3), and explore its links to existing models in the admixture graph literature (Sec. 2.4), before introducing the Gaussian-Coalescent model (Sec. 2.5) and showing how it differs from previous PCM approaches (Sec. 2.6). Most of the mathematical proofs are in the Supplementary Material (SM). We then study the GC model in Section 3, detailing its implementation (Sec. 3.1),

its behavior on simulated data (Sec. 3.2) and on wild tomato floral traits (Sec. 3.3). Section 4 concludes with a discussion and some recommendations.

2. Continuous trait evolution model with ILS

In this section, we define a process-based model for the evolution of a polygenic trait on a phylogeny (Sec. 2.1), and then show how to compute the expectation vector and the variance matrix from this process on a phylogeny (Sec. 2.2). The core Gaussian-Coalescent (GC) model is summarized in Section 2.5. Sections 2.3, 2.4 and 2.6 (marked with *) describe some of the properties of the model and its links with previous approaches from population genetics and phylogenetic comparative methods, and can be skipped to focus on the definition and use of the model.

Throughout, we use the standard definition of the variance matrix: the variance of the trait measured on one individual is defined as the expected squared deviation from the expected value of the trait under the model, and the covariance between the trait of two individuals is the expected value of the product of their deviations from their respective expected value. The phylogenetic variance matrix (also called the variance - covariance matrix) contains the trait covariance for each pair of measured individuals.

2.1. Polygenic evolution model on a phylogeny

We start by describing a model for the evolution of a polygenic trait X within a single population, and then show how to extend the model to a phylogeny. We refer to Felsenstein (2004, Chap. 24) for background on quantitative characters.

2.1.1. Polygenic evolution model in one population. We assume that the trait is determined by L genetic loci, each having an infinitesimally small effect $Y^{(l)}$ and evolving according to the following process. For simplicity, we also assume haploid individuals, each carrying one copy of each locus.

Definition 1 (Lévy evolution of the effect of one locus l). Let l be a locus, with a gene genealogy G_l . We assume that G_l is drawn from the coalescent process (backwards in time) and, as in Mendes et al. (2018), that the evolution of the locus effect $Y^{(l)}$ along G_l (forward in time) is independent of the realization of the coalescent process that generated G_l , as under neutral evolution in the absence of selection. The effect $Y^{(l)}$ is assumed to evolve according to a Lévy process $\mathcal{D}_\theta$ with parameter θ along G_l , as in Landis et al. (2013). Lévy processes have independent stationary increments: for any $t, u \geq 0$, $Y_{u+t}^{(l)} - Y_u^{(l)}$ is independent of $Y_u^{(l)}$ and has distribution $\mathcal{D}_\theta(t)$. Then, within each edge in G_l , the change in $Y^{(l)}$ over t units of time has distribution $\mathcal{D}_\theta(t)$, and the process splits into two independent processes starting at the same value at each node of the tree G_l . We also assume that the process has mean 0 (centered) and finite variance σ_θ^2 at time 1, so that $\mathcal{D}_\theta(t)$ is also centered for any t , and with variance $\sigma_\theta^2 t$.

Below are two common examples of processes that satisfy Def. 1.

Example 1 (Brownian motion). The Brownian motion with variance rate σ^2 is the most common example of a centered Lévy process, with $\sigma_\theta^2 = \sigma^2$. A Brownian motion with a trend is also a Lévy process, but not centered.

Example 2 (Mutational model). A traditional mutational model assumes that changes occur in $Y^{(l)}$ via mutations, with mutation rate μ per locus per generation, and individual mutational effects (jumps in $Y^{(l)}$) drawn from a distribution with mean 0 and variance σ_m^2 (used in e.g., Schraiber and Landis, 2015). Then $Y^{(l)}$ evolves according to a compound Poisson process, which is a centered Lévy process with variance rate $\sigma_\theta^2 = \mu\sigma_m^2$.

Other examples include processes studied by Landis et al. (2013) like the variance-gamma process and the symmetric α -stable processes (including the Cauchy process), although some α -stable processes have infinite variance (at any time t), so that they do not enter in our present

setting (Schraiber and Landis, 2015). Assuming additive effects, this locus effect model leads to the following model for the trait X .

Definition 2 (Polygenic model in one population). We assume that each locus l follows the model in Def. 1, and has an effect $Y^{(l)}$ on the trait, acting additively across loci:

$$X = \sum_{l=1}^L Y^{(l)}.$$

We also assume loci are independent. This corresponds to free recombination between loci (for independent gene trees) and selective neutrality (for independent effects given the gene trees, e.g. Bulmer, 1971). Then the phenotype X is itself a Lévy process as a sum of independent Lévy processes, with a finite rate

$$\sigma_L^2 = L\sigma_\theta^2.$$

2.1.2. Phylogenetic model. We now turn to the case when several populations are observed. These populations are related by a given phylogeny $\mathcal{N}$, which may have reticulations, and we define a model for the evolution of the trait on this phylogeny.

Definition 3 (Phylogenetic model for one locus). Let $\mathcal{N}$ be a known phylogenetic network specifying the relationships between the populations from which individuals are sampled, such that each of its edges e has length ℓ_e in *coalescent units*. Let l be a given locus, with a gene genealogy G_l drawn from the network multispecies coalescent process. As in Def. 1, we assume that the evolution of the locus effect $Y^{(l)}$ along G_l is independent of the realization of the coalescent process that generated G_l . Further, we assume that there is a unique ancestral population. A standard choice is the most recent common ancestor (MRCA) of all sampled populations when $\mathcal{N}$ is a tree, or the lowest stable ancestor more generally (Steel, 2016). Without loss of generality, we may fix this ancestral population to be the root ρ of $\mathcal{N}$, possibly attached to a single “root edge”. The evolution of the locus effect on $\mathcal{N}$ is then as follows.

- At the root population ρ , the effect $Y^{(l)}$ of ancestral individuals in G_l are drawn independently from a common distribution $\mathcal{P}_0$, with mean m_0/L and variance v_0/L .
- Along the gene tree G_l , $Y^{(l)}$ evolves according to a centered Lévy process with finite variance as in Def. 1.
- The parameter $\sigma_{\theta_e}^2$ of the process may differ across branches of $\mathcal{N}$.

Remark 1 (Population size and time in generations). Since the phylogeny $\mathcal{N}$ has branch lengths in coalescent units, $\sigma_{\theta_e}^2$ is a variance rate per coalescent unit so that $\sqrt{\sigma_{\theta_e}^2 \ell_e}$ has the same unit as the trait being studied (for instance, if the trait is a corolla diameter in millimeters, then $\sigma_{\theta_e}^2 \ell_e$ would be in mm^2). If one has information on the haploid effective population size $2N(e)$ in branch e , then the branch length in generations can be recovered as $\kappa_e = 2N(e)\ell_e$, and the rate of evolution per generation is given by $\sigma_{\theta_e}^2/(2N(e))$. In the following, we make no assumption on effective population sizes, which can vary over time. The rate $\sigma_{\theta_e}^2$ can vary across branches, to account for an accelerating or decelerating rate of evolution per coalescent unit. If population sizes are known, different assumptions on the rate of evolution in number of generations can be used.

Combining Defs. 2 and 3, we get an evolutionary model for the measured polygenic trait X .

Definition 4 (Phylogenetic polygenic model). A trait X follows a phylogenetic polygenic model on a phylogenetic network $\mathcal{N}$ if it is the sum of independent additive effects $Y^{(l)}$ from L loci, each following the model of Def. 2 on $\mathcal{N}$. In particular, the distribution $\mathcal{P}_0$ of X in the root population ρ has mean m_0 and variance v_0 , and the rate of evolution on each branch e of $\mathcal{N}$ is given by $\sigma_{L,e}^2 = L\sigma_{\theta_e}^2$.

Definition 5 (Equilibrium variance ratio). We define the equilibrium variance ratio as

$$(1) \quad \lambda = \frac{v_0}{\sigma_{L,\rho}^2},$$

where $\sigma_{L,\rho}^2 = L\sigma_{\theta,\rho}^2$ is the equilibrium trait variance, from an ancestral lineage that evolved according to a Lévy process with variance $\sigma_{\theta,\rho}^2$ at each locus (Lynch and Hill, 1986). The root population ρ is at equilibrium when $\lambda = 1$.

To derive the expectation and variance matrix of the trait values from the process in Def. 4, we will need coalescence probabilities and expected times, as recalled below (from Mendes and Hahn 2018, Eq. (A.7) in Supplementary Material on dryad Mendes and Hahn, 2017).

Lemma 1 (Coalescent probability and shared time). *For a time ℓ in coalescent units, define:*

$$(2) \quad q(\ell) = 1 - e^{-\ell} \quad \text{and} \quad r(\ell) = 1 - \frac{1 - e^{-\ell}}{\ell}.$$

Two individuals from the same population coalesce within time ℓ with probability $q(\ell)$. In other words $q(\ell) = \mathbb{P}[T < \ell]$ where T is their random coalescent time. These two individuals share a common ancestor for an average of $\ell r(\ell)$ coalescent units within the past ℓ coalescent units, that is: $\mathbb{E}[(\ell - T)_+] = \ell r(\ell)$. Both $q(\ell)$ and $r(\ell)$ are between 0 and 1, are 0 at $\ell = 0$, and tend to 1 when $\ell = \infty$.

2.2. Covariances on a species network for a single locus effect

In this section, we derive explicit formulas to compute the covariance between the traits measured at various nodes and individuals in the phylogeny, assuming the model of Def. 3 to focus on a single locus effect Y . From Def. 4 and by additivity of variances for independent variables, the moments for the polygenic trait X simply follow. Compared to Mendes et al. (2018) and subsequent papers based on their calculations (e.g., Hibbins et al., 2023), we clarify the conditioning and parameters when taking expectations, variances and covariances, and we explicitly model ancestral polymorphism.

In this section, we state our results and provide informal justifications. The formal proofs rely on the definition of a graphical model, and are in SM Section A.

2.2.1. Covariances on a network for a single locus effect. We assume that Y evolves according to the model in Def. 3 on a species phylogeny $\mathcal{N}$, which may have reticulations. To specify the full variance-covariance matrix of Y , we use the following quantities.

Definition 6 (Expectation, variance and covariances). Take any two nodes u and v in the phylogeny $\mathcal{N}$. We write E_u the expectation and Φ_u the total variance of the trait Y of a random single individual in the given population u . The covariance between two different random individuals taken from populations u and v is denoted by $\Omega_{u,v}$ (when $u = v$, we take two different individuals from the same population u). To simplify notations, we omit ρ , but all expectations, variances and covariances are calculated conditional on the trait distribution in ρ .

These notations and their link to the coalescent are illustrated in Figure 1. The following result shows that $\mathbf{E}$, Φ and Ω can be computed recursively with one preorder traversal of the phylogeny.

Proposition 2 (Recursion formula). *Let $\mathcal{N}$ be a species phylogeny. For an edge e in $\mathcal{N}$, let ℓ_e denote its length in coalescent units. Assume that Y evolves according to the model of Def. 3 on $\mathcal{N}$, with increments of finite variance rate $\sigma_{\theta,e}^2$. The mean and variance of Y among individuals in ρ , the root population, are denoted by m_0 and v_0 . Then for any node u of $\mathcal{N}$,*

$$(3) \quad E_u = m_0.$$

At the root, $\Phi_\rho = v_0$ and $\Omega_{\rho,\rho} = 0$. Let u and v be two nodes in the network, such that v is not a descendant of u . Such is the case if v is listed before u in preorder. If u is a tree node with parent p and parent edge e then

$$(4a) \quad \Phi_u = \sigma_{\theta,e}^2 \ell_e + \Phi_p$$

$$(4b) \quad \Omega_{u,u} = \sigma_{\theta,e}^2 \ell_e r(\ell_e) + \Phi_p q(\ell_e) + \Omega_{p,p}(1 - q(\ell_e))$$

$$(4c) \quad \Omega_{u,v} = \Omega_{p,v}.$$

More generally, if u has m parents $p_1, \dots, p_m$, with edge $e_k = (p_k, u)$ of variance rate $\sigma_{\theta_k}^2$, length ℓ_k in coalescent units and inheritance γ_k , then

$$(5a) \quad \Phi_u = \sum_{k=1}^m \gamma_k \left(\sigma_{\theta_k}^2 \ell_k + \Phi_{p_k} \right)$$

$$(5b) \quad \Omega_{u,u} = \sum_{k=1}^m \gamma_k^2 \left(\sigma_{\theta_k}^2 \ell_k r(\ell_k) + \Phi_{p_k} q(\ell_k) + \Omega_{p_k, p_k} (1 - q(\ell_k)) \right) + 2 \sum_{k < k'} \gamma_k \gamma_{k'} \Omega_{p_k, p_{k'}}$$

$$(5c) \quad \Omega_{u,v} = \sum_{k=1}^m \gamma_k \Omega_{p_k, v}.$$

Proof. The formal proof is in SM Section A. Informally, in the tree case, for a single individual in a single population, (3) and (4a) directly follow from the assumption of centered Lévy increments with finite variance $\sigma_{\theta_e}^2$, while (4c) comes from the conditional independence structure given by the phylogeny along with the preorder of the nodes. In (4b), the three terms can be interpreted using Lemma 1. The first term is just the variance of the process $\sigma_{\theta_e}^2$ times $\ell_e r(\ell_e)$, which is the expected time that the two different individuals from population u share a common ancestor within the branch of length ℓ_e . Forward in time, the two individuals shared the same trait for this amount of time, and then split, following independent processes. If these two individuals do coalesce within the branch (backwards in time), then they inherit from a single individual in the parent population p with variance Φ_p . This happens with probability $q(\ell_e)$, giving rise to the second term. Finally, with probability $1 - q(\ell_e)$, the two individuals do not coalesce within the branch, and inherit from the covariance $\Omega_{p,p}$ of two distinct individuals within parent population p , which gives the third term. General formulas (5) have similar interpretations, but allow for node u to have several parent populations. $\square$

The next proposition states that the recursion formulas of Prop. 2 are robust to sampling: the covariance between two individuals sampled from two populations does not depend on whether a third individual is sampled or not. On the network in Figure 1 for example, $\Omega_{A,B}$ is the same whether E is sampled or not. This property, which directly stems from our process-based definition of the model, makes it different from the formulas in Mendes et al. (2018), that are not robust to sub-sampling in general (see Section 2.6).

Proposition 3 (Sampling stability). *Assume a model on a network $\mathcal{N}$ such that the covariances are given by Prop. 2 above. For a taxon t , define $\mathcal{N}_{-t}$ the pruned subnetwork of $\mathcal{N}$ with t removed, but keeping the same root (possibly leaving the root connected by a single branch to the most recent stable ancestor of the remaining leaves) and not suppressing degree-2 nodes. Then, for any population nodes u and v that remain in $\mathcal{N}_{-t}$ (possibly with $u = v$), Φ_u and $\Omega_{u,v}$ given by Prop. 2 are the same when the recursion is applied to $\mathcal{N}_{-t}$ as when it is applied to the full network $\mathcal{N}$. Further, if the rate σ_{θ}^2 is constant across branches, then the non-root degree-2 nodes can be suppressed from $\mathcal{N}_{-t}$ without altering the distribution of the remaining nodes.*

The proof of this result is presented SM Section A. Note that it is generally useful to allow for degree-2 nodes in the phylogeny to allow for changes in evolutionary parameters, such as bottlenecks or other changes in population sizes, or changes in the rate of phenotypic evolution. These parameters $N(e)$ and $\sigma_{\theta_e}^2$ can vary across edges but are fixed within an edge e . To model variation in N or σ_{θ}^2 within an edge e from $u \rightarrow v$, one may subdivide e into edges e_1 from $u \rightarrow w$ and e_2 from $w \rightarrow v$, where w is a new node of degree-2. This brings more flexibility in the model, by allowing for $N(e_1) \neq N(e_2)$ or $\sigma_{\theta_{e_1}}^2 \neq \sigma_{\theta_{e_2}}^2$.

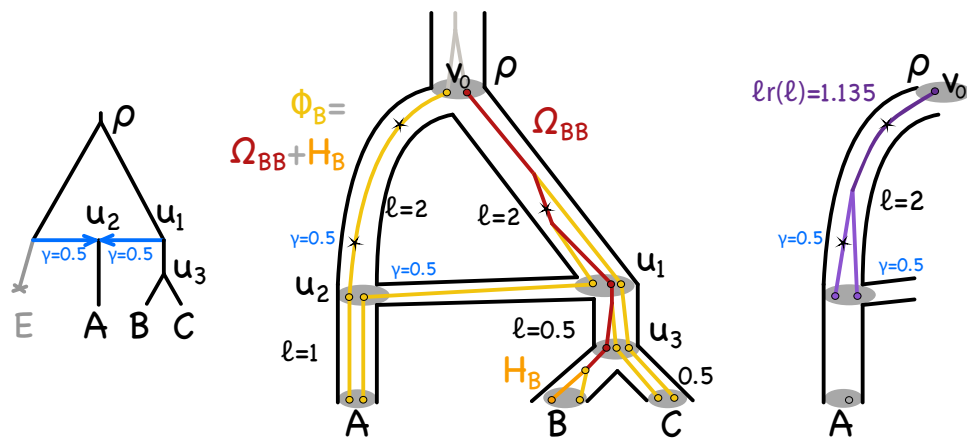


Figure 1 – Network to illustrate the variances Φ_u (for 1 random individual from u) and covariances $\Omega_{u,v}$ (for 2 random individuals from u and v) conditional on the ancestral population ρ . Expectations and variances are taken over all the possible gene trees.

Left: population history with u_2 arising from hybrid speciation with proportion $\gamma = 0.5$ of its genome from each parent, and E either extinct or unsampled.

Middle: phylogenetic network (black) for A , B and C with 2 individuals sampled from each, and an example gene tree for one sampled locus (colored). The locus effect evolves on the gene tree according to a centered Lévy process (Def. 1), such as a compound Poisson process with mutations (black stars) appearing at rate μ and adding an effect of variance σ_m^2 on the trait, with total variance rate $\sigma^2 = \mu\sigma_m^2$. Ancestral populations, in preorder, are ρ , u_1 , u_2 and u_3 . $\Omega_{B,B}$ is the expected time of shared ancestry for 2 individuals randomly sampled from B (in red in the example gene tree) up to ρ , scaled by the evolutionary rate σ^2 , plus a term that depends on the root variance v_0 . Φ_B is the scaled expected time from a random individual in B up to ρ , plus v_0 . The difference $H_B = \Phi_B - \Omega_{B,B}$ is the expected within-population variance within B (Lemma 7), and is the scaled expected time between the individual in B and the first coalescent event (orange line), plus a term that depends on v_0 . In the example gene tree, the 2 individuals from A do not coalesce along the lineage from u_2 to A , and their ancestors in u_2 are inherited from different parental populations.

Right: alternative scenario, in which both individuals in u_2 are inherited from the same parental population. The formulas involve $\ell_r(\ell)$, the expected time of shared ancestry within a lineage of ℓ coalescent units (dark purple in the example gene tree).

Example: using the recursive (co)variance calculations on the network in Figure 1

We illustrate Prop. 2 assuming a constant evolutionary rate σ_θ^2 across branches.

node	parents (parameters)	variance	covariances
ρ	initialize	$\Phi_\rho = v_0$	$\Omega_{\rho,\rho} = 0$
u_1	ρ ($\ell = 2$)	$\Phi_1 = 2\sigma_\theta^2 + v_0$	$\Omega_{1,1} = 2r(2) \sigma_\theta^2 + q(2) \Phi_\rho$ $= 1.135 \sigma_\theta^2 + 0.865 v_0$
	use (4)		$\Omega_{1,\rho} = 0$
u_2	ρ ($\gamma = 0.5, \ell = 2$)	$\Phi_2 = 0.5(2\sigma_\theta^2 + \Phi_\rho)$	$\Omega_{2,2} = 0.568 \sigma_\theta^2 + 0.432 v_0$
	u_1 ($\gamma = 0.5, \ell = 0$) use (5)	$+ 0.5\Phi_1$ $= 2\sigma_\theta^2 + v_0$	$\Omega_{2,1} = 0.568 \sigma_\theta^2 + 0.432 v_0$ $\Omega_{2,\rho} = 0$
u_3	u_1 ($\ell = 0.5$)	$\Phi_3 = 2.5 \sigma_\theta^2 + v_0$	$\Omega_{3,3} = 1.582 \sigma_\theta^2 + 0.918 v_0$
	use (4)		$\Omega_{3,2} = \Omega_{1,2} = 0.568 \sigma_\theta^2 + 0.432 v_0$ $\Omega_{3,1} = \Omega_{3,\rho} = 0$
A	u_2 ($\ell = 1$)	$\Phi_A = 3 \sigma_\theta^2 + v_0$	$\Omega_{A,A} = 1.841 \sigma_\theta^2 + 0.791 v_0$
	use (4)		$\Omega_{A,3} = \Omega_{A,2} = 0.568 \sigma_\theta^2 + 0.432 v_0$ $\Omega_{A,1} = \Omega_{A,\rho} = 0$
B	u_3 ($\ell = 0.5$)	$\Phi_B = 3 \sigma_\theta^2 + v_0$	$\Omega_{B,B} = 2.050 \sigma_\theta^2 + 0.950 v_0$
	use (4)		$\Omega_{B,A} = \Omega_{B,2} = 0.568 \sigma_\theta^2 + 0.432 v_0$ $\Omega_{B,3} = 1.582 \sigma_\theta^2 + 0.918 v_0$ $\Omega_{B,1} = \Omega_{B,\rho} = 0$
C	u_3 ($\ell = 0.5$)	same as B	and $\Omega_{C,B} = 1.582 \sigma_\theta^2 + 0.918 v_0$

Regardless of σ_θ^2 and v_0 , we get that $\Omega_{A,A} < \Omega_{B,B}$. This is because A is below a reticulation: 2 individuals in A have reduced covariance as one might inherit from gene flow and the other not. From Lemma 7, this implies that the within-population variance is expected to be greater in A than in B . Here we obtain $\Phi_A = \Phi_B = \Phi_C$ because the network is time-consistent (all paths from ρ to A have the same length) and ultrametric (all paths from ρ to A , B or C have the same length).

2.3. Additivity and branch length transformation*

In this section, we explore two interesting properties of the model, and show that it can be seen as a sum of two independent Brownian motions on phylogenies with modified branch lengths. This section can be skipped to focus on the definition and use of the model.

2.3.1. Additivity of mutation and drift. We can decompose the effect of one locus Y as the sum of two independent processes: one representing trait evolution through molecular change after ρ (modeled by a Lévy process), and one representing pure drift through the coalescent. For example, all terms calculated in the example on network in Figure 1 are linear combinations of σ_θ^2 (the Lévy contribution) and of v_0 (from drift and ancestral polymorphism in ρ). Indeed, the recursion formulas of Prop. 2 can be re-written as follows.

Corollary 4 (Additivity). *Using the same model and notations as in Prop. 2, we can decompose*

$$Y = Y^{(\mathcal{L})} + Y^{(0)}$$

with $Y^{(\mathcal{L})}$ and $Y^{(0)}$ independent, $Y^{(\mathcal{L})}$ the process following Def. 3 with $m_0 = 0$ and $v_0 = 0$, and $Y^{(0)}$ the process following Def. 3 with the same m_0 and v_0 as Y , but $\sigma_{\theta_e}^2 = 0$ on all branches. In other words, $Y^{(\mathcal{L})}$ captures the evolutionary Lévy process after ρ , while $Y^{(0)}$ only captures the effect of the ancestral population through pure drift.

Then $Y^{(\mathcal{L})}$ has expectation $\mathbf{0}$ and variances $\Phi^{(\mathcal{L})}$ and $\Omega^{(\mathcal{L})}$; $Y^{(0)}$ has expectation $m_0 \mathbf{1}$ and variances $\mathbf{1}$ and $\Omega^{(0)}$; and we have:

$$(6) \quad \begin{cases} \Phi = \Phi^{(\mathcal{L})} + v_0 \mathbf{1} \\ \Omega = \Omega^{(\mathcal{L})} + v_0 \Omega^{(0)} \end{cases}$$

with $\mathbf{1}$ the vector of ones. At the root $\Phi_\rho^{(\mathcal{L})} = \Omega_{\rho,\rho}^{(\mathcal{L})} = \Omega_{\rho,\rho}^{(0)} = 0$. Then we have by preorder recursion:

$$(7a) \quad \Phi_u^{(\mathcal{L})} = \sum_{k=1}^m \gamma_k \left(\sigma_{\theta_k}^2 \ell_k + \Phi_{p_k}^{(\mathcal{L})} \right)$$

$$(7b) \quad \Omega_{u,u}^{(\mathcal{L})} = \sum_{k=1}^m \gamma_k^2 \left(\sigma_{\theta_k}^2 \ell_k r(\ell_k) + \Phi_{p_k}^{(\mathcal{L})} q(\ell_k) + \Omega_{p_k,p_k}^{(\mathcal{L})} (1 - q(\ell_k)) \right) + 2 \sum_{k < k'} \gamma_k \gamma_{k'} \Omega_{p_k,p_{k'}}^{(\mathcal{L})}$$

$$(7c) \quad \Omega_{u,v}^{(\mathcal{L})} = \sum_{k=1}^m \gamma_k \Omega_{p_k,v}^{(\mathcal{L})}$$

and

$$(8a) \quad \Omega_{u,u}^{(0)} = \sum_{k=1}^m \gamma_k^2 \left(q(\ell_k) + \Omega_{p_k,p_k}^{(0)} (1 - q(\ell_k)) \right) + 2 \sum_{k < k'} \gamma_k \gamma_{k'} \Omega_{p_k,p_{k'}}^{(0)}$$

$$(8b) \quad \Omega_{u,v}^{(0)} = \sum_{k=1}^m \gamma_k \Omega_{p_k,v}^{(0)}.$$

Proof. The additivity and recursion formulas directly stem from Equations (5a)-(5c) of Prop. 2. $\square$

This decomposition means that we can separate the evolution process of Def. 3 into the two independent contributions of the trait evolution $Y^{(\mathcal{L})}$ and of the random drift $Y^{(0)}$. Also, (8) clarifies that v_0 impacts the model via $\Omega^{(0)}$, which solely depends on the network's branch lengths in coalescent units.

2.3.2. Generic branch length transformation. In traditional PCMs, models such as the Ornstein-Uhlenbeck or the Early Burst are commonly fitted using an equivalent BM on a tree with modified branch lengths (Ho and Ané, 2014). In this section, we show that our process on a network can also be re-written as the sum of two Brownian motions on modified phylogenies. Although the general formulas derived here have no direct computational use, they are used in Section 2.4 to link the model to a population genetics framework, and in Section 3.1 for an efficient implementation when the phylogeny is a tree.

Ignoring ILS, the Brownian motion with constant variance rate σ^2 can be used along a species network (e.g., Pickrell and Pritchard, 2012). Under the BM model, the vector of traits at the tips is Gaussian, with a variance-covariance matrix $\sigma^2 \mathbf{C}$ that can be computed by traversing the tree in a preorder as (Bastide et al., 2018):

$$(9) \quad \begin{cases} C_{uu} = \sum_{k=1}^m \gamma_k^2 (C_{p_k p_k} + \ell_k) + 2 \sum_{k < k'} \gamma_k \gamma_{k'} C_{p_k p_{k'}} \\ C_{uv} = \sum_{k=1}^m \gamma_k C_{p_k v} \end{cases}$$

for any node u with parent nodes p_k , parental branches of length ℓ_k and inheritance probabilities γ_k ($1 \leq k \leq m$), and v any node that is above u in the preorder. These propagation equations bear similarities with the equations of Cor. 4. Indeed, we can see the variance matrices $\Omega^{(\mathcal{L})}$ and $\Omega^{(0)}$ as formally emerging from two different Brownian motions on the network, with rescaled branch lengths (see SM Section A for the proof):

Proposition 5 (BM on a rescaled network). *The variance matrix $v_0 \Omega^{(0)}$ on network $\mathcal{N}$ is the same as under the BM with variance rate v_0 on a network with $\mathcal{N}$'s topology and rescaled branch lengths:*

$$(10) \quad L_e^{(0)} = q(\ell_e) (1 - \Omega_{p,p}^{(0)}),$$

for edge e with parent node p . Similarly, assuming a constant rate σ_θ^2 across branches for simplicity, so that the scaled covariances and variances $\tilde{\Omega}^{(\mathcal{L})} := \Omega^{(\mathcal{L})}/\sigma_\theta^2$ and $\tilde{\Phi}^{(\mathcal{L})} := \Phi^{(\mathcal{L})}/\sigma_\theta^2$ do not depend on σ_θ^2 , then $\Omega^{(\mathcal{L})}$ is the same covariance structure as under a BM with variance rate σ_θ^2 and rescaled branch lengths:

$$(11) \quad L_e^{(\mathcal{L})} = \ell_e - q(\ell_e) (1 - \tilde{\Phi}_p^{(\mathcal{L})} + \tilde{\Omega}_{p,p}^{(\mathcal{L})}).$$

Further, both transformations (10) and (11) are additive: for any nodes $u \rightarrow v \rightarrow w$ along one lineage such that v is a tree node, then $L_{uw} = L_{uv} + L_{vw}$, where L_{uv} is the transformed branch length of the combined edge going from u to w .

The computation of the rescaled branch lengths involves one preorder traversal of the tree, but these lengths do not depend on the parameters σ_θ^2 or v_0 . This transformation cannot be used in general as a computation technique (except in some special cases explored below), but ensures that the structure of the covariance matrix can be framed as emerging from a BM on the phylogeny. Note that the constant rate assumption is only technical. It leads to rescaled branch lengths $L_e^{(\mathcal{L})}$ (and $\tilde{\Omega}^{(\mathcal{L})}$) that do not depend on σ_θ^2 . In SM Section A (Prop. S3), we describe a branch length transformation that depends on the rates $\sigma_{\theta_e}^2$ but can still cast $\Omega^{(\mathcal{L})}$ as emerging from a BM with varying rates on the phylogeny.

2.3.3. Special case of a phylogenetic tree. When the phylogeny is a tree, the recursion formulas and branch length transformations lead to closed-form expressions.

Proposition 6 (Closed-form expression in the tree case). *When the phylogeny is a tree, using the same model and notations as in Cor. 4, we get, for any two nodes u and v :*

$$(12) \quad \Omega_{uv}^{(0)} = (1 - e^{-t_{uv}})$$

where t_{uv} is the length of the unique path between the root and the MRCA of u and v (in coalescent units). Using (10), this leads to the following explicit branch length transformation:

$$(13) \quad L_u^{(0)} = q(\ell_u) e^{-t_{pa(u)}},$$

where ℓ_u and $L_u^{(0)}$ are the original and transformed lengths of the branch going into u , and $t_{pa(u)}$ is the length of the path between the root and the parent of u . Similarly, assuming for convenience a constant rate σ_θ^2 on all branches:

$$(14a) \quad \Phi_u^{(\mathcal{L})} = \sigma_\theta^2 t_u$$

$$(14b) \quad \Omega_{u,v}^{(\mathcal{L})} = \sigma_\theta^2 (t_{uv} - (1 - e^{-t_{uv}})).$$

Using (11), this leads to the following explicit branch length transformation:

$$(15) \quad L_u^{(\mathcal{L})} = \ell_u - q(\ell_u) e^{-t_{pa(u)}}.$$

The proof is presented in SM Section A. Note that both transforms yield positive lengths, as $0 \leq 1 - e^{-x} \leq x$ for any $x \geq 0$.

2.4. Phylogenetic model for the mean trait in populations*

This section explores links between our model and established models in population genetics, e.g. for allelic frequencies. It can be skipped to focus on the practical use of the model.

2.4.1. Mean population trait evolution on a network. The previous section tracks the trait expectation and variance in individuals sampled from each population. Another way to describe the model is to track the evolution of the trait distribution in a population, summarized by its mean (across all individuals in the population) and its within-population variance.

More precisely, define M_u as the average trait among all individuals in population u for a given locus and V_u its variance among individuals in u , that is, V_u is the within-population variance. Also define the expected within-population variance as $H_u = \mathbb{E}[V_u]$ at node u . Then the moments of Y are linked to those of M_u and V_u by the following formulas.

Lemma 7 (Population mean and variance). Write $\mathbf{M}$ and $\mathbf{V}$ the vectors of trait population means and variances at all nodes as above. Then:

$$(16a) \quad \text{Var} [\mathbf{M}] = \mathbf{\Omega}$$

$$(16b) \quad \mathbf{H} = \mathbb{E} [\mathbf{V}] = \mathbf{\Phi} - \text{diag}(\mathbf{\Omega})$$

where $\text{diag}(\mathbf{\Omega})$ is the vector made of the diagonal elements of $\mathbf{\Omega}$.

The proof of this lemma is in SM Section A. By (16a), the covariance $\text{Cov} [M_u; M_v]$ between the trait means of two populations is equal to the covariance $\Omega_{u,v}$ between two different individuals randomly drawn from these populations. From (16b) we see that Φ_u , the trait variance of an individual from a population u , is the sum of the covariance $\Omega_{u,u}$ between two individuals from u plus the expected within-population variance of the trait H_u .

2.4.2. Drift model for allele frequencies. A special case of interest is when Y represents the genotype of an individual at a given locus, with $Y = 0$ if the individual has the major allele, and 1 otherwise. In general, a classical Markov model of evolution for Y is not a Lévy process, as the increments are not independent. However, the special case with no mutations along the network can be represented with our model, setting $\sigma_\theta^2 = 0$. This special case corresponds to the “infinite site assumption” by which any new mutation affects a new locus so each locus displays at most one mutated allele, and further assuming that all mutations occurred prior to the root population ρ . Then, using notations from Cor. 4, $Y = Y^{(0)}$ from random drift due to ILS. The minor allele frequency p_u in population u is then the average of Y over all individuals in u , which is exactly M_u . As Y can only be 0 or 1, $Y^2 = Y$ and the population variance has a simple expression:

$$V_u = p_u(1 - p_u)$$

so that $2V_u$ is equal to the heterozygosity in population u . In that case, the expectation and variance in the root population are $m_0 = p_0$ and $v_0 = p_0(1 - p_0)$, and the covariance matrix between allele frequencies at all nodes is given by the matrix $p_0(1 - p_0)\mathbf{\Omega}^{(0)}$, as computed from Cor. 4. The total variance $\mathbf{\Phi}$ is constant equal to $p_0(1 - p_0)$, while the expected half heterozygosity is $H_u = \mathbb{E} [p_u(1 - p_u)] = p_0(1 - p_0)(1 - \Omega_{u,u}^{(0)})$. The expected f_2 statistic used in population genetics can then be expressed, on a tree edge e going from v to u , as:

$$F_2(u, v) = \mathbb{E} [(p_u - p_v)^2] = \Omega_{u,u}^{(0)} + \Omega_{v,v}^{(0)} - 2\Omega_{u,v}^{(0)} = (1 - e^{-\ell_e})\mathbb{E} [p_v(1 - p_v)].$$

We recover the formulas from Patterson et al. (2012) when conditioning on p_v . If the phylogeny is a tree, $\mathbf{\Omega}^{(0)}$ is also identical to the covariance used in Mary-Huard and Balding (2023).

These equations correspond to the population genetics framework that aims at recovering admixture graphs from frequencies of biallelic markers, using tools such as TreeMix (Pickrell and Pritchard, 2012) or admixtools (Maier et al., 2023; Patterson et al., 2012). In particular, if we assume that the minor allele frequencies $(p_u)_u$ are jointly Gaussian, and that branch lengths in coalescent units are small so that $1 - q(\ell) \approx 1$ for all branches, then the propagation equations from Cor. 4 for the covariance become:

$$\Omega_{u,u}^{(0)} \approx \sum_{k=1}^m \gamma_k^2 \left(\Omega_{p_k,p_k}^{(0)} + q(\ell_k) \right) + 2 \sum_{k < k'} \gamma_k \gamma_{k'} \Omega_{p_k,p_{k'}}^{(0)} \quad ; \quad \Omega_{u,v}^{(0)} = \sum_{k=1}^m \gamma_k \Omega_{p_k,v}^{(0)}.$$

We recognize then the propagation formulas derived in Pickrell and Pritchard (2012) and used in many statistical tools such as poolstat (Gautier et al., 2022) or AdmixtureBayes (Nielsen et al., 2023). These formulas also match (9), so that using this approximation, the model for the allele frequency is equivalent to a BM on the network, with branch lengths rescaled as $q(\ell) = 1 - e^{-\ell}$.

As noted in Patterson et al. (2012), this rescaling of branch lengths is not additive: taking tree nodes $u \rightarrow v \rightarrow w$ along one lineage, with $\ell_{uw} = \ell_{uv} + \ell_{vw}$, the rescaled $q(\ell_{uw}) = 1 - e^{-(\ell_{uv} + \ell_{vw})}$ is different from $q(\ell_{uv}) + q(\ell_{vw}) = 1 - e^{-\ell_{uv}} + 1 - e^{-\ell_{vw}}$. This is because the approximation ignores the fact that the drift decreases as we get away from the root.

Prop. 5 and Prop. 6 provide an exact branch length transformation, as demonstrated in the following corollary.

Corollary 8 (Allele frequency evolution as a Brownian motion). *The variance-covariance structure of the vector of allele frequencies p_u on the network is the same as the one obtained from a Brownian motion with variance $p_0(1 - p_0)$ on the network, with branch lengths rescaled to*

$$L_e^{F_2} = q(\ell_e) (1 - \Omega_{v,v}^{(0)})$$

for edge e , where v is the parent node of e . This transformation is additive, and transformed lengths can be seen as measuring “ F_2 -units” (Gautier et al., 2022; Patterson et al., 2012).

If the phylogeny is a tree, then $\Omega_{u,v}^{(0)} = 1 - e^{-t_{uv}}$ and the transformation becomes explicit:

$$L_u^{F_2} = (1 - e^{-\ell_u}) e^{-t_{pa(u)}},$$

writing $pa(u)$ the unique parent node of u , and identifying an edge by its child node u . The back transformation from F_2 -units to coalescent units is then:

$$\ell_u = \ln(1 - T_{pa(u)}^{F_2}) - \ln(1 - T_u^{F_2})$$

with $T_u^{F_2}$ the transformed length of the path from the root to u .

The classical approximation from Pickrell and Pritchard (2012) then amounts to assuming constant expected heterozygosity along the phylogeny, that is, setting $H_u = \mathbb{E}[p_u(1 - p_u)] = p_0(1 - p_0)(1 - \Omega_{u,u}^{(0)}) \approx p_0(1 - p_0)$. This leads to $L_e^{F_2} \approx q(\ell_e)$ in Cor. 8.

2.5. The Gaussian-Coalescent model

We summarize here the core GC model for a univariate trait. The computations exploit the recursion formulas of Section 2.2 on a generic phylogenetic network, and we further derive a closed form for the GC variance matrix on a phylogenetic tree. In practice, we are interested in the distribution of the trait $\mathbf{X} = (X_1, \dots, X_n)$ measured on n individuals sampled from the populations at the leaves of a phylogeny, which may be a tree or a network with reticulations. In many applications, only one individual is sampled from each population but we do not make this assumption here.

2.5.1. Gaussian limit and the Gaussian-Coalescent model. In this section, we assume that X is a polygenic trait evolving according to Def. 4. Prop. 2 allows us to specify the expectation and variance for the vector $\mathbf{X}$ of measured traits.

Corollary 9 (Covariance of sampled traits). *Let $\mathbf{X} = (X_1, \dots, X_n)$ be the trait measured on individuals sampled from the tips of a species phylogeny. Assume that there are n_s different populations, and that $n_{\text{ind},u}$ individuals are sampled from population u , with $\sum_{u=1}^{n_s} n_{\text{ind},u} = n$. We re-write this vector as $\mathbf{X} = (X_{u,i})_{1 \leq u \leq n_s; 1 \leq i \leq n_{\text{ind},u}}$, where $X_{u,i}$ denotes the trait for individual i in population u . Then, for any two populations u and v and distinct individuals i and j , we have:*

$$\mathbb{E}[X_{u,i}] = m_0 \quad ; \quad \text{Var}[X_{u,i}] = L\Phi_u \quad ; \quad \text{Cov}[X_{u,i}; X_{v,j}] = L\Omega_{u,v},$$

where Φ and Ω are computed from Prop. 2. Using $\mathbf{H} = \Phi - \text{diag}(\Omega)$ as in (16), the variance of $\mathbf{X}$ is the sum of the between-population variance Ω and the expected within-population variance $\mathbf{H}$:

$$\text{Cov}[X_{u,i}; X_{v,j}] = L(\Omega_{u,v} + \mathbf{1}_{u=v, i=j} H_u).$$

In general, the distribution of $\mathbf{X}$ is not Gaussian. However, from the central limit theorem, the vector of traits $\mathbf{X}$ can be approximated by a Gaussian vector when L is large.

Proposition 10 (Gaussian limit). *When the number of loci L goes to infinity, the trait vector $\mathbf{X}$, with elements $X_{u,i} = \sum_{l=1}^L Y_{u,i}^{(l)}$ across nodes $1 \leq u \leq n_s$ and individuals $1 \leq i \leq n_{\text{ind},u}$, converges in distribution towards a Gaussian vector $\mathbf{Z}$.*

This proposition is a direct consequence of the multivariate central limit theorem, as the polygenic trait is the sum of independent identically distributed random vectors $\mathbf{Y}^{(l)}$ with finite variance matrix. It allows us to fit a statistical model to a vector of observed trait $\mathbf{X}$, using the Gaussian approximation. In all statistical analyses below, we use the “Gaussian-Coalescent”

model, which approximates the coalescent-based process by a Gaussian vector with an expectation and a covariance matrix as in Prop. 10. This joint Gaussian approximation was already used in previous work (Hibbins et al., 2023; Mendes et al., 2018). We note that it remains an approximation even if all loci follow a Brownian motion, as demonstrated in SM Section A (Remark S1) and illustrated in SM Section B.

2.5.2. Closed-form expression for species trees. In the special case when the species phylogeny $\mathcal{N}$ is a tree, a direct consequence of Cor. 4 and Prop. 6 is the analytical formulation of the covariance matrix, beyond the recursion formula, in the following proposition. We use it later to compare our model with ILS to other models, without ILS or from alternate approaches taken by other authors.

Proposition 11. Assume a phylogenetic polygenic evolution model on a species tree with a finite and constant variance rate (per coalescent unit) $\sigma_\theta^2 = \sigma_L^2/L$ for the Lévy distribution of increments at each locus, and trait distribution $\mathcal{P}_0$ at the root ρ with mean m_0 and variance v_0 . If t_u and t_{uv} denote the path and shared path lengths in coalescent units, then the conditional trait means and conditional (co)variances between individuals from population(s) u and v , conditional on ρ , are $\mathbb{E}[X_{u,i}] = m_0$ and

$$(17) \quad \Phi_u = \sigma_L^2 t_u + v_0,$$

$$(18) \quad \Omega_{u,v} = \sigma_L^2 (t_{uv} - (1 - e^{-t_{uv}})) + v_0 (1 - e^{-t_{uv}}).$$

Writing $\lambda = v_0/\sigma_L^2$ as in (1), we get:

$$(19) \quad \Phi_u = \sigma_L^2(t_u + \lambda) \quad \text{and} \quad \Omega_{u,v} = \sigma_L^2(t_{uv} + (\lambda - 1)(1 - e^{-t_{uv}})).$$

Remark 2 (interpretation). Using notations from (2), we can rewrite (18) as:

$$\Omega_{u,v} = \sigma_L^2 t_{uv} r(t_{uv}) + v_0 q(t_{uv}) = \sigma_L^2 \mathbb{E}[(t_{uv} - T_{uv})_+] + v_0 \mathbb{P}[T_{uv} < t_{uv}]$$

where T_{uv} is the random coalescent time between two individuals, one from population u and the other from population v , that are allowed to coalesce only starting from the branch above their most recent common ancestor. In $\Omega_{u,v}$, the first term is the expected shared time of evolution since the root scaled by the evolutionary rate σ_L^2 , and the second term is the ancestral variance v_0 times the probability that they have coalesced during the allowed time t_{uv} . Up to a constant, this formula has the same general form as eq. (11) from Schraiber et al. (2024), who assumed a compound Poisson model of evolution on L loci and $v_0 = 0$.

Remark 3 (within-population variance). The expected within-population variance of the trait is

$$(20) \quad H_u = \Phi_u - \Omega_{u,u} = \sigma_L^2(1 - e^{-t_u}) + v_0 e^{-t_u}.$$

It is equal to the equilibrium variance σ_L^2 when $\lambda = 1$, and otherwise tends to it when t_u goes to infinity, that is, for a deep tree, or a tree with a root being pushed back in time with a single long branch. Our model hence recovers the expected equilibrium variance (Lynch and Hill, 1986).

Remark 4 (root variance at equilibrium). When we assume that the root variance is at equilibrium, i.e. $v_0 = \sigma_L^2$ or $\lambda = 1$, then (19) and (20) give:

$$(21) \quad \Omega_{u,v} = \sigma_L^2 t_{uv} \quad \text{and} \quad H_u = \sigma_L^2.$$

The process is then the same as the standard BM on the tree in coalescent units, but with added within-population variation, with a variance at each individual increased by one coalescent unit compared to the covariance between individuals from the same population.

2.5.3. Equivalence to the BM with transformed branch lengths. In this section we continue to assume that a $\mathcal{N}$ is a tree. We can see present-day populations as internal nodes, with extra branches leading to each sampled individual, initially of length 0 if there is no within-species variation. The covariance of our model with ILS is then equivalent to that of a standard BM with

variance σ_L^2 on this extended tree, with length $\ell_u(\lambda)$ for the internal edge to population u and $\ell_{u,i}(\lambda)$ for the external edge to individual i sampled from u , given by

$$\begin{cases} \ell_u(\lambda) = \ell_u + (\lambda - 1)(q(t_u) - q(t_{pa(u)})) & = \ell_u + (\lambda - 1)(1 - e^{-\ell_u})e^{-t_{pa(u)}} \\ \ell_{u,i}(\lambda) = 1 + (\lambda - 1)(1 - q(t_u)) & = 1 + (\lambda - 1)e^{-t_u}. \end{cases}$$

For internal branches, the transformation directly stems from (13) and (15) of Prop. 6. These give the covariance structure between different individuals, that exactly matches Ω . For external branches, the transformation accounts for the within-population variation $\mathbf{H} = \Phi - \text{diag}(\Omega)$, and stems from (14) of Prop. 6.

If $\lambda = 1$ (root at equilibrium), then the internal branch lengths are unchanged, and the tip branches are assigned one coalescent unit: the process reduces to a standard BM with within-population variation. Increasing or decreasing λ leads to an increase or decrease of all branch lengths.

This equivalence to a standard BM allows us to use the three-point structure algorithm of Ho and Ané (2014). Like for other models such as the Ornstein-Uhlenbeck or the Early Burst, this tree transformation can be combined with an extra non-heritable independent within-species variation by lengthening the external branches, as implemented in `phyloIm` (Ho and Ané, 2014).

2.5.4. Population size as inverse phylogenetic signal. In the case when $\mathcal{N}$ is a tree as earlier, and assuming a constant haploid effective population size $2N$ over $\mathcal{N}$, we study in SM Section C the limits in very small and very large populations. Under no ILS ($N \rightarrow 0$ and immediate coalescence) the model is equivalent to a BM model on the original tree; while under maximum ILS ($N \rightarrow \infty$ and all coalescences above ρ) the model is equivalent to a single population (or star tree) with independent individuals. In short, the population size N can be seen as an inverse measure of “phylogenetic signal”: the larger it is, the weaker the correlations between populations (see also Mendes et al. 2018).

If branch lengths were known in number of generations, then we could imagine fitting N to the continuous traits at hand, so that our model would adapt to the specific level of coalescence affecting the genes controlling the traits.

2.6. Comparison with alternate approaches accounting for ILS*

In this section, we explore similarities and differences with previous approaches that account for ILS using the theoretical distribution of gene trees from the coalescent. We did not consider methods that use an empirical distribution of gene trees as input, instead of a single species phylogeny. This section can be skipped to focus on the use of the GC model.

2.6.1. The C^* matrix. Mendes et al. (2018) proposed the first approach accounting for ILS, in the tree case. As here, they considered a phenotype controlled by many independent and additive loci. They assumed an ultrametric tree (with all taxa at the same distance from the root), and one individual per taxon. Their trait covariance is obtained by averaging the covariance matrices associated with gene trees, weighted by their probability under the coalescent process. For each gene tree, the associated matrix corresponds to the covariance conditional on the ancestral value at the root of the tree, that is, at the most ancient coalescent event between the sampled individuals. When deriving the BM covariance on a tree, this root value is taken as known (fixed to 0 in Mendes et al., 2018, S.13), or as a parameter to be estimated. Each gene tree covariance then corresponds to conditioning on the locus effect at a random time, and these times differ across loci. In contrast, our model conditions on a fixed ancestral population, in which ancestral polymorphism is modeled via the distribution $\mathcal{P}_0$.

Mendes et al. (2018) found the following general expression for their covariance with ILS, in their equations (S.14) for the variance case $u = v$ and (S.17) for the covariance case $u \neq v$:

$$(22) \quad C_{uv}^* = L\sigma_\theta^2 \left[t_{uv} + (1 + \mathbf{1}_{u=v}) - \sum_{k=2}^n \frac{2}{k} p_k \right]$$

where n is the number of populations (tips) in the species tree, $\sigma_\theta^2 = \mu\sigma_m^2$ is the compound Poisson model with rate μ and mutation variance σ_m^2 , and p_k is the probability that k lineages of the gene tree enter the population of the MRCA of all n populations (going back in time). For example, if the tree has $n = 3$ populations and an internal edge length of t coalescent units, then $p_2 = (1 - e^{-t})$ and $p_3 = e^{-t}$. In general, p_k depends on the tree size (number of sampled populations n), topology and edge lengths in coalescent units. When $u = v$, C_{uu}^* is the phenotype variance of one individual in population u , similar to Φ_u in Prop. 11.

Remark 5 (impact of taxon sampling). In the model derived by Mendes et al. (2018), the covariance between two populations depends on all other populations sampled in the tree, to know the MRCA of the entire sample in each gene tree. This means that adding or removing a tip somewhere in the tree impacts the covariances of all the other populations (see Sec. 2.6.2 below). This is not the case with the GC model provided that the root population is not affected by the taxon addition or removal (as e.g., in Figure 2), as can be seen directly in Prop. 11 on a tree, and by Prop. 3 generally. The model derived by conditioning on a given root ρ enjoys the following property: if S_1 is a “large” sample of individuals and populations, and if S_2 is a subset of individuals and/or populations from S_1 , then the model defined for $(X_i)_{i \in S_1}$ on the full S_1 reduced to data on S_2 is equivalent to the model defined directly for $(X_j)_{j \in S_2}$ on S_2 only.

Remark 6 (3-taxon tree with root at equilibrium). On a 3-taxon species tree with an internal branch length of t coalescent units, (22) simplifies to:

$$C_{uv}^* = L\sigma_\theta^2 \left[t_{uv} + \mathbf{1}_{u=v} + \frac{1}{3}e^{-t} \right].$$

Taking $\lambda = 1$ in (19), the GC model leads to

$$\Phi_u = L\sigma_\theta^2(t_u + 1) \quad \text{and} \quad \Omega_{u,v} = L\sigma_\theta^2 t_{uv}.$$

When there is only one individual per population, the variance matrix of the GC model is then:

$$\Phi_u \mathbf{1}_{u=v} + \Omega_{u,v} \mathbf{1}_{u \neq v} = L\sigma_\theta^2 [t_{uv} + \mathbf{1}_{u=v}].$$

In this specific case, the two formulas agree up to a uniform term $\frac{1}{3}e^{-t}$, that goes to zero when t goes to infinity. Note that the actual value of this term does not matter to estimate parameter values: both covariance matrices lead to the same estimated ancestral state at the root and the same variance rate (but differ in their estimated standard errors). Our examples below show that this agreement between C^* and our GC model covariance does not generalize to more than 3 taxa.

2.6.2. Impact of taxon sampling: numerical experiments. Hibbins et al. (2023), who coined the name ‘ C^* ’, extended the approach of Mendes et al. (2018), which takes C^* as the average of covariance matrices across gene trees, and implemented it in the R package *seastaR*. They consider gene trees generated from the coalescent on a species tree (as in Mendes et al., 2018) or observed from a sample of loci. Their coalescent-based species-tree method matches that of Mendes et al. (2018) on 3 taxa, but differs on 4 or more taxa, as they break down the full tree into 3-taxon subtrees and then calculate C^* from (22) on each triplet.

To compare with the GC model in the tree case, we considered the phylogenies in Figure 2 with focal populations A and C , and n sampled populations below their MRCA. For each phylogeny, we calculated the coalescent-based C^* matrix implemented in *seastaR* (Hibbins et al., 2023, function `get_full_matrix`), and the GC covariance matrix (Prop. 11) assuming equilibrium variance in the root population ($v_0 = \sigma_L^2$, based on Remark 6). For both, we assumed one individual per taxon (as required by C^*) and calculated the *scaled* covariance matrix, omitting the variance rate factor $\sigma_L^2 = L\sigma_\theta^2$.

We first considered the trait variance, at A and C in the phylogeny in Figure 2 (left) in which C is sister to a clade containing A . Figure 3 (left) shows these variances as a function of the clade’s age and size. One would expect $\text{Var}[X_A]$ and $\text{Var}[X_C]$ to be equal, because A and C are at the same distance from the root (fixed to $T = 30$ or 3). Also, one would expect these variances

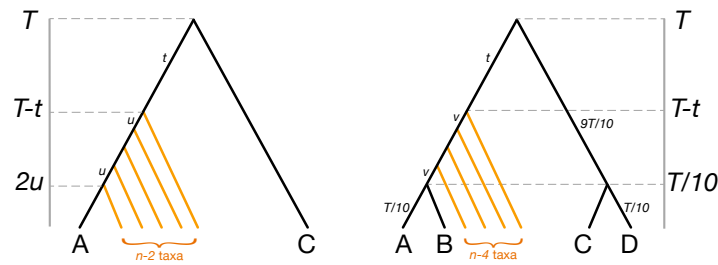


Figure 2 – Species trees used to quantify the effect of taxon sampling on ILS-based methods, each having n taxa at distance T from the root. The gray axis indicates time before present (from 0 at the tips to T at the root). Branches are annotated by their length. **Left:** tree used in Figure 3, with $n - 2$ unlabeled taxa (shown with orange branches) forming a clade with A. This clade's MRCA is at distance t from the root, with successive speciation events separated by branches of length $u = (T - t)/(n - 1)$, such that A is at distance T from the root. **Right:** tree used in Figure 4, with $n - 4$ unlabeled taxa (orange) forming a clade with A and B. This clade is at distance t from the root, with successive speciation events separated by branches of length $v = (9T/10 - t)/(n - 4)$, such that all species are at distance T from the root, and the MRCA of A & B is at distance $9T/10$ from the root.

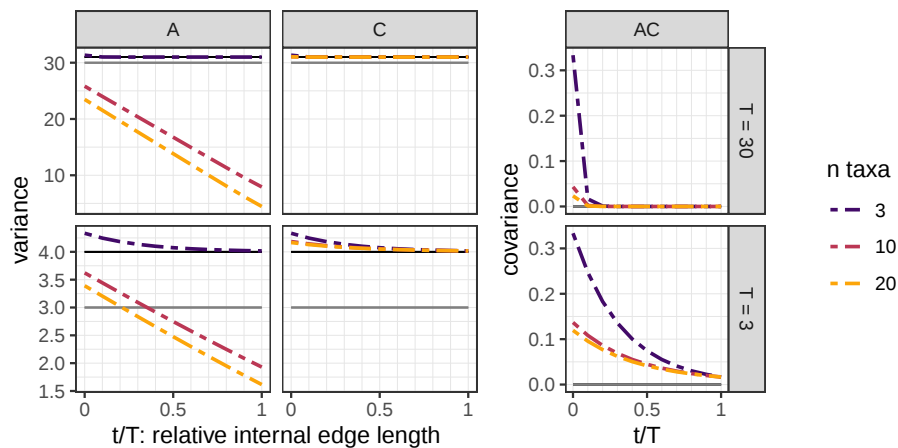


Figure 3 – Effect of taxon sampling and clade's distance to the root on the variance and covariance calculated by the ILS-based C^* (Hibbins et al., 2023; Mendes et al., 2018), on the species tree in Figure 2 (left) with branch lengths (T and t) in coalescent units. **Left:** trait variance at A and at C. **Right:** covariance between the traits at A and C. Colored dashed lines show C^* values. The value from the GC model is shown by a black horizontal line (at $T + 1 = 31$ or 4 for the variance, 0 for the covariance). The value from the standard BM model (without ILS) is indicated with a dark gray horizontal line (at $T = 30$ or 3 for the variance, 0 for the covariance).

to be independent of the clade size and age, to avoid a dependence on taxon sampling in this clade. These expectations hold for the GC model but not for C^* . When the full phylogeny has $n = 3$ taxa (A, its sister, and C) then $C_{AA}^* = C_{CC}^*$ holds, but its value depends on the divergence time between A and its sister taxon (very slightly when $T = 30$ and less slightly when $T = 3$ coalescent units) – in agreement with Remark 6. However, when A is in a larger clade ($n \geq 4$) then C^* assigns a significantly smaller variance to A than to C ($C_{AA}^* < C_{CC}^*$), with a younger clade age (larger t in Figure 3) leading to a larger gap.

Intuitively, this is because A is part of triplets $AB_1|B_2$ where B_1 and B_2 are orange taxa in Figure 2, whose MRCA is younger than the species tree root. In a gene tree, denote the age of the oldest coalescent event as $t_{ab_1b_2}$ between A, B_1 and B_2 , and t_{ab_1c} between A, B_1 and C. Then $t_{ab_1b_2}$ is typically younger: $\mathbb{E}t_{ab_1b_2} < \mathbb{E}t_{ab_1c}$. Conditional on the trait at the gene tree root, the variance of A calculated from the gene tree with B_1 and B_2 is only $t_{ab_1b_2}$, which excludes the

variation in the trait between t_{ab_1c} and $t_{ab_1b_2}$. Ignoring this ancestral variation by conditioning on the trait value at a more recent time causes a reduction in C_{AA}^* . This reduction does not affect C_{CC}^* , because C is not part of triplets whose MRCA is younger than the root of the species tree.

We then turned to phylogenetic covariances. Using the tree in Figure 2 (left), $\text{Cov}(X_A, X_C)$ does not depend on taxon sampling in the GC model, with a covariance of 0 because we condition on the MRCA population of A and C (rather than a more ancestral population). Based on C^* , the phylogenetic covariance between A and C depends on taxon sampling, especially when the total tree height is low ($T = 3$ coalescent units) or when taxon sampling adds speciations near the root (t small, Figure 3).

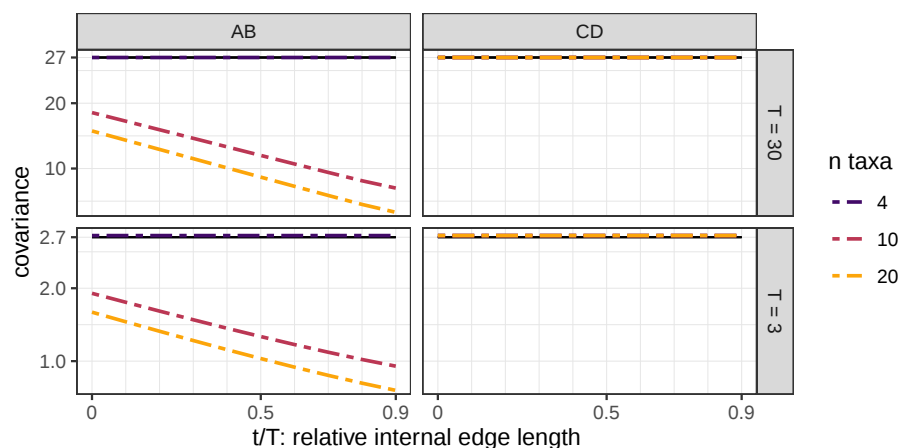


Figure 4 – Effect of taxon sampling and clade’s distance to the root on the covariance C^* of two sister pairs, on the species tree in Figure 2, right, with branch lengths (T and t) in coalescent units. **Left:** trait covariance between A and B . **Right:** trait covariance between C and D . The value from the GC model is shown by a black horizontal line. It is identical to the value from the standard BM model (without ILS), at $\frac{9}{10}T = 27$ (top) or 2.7 (bottom).

In Figure 2 (right), we considered a phylogeny with 2 similar sister pairs: AB and CD . As both pairs share the same divergence time, one would expect the covariances $\text{Cov}(X_A, X_B)$ and $\text{Cov}(X_C, X_D)$ to be equal and independent of taxon sampling within the species tree. This expectation holds for the GC model but not for C^* (Figure 4). When no other taxa are sampled than the 2 sister pairs ($n = 4$) then $C_{AB}^* = C_{CD}^*$ holds. But when A and B are part of a larger clade (orange in Figure 2) then C_{AB}^* is significantly smaller than C_{CD}^* , with a larger gap when the clade is younger (larger t).

Intuitively, this is because the AB pair is part of triplets $AB|B_2$ whose MRCA is younger than the species tree root, whereas all triplets containing the CD pair have their MRCA at the root. By conditioning on the trait at the younger root of gene trees, a shorter evolutionary path is considered to be shared between A and B , causing a reduction of their covariance compared to the CD pair.

Averaging over all triplets to get C^* could be partly responsible for some of the sensitivity of C^* to taxon sampling. To reduce this sensitivity, a future version of *seastaR* may use a subset of triplets only, such as those spanning the root of the tree (an idea by M. Hahn and M. Hibbins in their reviews, see Rogers 2026). We used here the released version of the software at the time of the study (v0.1.0#9c6cf78), which is also the version used in Hibbins et al. (2023).

Another avenue to reduce this sensitivity is to follow a procedure by Mo and Hahn (2026): simulate many gene trees from the species phylogeny, then average their associated BM covariance matrices using function `trees_to_vcv` in *seastaR*. This alternative C^* estimate is still sensitive to taxon sampling, and sometimes more so than `get_full_matrix` (see appendix D.1).

2.6.3. Introgression and ILS. To account for introgression in addition to ILS, Hibbins and Hahn (2021) extended the C^* method in the specific case of 3 populations, assuming one bidirectional

gene flow event between the outgroup population and one ingroup population. They obtained the phylogenetic trait covariance as a weighted average of C^* derived from 3 parent species trees: one representing no introgression, and two representing unidirectional introgression. Our polygenic model has the advantage of being applicable to any phylogenetic network on any number of populations and individuals, following the multispecies network coalescent exactly, and with the guarantee that given a fixed root population ρ , variances and covariances are independent of taxon sampling.

3. Impact on comparative methods

3.1. Phylogenetic regression and implementation

We implemented a phylogenetic regression whose residuals are Gaussian with a variance matrix structured as the GC, assuming a constant rate across branches. The regression takes as input: (i) a vector of univariate continuous traits $\mathbf{X} = (X_1, \dots, X_n)$ measured on n individuals sampled from the populations at the leaves of a phylogeny, with possibly several measured individuals per population; (ii) a phylogeny with branch length in coalescent units and (iii) a regression design matrix $\mathbf{R}$ with n rows and columns that may include an intercept (vector of ones), structured vectors for phylogenetic ANOVA, and/or covariates of interest. The phylogenetic regression models $\mathbf{X}$ using the covariance from the GC model in Cor. 9, denoted $\mathbf{C}_{GC}(\sigma_L^2, \lambda)$ here:

$$(23) \quad \mathbf{X} = \mathbf{R}\beta + \varepsilon \quad \text{with} \quad \varepsilon \sim \text{Normal}(\mathbf{0}, \mathbf{C}_{GC}(\sigma_L^2, \lambda)) .$$

This covariance can be computed from the recursion in Prop. 2, or using the closed form from Prop. 11 for a tree. The model parameters are the vector of coefficients β , the rate σ_L^2 per coalescent unit, and the ratio $\lambda = v_0/\sigma_L^2$ between the root variance and the rate.

For example, $\mathbf{X}$ may contain a measure of the constitutive defense response to herbivory, observed across multiple plants from various species and populations. If the design matrix $\mathbf{R}$ is reduced to a vector of ones, then β has only one coefficient (the intercept) $\beta_1 = m_0$, that represents the mean response in the ancestral population. To test for a negative correlation between constitutive defense and extra defense induced by prior herbivory, $\mathbf{R}$ may be extended with a second column containing induced defense values (as in Haak et al., 2014), with associated regression coefficient β_2 and $\beta_2 < 0$ providing evidence for a negative relationship between the two defense types. (See e.g., Harmon 2019 for other application examples of phylogenetic regression.) The design matrix $\mathbf{R}$ can also be constructed to allow for different mean values among different groups of populations, as in a typical ANOVA. This amounts to allowing for shifts in the trait mean, as is done, e.g., in Bastide et al. (2018) to test for transgressive evolution after hybridization events.

As in classical phylogenetic regressions (see e.g., Ho and Ané, 2014; Teo et al., 2023), the variance parameters are fitted using restricted maximum likelihood (REML) with a numerical optimization of λ , and β is estimated with maximum likelihood. Tests, confidence intervals and ancestral state reconstructions can then be performed conditionally on λ as in the classical setting (see e.g., Bastide et al., 2018). Similarly, classical penalized likelihood scores such as AIC can be used to compare this GC phylogenetic regression with other models. For example, one may compare the GC, BM and Ornstein-Uhlenbeck covariance models, or fit the GC model across multiple phylogenies with various reticulation events to test how introgression helps explain the trait data (Karimi et al., 2020; Morley et al., 2026; Teo et al., 2023). This model could also be used to infer the phylogeny from continuous traits, provided that data are combined across many independent traits. For instance, admixture graphs are commonly reconstructed from allele frequencies (e.g., Nielsen et al., 2023; Patterson et al., 2012; Pickrell and Pritchard, 2012), that can be modelled with the GC (see Section 2.4.2).

We implemented the GC model in the `julia` package `PhyloTraits` (Bastide et al., 2018) v.1.2.0, which can handle phylogenetic networks. We also implemented the GC model in the `R` package `phyloIm` (Ho and Ané, 2014) v2.7.0, which is restricted to species trees, but can estimate non-heritable trait variation in addition to the within-species heritable variation expected from

the GC model, as mentioned above. Both the `julia` and `R` implementations can also perform phylogenetic linear models (ANOVA and multiple regression) using predictors and the Gaussian-Coalescent model of phylogenetic correlation for residuals.

3.2. Estimation under ILS and gene flow

We performed simulations to quantify the impact of accounting for ILS using the GC model, versus ignoring ILS or using C^* from the `seastaR` package (in the tree case). We looked at the accuracy of the estimated rate of trait evolution, as in Mendes et al. (2018) and Hibbins et al. (2023).

3.2.1. Simulations settings. To simulate polygenic traits controlled by loci whose gene trees were generated under the coalescent, we used `PhyloCoalSimulations` (Fogg et al., 2023) v1.1.0. We generated each locus under a compound Poisson distribution with Laplace-distributed effects. Specifically, the number of mutations was generated from a Poisson process, and each mutational effect was drawn from a centered Laplace (symmetric exponential) distribution with shape $\sigma_m/\sqrt{2}$, which has mutational variance σ_m^2 . We used a large $L = 100$, from which the Gaussian approximation should be accurate.

For this simulation, we used the *Polemonium* phylogeny from Teo et al. (2022): a 17-taxon network with 3 reticulation events estimated in Rose (2021) and calibrated in Teo et al. (2023). From each taxon, we simulated $n_i = 20$ individuals. To simulate either low or high ILS, branch lengths in the phylogeny were rescaled to a total height of $T = 10$ (low ILS) or $T = 1$ (high ILS) coalescent units. Each locus effect had $\mu = 3/T$ mutations per coalescent unit, for an expected 3 mutations along gene trees from the root population to each individual. Mutation effects had mutational variance $\sigma_m^2 = 1/(L\mu) = T/(3 \times 100)$, such that the trait variance rate was $\sigma_L^2 = L\sigma_\theta^2 = L\mu\sigma_m^2 = 1$. The ratio $\lambda = v_0/\sigma_L^2 = v_0$ of the phenotype root variance over its equilibrium variance was varied in $\{0, 0.5, 1, 2\}$. As the phenotype's root variance is L times that of the locus effect's root variance, each locus was taken to have variance v_0/L in the root population. The mean in the root population was fixed to $m_0 = 0$. We simulated 100 independent replicates for each parameter combination. While 17 taxa is a small sample size to estimate the evolutionary variance rate, $n_i = 20$ individuals per population was expected to provide information on λ , or the ancestral trait variance v_0 at the root.

For each data set, we used the `gaussiancoalescent` model in `PhyloTraits` v1.2.0 (Bastide et al., 2018) to estimate the evolutionary variance rate σ_L^2 , either assuming variance equilibrium at the root ($\lambda = 1$, which was not necessarily the true value), or assuming the true variance at the root, or simultaneously estimating λ . We also analyzed each data set under a Brownian motion model, to estimate the evolutionary variance rate σ_L^2 simultaneously with a within-population variance σ_w^2 , assumed shared across all populations. Specifically, this model assumes a BM with rate σ_L^2 for the evolution of the population means, which are not observed. Conditional on these means, an observed individual value is assumed normally distributed centered at its population mean and with variance σ_w^2 . See Teo et al. (2023) for details. All the variances were estimated using the restricted maximum likelihood (REML) procedure, and we used a phylogenetic regression with a single intercept representing the root population mean value m_0 . Given the reticulate phylogeny and the multiple individuals per species, estimation with `seastaR` was not applicable.

3.2.2. Simulation results. Under low ILS, the evolutionary variance rate σ_L^2 was estimated without practical bias regardless of the estimation model: under a BM, or under a GC with λ fixed or estimated (Figure 5, top row). There was no noticeable gain to knowing the true value of λ . This is expected from theory for limit cases with no ILS (see Sections 2.5.4 and C). Further, precision increased greatly under any GC model compared the BM model. For example, the average root-mean-squared error (RMSE) on σ_L was 0.046 from the GC assuming $\lambda = 1$ versus 0.179 under the BM model. Under high ILS, σ_L^2 was estimated accurately only under the GC model with λ fixed to its true value (Figure 5, bottom row). Without knowledge of λ , fixing $\lambda = 1$ was the best option, with error dominated by the bias (RMSE of 0.136). Bias was smaller when λ was estimated, but precision was lower. Under the BM model, bias was similar to the GC assuming

$\lambda = 1$ but precision was lower, resulting in higher error overall (RMSE of 0.256). In short, the GC model fixing $\lambda = 1$ was the most accurate in all our simulation settings (in case λ is unknown).

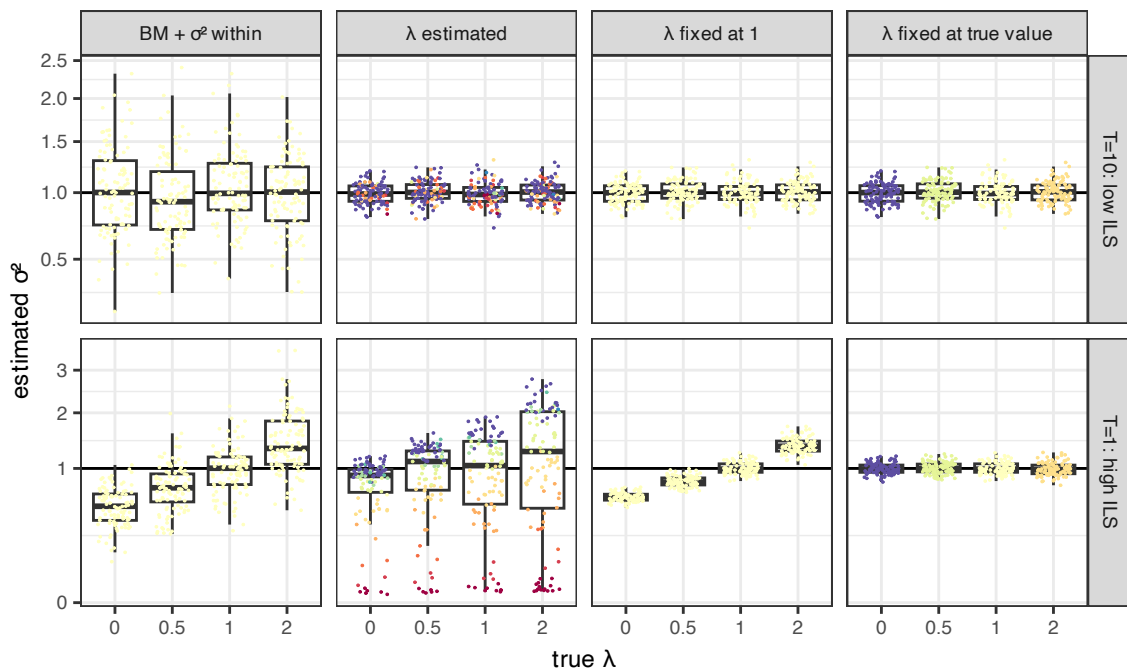


Figure 5 – Estimated variance rate $\hat{\sigma}_L^2$ (displayed on the square-root scale) from data simulated under the *Polemonium* 17-taxon network, 20 individuals per taxon, and traits controlled by $L = 100$ loci with compound Poisson-Laplace effects. The true rate was $\sigma_L^2 = 1$. Top: low ILS with network height of $T = 10$ coalescent units. Bottom: high ILS, with $T = 1$. The top and bottom rows use different limits for the vertical axis. Each column corresponds to one estimation method: the BM model with extra within-population variation, the GC model with λ estimated, fixed at 1, or fixed at its true value (from left to right). Colors indicate the estimated or fixed λ (or yellow for the BM model), with cold colors for low values and warm colors for large values.

The ancestral state, or trait mean in the root population m_0 was unbiased, from all estimation methods (Figure S6 in SM Section D). Its precision was much better under high than low ILS. This is expected from theory again (see Section 2.5.4), as limit cases with maximum ILS correspond to a lack of phylogenetic correlation with independent individuals. Informally, gene genealogies trace back to more individuals from the root population at high ILS, leading to more information about the root population under high than low ILS. At low ILS, precision in m_0 is driven by the number of populations and their phylogenetic correlation. At high ILS, precision in m_0 is driven by the total number of individuals and the trait variance at the root, v_0 .

Under GC, the estimation of λ was extremely imprecise, especially under low ILS. Under high ILS, estimates of λ were slightly more concentrated around the true value (Figure S7 in SM Section D). This is expected, as λ depends on the trait variance at the root, and because the trait data contain more information about the root population under high ILS.

Model selection also suggested that contemporary trait data carry little information about λ , at least in our simulation setting. Using a likelihood ratio test at level $\alpha = 0.05$, $\lambda = 1$ was rejected in only 1.1% of replicates (black dots in Figure S7). Using AIC, the GC model with fixed $\lambda = 1$ was favored 79% of the time, followed by the BM model 17% of the time. Estimating λ was deemed worthwhile only 4% of the time (Figure 6).

The ancestral trait variance v_0 , like λ , was poorly estimated (Figure S8), unless the GC model was used with λ fixed to its true value. However, the expected within-population variance H_u

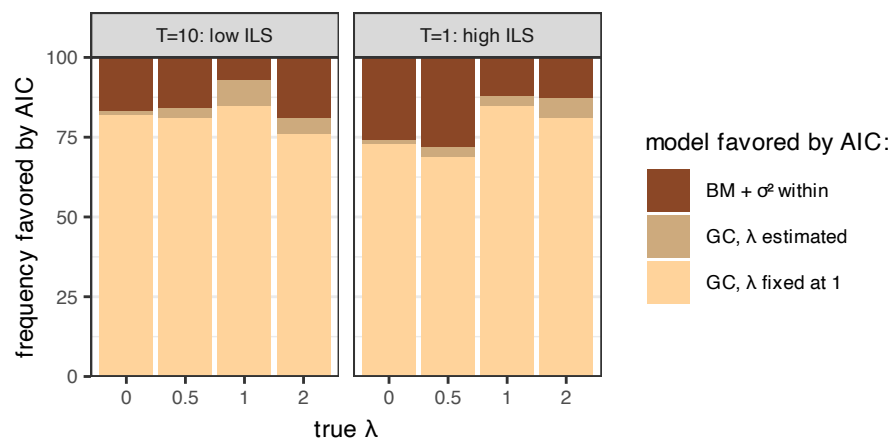


Figure 6 – Proportion of times that each model was favored by AIC, out of 100 replicate data sets, when comparing 3 models: Brownian motion with intra-specific variation, Gaussian-Coalescent (GC) with λ estimated simultaneously with σ_L^2 , and GC with $\lambda = 1$ fixed.

was estimated accurately, at both ILS levels. For a population u that is not below any reticulation, the expected variance within u is given by (20): it is close to σ_L^2 under low ILS and close to v_0 under high ILS. For a population u whose history is affected by one or more reticulation, H_u can be calculated recursively. Intuitively, reticulation increases H_u (at fixed ILS, or fixed distance T between u and the root population), especially if admixture involved divergent parent populations. We calculated H_u for each population u based on the true model parameters, then considered the average within-population variance h , averaged across the 17 populations. For each simulated data set, we repeated this calculation of the average expected within-population variance, but using the estimated parameters, to obtain an estimate $\hat{h}$ of h . When fitting with the BM model, we used the estimated within-species variance (assumed shared across populations) as $\hat{h}$. This within-population parameter h was well estimated at both ILS levels (Figure S9). From 20 individuals per population, this is not surprising. However, considering that $\hat{h}$ was calculated from very noisy estimates $\hat{\lambda}$, $\hat{v}_0$, and $\hat{\sigma}_L^2$ (noisy under high ILS) under GC, accurate estimation of $\hat{h}$ is still noteworthy.

Finally, we note that the “high ILS” setting was particularly challenging, as the entire network had a total height of only one coalescent unit. When reproducing this simulation study on the empirical tomato tree described below, which has a height of 6 coalescent units, results were in line with the “low ILS” setting above. In particular, the GC model with fixed $\lambda = 1$ received the most AIC support, and produced accurate estimation of σ_θ^2 , with an RMSE value of 0.120, much lower than the BM model (0.269), and almost equal to the best-case GC model with λ fixed to its true value (0.119). See SM Section D for details.

3.3. Floral traits in the wild tomato clade

3.3.1. Methods. We re-analyzed the wild tomato dataset studied in Hibbins et al. (2023), that includes several floral traits (corolla diameter, anther length and stigma length) compiled in Haak et al. (2014) and a phylogenetic tree from Pease et al. (2016), dated in coalescent units assuming a population size of 100 000 and one generation every two years as in Hamlin et al. (2020).

We first reproduced the analysis of Hibbins et al. (2023), who used two different triplets extracted from the phylogeny (see Figure 7): a “high ILS” triplet (tree height of 0.5 coalescent units) consisting of *S. galapagense* (LA0436), *S. cheesmaniae* (LA3124), and *S. pimpinellifolium* (LA1269); and a “low ILS” triplet (tree height of 6 coalescent units) with *S. pennellii* (LA3778), *S. pennellii* (LA0716), and *S. pimpinellifolium* (LA1589). IDs are accession numbers from the Tomato Genetics Resource Center. For each population (accession) and trait, we took the average over all the

individual measurements, and then fitted 4 models to each triplet: the simple BM, the *seastaR* method as in Hibbins et al. (2023), and the Gaussian-Coalescent model with λ fixed at either 0 or 1. We did not attempt to estimate λ on this first analysis, as the datasets were small (3 values). Note that we used a corrected version of *seastaR*, as the one used in Hibbins et al. (2023) had an implementation error¹ that led to unrealistically high estimates of the evolutionary rate (going as high as 2.7×10^5 , see Figure 6 in Hibbins et al. 2023). As above, the variances were estimated using the REML procedure.

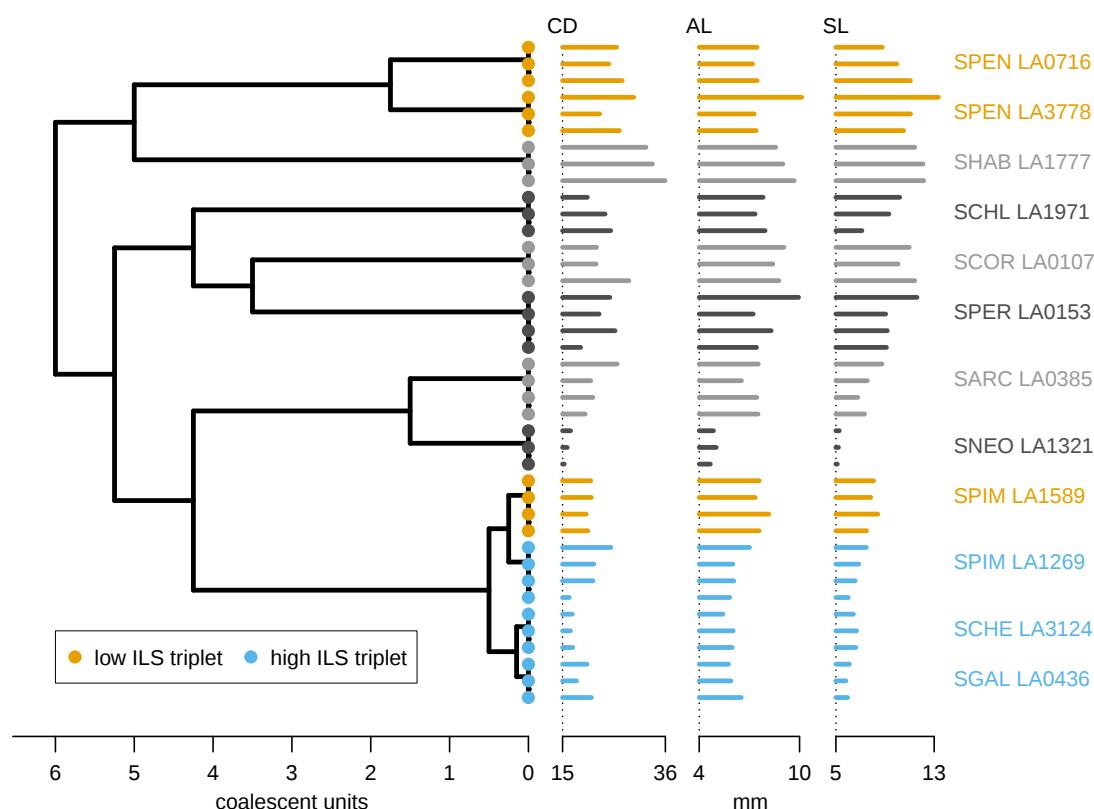


Figure 7 – Phylogenetic tree of the wild tomato populations extracted from Pease et al. (2016) scaled in coalescent units as in Hibbins et al. (2023), with individual measurements (in mm) of corolla diameter (CD), anther length (AL) and stigma length (SL) from Haak et al. (2014). Species from “low ILS” (orange) and “high ILS” (blue) triplets are colored.

We then used the “full” dataset, that included as many populations and individual trait measurements as possible. We excluded two species from the tree that did not have any trait data (*S. huaylasense* and *S. tuberosum*), and removed three species IDs from the trait dataset that were not in the tree (SCER, SCHM and SLYC). Some species contained several populations (accessions). When this was the case, and when we did not have any further information on the divergence time between the different populations of a given species, we kept only the population with the highest number of samples. The resulting tree had 12 populations, with between 3 and 4 measurements per population, a total height of 6 coalescent unit, and room for relatively low overall ILS (see Figure 7). We then compared the fit of seven models: the GC with λ either estimated, or fixed to a value of 0 or 1, either with or without additional within-population variation, and the BM with within-population variation, all permitted by the R implementation in *phyloIm*. We compared the AIC weights (wAIC, Wagenmakers and Farrell, 2004) of each model and their estimated evolutionary rates (using REML). For each model, we also computed the total expected

¹See pull request #7 on the *seastaR* repository: <https://github.com/larabreithaupt/seastaR/pull/7>.

within-population variance. For the BM with extra within-population variation, this was simply the within-population variance parameter σ_w^2 . For the GC models, it was σ_w^2 plus the evolutionary within-population heritable variance expected from the coalescent as computed in (20) with $t_u = T$ the total height of the tree.

3.3.2. Results. When studying the population mean traits on triplets, we found evolutionary variance rates qualitatively similar to those by Hibbins et al. (2023), with the BM model estimating $\hat{\sigma}_L^2$ generally higher for the “high ILS” triplet compared to methods taking ILS into account, but similar for the “low ILS” triplet (Figure 8). Further, in this setting where only triplets are used, our GC method with λ fixed to 1 (root at equilibrium) is equivalent to the *seastaR* model as explained in Remark 6, so that they give the exact same estimates $\hat{\sigma}_L^2$. When fixing λ to zero (i.e., root variance $v_0 = 0$), we found estimates that were in between, indicating that this model estimated a higher evolutionary variance compared to the model with $\lambda = 1$, but smaller compared to the BM, which accounted for no within-population variation at all.

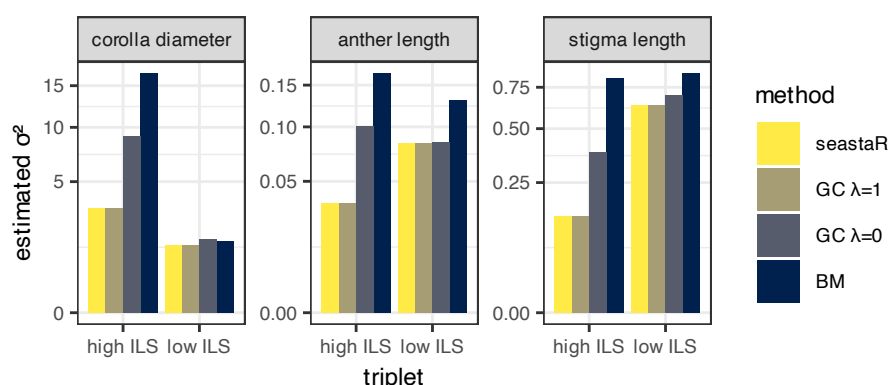


Figure 8 – Estimated variance rate $\hat{\sigma}_L^2$ (displayed on the square-root scale) from population means, for three floral traits and three triplets as in Hibbins et al. (2023) (“low ILS”, $T = 6$ and “high ILS”, $T = 0.5$). Each panel uses a different limit for the vertical axis. Colors indicate the statistical method used: *seastaR* from Hibbins et al. (2023), the GC model with λ fixed at 0 or 1, and the BM. Under low ILS, $\hat{\sigma}_L^2$ is similar across methods. Under high ILS, $\hat{\sigma}_L^2$ is generally higher for the BM model compared to ILS-based models.

When analyzing the full dataset with multiple values per population, we found that AIC favored the GC model with λ fixed and without extra non-evolutionary within-population variation (Figure 9). The exact value of λ seemed to have less impact, with $\lambda = 1$ slightly favored for the corolla diameter and stigma length and $\lambda = 0$ for the anther length. This indicates that the GC model fits these data better than the BM model with added within-population variation. In other words, the evolutionary within-population variation expected from the coalescent was sufficient to explain the trait variability within each population.

In most cases, the estimated evolutionary rate and the total within-population variation were highly similar across models, including the BM with extra within-population variation (Figure 10). For corolla diameter, methods that included extra non-evolutionary variance tended to estimate a lower evolutionary variance.

Note that, because the phylogeny is an ultrametric tree and we assumed a constant rate σ_L^2 , the within-population variance from the GC model, given in (20), is the same for all present-day populations. Further work could consider fitting a GC model with rates that differ in some clades or some branches, as permitted by Prop. 2, and as advocated recently in Mo and Hahn (2026).

4. Discussion

In this work, we describe a probability model for the evolution of a trait along a species phylogeny, in which the trait is controlled by one or several genes, that are themselves affected by

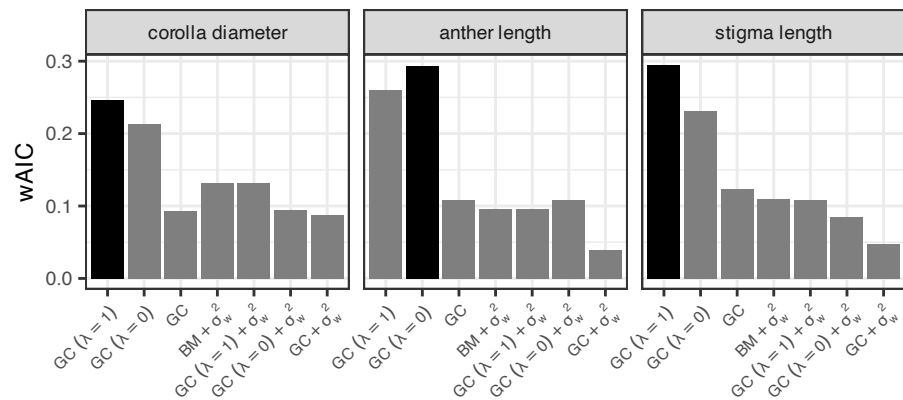


Figure 9 – AIC weight for each model fitted to the full tomato tree, for the three floral traits, keeping replicate measurements within each populations. Models are the GC with λ fixed to 0 or 1 or estimated without or with an extra non-evolutionary within-population variance parameter (σ_w^2), and the BM with within-population variance. The model in black is the one with the highest AIC value: it is always a GC model with fixed λ and no extra variation.

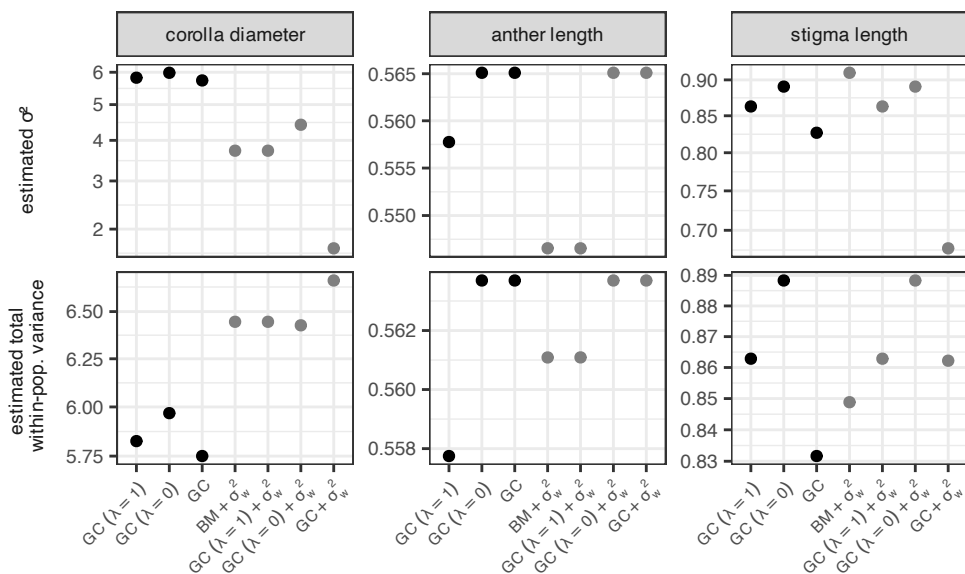


Figure 10 – Estimated evolutionary variance $\hat{\sigma}_L^2$ (top row) and total within-population variance for each model fitted to the full tomato tree (12 populations, 40 individuals). Variances are displayed on the square-root scale. Note that the scale for each panel is different, and often very narrow. Models are the GC with λ fixed to 0 or 1 or estimated and the BM, possibly with an extra non-evolutionary within-population variance parameter (σ_w^2 , gray). Estimates are consistent between methods.

incomplete lineage sorting, as modeled by the coalescent. If the phylogeny has reticulations, this process also includes the effect of gene flow, admixture, or hybridization. Our model accounts for ancestral polymorphism in the population at the root of the phylogeny, and implies an expected amount of heritable within-species variation in each present-day population. For tomato floral traits, this was sufficient to explain the within-population variation observed in the data.

Future work could survey other traits and groups to assess whether this pattern holds more generally.

Unlike prior methods accounting for ILS, our model is independent of taxon sampling, in the sense that the model for a sub-sample of populations remains unchanged when considering either the full or reduced phylogeny (given a fixed root population). Our model also applies to any phylogenetic network, of any topology and number of taxa, and applies to individual values for any number of individuals per taxon. It is implemented in two software packages (`phylo1m` in R and `PhyloTraits` in Julia) and can be fitted to data at no or little additional computational cost compared to standard PCM tools.

When the trait is polygenic controlled by many genes and if the effects of individual loci evolve with a finite variance rate, we showed that the distribution of the trait data is approximately Gaussian, leading to our Gaussian-Coalescent model.

In practice, there are several caveats for using our GC model. First, it requires a species phylogeny with branch lengths in coalescent units, which can be obtained as the ratio of number of generations over effective population size. Coalescent units can be estimated from the amount of gene tree discordance or from site pattern frequencies. While these estimates can be accurate at low values, such as between recently diverged populations, long coalescent units cannot be estimated accurately from genealogical discordance alone. This is because coalescence between individuals is extremely likely beyond 10 coalescent units, say, so that discordance disappears and becomes uninformative about the exact length of a branch in coalescent units, when it is long. For long coalescent times, accurate estimates may come from other units on branch lengths (substitutions, generations, calendar time) and assumptions on how these other units correlate with coalescent units. This issue has been considered for calibrating networks, although primarily to calibrate branch lengths proportional to time (Karimi et al., 2020; Tabatabaee et al., 2026; Xu and Ané, 2024).

Second, the GC model has one extra parameter compared to the Brownian motion: the trait variance in the root population. Ancestral polymorphism is typically not observed. Our simulations show that present-day data contain scant information about this ancestral variance, which is estimated with low precision. However, it is reasonable to assume that the root population was at equilibrium for the process (by fixing $\lambda = 1$). In simulations and in the tomato flower dataset, this choice was most often favored by AIC. It led to accurate estimates of the other parameters in most simulation scenarios. Under high ILS, it caused some bias in estimating the evolutionary rate, but much less bias and more accuracy than when using the BM model ignoring ILS.

Finally, our polygenic trait model assumes that the trait is the result of purely additive effects across genes, without epistatic interactions. Also, individuals are assumed haploids, carrying one copy of each locus. Our model readily applies to diploid individuals, if we further assume that the two copies of each locus have additive effects, like alleles across loci. Adding epistasis between loci and a model of dominance within loci would all be interesting future work. Valuable extensions also include relaxing the assumption that each locus evolves according to a Lévy process, modeling selection at the trait level such as with correlated loci (Bulmer, 1971), and considering the correlated evolution of several traits.

As described, our GC model should be used to analyze trait data at the individual level, because the GC variance includes within-population heritable variation. Averaging multiple values to estimate population means leads to a model violation, because taking the mean across individuals reduces the variance. This model violation also applies when using `seastaR`, which requires a single value per population. The tomato floral data analysis on triplets conducted here and in Hibbins et al. (2023) applies such trait averages within populations, so that its conclusions might be taken with caution. If measurement error, environmental variation or plasticity are expected to dominate genetic variance, then this practice could be beneficial despite its model violation. However, a sound alternative is to combine the GC model with extra non-heritable variation, as done here on the full tomato dataset.

Future work could build on our theoretical framework to explore using the GC model for a vast array of phylogenetic comparative methods, such as phylogenetic ANOVA and phylogenetic

regression, or the detection of shifts in evolutionary rates. Many traits are inherently multivariate (Clavel et al., 2015; Clavel and Morlon, 2020). Careful considerations of the biological mechanisms underlying the evolution of a multivariate polygenic trait would be required to extend the GC model to this setting. For instance, would all the traits be controlled by the same set of genes, or have only a subset in common? Building such a multivariate model could be an interesting direction for future work. Another valuable extension would consist in using observed gene trees, rather than gene trees expected from the coalescent, as pioneered by Hibbins et al. (2023), combined with conditioning on ρ as in the GC. A challenge is that gene trees are estimated with branch lengths in substitution per site, rather than calendar time or coalescent units. It may be necessary to account for substitution rate variation (if any) and to calibrate gene trees to embed them into the species phylogeny (tree or network) with comparable edge lengths, so as to identify the placement of the root population ρ in each gene tree. Recently, Mo and Hahn (2026) built on the C^* framework using a sample of gene trees with edge lengths in coalescent units, that can be embedded into the species tree. They used this embedding to map phenotypic rates from the species tree onto gene trees. Modifying their approach by slicing gene trees at ρ would permit conditioning on the root population and obtain a method that is stable to taxon sampling.

Acknowledgements

Preprint version 3 of this article has been peer-reviewed and recommended by Peer Community In Mathematical & Computational Biology (<https://doi.org/10.24072/pci.mcb.100819>; Rogers 2026).

Funding

This work was supported in part by the “International Emerging Action” program of the CNRS, and by the National Science Foundation through grants DMS-2023239 (to C.A.) and DMS-1929284 while C.A. was in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the “Theory, Methods, and Applications of Quantitative Phylogenomics” program.

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article. The authors declare the following non-financial conflict of interest: C.A. is a recommender for PCI Evolutionary Biology and for PCI Mathematical and Computational Biology.

Scripts and code availability

The GC model is available as part of the R package `phylo1m` (Ho and Ané, 2014), version 2.7.0; and of the Julia package `PhyloTraits` (Bastide et al., 2018), version 1.2.0. The code to reproduce the simulations, and the data and code to reproduce the analysis of the tomato floral traits are available at <https://github.com/cecileane/GCmodel-2026-data-code> and at Zenodo (Bastide and Ané, 2026).

A. Graphical model and proof of covariance formulas

A.1. Graphical model

In this section we express the one-locus model in Def. 3 as a formal graphical model. To this end, for each node in the network we consider two individuals drawn at random from the population that this node represents.

Definition S1 (Graphical model). Let $\mathcal{N}$ be phylogenetic network. For each node $u \in V(\mathcal{N})$, we define two random variables $Y_{u,1}$ and $Y_{u,2}$. The joint distribution of the vector $\mathbf{Y} = (Y_{u,i})_{u \in V(\mathcal{N}), i \in \{1,2\}}$ is defined by the following graphical model. At the root ρ of $\mathcal{N}$, we assume that $Y_{\rho,1}$ and $Y_{\rho,2}$ are independent and identically distributed according to a prior distribution $\mathcal{P}_0$.

Let u be a non-root node of $\mathcal{N}$, with $m \geq 1$ parents $p_1, \dots, p_m$, edge (p_k, u) of length ℓ_k in coalescent units, inheritance probability γ_k , and centered Lévy process $\mathcal{D}_{\theta_k}$. Further, let H_1 and H_2 be two independent random variables such that $\mathbb{P}[H_1 = k] = \mathbb{P}[H_2 = k] = \gamma_k$ for all $1 \leq k \leq m$. Define a random variable T as follows. If $H_1 \neq H_2$, then $T = +\infty$. If $H_1 = H_2 = k$, then T is drawn from an exponential distribution with mean 1, that is, $\mathbb{P}[T \leq \ell | H_1 = H_2 = k] = q(\ell)$. Finally, define a random variable J such that $\mathbb{P}[J = 1] = \mathbb{P}[J = 2] = 1/2$. The distribution of $(Y_{u,1}, Y_{u,2})$, conditional on all parent variables $(Y_{p_k,j})_{1 \leq k \leq m, j \in \{1,2\}}$ and on H_1, H_2, T and J is then defined as follows.

$$\begin{aligned}
 \text{(S1)} \quad & \begin{cases} Y_{u,i} = Y_{p_{H_i},i} + \epsilon_i \text{ for } i \in \{1,2\} \\ \epsilon_i \sim \mathcal{D}_{\theta_{H_i}}(\ell_{H_i}) \\ \epsilon_1, \epsilon_2 \text{ independent} \end{cases} & \text{if } H_1 \neq H_2 \text{ or } H_1 = H_2 = k \text{ and } T > \ell_k; \\
 \text{(S2)} \quad & \begin{cases} Y_{u,i} = Y_{p_k,J} + \delta_0 + \delta_i \text{ for } i \in \{1,2\} \\ \delta_0 \sim \mathcal{D}_{\theta_k}(\ell_k - T), \delta_i \sim \mathcal{D}_{\theta_k}(T) \\ \delta_0, \delta_1, \delta_2 \text{ independent} \end{cases} & \text{if } H_1 = H_2 = k \text{ and } T \leq \ell_k.
 \end{aligned}$$

For convenience, we denote by $C_{u,k}$ the random event $H_1 = H_2 = k$ and $T \leq \ell_k$, and by $\bar{C}_u$ the complement over all k : $H_1 \neq H_2$ or $H_1 = H_2$ and $T > \ell_{H_1}$. Note that, using the properties of the Lévy process, $\delta_0 + \delta_i \sim \mathcal{D}_{\theta_k}(\ell_k)$ in (S2) has the same distribution as ϵ_i in (S1) and does not depend on T .

Proposition S1. For any node u in $\mathcal{N}$, the random variables $Y_{u,1}$ and $Y_{u,2}$ in Def. S1 are exchangeable. Further, consider a non-root node u , any $i, j \in \{1,2\}$, and use notations in Def. S1. Conditional on H_1, H_2, T, J and on $H_i = k$, $Y_{u,i}$ is distributed as $Y_{p_k,j} + \epsilon$ where $\epsilon \sim \mathcal{D}_{\theta_k}(\ell_k)$ is independent of $Y_{p_k,j}$.

Proof. We prove the exchangeability by induction on nodes in $\mathcal{N}$, considering them in preorder (from root to leaves). By definition, $Y_{\rho,1}$ and $Y_{\rho,2}$ are independent and identically distributed, and hence are exchangeable. Let u be a node in $\mathcal{N}$. The exchangeability of $Y_{u,1}$ and $Y_{u,2}$ derives from the symmetry in $i \in \{1,2\}$ in (S1) and (S2), using the exchangeability of (H_1, H_2) , the induction hypothesis on p_{H_i} for (S1), and the symmetry of $\mathbb{P}[J = 1] = \mathbb{P}[J = 2]$ for (S2).

We now turn to the second claim and fix $i, j \in \{1,2\}$. Conditional on $H_i = k$ and on $\bar{C}_u$, $Y_{u,i} - Y_{p_k,i} \sim \mathcal{D}_{\theta_k}(\ell_k)$ from (S1), regardless of T, J and $H_{i'}$ for $i' \neq i$. The distribution of $Y_{u,i}$ conditional on $H_i = k$ and $\bar{C}_u$ is then identical to that of $Y_{p_k,j} + \epsilon$ by exchangeability of $Y_{p_k,1}$ and $Y_{p_k,2}$. Conditional on $C_{u,k}$ and on J , $Y_{u,i} = Y_{p_k,J} + \epsilon$ with $\epsilon \sim \mathcal{D}_{\theta_k}(\ell_k)$ from (S2) and properties of the Lévy process. Since $Y_{p_k,1}$ and $Y_{p_k,2}$ are exchangeable, the distribution of $Y_{u,i}$ conditional on $C_{u,k}$ and J is the same as that of $Y_{p_k,j} + \epsilon$, regardless of the value of J . Note that as a corollary, the claim holds conditional on $H_i = k$ only (e.g., integrating over T and J). $\square$

Proposition S2. Assume that Y evolves on $\mathcal{N}$ according to the model in Def. 3 for one locus. Consider two individuals sampled from each population u (ancestral or extant) in $\mathcal{N}$, with values $Y_{u,1}$ and $Y_{u,2}$. Then the joint distribution of $(Y_{u,i})_{u \in V(\mathcal{N}), i \in \{1,2\}}$ is equal to the joint distribution given by the graphical model in Def. S1.

Proof. We use induction on the nodes in $\mathcal{N}$, listed in a preorder, that is: the root comes first and any node u must be listed after all its parents have already been listed. We note $u < v$ if u is listed

before v . We will prove by induction on v , that the joint distribution of $(Y_{u,i})_{u \leq v, i \in \{1,2\}}$ (up to v), is equal under Def. 3 and Def. S1. By definition, $Y_{\rho,1}$ and $Y_{\rho,2}$ are independent and identically distributed from $\mathcal{P}_0$ in both models.

Now assume that the joint distribution up to v are identical, and take u the following node in the preorder. Take the two individuals 1 and 2 in population u , and define G_p as the tree for these two individuals, for the locus of interest, truncated to the parts that evolved within the m parent populations (p_k, u) for $k = 1, \dots, m$. This gene tree includes the identity H_i of the parent population that individual $i \in \{1, 2\}$ was inherited from (at the locus being considered). For example, if individual 1 came from p_1 , then $H_1 = 1$. By definition of the network multispecies coalescent process (Degnan, 2018; Fogg et al., 2023), $\mathbb{P}[H_i = k] = \gamma_k$, and these variables coincide with the H_i introduced in Def. S1. Individuals 1 and 2 can either coalesce before reaching a parent of u , or not coalesce, in which case G_p consists of a “forest” of two separate gene edges. We denote by T the coalescent time (starting from u) between individuals 1 and 2 at that locus. T has an exponential distribution, and coincide with the variable introduced in Def. S1.

Individuals 1 and 2 do not coalesce before reaching a parent of u if either $H_1 \neq H_2$ (they came from different parents) or if $H_1 = H_2 = k$ and $T > \ell_k$, that is, if event $\bar{C}_u$ from Def. S1 occurs. In this case, each individual 1 and 2 chooses its own ancestor in its own parent population p_{H_1} and p_{H_2} , with traits $Y_{p_{H_1},1'}$ and $Y_{p_{H_2},2'}$, where i' the ancestor individual of i . From these ancestors, they evolve as the Lévy process, so that $Y_{u,i} = Y_{p_{H_i},i'} + \epsilon_i$, with $\epsilon_i \sim \mathcal{D}_{\theta_{H_i}}(\ell_{H_i})$ independent from the rest of the variables. Individual i' might not be sampled at node p_{H_i} , as we assumed that only two individuals were sampled per population. However, since the backward-in-time coalescent process samples the ancestor of i' at random from the parent population p_{H_i} , this ancestor's trait has the same distribution as $Y_{p_{H_i},i} + \epsilon_i$, with $i \in \{1, 2\}$. We then recover exactly (S1) from Def. S1.

Finally, individuals 1 and 2 coalesce if $H_1 = H_2 = k$ and $T \leq \ell_k$, i.e., event $C_{u,k}$ from Def. S1 is true. In this case they both have the same ancestor, picked at random in parent population p_k . Without loss of generality, we can take this ancestor to be J , with $\mathbb{P}[J = 1] = \mathbb{P}[J = 2] = 1/2$. Then the two individuals diverge after a time $\ell_k - T$ from p_k , and, according to the Lévy model of evolution, we get, for $i \in \{1, 2\}$: $Y_{u,i} = Y_{p_k,J} + \delta_0 + \delta_i$, with $\delta_0 \sim \mathcal{D}_{\theta_k}(\ell_k - T)$, $\delta_i \sim \mathcal{D}_{\theta_k}(T)$, and δ_0 , δ_1 and δ_2 mutually independent and independent from other variables. We hence recover (S2) from Def. S1.

To summarize, the joint distribution of all trait variables are equal at all nodes up to u in preorder, which ends the proof. $\square$

Remark S1 (Gaussian marginals). In the special case when $\mathcal{D}_{\theta_k}$ is a centered Gaussian distribution, the graphical model is not jointly Gaussian, due to the random variables H_i and T that are not Gaussian. Marginally, when considering the locus effect of a single randomly sampled individual, $Y_{u,i}$ is not generally Gaussian either. See appendix B for examples. However, as $\epsilon_0 + \epsilon_i \sim \text{Normal}(0, \sigma_{\theta_k}^2 \ell_k)$ does not depend on T , then $Y_{u,i}$ is Gaussian conditional on its ancestor $Y_{v,J}$ if u is a tree node with unique parent v . Consequently, $Y_{u,i}$ is Gaussian conditional on the locus gene tree history, with variance $v_0 + \sum_e \sigma_{\theta_e}^2 \ell_e$ where the sum is over edges e that the gene tree evolved in. For example, if $\mathcal{N}$ is a tree, then each $Y_{u,i}$ is marginally Gaussian, even though the full vector of all variables is not a Gaussian random vector. This property is maintained if the network is time-consistent in variance units $\sigma^2 \ell$, that is, if all paths from the root to u have equal lengths.

Remark S2 (Mutational model). When $\mathcal{D}_{\theta_k}$ is defined by a compound Poisson process, the model is similar to that described in Schraiber and Landis (2015) (who focused on a single evolving population), with the same propagation rules. Here we model ancestral polymorphism with distribution $\mathcal{P}_0$ in the root population ρ , whereas Schraiber and Landis (2015) assume that the trait of all individuals in ρ is fixed at zero.

A.2. Proofs of main results

In this section, we assume that the increment distribution has a finite variance rate, and we present the proofs delayed to this appendix. As before, we omit ρ to simplify notations, but all expectations, variances and covariances are calculated conditional on the trait distribution in ρ .

Proof of Proposition 2 (Recursion formula). Thanks to Prop. S2, the quantities in Prop. 2 are the mean $E_u = \mathbb{E}[Y_{u,i}]$, covariances $\Omega_{u,v} = \text{Cov}[Y_{u,1}; Y_{v,2}]$ and variances $\Phi_u = \text{Var}[Y_{u,1}]$, all conditional on the distribution $\mathcal{P}_0$ in the root population ρ , where $\mathbf{Y} = (Y_{u,i})_{u \in V(\mathcal{N}), i \in \{1,2\}}$ are the random variables defined by the graphical model in Def. S1.

As (4) is a special case for a tree node with $m = 1$ parents, we only need to prove (5).

Consider node u and assume that (3)-(5) hold for its parent nodes. Using notations from Def. S1 we get by Prop. S1 that

$$(S3) \quad \mathbb{E}[Y_{u,i} \mid H_i = k, H_1, H_2, T, J] = \mathbb{E}[Y_{p_k,1}] = m_0$$

where the second equality comes from (3) at parent node p_k . This is a stronger version of (3) at u , since it implies that $\mathbb{E}[Y_{u,i}] = \mathbb{E}[m_0] = m_0$, that is, (3) at u . For variances and covariances, we condition again on the partial gene tree history $G_u = \{H_1, H_2, T, J\}$ describing how u inherits from its parents. First,

$$\text{Var}[Y_{u,i}] = \mathbb{E}[\text{Var}[Y_{u,i} \mid G_u]] + \text{Var}[\mathbb{E}[Y_{u,i} \mid G_u]].$$

By (S3), the second term is 0. For the first term, using Prop. S1, $\mathbb{P}[H_i = k] = \gamma_k$ and the variance of the Lévy process, we get:

$$\text{Var}[Y_{u,i}] = \sum_{k=1}^m \gamma_k (\sigma_{\theta_k}^2 \ell_k + \text{Var}[Y_{p_k,i}])$$

as claimed in (5a). Next, for individuals 1 and 2 in u , we condition again on the partial gene tree history $G_u = \{H_1, H_2, T, J\}$ at u :

$$\text{Cov}[Y_{u,1}; Y_{u,2}] = \mathbb{E}[\text{Cov}[Y_{u,1}; Y_{u,2} \mid G_u]] + \text{Cov}[\mathbb{E}[Y_{u,1} \mid G_u]; \mathbb{E}[Y_{u,2} \mid G_u]].$$

By (S3) again, the second term is 0. For the first term, we condition more specifically on $C_{u,k}$ for $k = 1, \dots, m$ or their complement $\bar{C}_u$. If $\bar{C}_u$ occurs, then from (S1) the increments $Y_{u,i} - Y_{p_{H_i},i}$ are independent ($i = 1, 2$) and their correlation is 0 conditional on G_u and $\bar{C}_u$. If instead $C_{u,k}$ occurs, then from (S2) the increments are $Y_{u,i} - Y_{p_k,i} = \delta_0 + \delta_i$ so their covariance is $\text{Var}[\delta_0] = \sigma_{\theta_k}^2 (\ell_k - T)$. Therefore, we get

$$\begin{aligned} \text{Cov}[Y_{u,1}; Y_{u,2}] &= \mathbb{E} \left[\sum_{k=1}^m (\sigma_{\theta_k}^2 (\ell_k - T) + \text{Var}[Y_{p_k,1}]) \mathbf{1}_{C_{u,k}} \right. \\ &\quad \left. + \sum_{k=1}^m \text{Cov}[Y_{p_k,1}; Y_{p_k,2}] \mathbf{1}_{H_1=H_2=k, T > \ell_k} + \sum_{k \neq k'} \text{Cov}[Y_{p_k,1}; Y_{p_{k'},2}] \mathbf{1}_{H_1=k, H_2=k'} \right] \end{aligned}$$

from which (5b) follows. Finally, if v is not a descendant of u , then the increment between $Y_{u,i}$ and its parent is independent of $Y_{v,j}$. Conditioning on G_u and using (S3) again, we get

$$\text{Cov}[Y_{u,i}; Y_{v,j}] = \mathbb{E} \left[\sum_{k=1}^m \text{Cov}[Y_{p_k,i}; Y_{v,j} \mid G_u] \mathbf{1}_{H_i=k} \right] = \sum_{k=1}^m \gamma_k \text{Cov}[Y_{p_k,i}; Y_{v,j}]$$

which gives (5c) and finishes the proof. $\square$

Next, we turn to proving Prop. 3.

Proof of Proposition 3 (Sampling stability). If networks are allowed to have nodes of degree 2 and if we do not seek to suppress any degree-2 node in $\mathcal{N}_{-t}$ when pruning taxon t from $\mathcal{N}$, then Prop. 3 follows directly from Prop. 2, as the recursion formulas for a given node only depends on its parent nodes, and not its sibling nodes.

Now assume a constant variance rate σ_θ^2 on all the branches. We will show that this assumption allows for the suppression of a node v along a lineage $u \rightarrow v \rightarrow w$, if v is of degree-2 (u and w are its only neighbors), by fusing (u, v) and (v, w) into a single edge of length $\ell_{uw} = \ell_{uv} + \ell_{vw}$ coalescent units, and inheritance $\gamma_{uw} = \gamma_{vw}$.

Take any leaf taxon t in the network. Being a leaf, it has a unique parent v . First assume that v is a tree node. If v is the root ρ , then we just keep this node, so that the conditioning used for all expectations is not affected. Otherwise, v has a unique parent u . If v has more than two children in $\mathcal{N}$, then v is still a tree node with at least 2 children in $\mathcal{N}_{-t}$, and the formulas of Prop. 2 are unchanged in $\mathcal{N}_{-t}$. If v has exactly two children in $\mathcal{N}$, t and w , then v has degree two after removing t , with parent u and child w . Node w can be a leaf or internal, possibly hybrid. Assume that w has $m + 1$ parents in $\mathcal{N}$, that are v and $(p_k)_{1 \leq k \leq m}$ (if w is a tree node then $m = 0$). Then from (5a), (5b) and (5c) in the full network:

$$\begin{aligned}\mathbb{V}\text{ar}[Y_{w,i}] &= \gamma_{vw} \left(\sigma_\theta^2 \ell_{vw} + \mathbb{V}\text{ar}[Y_{v,i}] \right) + \sum_{k=1}^m \gamma_k \left(\sigma_\theta^2 \ell_k + \mathbb{V}\text{ar}[Y_{p_k,i}] \right) \\ &= \gamma_{vw} \left(\sigma_\theta^2 \ell_{vw} + \sigma_\theta^2 \ell_{uv} + \mathbb{V}\text{ar}[Y_{u,i}] \right) + \sum_{k=1}^m \gamma_k \left(\sigma_\theta^2 \ell_k + \mathbb{V}\text{ar}[Y_{p_k,i}] \right) \\ &= \gamma_{uw} \left(\sigma_\theta^2 \ell_{uw} + \mathbb{V}\text{ar}[Y_{u,i}] \right) + \sum_{k=1}^m \gamma_k \left(\sigma_\theta^2 \ell_k + \mathbb{V}\text{ar}[Y_{p_k,i}] \right)\end{aligned}$$

where $\gamma_{uw} = \gamma_{vw}$ and $\ell_{uw} = \ell_{uv} + \ell_{vw}$ are for the fused edge (u, w) after suppressing v . Similarly:

$$\begin{aligned}\mathbb{C}\text{ov}[Y_{w,1}; Y_{w,2}] &= \sum_{k=1}^m \gamma_k^2 \left(\sigma_\theta^2 \ell_k r(\ell_k) + \mathbb{V}\text{ar}[Y_{p_k,1}] q(\ell_k) + \mathbb{C}\text{ov}[Y_{p_k,1}; Y_{p_k,2}] (1 - q(\ell_k)) \right) \\ &\quad + 2 \sum_{k < k' \leq m} \gamma_k \gamma_{k'} \mathbb{C}\text{ov}[Y_{p_k,1}; Y_{p_{k'},2}] + (\alpha) + (\beta)\end{aligned}$$

where

$$(\alpha) = \gamma_{vw}^2 \left(\sigma_\theta^2 \ell_{vw} r(\ell_{vw}) + \mathbb{V}\text{ar}[Y_{v,1}] q(\ell_{vw}) + \mathbb{C}\text{ov}[Y_{v,1}; Y_{v,2}] (1 - q(\ell_{vw})) \right)$$

and

$$(\beta) = 2 \sum_{k=1}^m \gamma_k \gamma_{vw} \mathbb{C}\text{ov}[Y_{p_k,1}; Y_{v,2}].$$

As v is a tree node, $\mathbb{C}\text{ov}[Y_{p_k,1}; Y_{v,2}] = \mathbb{C}\text{ov}[Y_{p_k,1}; Y_{u,2}]$ for any k , and v can be replaced by u in (β) . Also, still using that v is a tree node:

$$\begin{aligned}(\alpha) &= \gamma_{vw}^2 \left(\sigma_\theta^2 \ell_{vw} r(\ell_{vw}) + \mathbb{V}\text{ar}[Y_{v,1}] q(\ell_{vw}) + \mathbb{C}\text{ov}[Y_{v,1}; Y_{v,2}] (1 - q(\ell_{vw})) \right) \\ &= \gamma_{vw}^2 \left(\sigma_\theta^2 \ell_{vw} r(\ell_{vw}) \right. \\ &\quad \left. + [\sigma_\theta^2 \ell_{uv} + \mathbb{V}\text{ar}[Y_{u,1}]] q(\ell_{vw}) \right. \\ &\quad \left. + [\sigma_\theta^2 \ell_{uv} r(\ell_{uv}) + \mathbb{V}\text{ar}[Y_{u,1}] q(\ell_{uv}) + \mathbb{C}\text{ov}[Y_{u,1}; Y_{p,2}] (1 - q(\ell_{uv}))] (1 - q(\ell_{vw})) \right) \\ &= \gamma_{vw}^2 \left(\sigma_\theta^2 [\ell_{vw} r(\ell_{vw}) + \ell_{uv} q(\ell_{vw}) + \ell_{uv} r(\ell_{uv}) (1 - q(\ell_{vw}))] \right. \\ &\quad \left. + \mathbb{V}\text{ar}[Y_{u,1}] [q(\ell_{vw}) + q(\ell_{uv}) (1 - q(\ell_{vw}))] \right. \\ &\quad \left. + \mathbb{C}\text{ov}[Y_{u,1}; Y_{p,2}] (1 - q(\ell_{uv})) (1 - q(\ell_{vw})) \right) \\ &= \gamma_{uw}^2 \left(\sigma_\theta^2 \ell_{uw} r(\ell_{uw}) + \mathbb{V}\text{ar}[Y_{u,1}] q(\ell_{uw}) + \mathbb{C}\text{ov}[Y_{u,i}; Y_{p,j}] (1 - q(\ell_{uw})) \right)\end{aligned}$$

and we recover formulas (5a), (5b) and (5c) applied to the branch going directly from u to w , skipping node v of degree two, that can be suppressed, yielding subnetwork $\mathcal{N}_{-t}$. This ends the proof in the case when v is a tree node.

Now assume that v is a hybrid node. If it has two or more children, then it is still a hybrid node with at least 1 child in $\mathcal{N}_{-t}$, and the formulas of Prop. 2 are again unchanged in $\mathcal{N}_{-t}$. Otherwise t is the only child of v in $\mathcal{N}$, and when t is pruned, v is removed along with the edges from v 's parents p_k to v . Then, each p_k in the subnetwork is either of degree more than two, in which cases it remains, or exactly two, in which case it can be suppressed using the same derivations as above, or exactly one, i.e. it is the root ρ , that we preserve in the subnetwork. This ends the proof, as suppressing all nodes of degree exactly two in the subnetwork leads to the same propagation formulas on the subnetwork. $\square$

Proof of Proposition 5 (BM on a rescaled network). We re-write covariances in Cor. 4 so that they have the form in (9). To prove (10), we write:

$$\begin{aligned}\Omega_{u,u}^{(0)} &= \sum_{k=1}^m \gamma_k^2 \left(q(\ell_k) + \Omega_{p_k,p_k}^{(0)} (1 - q(\ell_k)) \right) + 2 \sum_{k < k'} \gamma_k \gamma_{k'} \Omega_{p_k,p_{k'}}^{(0)} \\ &= \sum_{k=1}^m \gamma_k^2 \left(\Omega_{p_k,p_k}^{(0)} + q(\ell_k) (1 - \Omega_{p_k,p_k}^{(0)}) \right) + 2 \sum_{k < k'} \gamma_k \gamma_{k'} \Omega_{p_k,p_{k'}}^{(0)} \\ &= \sum_{k=1}^m \gamma_k^2 \left(\Omega_{p_k,p_k}^{(0)} + L_{e_k}^{(0)} \right) + 2 \sum_{k < k'} \gamma_k \gamma_{k'} \Omega_{p_k,p_{k'}}^{(0)},\end{aligned}$$

which takes form (9) using the rescaled lengths $L^{(0)}$. The proof of (11) is similar.

Next, take nodes $u \rightarrow v \rightarrow w$ along one lineage, such that v is a tree node and $\ell_{uv} + \ell_{vw} = \ell_{uw}$. Then:

$$\begin{aligned}L_{uv}^{(0)} + L_{vw}^{(0)} &= q(\ell_{uv}) (1 - \Omega_{u,u}^{(0)}) + q(\ell_{vw}) (1 - \Omega_{v,v}^{(0)}) \\ &= q(\ell_{uv}) (1 - \Omega_{u,u}^{(0)}) + q(\ell_{vw}) (1 - (1 - e^{-\ell_{uv}} + \Omega_{u,u}^{(0)} e^{-\ell_{uv}})) \quad \text{from Cor. 4, as } v \text{ is a tree node;} \\ &= (1 - \Omega_{u,u}^{(0)}) (1 - e^{-\ell_{vw} - \ell_{uv}}) \\ &= L_{uw}^{(0)}\end{aligned}$$

so that the branch length transformation (10) is additive. Similarly:

$$\begin{aligned}L_{uv}^{(\mathcal{L})} + L_{vw}^{(\mathcal{L})} &= \ell_{uv} - q(\ell_{uv}) (1 - \tilde{\Phi}_u^{(\mathcal{L})} + \tilde{\Omega}_{u,u}^{(\mathcal{L})}) + \ell_{vw} - q(\ell_{vw}) (1 - \tilde{\Phi}_v^{(\mathcal{L})} + \tilde{\Omega}_{v,v}^{(\mathcal{L})}) \\ &= \ell_{uv} + \ell_{vw} - q(\ell_{uv}) (1 - \tilde{\Phi}_u^{(\mathcal{L})} + \tilde{\Omega}_{u,u}^{(\mathcal{L})}) \\ &\quad - q(\ell_{vw}) (1 - (\ell_{uv} + \tilde{\Phi}_u^{(\mathcal{L})}) + \ell_{uv} r(\ell_{uv}) + \tilde{\Phi}_u^{(\mathcal{L})} q(\ell_{uv}) + \tilde{\Omega}_{u,u}^{(\mathcal{L})} (1 - q(\ell_{uv}))) \quad \text{from Cor. 4;} \\ &= \ell_{uw} - q(\ell_{uw}) (1 - \tilde{\Phi}_u^{(\mathcal{L})} + \tilde{\Omega}_{u,u}^{(\mathcal{L})}) \\ &= L_{uw}^{(\mathcal{L})},\end{aligned}$$

where we used the relationship: $\ell r(\ell) = \ell - \frac{\ell}{e} (1 - e^{-\ell}) = \ell - q(\ell)$. $\square$

The following extends (11) in Prop. 5 to the case when the rate is not constant. It shows that $\Omega^{(\mathcal{L})}$ can still be seen as a BM on the phylogeny, although the transformed branch lengths now depend on the branch parameters $\sigma_{\theta_e}^2$.

Proposition S3 (BM on a rescaled network - variable rate). *The variance matrix $\Omega^{(\mathcal{L})}$ is the same as the one obtained from a BM on the phylogeny, with branch-specific variance parameters $\sigma_{\theta_e}^2$ and rescaled branch lengths:*

$$L_e^{(\mathcal{L})} = \ell_e - q(\ell_e) \left(1 + \frac{\Omega_{p,p}^{(\mathcal{L})} - \Phi_p^{(\mathcal{L})}}{\sigma_{\theta_e}^2} \right).$$

Proof. As in the proof of Prop. 5, we simply re-arrange terms in (7) from Cor. 4 to recover form (9). $\square$

Proof of Proposition 6 (Closed-form expression in the tree case). We use the recursions from Cor. 4 to prove (12) and (14) by induction on nodes u, v listed in preorder. At the root, $t_{\rho\rho} = 0$, so that $\Phi_{\rho}^{(\mathcal{L})} = \Omega_{\rho,\rho}^{(\mathcal{L})} = \Omega_{\rho,\rho}^{(0)} = 0$ and (12) and (14) hold.

Now assume that (12) and (14) hold for all nodes listed before u in preorder, and for its parent node p in particular. To prove (14a), we use (7a), that gives $\Phi_u^{(\mathcal{L})} = \sigma_{\theta}^2 \ell_u + \Phi_p^{(\mathcal{L})}$ where ℓ_u is the length of the branch from p to u in coalescent units. By induction, $\Phi_p^{(\mathcal{L})} = \sigma_{\theta}^2 t_p$ so we get $\Phi_u^{(\mathcal{L})} = \sigma_{\theta}^2 t_u$ from $t_u = t_p + \ell_u$. To prove (12) and (14b), first consider a population $v \neq u$ listed before u . From (7c) and (8a), we get $\Omega_{u,v}^{(\mathcal{L})} = \Omega_{p,v}^{(\mathcal{L})}$ and $\Omega_{u,v}^{(0)} = \Omega_{p,v}^{(0)}$. As v is not a descendant of u by assumption, $t_{uv} = t_{pv}$, and we get (12) and (14b) by induction. Now consider the case $u = v$. Applying (8a) and by induction, we get $\Omega_{u,u}^{(0)} = q(\ell_u) + (1 - e^{-t_{pp}})(1 - q(\ell_u)) = 1 - e^{-t_{pp} - \ell_u}$, which is precisely (12) as $t_{uu} = t_{pp} + \ell_u$. Finally, applying (7b) and by induction, we get $\Omega_{u,u}^{(\mathcal{L})} = \sigma_{\theta}^2 \ell_u r(\ell_u) + \sigma_{\theta}^2 t_p q(\ell_u) + \sigma_{\theta}^2 (t_{pp} - (1 - e^{-t_{pp}})(1 - q(\ell_u)))$. Using the relationships $\ell_u r(\ell_u) = \ell_u - \frac{\ell_u}{\ell_u}(1 - e^{-\ell_u}) = \ell_u - q(\ell_u)$, and again $t_{uu} = t_{pp} + \ell_u$, we get (14b) as expected.

The branch length transformations (13) and (15) then directly follow from Prop. 5. $\square$

A.3. Mean population trait evolution on a network

Recall that $M_u = \mathbb{E}_i[Y_{u,i}]$ denotes the mean trait among all individuals i in population u for a given locus, and $V_u = \text{Var}_i[Y_{u,i}] = \mathbb{E}_i[(Y_{u,i} - M_u)^2]$ its variance within population u . Then the moments of $Y_{u,i}$ are linked to those of M_u and V_u as claimed in Lemma 7, which we prove below.

Proof of Lemma 7. We obtain (16) by re-writing the variances and covariances in a different order. In the following, we use $\mathbb{E}_{\text{coal}}$ to specify that we take the expectation over the coalescent process, and $\mathbb{E}_i$ for the expectation over individuals within a population. With no index, the expectation is taken over the whole process. First, it is easy to see that

$$\mathbb{E}_{\text{coal}}[M_u] = \mathbb{E}_{\text{coal}}[\mathbb{E}_i[Y_{u,i}]] = \mathbb{E}[Y_{u,i}] = m_0.$$

For the covariance, we get:

$$\begin{aligned} \text{Cov}_{\text{coal}}[M_u; M_v] &= \mathbb{E}_{\text{coal}}[(M_u - \mathbb{E}_{\text{coal}}[M_u])(M_v - \mathbb{E}_{\text{coal}}[M_v])] && \text{by definition of the covariance;} \\ &= \mathbb{E}_{\text{coal}}[(\mathbb{E}_i[Y_{u,i}] - m_0)(\mathbb{E}_i[Y_{v,i}] - m_0)] && \text{by definition of } M_u; \\ &= \mathbb{E}_{\text{coal}}[\mathbb{E}_{i,j}[(Y_{u,i} - m_0)(Y_{v,j} - m_0)]] && \text{sampling } i \text{ and } j \text{ independently;} \\ &= \text{Cov}[Y_{u,i}; Y_{v,j}] = \Omega_{u,v} && \text{by definition of the covariance.} \end{aligned}$$

Using similar techniques, we get for the variance:

$$\begin{aligned} \mathbb{E}_{\text{coal}}[V_u] &= \mathbb{E}_{\text{coal}}[\mathbb{E}_i[(Y_{u,i} - M_u)^2]] \\ &= \mathbb{E}_{\text{coal}}[\mathbb{E}_i[(Y_{u,i} - m_0 + m_0 - M_u)^2]] \\ &= \mathbb{E}_{\text{coal}}[\mathbb{E}_i[(Y_{u,i} - m_0)^2]] + \mathbb{E}_{\text{coal}}[\mathbb{E}_i[(M_u - m_0)^2]] - 2\mathbb{E}_{\text{coal}}[\mathbb{E}_i[(Y_{u,i} - m_0)(M_u - m_0)]] \\ &= \text{Var}[Y_{u,i}] + \text{Var}[M_u] - 2\mathbb{E}_{\text{coal}}[(M_u - m_0)(\mathbb{E}_i[Y_{u,i}] - m_0)] \\ &= \text{Var}[Y_{u,i}] + \text{Var}[M_u] - 2\mathbb{E}_{\text{coal}}[(M_u - m_0)(M_u - m_0)] \\ &= \text{Var}[Y_{u,i}] - \text{Var}[M_u] = \Phi_u - \Omega_{u,u}, \end{aligned}$$

which ends the proof. $\square$

B. Non-Gaussian joint distribution

We give here a numerical illustration of Remark S1, which notes that the trait distribution may not be Gaussian given a finite number of loci, even if increments have a Gaussian distribution.

We simulated a trait under the network in Figure S1, controlled by $L = 1$ locus whose effect evolved as a Brownian motion with rate $\sigma_\theta^2 = 1$, starting with variance $v_0 = 0.1$ at the root ρ ($\lambda = 0.1$). Some of these settings are fairly extreme to obtain non-Gaussian distributions, to be 'far' from the Gaussian limit (Prop. 10). Namely, we chose a small $L = 1$, and a phylogeny that is not time-consistent (in variance units) to obtain non-Gaussian marginals. For example, the two paths from the root to the hybrid node have very different lengths: 1.4 and 0.1.

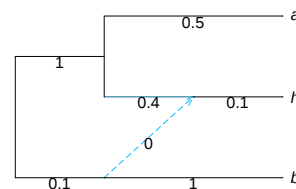


Figure S1 – Network with branch lengths showing coalescent units, used to simulate the traits shown in Figure S2. Both hybrid edges were assigned inheritance $\gamma = 0.5$. In each population, 2 individuals were simulated.

We simulated the trait for 2 individuals per population 100,000 times (Figure S2). As population a is not below any reticulation, each individual a_1 and a_2 has a trait that is marginally Gaussian (histograms not shown). But their joint distribution is clearly not bivariate Gaussian, with non-elliptical contours (Figure S2 left). For individuals h_1, h_2 in population h , their joint distribution is even more clearly non-Gaussian (Figure S2, middle). This is due to the additional variation in gene trees inherited from the different parents at the reticulation. For individual h_1 (or h_2) taken individually, its trait is not Gaussian (Figure S2, right), although mildly so.

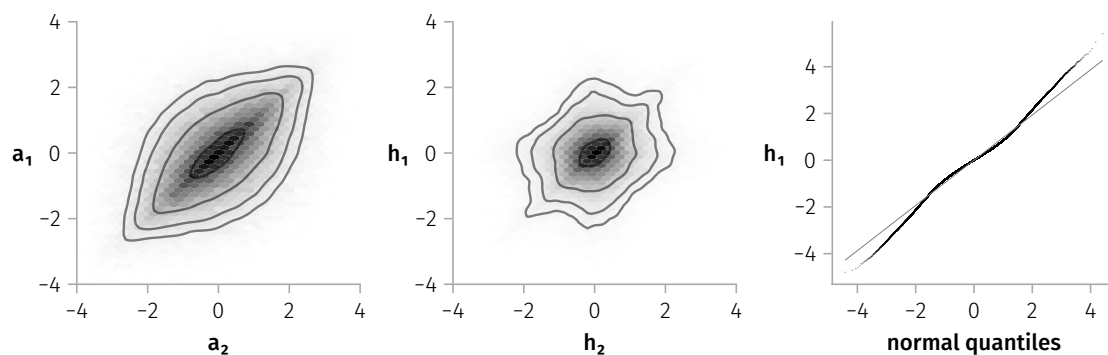


Figure S2 – Joint trait distribution from $L = 1$ locus for two individuals a_1 and a_2 from population a (left panel), or h_1 and h_2 from population h (middle panel), under the network in Figure S1. For each of a_1 and a_2 , the trait followed a Gaussian distribution marginally (not shown). For h_1 (and h_2 , by symmetry), the trait did not follow a Gaussian distribution marginally, as its quantiles do not align with those of a normal distribution (right panel).

C. Limits in small or large populations

We assume here that $\mathcal{N}$ is a tree, and study the limits of the GC and C^* covariances when the haploid effective population size $2N$, assumed constant over $\mathcal{N}$, becomes very small or very large. These two limits correspond to “no ILS” with immediate coalescence when $N \rightarrow 0$, and “maximum ILS” with no coalescence until the root ρ when $N \rightarrow \infty$.

For the GC covariance, we can re-write (18) and (20) from the main text with time in generations instead of coalescent units. We get

$$\Omega_{u,v} = 2N\sigma_{L,\text{gen}}^2 \left(\frac{s_{uv}}{2N} - (1 - e^{-\frac{s_{uv}}{2N}}) \right) + v_0 (1 - e^{-\frac{s_{uv}}{2N}}) \quad \text{and} \quad H_u = 2N\sigma_{L,\text{gen}}^2 (1 - e^{-\frac{s_u}{2N}}) + v_0 e^{-\frac{s_u}{2N}}$$

where $\sigma_{L,\text{gen}}^2 = \sigma_L^2/(2N)$ is the evolution rate per generation, $s_{uv} = 2Nt_{uv}$ is the shared evolutionary time between u and v in generations, and $s_u = s_{uu} = 2Nt_u$ is the number of generations from the root to u . It is then possible to study the limits of these equations when the population becomes very large or very small, either fixing $\sigma_{L,\text{gen}}^2$ or σ_L^2 :

- Under no ILS ($N \rightarrow 0$), the model is equivalent to a BM model on the original tree as

$$\Omega_{u,v} \xrightarrow{N \rightarrow 0} \sigma_{L,\text{gen}}^2 s_{uv} = \sigma_L^2 \frac{s_{uv}}{2N} \quad \text{and} \quad H_u \xrightarrow{N \rightarrow 0} (2N)\sigma_{L,\text{gen}}^2 = \sigma_L^2,$$

with an added within population variation of σ_L^2 when the rate in coalescent unit is fixed rather than the rate in generations. Note that the variance within the root population v_0 is “forgotten” at the limit.

- Under maximum ILS ($N \rightarrow \infty$), the model is equivalent to a single population (or star tree) with independent individuals, as

$$\Omega_{u,v} \xrightarrow{N \rightarrow +\infty} 0 \quad \text{and} \quad H_u \xrightarrow{N \rightarrow +\infty} \sigma_{L,\text{gen}}^2 s_u + v_0 = \frac{\sigma_L^2}{2N} s_u + v_0.$$

Note that, when varying N , the total variance of an individual sampled from population u remains unchanged, equal to $\Phi_u = \sigma_{L,\text{gen}}^2 s_u + v_0$, with time s_u in generations.

For the C^* matrix, we first write (22) from the main text with time in generations:

$$C_{uv}^* = \sigma_{L,\text{gen}}^2 \left[s_{uv} + 2N(1 + \mathbf{1}_{u=v}) - 2N \sum_{k=2}^n \frac{2}{k} p_k \right],$$

and change N to vary the level of ILS in the method by Mendes et al. (2018).

- Under no ILS ($N \rightarrow 0$), C^* converges to the simple BM covariance matrix.
- Under maximum ILS ($N \rightarrow \infty$), the covariances and variances in C^* increase indefinitely. This is because the lineages never coalesce, and the process is conditioned on the individual's locus effect at the last coalescent event, which is infinitely far in the past.

Note that Mendes et al. (2018) characterize “no ILS” and “maximum ILS” differently than by varying N (see Section 2.3.1 of their Appendix). For “no ILS”, they set $p_2 = 1$ and $p_k = 0$ for $k > 2$, which corresponds to no ILS after the root ρ in $\mathcal{N}$ and a regular coalescent process before ρ , in which case their formulas converge to the simple BM with an extra term of $2N\sigma_{L,\text{gen}}^2$ for the variance of the tip traits. For “maximum ILS”, they set $p_n = 1$ and $p_k = 0$ for $k < n$, which corresponds to no coalescent event after the root and a regular coalescent process before ρ , in which case their formulas converge to a simple BM with an extra variance $2N\sigma_{L,\text{gen}}^2$ at the tips, and an extra covariance term $2N\sigma_{L,\text{gen}}^2(1 - 2/n)$ for all pairs of tips.

D. Supplementary figures

D.1. Numerical experiments with *seastaR*

D.1.1. Methods. As detailed in Section 2.6, the package *seastaR* (version 0.1.0 at commit [b23476f](#)) is based on Hibbins et al. (2023) and Mendes et al. (2018), and contains two main functions:

- `get_full_matrix` takes a species tree as input, and implements the formulas of Mendes et al. (2018) on triplets. For two tips in the tree, their covariance is computed as the average covariance of these two tips across all the triplets containing them.
- `trees_to_vcv` takes as an input a set of gene trees relating the populations under study, with branch lengths assumed to be on the same scale (such as coalescent units). It calculates a covariance matrix from each gene tree and returns their average.

In Section 2.6.2, we focused on `get_full_matrix` because it takes the same input as the GC model: the species phylogeny. However, `get_full_matrix` is based on theory for 3-taxon trees, it has not been thoroughly evaluated for trees with more than three species; and it currently uses all triplets from the tree including those that do not span the root. An alternative approach recently described in Mo and Hahn (2026) can apply to a species phylogeny with more than three taxa:

- (1) simulate N_g gene trees from the (network) multispecies coalescent within the input species phylogeny; and
- (2) use these simulated gene trees as input to `trees_to_vcv`.

Currently, this approach is not natively supported by *seastaR*. It requires some light engineering to call an external program to simulate gene trees in step (1).

We reproduced the numerical experiments of Section 2.6.2 using this procedure, simulating $N_g = 1\,000$ gene trees for each species tree used the *julia* package `PhyloCoalSimulations.jl` (v.1.1.0, Ané et al., 2025; Fogg et al., 2023). We also tracked the computation time of each covariance calculation method, after gene trees were simulated (that is, excluding the time to run step (1) above). All scripts are available on the companion repository <https://github.com/cecileane/GCmodel-2026-data-code>.

D.1.2. Results. Figures S3 and S4 reproduce the analyses presented in the main text on Figures 3 and 4, with added results using the `trees_to_vcv` procedure (colored full lines). First, and as expected from Mendes et al. (2018), both *seastaR* procedures give the same results on 3-taxon trees (Figure S3, dark purple). The difference between `get_full_matrix` and `trees_to_vcv` appears for larger trees, and deepens when the number of taxa increases.

Second, the linear decrease with t for the variance of A or the covariance between A and B (Figure S3 and S4, left-most panel) mostly disappears with the `trees_to_vcv` method. This pathological behavior aligns with the problem mentioned by M. Hahn and M. Hibbins in their reviews (see Rogers, 2026) that `get_full_matrix` should only average over triplets that span the root. If a future version of *seastaR* implemented this modification, this would not change the behavior of the method for the variance of C or the covariance of A and C (Figure S3, middle and right panels), as all the triplets containing C and two other tips span the root in the left tree of Figure 2. The same goes for the covariance of C and D in the right panel of Figure S4. Yet, `get_full_matrix` and `trees_to_vcv` still return different C^* matrices in this case, so further improvements are needed for the two methods to return matching results.

The variances and covariances given by `trees_to_vcv` still depend on both the number of “extra” taxa and on the relative internal height t . This dependence is even stronger than that of `get_full_matrix` for (co)variances involving taxon C (for which all triplets span the root, see above). This is particularly visible, for instance, for the covariance of A and C (Figure S3, right panel), that sees a sharp increase when t goes to 0. This could be a consequence of the conditioning of `trees_to_vcv` on the random coalescent time of each gene tree, as a smaller internal branch length leads to an average longer time of shared evolution between A and C. In contrast, the covariance between A and C remains constant and equal to zero for the GC method that condition on the species tree root population, and does not average over covariances computed

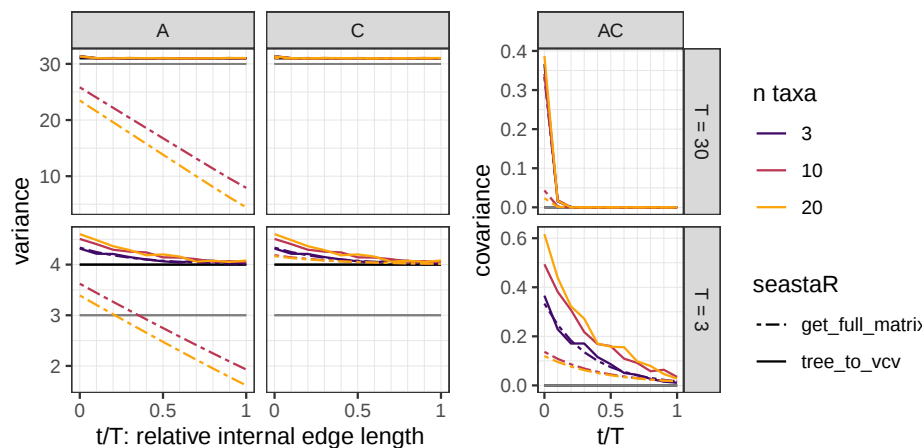


Figure S3 – Effect of taxon sampling and clade's distance to the root on the variance and covariance calculated by various methods on the species tree in Figure 2 (left) with branch lengths (T and t) in coalescent units. **Left:** trait variance at A and at C. **Right:** covariance between the traits at A and C. Colored dashed lines show C^* values from `get_full_matrix`, as in Figure 3. Colored full lines show values from `trees_to_vcv` on $N_g = 1000$ simulated gene trees. The value from the GC model is shown by a black horizontal line (at $T + 1 = 31$ or 4 for the variance, 0 for the covariance). The value from the standard BM model (without ILS) is indicated with a dark gray horizontal line (at $T = 30$ or 3 for the variance, 0 for the covariance).

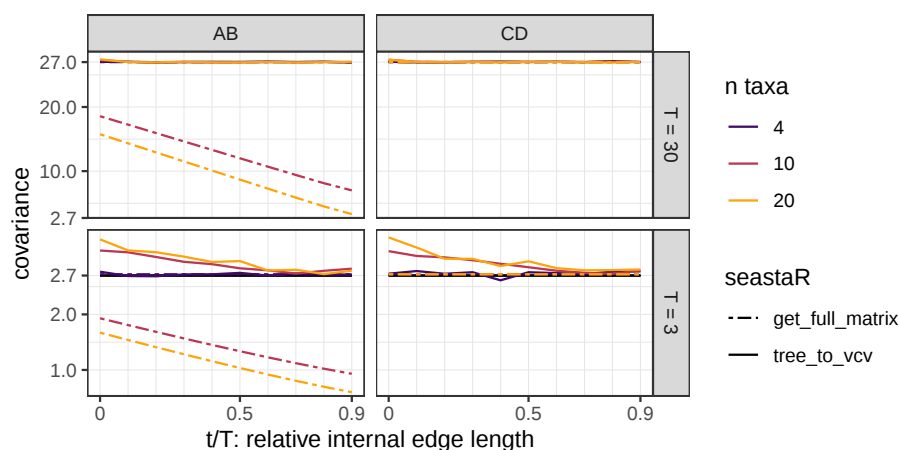


Figure S4 – Effect of taxon sampling and clade's distance to the root on the covariance C^* of two sister pairs, on the species tree in Figure 2, right, with branch lengths (T and t) in coalescent units, as in Figure 4. **Left:** trait covariance between A and B. **Right:** trait covariance between C and D. The value from the GC model is shown by a black horizontal line. It is identical to the value from the standard BM model (without ILS), at $\frac{9}{10}T = 27$ (top) or 2.7 (bottom).

using different conditioning times. A similar pattern can be observed for the covariance of C and D (Figure S4, right panel), where `get_full_matrix` gives a constant covariance that is equal to the covariance computed by the GC method, but the covariance computed with `trees_to_vcv` exhibits some variations both in n and t .

Finally, the two-step procedure using `trees_to_vcv` gives noisy results, especially for larger species trees. This is expected from the procedure itself, which samples a finite number of gene trees (here, $N_g = 1000$). Further analyses could simulate more trees, for instance scaling N_g with the number of populations n , and `trees_to_vcv` should ideally output standard errors so the user

can decide if N_g is sufficiently high, and obtain error bars. Such analyses could however be time consuming, on large phylogenies. Figure S5 shows that the method based on `trees_to_vcv`, and, to a lesser extent, the `get_full_matrix` method, are both much slower than the traditional BM variance calculation method (using `vcv` from the `ape` package) or our GC approach implemented in `phylolm`, and that their timing performances sharply worsen with the number of species. For instance, computing a single covariance matrix for a tree with 20 species takes on average about 1 millisecond for `ape::vcv` and `phylolm` (GC), but about 0.7 second for `get_full_matrix` and around 24 seconds for `trees_to_vcv` (after simulating $N_g = 1\,000$ gene trees).

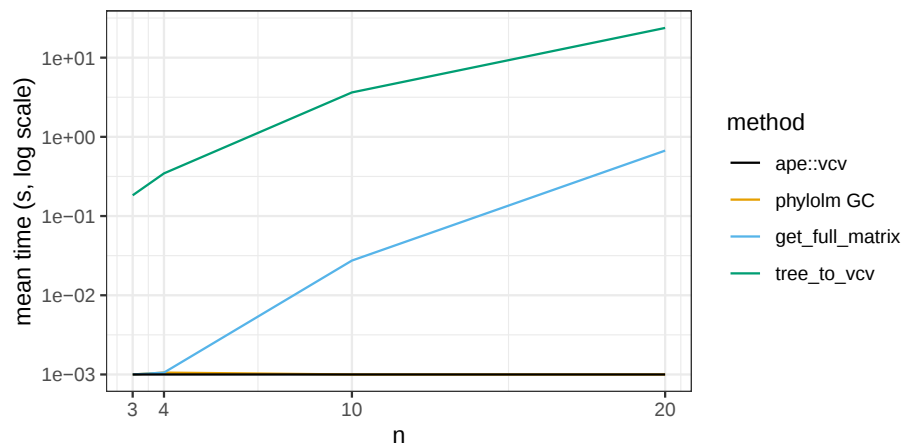


Figure S5 – Average time, in seconds, to compute the correlation matrix for one species tree with n species. For the method based on `trees_to_vcv`, the time to simulate gene trees is not included. Times below the R precision to measure CPU time (1ms) or measured as zeros were arbitrary set to 0.001 seconds.

D.2. Simulations on a network

The following figures show the estimation accuracy of various parameters of our Gaussian-Coalescent model from simulations. Data were simulated under a 17-taxon network with 3 reticulations, 20 individuals per taxon, and traits controlled by $L = 100$ loci with compound Poisson-Laplace effects (see details in Section 3.2). Branch lengths in the network were rescaled for a total height of $T = 10$ for low ILS, and $T = 1$ for high ILS. The true evolutionary variance rate was fixed at $\sigma_L^2 = 1$. We varied the ancestral trait variance in the root population $v_0 = \lambda\sigma_L^2 = \lambda$.

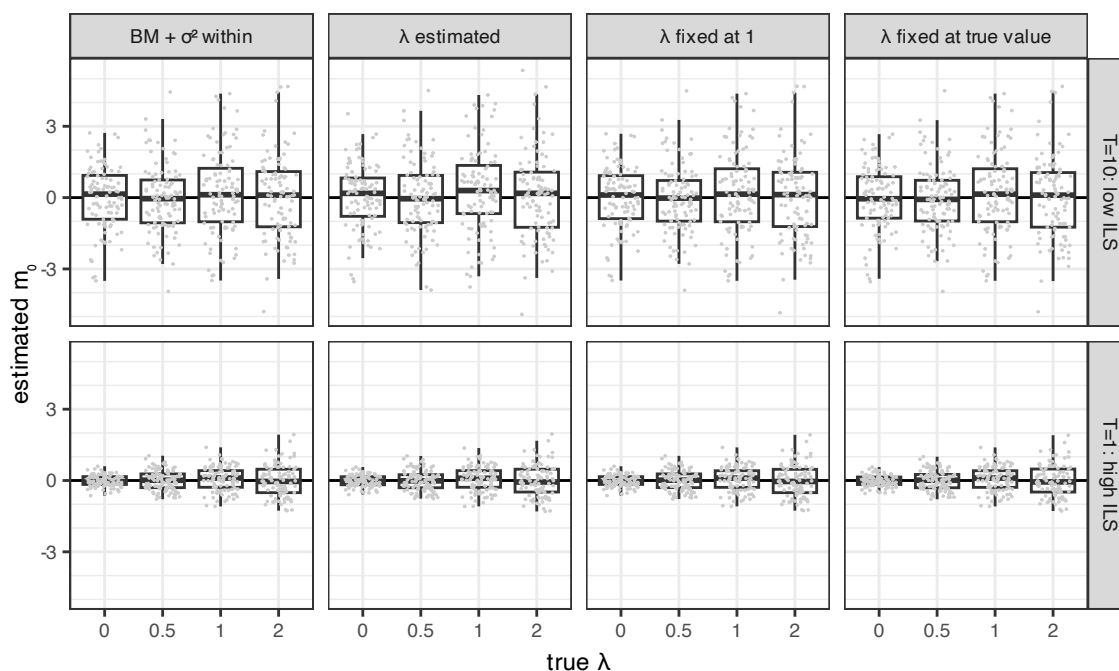


Figure S6 – Estimated mean of the ancestral population, whose true value was $m_0 = 0$. Phylogenetic correlation is higher under low ILS (top) than under high ILS (bottom), resulting in less precision. Under high ILS, estimation error increases with λ because it is affected by the root variance $v_0 = \lambda$.

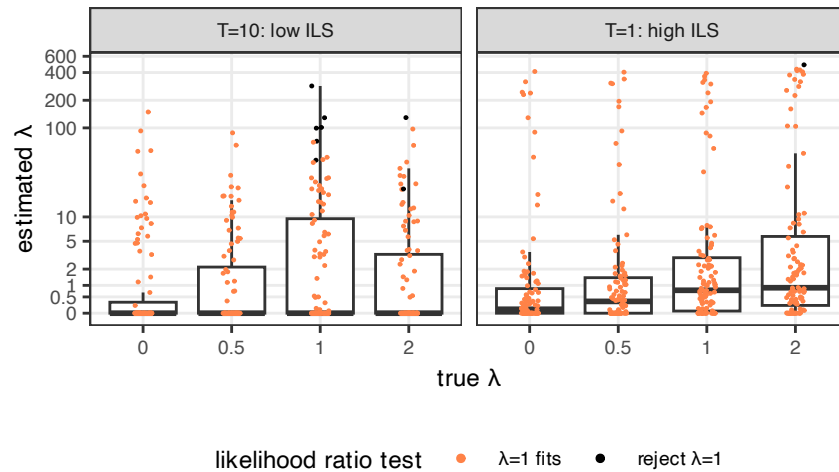


Figure S7 – Estimated $\hat{\lambda}$, displayed on the $\log(\cdot + 1)$ scale, when σ_L^2 and λ were estimated simultaneously. Under low ILS (left), the trait data contain very little information about λ . Estimates are slightly more concentrated under high ILS (right). Under both low and high ILS, the null hypothesis that $\lambda = 1$ is rejected (black dots) very rarely by a likelihood ratio test at the 5% level.

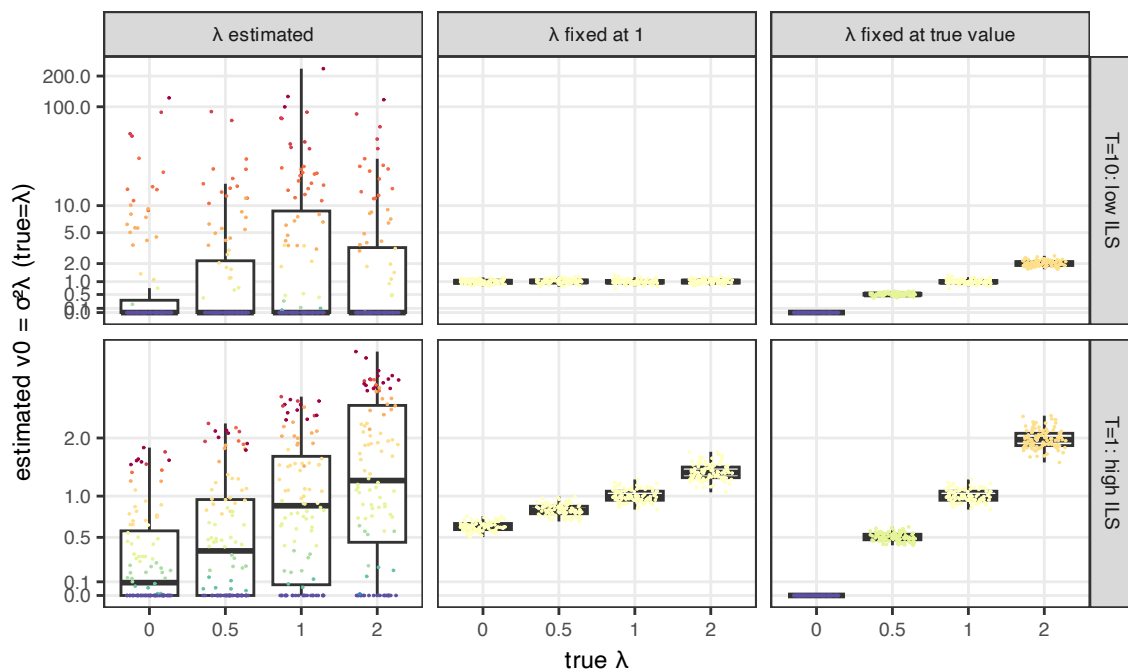


Figure S8 – Estimated ancestral state variance $\hat{v}_0 = \hat{\sigma}_L^2 \hat{\lambda}$, displayed on the $\log(\cdot + 1)$ scale. Note that the top row (low ILS) and bottom row (high ILS) use different limits for the vertical axis. If λ is known correctly (right), v_0 is well estimated. But if λ is unknown, as is realistic in practice, the trait data contain little information about v_0 , especially under low ILS. Colors indicate the estimated (or fixed) λ , with cold (blue) colors for low values and warm (red) colors for large values.

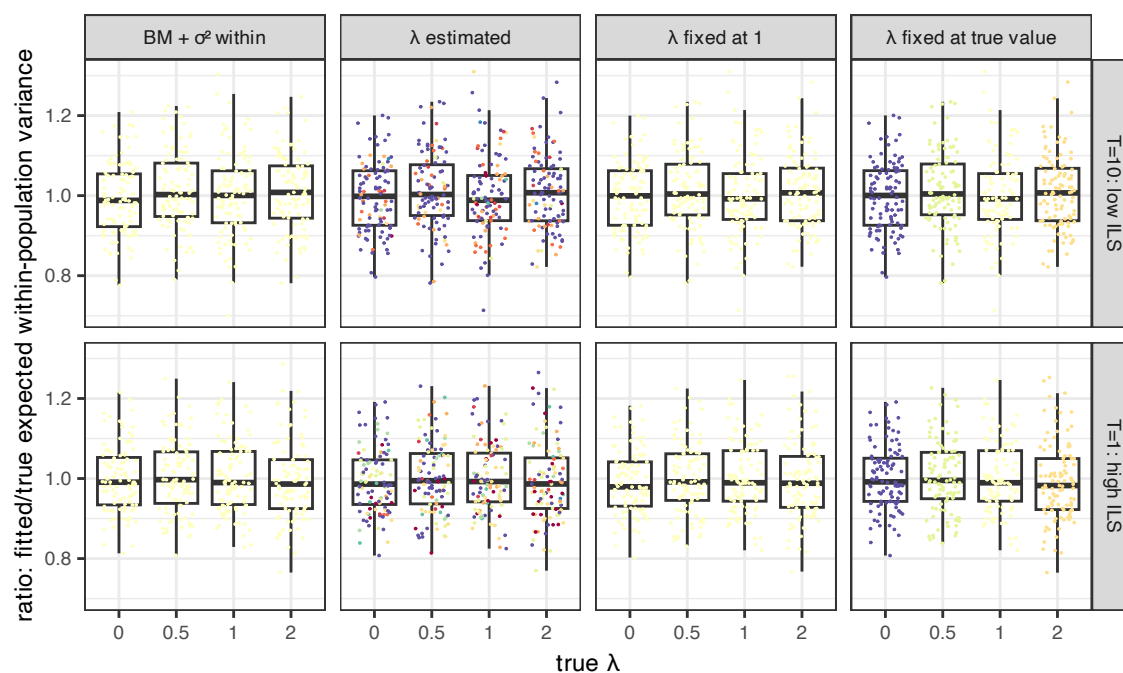


Figure S9 – Ratio $\hat{h}/h$, where h is the average expected within-species variance: expected variance H_u in population u , averaged across all 17 populations. For u not below any reticulation, H_u is given in (20). Generally, H_u can be calculated recursively, using the true parameters. When fitting the data with the GC model, $\hat{h}$ is calculated similarly but using the estimated parameters $\hat{\lambda}$ and $\hat{\sigma}_L^2$. When fitting the data with the BM model, $\hat{h}$ was set to the estimated within-species variance.

D.3. Simulations on the tomato tree

D.3.1. Setting. We reproduced the simulations described in Section 3.2, but using the empirical tomato tree from Section 3.3 instead of the scaled *Polemonium* network. More precisely, we took the tree that was used in the “full dataset” tomato analysis, with 12 populations and a total height of $T = 6$ coalescent units (Figure 7), and kept only 3 individual per populations. We then simulated polygenic traits on the tree using the same set-up as described in Section 3.2, except that we simulated $n_i = 3$ individuals per population.

D.3.2. Results. Results on the tomato tree were on par with those obtained under the “low ILS” setting above. The evolutionary variance rate σ_L^2 was well estimated by all models (Figure S10, top row), but with a higher precision for GC models compared to the BM model: the root mean square error (RMSE) for the GC with λ estimated, fixed to 1 or fixed to its true value were, respectively, equal to 0.127, 0.120 and 0.119, versus 0.269 for the BM. There was again no noticeable gain to knowing the true value of λ . The estimate of the mean m_0 in the ancestral root population was unbiased for all methods (Figure S10, bottom row).

Under the GC model, the estimation of λ was again extremely imprecise (Figure S11). Using a likelihood ratio test at level $\alpha = 0.05$, $\lambda = 1$ was rejected in only 3 of the 400 replicates in all simulation settings (black dots in Figure S11). Using AIC, the GC model with fixed $\lambda = 1$ was favored 80% of the time, followed by the BM model 15% of the time. Estimating λ was deemed worthwhile only 4% of the time (Figure S12).

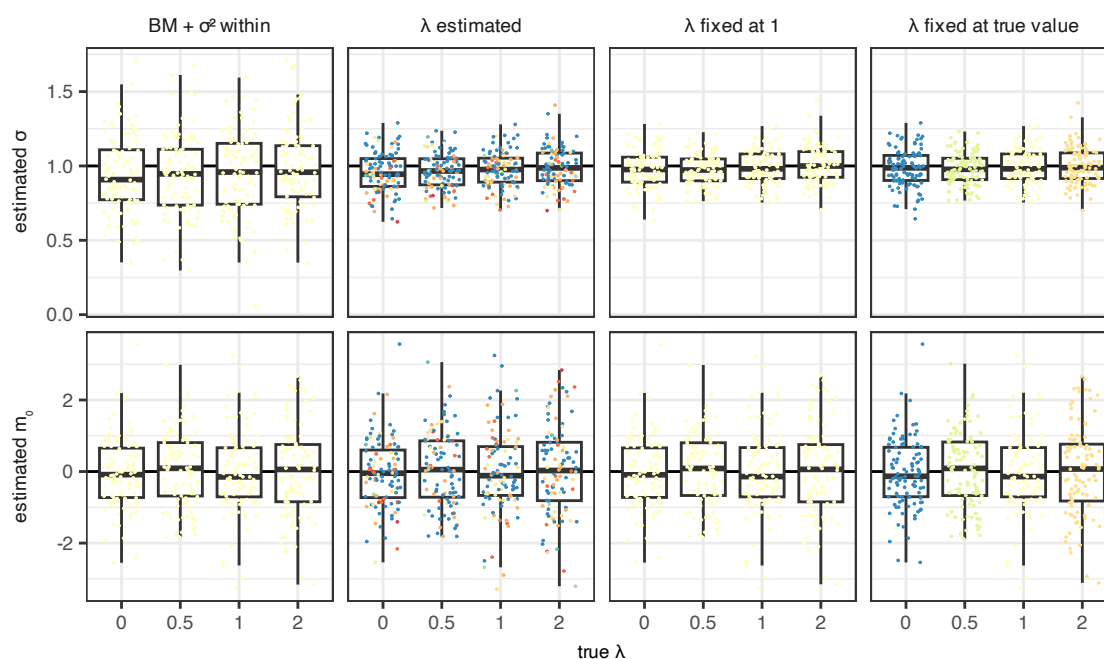


Figure S10 – Estimated standard error rate $\sqrt{\hat{\sigma}_L^2}$ (top) and expectation of the ancestral population $\hat{m}_0$ (bottom) from data simulated under the tomato 12-taxon tree, 3 individuals per taxon, and traits controlled by $L = 100$ loci with compound Poisson-Laplace effects. The true parameters were $\sigma_L^2 = 1$ and $m_0 = 1$. Each column corresponds to one estimation method: the BM model with extra within-population variation, the GC model with λ estimated, fixed at 1, or fixed at its true value (from left to right). Colors indicate the estimated or fixed λ (or yellow for the BM model), with cold colors for low values and warm colors for large values.

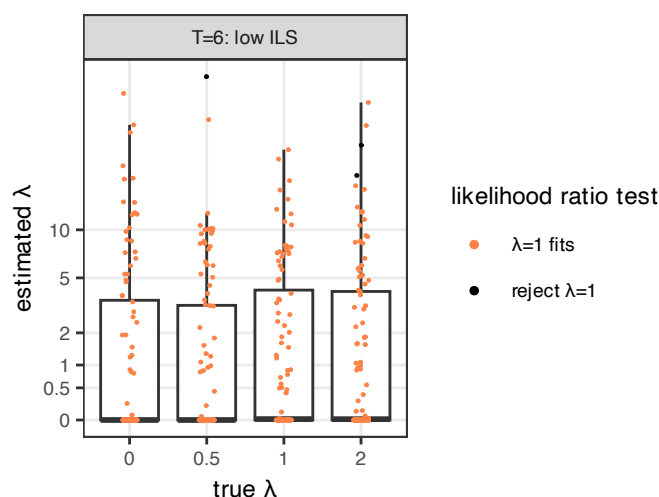


Figure S11 – Estimated $\hat{\lambda}$, displayed on the $\log(\cdot + 1)$ scale, when σ_L^2 and λ were estimated simultaneously. The trait data contain very little information about λ . The null hypothesis that $\lambda = 1$ is rejected (black dots) very rarely by a likelihood ratio test at the 5% level.

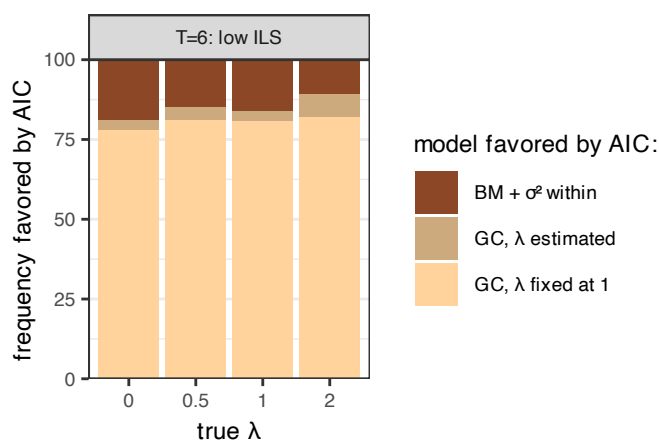


Figure S12 – Proportion of times that each model was favored by AIC, out of 100 replicate data sets, when comparing 3 models: Brownian motion with intra-specific variation, GC with λ estimated simultaneously with σ_L^2 , and GC with $\lambda = 1$ fixed.

References

- Ané C, Fogg J, Justison J, and contributors (2025). *PhyloCoalSimulations.jl*. Version 1.1.0. URL: <https://github.com/JuliaPhylo/PhyloCoalSimulations.jl>.
- Avisé JC, Robinson TJ (2008). Hemipecty: A New Term in the Lexicon of Phylogenetics. *Systematic Biology* **57**, 503–507. <https://doi.org/10.1080/10635150802164587>.
- Bastide P, Ané C (2026). Code for "Trait evolution with incomplete lineage sorting and gene flow: the Gaussian Coalescent model". Version v3. <https://doi.org/10.5281/zenodo.22083886>.
- Bastide P, Solís-Lemus C, Kriebel R, William Sparks K, Ané C (2018). Phylogenetic Comparative Methods on Phylogenetic Networks with Reticulations. *Systematic Biology* **67**, 800–820. <https://doi.org/10.1093/sysbio/syy033>.
- Bulmer MG (1971). The Effect of Selection on Genetic Variability. *The American Naturalist* **105**, 201–211. <https://doi.org/10.1086/282718>.
- Clavel J, Escarguel G, Merceron G (2015). mvMORPH : an R package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution* **6**. Ed. by Timothée Poisot, 1311–1319. <https://doi.org/10.1111/2041-210X.12420>.
- Clavel J, Morlon H (2020). Reliable Phylogenetic Regressions for Multivariate Comparative Data: Illustration with the MANOVA and Application to the Effect of Diet on Mandible Morphology in Phyllostomid Bats. *Systematic Biology* **69**, 927–943. <https://doi.org/10.1093/sysbio/syaa010>.
- Degnan JH (2018). Modeling Hybridization Under the Network Multispecies Coalescent. *Systematic Biology* **67**, 786–799. <https://doi.org/10.1093/sysbio/syy040>.
- Degnan JH, Rosenberg NA (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* **24**, 332–340. <https://doi.org/10.1016/j.tree.2009.01.009>.
- Felsenstein J (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *American journal of human genetics* **25**, 471–492.
- Felsenstein J (1985). Phylogenies and the comparative method. *The American Naturalist* **125**, 1–15. <https://doi.org/10.1086/284325>.
- Felsenstein J (2004). *Inferring Phylogenies*. Sunderland, Massachusetts: Sinauer Associates. 664 pp.
- Fogg J, Allman ES, Ané C (2023). PhyloCoalSimulations: A simulator for network multispecies coalescent models, including a new extension for the inheritance of gene flow. *Systematic Biology* **72**, 1171–1179. <https://doi.org/10.1093/sysbio/syad030>.
- Gautier M, Vitalis R, Flori L, Estoup A (2022). *f*-Statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package poolstat. *Molecular Ecology Resources* **22**, 1394–1416. <https://doi.org/10.1111/1755-0998.13557>.
- Guerrero RF, Hahn MW (2018). Quantifying the risk of hemiplasy in phylogenetic inference. *Proceedings of the National Academy of Sciences* **115**, 12787–12792. <https://doi.org/10.1073/pnas.1811268115>.
- Haak DC, Ballenger BA, Moyle LC (2014). No evidence for phylogenetic constraint on natural defense evolution among wild tomatoes. *Ecology* **95**, 1633–1641. <https://doi.org/10.1890/13-1145.1>.
- Hamlin JAP, Hibbins MS, Moyle LC (2020). Assessing biological factors affecting postspeciation introgression. *Evolution Letters* **4**, 137–154. <https://doi.org/10.1002/evl3.159>.
- Harmon LJ (2019). *Phylogenetic Comparative Methods: Learning From Trees*. Ed. by Luke J. Harmon. Version 1. Center for Open Science. [EcoEvoRxiv preprint]. 196 pp. <https://doi.org/10.32942/osf.io/e3xnr>.
- Hibbins MS, Gibson MJ, Hahn MW (2020). Determining the probability of hemiplasy in the presence of incomplete lineage sorting and introgression. *eLife* **9**, e63753. <https://doi.org/10.7554/eLife.63753>.
- Hibbins MS, Breithaupt LC, Hahn MW (2023). Phylogenomic comparative methods: Accurate evolutionary inferences in the presence of gene tree discordance. *Proceedings of the National Academy of Sciences* **120**, e2220389120. <https://doi.org/10.1073/pnas.2220389120>.

- Hibbins MS, Hahn MW (2021). The effects of introgression across thousands of quantitative traits revealed by gene expression in wild tomatoes. *PLOS Genetics* **17**, e1009892. <https://doi.org/10.1371/journal.pgen.1009892>.
- Ho LST, Ané C (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology* **63**, 397–408. <https://doi.org/10.1093/sysbio/syu005>.
- Karimi N, Grover CE, Gallagher JP, Wendel JF, Ané C, Baum DA (2020). Reticulate Evolution Helps Explain Apparent Homoplasmy in Floral Biology and Pollination in Baobabs (*Adansonia*; Bombacoideae; Malvaceae). *Systematic Biology* **69**, 462–478. <https://doi.org/10.1093/sysbio/syz073>.
- Kingman JFC (1982). On the genealogy of large populations. *Journal of Applied Probability* **19**, 27–43. <https://doi.org/10.2307/3213548>.
- Landis MJ, Schraiber JG, Liang M (2013). Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. *Systematic Biology* **62**, 193–204. <https://doi.org/10.1093/sysbio/sys086>.
- Lynch M, Hill WG (1986). Phenotypic evolution by neutral mutation. *Evolution* **40**, 915–935. <https://doi.org/10.1111/j.1558-5646.1986.tb00561.x>.
- Maddison WP (1997). Gene Trees in Species Trees. *Systematic Biology* **46**, 523–536. <https://doi.org/10.1093/sysbio/46.3.523>.
- Maier R, Flegontov P, Flegontova O, Isildak U, Changmai P, Reich D (2023). On the limits of fitting complex models of population history to f-statistics. *Elife* **12**, e85492. <https://doi.org/10.7554/elife.85492>.
- Mary-Huard T, Balding D (2023). Fast and accurate joint inference of coancestry parameters for populations and/or individuals. *PLOS Genetics* **19**, e1010054. <https://doi.org/10.1371/journal.pgen.1010054>.
- Mendes FK, Hahn MW (2017). Data from: Why concatenation fails near the anomaly zone. [Dryad dataset]. <https://doi.org/10.5061/dryad.j6000>.
- Mendes FK, Fuentes-González JA, Schraiber JG, Hahn MW (2018). A multispecies coalescent model for quantitative traits. *eLife* **7**, e36482. <https://doi.org/10.7554/eLife.36482>.
- Mendes FK, Hahn MW (2018). Why Concatenation Fails Near the Anomaly Zone. *Systematic Biology* **67**, 158–169. <https://doi.org/10.1093/sysbio/syx063>.
- Mo YK, Hahn MW (2026). Estimating the rate of quantitative trait evolution in the presence of gene tree discordance by calculating likelihoods across trees. *Evolutionary Journal of the Linnean Society* **5**, kzag004. <https://doi.org/10.1093/evolinnean/kzag004>.
- Morley L, Lehmborg E, Kong S (2026). Evaluating Phylogenetic Comparative Methods under Reticulate Evolutionary Scenarios. *Systematic Biology*, syag064. <https://doi.org/10.1093/sysbio/syag064>.
- Nielsen SV, Vaughn AH, Leppälä K, Landis MJ, Mailund T, Nielsen R (2023). Bayesian inference of admixture graphs on Native American and Arctic populations. *PLOS Genetics* **19**, 1–22. <https://doi.org/10.1371/journal.pgen.1010410>.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D (2012). Ancient Admixture in Human History. *Genetics* **192**, 1065–1093. <https://doi.org/10.1534/genetics.112.145037>.
- Pease JB, Haak DC, Hahn MW, Moyle LC (2016). Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLOS Biology* **14**, e1002379. <https://doi.org/10.1371/journal.pbio.1002379>.
- Pickrell JK, Pritchard JK (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics* **8**, e1002967. <https://doi.org/10.1371/journal.pgen.1002967>.
- Rannala B, Yang Z (2003). Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci. *Genetics* **164**, 1645–1656. <https://doi.org/10.1093/genetics/164.4.1645>.
- Revell LJ, Harmon LJ (2022). *Phylogenetic Comparative Methods in R*. Princeton, Oxford: Princeton University Press.

- Rogers A (2026). Evolution of a Quantitative Character in a Phylogenetic Network. *Peer Community in Mathematical and Computational Biology*, 100819. <https://doi.org/10.24072/pci.mcb.100819>.
- Rose JP (2021). Taxonomy and relationships within *Polemonium foliosissimum* (Polemoniaceae): untangling a clade of colorful and gynodioecious herbs. *Systematic Botany* **46**, 519–537. <https://doi.org/10.1600/036364421X16312067913372>.
- Schraiber JG, Edge MD, Pennell M (2024). Unifying approaches from statistical genetics and phylogenetics for mapping phenotypes in structured populations. *PLOS Biology* **22**, e3002847. <https://doi.org/10.1371/journal.pbio.3002847>.
- Schraiber JG, Landis MJ (2015). Sensitivity of quantitative traits to mutational effects and number of loci. *Theoretical Population Biology* **102**, 85–93. <https://doi.org/10.1016/j.tpb.2015.03.005>.
- Schrempf D, Szöllösi G (2020). The Sources of Phylogenetic Conflicts. In: *Phylogenetics in the Genomic Era*. Ed. by Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. No commercial publisher | Authors open access book, 3.1:1–3.1:23. URL: <https://hal.science/hal-02535482>.
- Steel M (2016). *Phylogeny: Discrete and Random Processes in Evolution*. Philadelphia, PA: Society for Industrial and Applied Mathematics, p. 302. <https://doi.org/10.1137/1.9781611974485>.
- Tabatabaee Y, Claramunt S, Mirarab S (2026). Coalescent-Based Branch Length Estimation Improves Dating of Species Trees. *Systematic Biology*, syag038. <https://doi.org/10.1093/sysbio/syag038>.
- Teo B, Rose JP, Bastide P, Ané C (2022). Supplementary material for: Accounting for within-species variation in continuous trait evolution on a phylogenetic network. <https://doi.org/10.5061/DRYAD.9GHX3FFKC>.
- Teo B, Rose JP, Bastide P, Ané C (2023). Accounting for within-species variation in continuous trait evolution on a phylogenetic network. *Bulletin of the Society of Systematic Biologists* **2**, 1–29. <https://doi.org/10.18061/bssb.v2i3.8977>.
- Wagenmakers EJ, Farrell S (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* **11**, 192–196. <https://doi.org/10.3758/BF03206482>.
- Wang Y, Cao Z, Ogilvie HA, Nakhleh L (2021). Phylogenomic assessment of the role of hybridization and introgression in trait evolution. *PLOS Genetics* **17**, e1009701. <https://doi.org/10.1371/journal.pgen.1009701>.
- Xu J, Ané C (2024). A consistent least-squares criterion for calibrating edge lengths in phylogenetic networks. [arXiv preprint]. <https://doi.org/10.48550/arXiv.2407.19343>.