Phylogenomic approaches to detecting and characterizing introgression

Mark S. Hibbins* and Matthew W. Hahn*,†

*Department of Biology and †Department of Computer Science, Indiana University, Bloomington, IN 47405

Abstract

Phylogenomics has revealed the remarkable frequency with which introgression occurs across the tree of life. These discoveries have been enabled by the rapid growth of methods designed to detect and characterize introgression from whole-genome sequencing data. A large class of phylogenomic methods makes use of data from one sample per species to infer introgression based on expectations from the multispecies coalescent. These methods range from simple tests, such as the *D*-statistic, to model-based approaches for inferring phylogenetic networks. Here, we provide a detailed overview of the various signals that different modes of introgression are expected leave in the genome, and how current methods are designed to detect them. We discuss the strengths and pitfalls of these approaches and identify areas for future development, highlighting the different signals of introgression and the power of each method to detect them. We conclude with a discussion of current challenges in inferring introgression and how they could potentially be addressed.

Introduction

The potential for hybridization and subsequent backcrossing between lineages—also known as introgression—has long been understood (Heiser 1949, Heiser 1973, Rieseberg and Wendel 1993, Dowling and Secor 1997). Recent hybridization often leads to clear genome-wide patterns in hybrid individuals because they are the result of reproduction between two previously isolated lineages or species. This allows for the detection of F_1, F_2, and early back-cross hybrids from limited sequence data (Nason and Ellstrand 1993, Miller 2000, Anderson and Thompson 2002). However, many generations of back-crossing can substantially reduce the number of loci retaining a history of hybridization, rendering more ancient hybridization events difficult to detect. As a result, until genome sequencing became widely available to biologists, it was often difficult to quantify patterns of introgression effectively and reliably. In part precipitated by the discovery of introgression between archaic human populations (Green et al. 2010, Huerta-Sanchez et al. 2014), the past decade has seen an explosive increase in the rate of discovery of reticulate evolution across the tree of life (Mallet et al. 2016, Taylor and Larson 2019). Although great efforts have been made in recent years to synthesize the biological implications of these discoveries (Hedrick 2013, Ellstrand et al. 2013, Harrison and Larson 2014, Racimo et al. 2015, Ottenburghs et al. 2017, Suarez-Gonzalez et al. 2018, Dagilis et al. 2021), comparatively little synthesis has been provided on the accompanying growth in methods used to detect and characterize introgression.

Modern studies of introgression are often predicated on "phylogenomic" datasets. These typically consist of whole-genome or whole-transcriptome sequencing data, collected from a single individual in at least three populations or species. Gene trees can be constructed from alignments of individual loci or non-overlapping genomic windows (neither of which necessarily contain protein-coding genes), resulting in a collection of thousands of tree topologies; most methods also require a species tree to be inferred from the same data. A common finding from phylogenomic studies is the ubiquity of gene tree discordance—topologies from different loci will disagree with both each other and with the inferred species tree (e.g. Pollard et al. 2006, Fontaine et al. 2015, Pease et al. 2016, Novikova et al. 2016, Edelman et al. 2019). Although the gene tree topologies from neighboring loci are more likely to be similar (Slatkin and Pollack 2006), discordance occurs even between neighboring loci, as recombination uncouples the history of flanking genomic windows.

It is often difficult to uncover the processes leading to discordance at a single locus. When many loci are sampled in a phylogenomic framework, it becomes possible to learn about the general factors causing discordance in a dataset, allowing for introgression to be distinguished from other processes that generate gene tree heterogeneity. Data from a rooted triplet of species—or an unrooted quartet—is the minimum requirement to carry out powerful tests for introgression using genome-scale datasets. Importantly, this can be done using only a single haploid sequence per species (here, we use the term "species" loosely to refer to any lineage or population which shows evidence of historical long-term isolation from other such lineages) and without strong assumptions about neutrality. The robustness to non-neutral processes in some methods occurs

because much of the genealogical signal of introgression is not mimicked by selection (Przeworski et al. 1999, Williamson and Orive 2002, Vanderpool et al. 2020). Phylogenomic methods include the *D* statistic (also known as the ABBA-BABA test; Green et al. 2010, Durand et al. 2011), its numerous analogs and extensions (see below), methods based on pairwise sequence divergence such as the D_3 statistic (Hahn and Hibbins 2019), and phylogenetic network approaches such as those implemented in *PhyloNet* (Than et al. 2008, Wen et al. 2018), *SNaQ* (Solís-Lemus and Ané 2016), and *SpeciesNetwork* (Zhang et al. 2018).

In this review, we focus on phylogenomic methods for studying introgression, most of which are based on the multispecies coalescent model. We provide a detailed overview of the signals that various introgression scenarios are expected to leave in the genome, highlighted by a small simulation study, and the methods that are designed to detect these signals. We discuss common misuses and misinterpretations of these methods, and provide recommendations for best-use practices. Based on these results, we identify areas for future theoretical and methodological advancement, as well as the challenges that remain for visualizing and interpreting current methods.

Biological processes that generate gene tree heterogeneity

We begin our discussion of phylogenomic methods with the simplest possible sampling scheme: genomic data from a single sampled haploid individual from each of three focal species and an outgroup. By "genomic data" we mean data sampled from many loci across the genome, often with the standard assumption of no intra-locus recombination and free inter-locus recombination. This data structure will hereafter be referred to as a quartet or rooted triplet. For three ingroup species, *P1*, *P2*, and *P3*, and an outgroup species, *O*, there are three possible tree topologies describing how they can be related: (((*P1*,*P2*),*P3*),*O*), (((*P2*,*P3*),*P1*),*O*), or (((*P1*,*P3*),*P2*),*O*) (Figure 1). In addition to a single phylogeny describing the evolutionary history of the quartet, trees can be constructed for each individual locus. The frequencies of each topology across loci are referred to as gene tree frequencies, even when they do not come from protein-coding genes. This heterogeneity in both the topology and branch lengths of gene trees is caused by two different biological processes: incomplete lineage sorting and introgression. Below we describe the expected effects of both processes in order to understand how tests for introgression work.

Incomplete lineage sorting as a null hypothesis for tests of introgression

The phenomenon of incomplete lineage sorting (ILS), in which two or more lineages fail to coalesce in their most recent ancestral population (looking backwards in time), can result in individual gene trees that are discordant with the species history (Figure 1). Phylogenomic methods must account for this phenomenon to make accurate inferences about introgression. Discordant gene trees occur because, when ILS occurs, it becomes possible for the order of coalescent events to differ from the order of splits in the species phylogeny (Figure 1, top right panel). Gene tree discordance due to ILS is very common in modern phylogenomic datasets (e.g. Pollard et al. 2006, Fontaine et al. 2015, Pease et al. 2016, Novikova et al. 2016, Copetti et al. 2017, Wu et al. 2018a; Edelman et al. 2019) and can arise within phylogenies that contain no introgression events. Because both ILS and introgression can generate many of the same

genealogical patterns, it is essential to incorporate ILS into the null hypothesis of tests for introgression.

Fortunately, the effects of the parameters mostly likely to influence the probability of ILS—time between speciation events and ancestral population size—are well understood from the multispecies coalescent (MSC) model (Hudson 1983, Tajima 1983, Pamilo and Nei 1988). For a rooted triplet, the probability that the two sister lineages (e.g. *P1* and *P2* in Figure 1) coalesce in their most recent common ancestral population is given by the formula $1 - e^{-\tau}$, where τ is the length of this internal branch in units of 2*N* generations (sometimes referred to as "coalescent units"). Conversely, the probability of ILS (i.e. that they do not coalesce) is $e^{-\tau}$. If ILS occurs, all three lineages (*P1*, *P2*, and *P3*) enter their joint ancestral population. Within this population the coalescent events happen at random, such that lineages leading to each pair of species have a 1/3 chance of coalescing first. This means that the two discordant gene tree topologies are expected to be equal in frequency (Figure 1, top right), with probabilities of $1/3\, e^{-\tau}$ each. In addition, the concordant tree topology can be produced either by lineage sorting with probability $1 - e^{-\tau}$ or incomplete lineage sorting with probability $1/3\, e^{-\tau}$ (Figure 1, top left). This guarantees that the concordant tree topology will always be at least as frequent as the two discordant trees (Figure 1, top row). These expectations under ILS form the null hypothesis for tests of introgression based on gene tree frequencies.

In addition to gene tree frequencies, ILS affects expected coalescence times, and therefore sequence divergence, between pairs of species. In any population, the expected times to coalescence depends on how many lineages are present (Kingman 1982, Hudson 1983, Tajima 1983). If three lineages are present, the first coalescence is expected to occur $2/3\, N$ generations in the past. After this first coalescence—or if only two lineages were present to begin with—the next coalescence is expected a further 2*N* generations in the past. These expectations are equally applicable to current populations as to ancestral populations, but coalescence cannot occur until the lineages under consideration are in a common population. Therefore, expected coalescence times between species always have the time of speciation included as a constant, no matter how far back lineage-splitting occurred (Gillespie and Langley 1979).

For example, the time to coalescence between species *P1* and *P2* in Figure 1 is expected to be 2*N* generations prior to their speciation event. If this coalescent event happens in their most recent common ancestral population (i.e. lineage sorting), then the next coalescent event occurs between the resulting single lineage and the lineage leading to *P3* in the common ancestral population of all three species (Figure 1, bottom row). This event is again 2*N* generations prior to the speciation event between *P3* and the common ancestor of *P1*+*P2*. If ILS occurs, then the first coalescence (regardless of which lineages are involved) occurs $2/3\, N$ generations prior to this same speciation event, and the second coalescence 2*N* generations before this. Note that, if we condition on lineage sorting having occurred, the expected coalescence times become slightly more complicated (see Mendes and Hahn 2018, Hibbins and Hahn 2019 for exact expectations)

The two pairs of non-sister lineages in a rooted triplet (*P1* and *P3* or *P2* and *P3* in Figure 1) can coalesce at one of two times, depending on whether they are the first or second pair to coalesce in a gene tree (there can only be a discordant topology if they are the first to coalesce). Owing to

the symmetry of gene tree topology shapes and frequencies, these times are equivalent across loci, leading to the null expectation under ILS that genome-wide divergence between both pairs of non-sister taxa should be equal (Figure 1, bottom row). Finally, each of these coalescence times is expected to follow a unimodal exponential distribution under ILS alone (Hudson 1983, Tajima 1983).

The effects of introgression on gene trees

Introgression between two lineages occurs when an initial hybridization event is followed by back-crossing into one or both of the parental lineages. Hybridization itself—the creation of a hybrid individual—is generally not sufficient to be called introgression, though polyploid or homoploid hybrid species will be identified by many of the same tests described here (e.g. Meng and Kubatko 2009; Blischak et al. 2018; Folk et al. 2018). Similarly, horizontal gene transfer will also generate discordant gene trees, but introgression is generally distinguished from this process by the requirement that there be mating between the hybridizing lineages in order to be considered introgression. This mating requirement means that phylogenetically distant species are unlikely to be closely related at individual loci due to introgression. Horizontal gene transfer, on the other hand, can produce highly discordant topologies that can only be produced by the interspecific exchange of genetic material (e.g. Knowles et al. 2018).

There are a large number of different introgression scenarios, each with a different effect on the underlying gene trees. While there are well-developed mathematical tools that describe the effects of introgression on gene tree topologies (e.g. the multispecies network coalescent; reviewed in Degnan 2018, Elworth et al. 2019), we generally do not need the predictions from these models to test for the presence of introgression (with some exceptions discussed below). Instead, because our tests are often simply looking for a rejection of the ILS-only model (see previous section for a description of expected patterns under ILS alone), a general understanding of the key outcomes of introgression will be sufficient. Figure 2 summarizes the scenarios involving introgression that are most commonly encountered.

As a first key distinction, introgression can occur either between sister lineages (events 1 and 2 in Figure 2A) or non-sister lineages (events 3, 4, and 5 in Figure 2A). As a general rule, introgression between sister lineages should increase the proportion of concordant gene trees relative to the case of ILS alone. To see why this is, consider introgression event 1 in Figure 2: gene flow after speciation between *P1* and *P2* effectively increases τ, the length of the internal branch separating these two lineages from their common ancestor with *P3*. This is because *P1* and *P2* can now be more closely related at introgressed loci than in the species phylogeny. As discussed in the previous section, the rate of ILS is inversely proportional to the value of τ. Loci with an introgressed history therefore have a reduced probability of ILS because of the increased time for *P1* and *P2* to coalesce. While there are some exceptions to this rule—all of which involve introgression between sister lineages on an internal branch of the species tree (i.e. event 2 in Figure 2; Solís-Lemus et al. 2016, Long and Kubatko 2018, Jiao and Yang 2020)—in no cases should gene flow between sister lineages result in one discordant topology becoming more common than the other discordant topology.

Because an increase in concordant topologies can also be generated under an ILS-only model with a longer internal branch in the species tree, gene tree frequencies alone cannot tell us whether introgression has occurred between sister lineages. Note, however, that loci with a history of introgression can have a different distribution of branch lengths in this scenario than expected under ILS alone: the coalescence times are more recent than expected under ILS for either event 1 or 2 (Figure 2B). Our ability to determine whether the distribution of branch lengths is due to a history of introgression partly depends on whether gene flow is continuously occurring after speciation or occurs as a single pulse of hybridization and backcrossing at a period considerably after speciation: pulses of introgression following secondary contact between species will almost always be easier to detect (see section on "*Detecting introgression using coalescence times*"). Using only a single haploid sequence from each species, we also cannot determine the direction of gene flow between sister lineages; this is why we have drawn events 1 and 2 as bidirectional introgression. In order to make this determination between sister species we must use population genetic methods (e.g. Schrider et al. 2018).

When introgression occurs between non-sister lineages (events 3, 4, and 5 in Figure 2A) then one discordant tree topology can become more common than the other. The resulting asymmetry in discordant tree topologies is one of the clearest signals of introgression. In both events 3 and 4 we expect loci that have introgressed to be more likely to have a gene tree topology placing *P2* and *P3* sister to one another: ((*P2*,*P3*),*P1*) (Figure 2C). While not all loci following this introgression history will have this discordant topology, the extended period of shared history between *P2* and *P3* makes it more likely for these lineages to coalesce. In general, the strength of the asymmetry in discordant topologies will depend on the net rate, timing, and direction of introgression (Durand et al. 2011; Martin et al. 2015; Zheng and Janke 2018), as well as the absence of introgression between the other non-sister pair (in which case the other discordant topology would also go up in frequency). Although the same discordant topology will be produced in excess by events 3 and 4 (Figure 2C), note that the resulting branch lengths will differ on average between the two. This difference makes it possible to determine the main direction of introgression between non-sister taxa (see below). Note that while we have drawn gene flow as unidirectional to highlight the fact that this distinction can be made, bidirectional gene flow between these lineages is equally biologically plausible.

Detecting introgression using gene tree frequencies

The D *statistic*

A widely used method for inferring introgression is the *D* statistic, or—perhaps because there are already so many *D*'s in use—what is commonly referred to as the ABBA-BABA test (Green et al. 2010). The statistic quantifies biallelic site patterns produced by introgression between non-sister taxa as a proxy for gene tree frequencies. Because it is just using site patterns, it avoids the need to infer gene trees from individual blocks of the genome; the test was originally formulated to test for evidence of gene flow between Neanderthals and archaic humans (Green et al. 2010, Durand et al. 2011), where reconstructing full gene trees may not have been feasible. Possibly as a result of this minimal requirement, it is the most commonly used test for introgression (Dagilis et al. 2021).

The *D* statistic counts the occurrence of two configurations of shared derived alleles across three species and an outgroup. Assuming the species tree (((*P1*,*P2*),*P3*)*O*), and denoting the ancestral allele as "A" and the derived allele as "B," there are two parsimony-informative patterns of discordant sites. The pattern “ABBA” represents sites where *P2* and *P3* share a derived allele, while *P1* and the outgroup have the ancestral allele. The pattern “BABA” represents sites where *P1* and *P3* share a derived allele, to the exclusion of *P2* and the outgroup (Figure 3). For clarity, note that sites supporting the species topology would have the pattern BBAA; however, these are not used in this statistic.

The *D* statistic assumes an infinite sites model, meaning that the two discordant site patterns can only arise via single mutations on the internal branches of discordant gene trees (Figure 3, blue dots/branches). Under this assumption, the frequencies of ABBA and BABA site patterns summed across many genomic loci are expected to reflect the frequencies of underlying gene trees. If the number of ABBA and BABA sites differ significantly, then an asymmetry in gene tree topologies is inferred, with introgression occurring between the species sharing the derived state more frequently. Figure 3 depicts the scenario when the site pattern ABBA is more common, implying introgression between *P2* and *P3*.

To make it comparable across studies, the value of the *D* statistic is typically reported after normalization using the sum of ABBA and BABA pattern counts, giving the following formula:

$$D = \frac{ABBA - BABA}{ABBA + BABA}$$

where ABBA and BABA represent the number of sites of each type. This statistic has an expected value of $D = 0$ if there is no gene flow (see “High ILS” simulation condition; Supplementary Figures 2, 3). When used as a whole-genome test of introgression between non-sister taxa, the *D*-statistic is robust under many different scenarios (Zheng and Janke 2018, Kong and Kubatko 2021), but can be affected by certain forms of ancestral population structure (Slatkin and Pollack 2008, Durand et al. 2011, Lohse and Frantz 2014) (see section entitled "*Distinguishing introgression from ancestral population structure*" for more discussion of this issue).

Despite the widespread popularity and relative robustness of *D*, there are several important considerations and limitations to its use, some of which are often overlooked. The first of these concerns how to properly test the null hypothesis that $D = 0$. The expected site pattern counts of the *D*-statistic can easily be calculated, so it may be tempting to use a parametric test for differences. However, such tests assume that individual observations represent independent samples: this assumption is violated because closely spaced sites often share the same underlying local genealogy, making them non-independent. The pseudoreplication that results from treating all sites independently leads to inaccurate *p*-values. The solution to this issue is to use a block-bootstrap (or block-jackknife) approach to estimate the sample variance and then to calculate the *p*-value (Green et al. 2010). This approach correctly accounts for correlations within blocks of adjacent sites.

Although formulated as a single genome-wide test, there are cases where the *D*-statistic has been applied to look for introgression in smaller genomic windows (e.g. Kronforst et al. 2013, Zhang et al. 2016, Wu et al. 2018b, Grau-Bové et al. 2020). However, the genome-wide expectation under ILS alone that $D = 0$ does not hold true for smaller genomic windows. Since a single non-recombining locus contains a single genealogy by definition, it is only capable of generating one parsimony-informative biallelic site pattern (again assuming an infinite sites mutation model). The consequence is that the value of *D* at a single locus can only be +1, 0, or -1, depending on the local genealogy (i.e. only ABBA, BBAA, or BABA). Therefore, even in ILS-only scenarios, there will be regions of the genome with extreme values of *D*, either positive or negative. This situation is more likely to occur in regions of low recombination, as in these regions even large genomic windows may only contain a small number of independent genealogies. Highlighting this problem, Martin et al. (2015) found that the variance of *D* is inflated in regions of low recombination, resulting in an excess of false positives if tests were to be performed on a per-window basis. Similar caution is warranted when applying *D* to inversions, as the entire inversion can act as a single locus (cf. Fuller et al. 2018). For these reasons, while it may be informative to plot the value of the *D* statistic along chromosomes, tests using *D* should be applied only to whole genomes, or at least to genomic regions that are sufficiently large to guarantee sampling a large number of underlying genealogies.

The *D*-statistic does not provide any information about introgression other than its presence or absence. While its value does increase with the proportion of introgressed loci, it is not a good estimator of this quantity, tending to greatly overestimate the true value (Martin et al. 2015, Hamlin et al. 2020). In addition, the sign of *D* is sometimes interpreted as providing information on the direction of introgression, though it can only identify which taxa are involved, and not the donor and recipient populations. For example, a significant *D* statistic implying introgression between *P1* and *P3* could involve the $P3 \rightarrow P1$ direction, the $P1 \rightarrow P3$ direction, or some combination of the two. *D* has more power to detect introgression in the $P3 \rightarrow P1$ direction (see simulation conditions "P1 into P3" and "P3 into P1"; Supplementary Figures 2,3), but can detect it in either direction. Lastly, the *D* statistic is agnostic to the timing of introgression (as long as it is post-speciation) and may yield a positive result under a variety of scenarios, including instantaneous "pulses" of introgression, hybrid speciation/admixed population formation, or gene flow over continuous periods of time.

Overall, the *D* statistic is a very reliable genome-wide test for introgression, but alternative methods are needed to infer more details about any detected introgression events.

Inferring the rate and direction of introgression using derived allele counts

Many researchers are interested not only in the presence or absence of introgression, but in quantifying its magnitude and in identifying the donor and recipient populations. The "rate" of introgression can generally be taken to mean one of two things. In the context of phylogenomic approaches and phylogenetic networks, the rate refers to the proportion of the genome that originates from a history of introgression. This is also sometimes referred to as the "inheritance probability" or "admixture proportion." Alternatively, in the isolation-with-migration (IM) framework, the rate refers to the movement of migrant individuals over continuous time

(Wakeley and Hey 1998, Nielsen and Wakeley 2001). In this and following sections, we will take the "rate" to have the first meaning.

Accurate estimates of the rate and direction can be obtained by considering additional biallelic site patterns to ABBA and BABA. Many such methods exist, and discussing them at length is unnecessary for the scope of our review; here we simply mention a few of these approaches and direct readers to the relevant literature. As mentioned earlier, simply using the *D* statistic does not provide an unbiased estimation of the rate of introgression (Martin et al. 2015, Hamlin et al. 2020). A recently proposed extension of *D* called D_p adds the counts of BBAA sites to the denominator to form:

$$D_p = \left| \frac{ABBA - BABA}{BBAA + ABBA + BABA} \right|$$

Taking the degree of asymmetry as a fraction of the total number of parsimony-informative biallelic sites brings D_p conceptually closer to estimating a genome-wide introgression proportion. The statistic tends to slightly underestimate the true rate of introgression (Supplementary Figure 5)—and its accuracy is affected by the direction of introgression—but it scales linearly with the rate of introgression and has better precision for lower true amounts of introgression (Hamlin et al. 2020).

Another common approach is to compare the observed value of an introgression test statistic to the value that would be expected under a scenario where the entire genome was introgressed. The F_4-ratio or α (Green et al. 2010, Patterson et al. 2012, Peter 2016) and f_d (Martin et al. 2015) statistics take this approach. The α statistic requires data from five samples and assumes an admixed population with two parent populations, while f_d assumes complete homogenization of allele frequencies under total introgression, making it applicable to a quartet. *HyDe* (Blischak et al. 2018, Kubatko and Chifman 2019) estimates the rate in a similar way under a hybrid speciation scenario using linear combinations of derived site patterns. The assumptions of F_4 and *HyDe* are somewhat restrictive and are not likely to be reflective of the majority of introgression in nature (Schumer et al. 2014). However, *HyDe* gives highly accurate estimates of the rate of introgression when its assumptions about hybridization are met, and still provides reasonable estimates for the rate when these assumptions are violated (Kong and Kubatko 2021).

Unless additional assumptions are made, there is not enough information contained in the frequency of gene tree topologies (i.e. site pattern counts) alone to estimate the direction of introgression in a quartet or rooted triplet. However, if a sample is obtained from a fifth species (Eaton and Ree 2013, Pease and Hahn 2015) or if polymorphism data is available for the quartet (Martin and Amos 2020), then it is possible to infer the direction of introgression. The "partitioned *D*-statistics" of Eaton and Ree (2013) were the first attempt to infer the direction of introgression in a five-taxon phylogeny. Unfortunately, redundant site pattern counts make the results of this directionality test uninterpretable. The D_{FOIL} method of Pease and Hahn (2015) resolves this problem by setting up a system of four *D* statistics, explicitly testing each of the 16 possible introgression events and directions. D_{FOIL} assumes that the 5-taxon phylogeny is symmetric, with two pairs of sister species. In this particular configuration of species it becomes

possible to polarize introgression events because the direction of introgression affects relationships between the donor and both the recipient species and its sister taxon. Unfortunately, D_{FOIL} does not work if the species tree is an asymmetric, or "caterpillar," tree.

Martin and Amos (2020) introduced an approached called the "*D* frequency spectrum," or D_{FS} for short, which makes use of multiple sampled individuals per lineage. D_{FS} estimates the *D* statistic in each bin of the joint derived allele frequency spectrum constructed for the two sister taxa in a quartet. The shape of the D_{FS} is expected to be affected by the direction, rate, and timing of introgression in predictable ways, allowing inferences about these quantities to be made. The shape of the D_{FS} is also affected by demographic history and changes under more complex introgression scenarios, so it will typically be necessary to perform simulations to explicitly test different introgression scenarios with this approach (Martin and Amos 2020).

Inferring introgression events from reconstructed gene trees

While methods based on site patterns and allele frequencies can be powerful, there are also fundamental limitations to the kinds of data they can be applied to. First, as mentioned earlier, a key assumption of the *D* statistic is an infinite sites model of mutation. When applied to closely related, extant species, this assumption is likely to hold. However, with increasing divergence times it becomes more likely that ABBA and BABA site patterns can accumulate due to convergent substitutions, and thus will no longer reflect underlying gene tree topologies. While this is not an issue for detecting introgression if convergent substitutions accumulate at the same rate on all branches of the species tree, it can potentially lead to false positives if there is variation in substitution rates among samples. For this reason, site patterns may not be a reliable way to test for introgression between more distantly related extant species, or along branches deeper in a species tree. Second, as the number of sampled species increases, the number of possible trees and quartets increases super-exponentially (Felsenstein 2004). This makes it impractical to apply quartet-based methods to trees with many taxa.

A solution to these problems is to estimate gene tree topologies directly, as many different methods can be used to accurately infer the topology at a locus. Once gene trees have been reconstructed from a large number of loci, the counts of discordant topologies can be used in much the same way as ABBA and BABA sites are in the *D* test. In fact, Huson et al. (2005) proposed such a test comparing alternate tree topologies in a triplet, using a statistic they called Δ. Significance in genome-scale datasets can be evaluated by bootstrap-sampling the estimated gene trees (Vanderpool et al. 2020) or by assuming a χ^2 distribution (Suvorov et al. 2021), with $\Delta = 0$ again representing the null hypothesis under ILS alone. While Δ has greater potential to be affected by sources of technical error such as systematic bias in gene tree inference—and may have limited power to detect very ancient introgression—it has the advantage of being more robust to the infinite-sites assumption and allows for testing of introgression along deep, internal branches of a phylogeny, while maintaining power comparable to *D* for more recent introgression scenarios (Supplementary Figure 3). Therefore, Δ represents a straightforward way to test for introgression using a small number of additional assumptions.

Estimated gene trees can also be used as input to phylogenetic network methods. These methods construct a likelihood or pseudolikelihood function that is explicitly derived from a phylogenetic network model, for which parameters can then be estimated using either maximum likelihood or Bayesian approaches. The program *PhyloNet* has methods that infer networks directly from gene tree topologies using either maximum likelihood (*InferNetwork_ML*, Yu et al. 2014) or maximum pseudolikelihood (*InferNetwork_MPL*, Yu and Nakhleh 2015). Similarly, *SNaQ* (Solís-Lemus and Ané 2016) estimates a network with reticulation edges via maximum pseudolikelihood using quartet concordance factors (Baum 2007)—essentially just the counts of the three possible unrooted tree topologies. We will discuss phylogenetic network methods in more detail later, in the section entitled "*Likelihood methods for detecting introgression.*"

Detecting introgression using coalescence times

While much can be learned about introgression from the frequency of gene tree topologies alone, including additional information about the distribution of coalescence times can lead to much richer inferences. Some advantages of including coalescence times include more flexibility in inferring introgression between non-sister species, detection of introgression between sister taxa, and distinguishing introgression from ancestral population structure. In the following sections we expand on the expected effects of introgression on coalescence times and branch lengths, followed by a description of how this information is used in concert with gene tree frequencies to make inferences about introgression.

Detecting introgression using signals of pairwise divergence

Just as was the case for gene tree topologies, it is possible to make inferences about introgression by studying violations of expected patterns of pairwise coalescence times under an ILS-only model. As previously mentioned, one of these expected patterns is a symmetry in coalescence times between the two pairs of non-sister taxa in a quartet (Figure 1, bottom). If one pair of non-sister taxa has more recent coalescence times on average than the other, post-speciation introgression between that pair is a likely explanation. Coalescence times can be approximated using simple measures of pairwise sequence divergence, assuming an infinite sites model (or at least that genetic distance is proportional to coalescence time). Therefore, one of the simplest ways to test for introgression is to test for an asymmetry in pairwise sequence divergence. This logic has been informally applied to test for introgression (Brandvain et al. 2014) and has recently been formalized in several test statistics including D_3 (Hahn and Hibbins 2019) and the branch-length test (Suvorov et al 2021). D_3 is straightforward, and has the following definition (changed from the original to be consistent with the notation used here):

$$D_3 = \frac{d_{P2P3} - d_{P1P3}}{d_{P2P3} + d_{P1P3}}$$

Where d denotes the genetic distance between the specified populations. This statistic takes the same general form as the D-statistic, where the relevant difference in the numerator is normalized by the sum of the two values in the denominator. Like the D-statistic, significance of D_3 can be evaluated using a block-bootstrap. A major advantage of D_3 over site-pattern based tests is that it does not require data from an outgroup—it only needs one haploid sequence from

three ingroup species. As with D, D_3 can only detect introgression between non-sister lineages, and has comparable power under this scenario (Supplementary Figure 3).

Characterizing introgression using reconstructed gene trees with branch lengths

Using pairwise divergences between only non-sister taxa ignores information about the full distribution of coalescence times within different gene tree topologies. More information is contained within these branch lengths, allowing for estimation of the timing and direction of introgression in a quartet. As with pairwise measures, we assume that branch lengths from gene trees are a good proxy for coalescence times. However, branch lengths can be affected by other factors such as mutation rate variation, selection, and/or sequencing error. Care must therefore be taken when applying all methods that use this information, including the likelihood methods described later. Despite these caveats, a number of signals appear to be robust to many perturbing factors.

Because introgressing taxa can coalesce via either introgression (Figure 4A, blue) or speciation (Figure 4A, red) depending on the history at a locus, a bimodal distribution arises when coalescence times are measured across loci (Figure 4A). This distribution is not expected under ILS alone, and can therefore be used to test for introgression. In addition, the more recent peak provides information about the timing of introgression, while the frequency of gene trees under this peak compared to the older peak provides information on the rate of introgression. This approach to characterizing introgression is implemented in the software *QuIBL* (Quantifying Introgression via Branch Lengths; Edelman et al. 2019).

The direction of introgression uniquely affects the coalescence times of the non-sister pair of species uninvolved in introgression (Figure 2C, Figure 4B). For example, the direction of introgression between *P2* and *P3* has predictable effects on the coalescence time between *P1* and *P3*. When introgression occurs from *P3* into *P2* (Figure 4B, left), *P2* traces its ancestry through the *P3* lineage at introgressed loci (note that while the direction of introgression is typically described forward in time, the coalescent process occurs backwards in time). Because of this, divergence between *P1* and *P3* is unchanged by introgression in this direction. By contrast, when introgression is from *P2* into *P3* (Figure 4B, right), *P3* traces its ancestry through the *P2* lineage at introgressed loci. This allows *P3* to coalesce with *P1* earlier than it normally would, which decreases the divergence between *P1* and *P3*.

These genealogical processes lead to general predictions that can be used to infer the primary direction of introgression between taxa. Gene trees that are concordant with the species tree can be used as a baseline for the expected amount of *P1-P3* divergence; although these trees can arise from ILS at introgressed loci, the effect of the direction will not be manifest since they are concordant. By comparing this baseline divergence to the amount of *P1-P3* divergence in gene trees consistent with a history of introgression, the direction of introgression can be inferred. Lower *P1-P3* divergence in the latter class of trees provides evidence for *P2* → *P3* introgression, but does not necessarily rule out the other direction (i.e. there could simply be less gene flow in the other direction). Alternatively, if *P1-P3* divergence is the same in both topologies, then

introgression is primarily *P3* → *P2*. This logic to polarizing introgression is used by the D_2 statistic (Hibbins and Hahn 2019) and the *DIP* method (Forsythe et al. 2020).

Finally, *PhyloNet*'s *InferNetwork_ML* method (Yu et al. 2014) is able to infer phylogenetic networks with reticulation edges (i.e. discrete introgression events) from gene trees with branch lengths using maximum likelihood. See the section "*Likelihood methods for detecting introgression*" for a more detailed discussion.

Distinguishing introgression from ancestral population structure

In addition to being generated by introgression, asymmetric gene tree topology frequencies can arise from certain kinds of ancestral population structure (Slatkin and Pollack 2008, Durand et al. 2011, Lohse and Frantz 2014). The scenario that generates asymmetries imagines that the population ancestral to all three species is split into at least two subpopulations, such that the ancestors of *P3* are more closely related to either the ancestors of *P1* or *P2* (but not both) (Supplementary Figure 1A). Because the gene tree topologies in this ancestral species will be skewed toward relationships joining *P3* and one of the sister lineages, this scenario can lead to a significant asymmetry in gene tree topologies even in the absence of post-speciation introgression (Durand et al. 2011). This will also result in a slight asymmetry of genome-wide pairwise divergence times, since the more common discordant tree will contribute more to the average value. All of this means that ancestral structure can result in false positives when testing for introgression using simple patterns of asymmetry.

Fortunately, while these two scenarios are indistinguishable using only gene tree topologies alone, they are distinguishable when using the distribution of branch lengths. Under ancestral population structure, divergence between the sister taxa in whichever discordant gene tree becomes more frequent will be higher than it would be under introgression. Lohse and Frantz (2014) incorporated the expected branch length differences in these two models into a maximum likelihood framework, which was then used to confirm the signal of human-Neanderthal introgression that was originally uncovered by the *D*-statistic. Additionally, ancestral population structure is not expected to result in a bimodal distribution of coalescence times. This means that methods capable of detecting two peaks of coalescence, such as *QuIBL* and *PhyloNet*-based methods that use trees with branch lengths or sequence data directly (and possibly other likelihood methods), should also be robust to the effects of population structure.

Detecting introgression between sister species

Introgression between sister species is very difficult to detect using a single haploid sequence from each species. The classic asymmetry patterns described in previous sections do not apply in this scenario, either for gene tree topologies or coalescence times. While introgression between sister species should lead to an increased variance in coalescence times compared to an ILS-only model, this signal is easily confounded by other processes such as non-equilibrium demography or linked selection (Cruickshank and Hahn 2014; Roux et al. 2016; Sethuraman et al. 2019). These limitations have typically been addressed by combining two alternative sources of information: 1) polymorphism data for the two introgressing species, and 2) local reductions in between-species divergence relative to a genome-wide baseline.

Most available methods for inferring introgression between sister taxa are not phylogenomic in multiple senses: they typically require polymorphism data, they often identify locally introgressed regions rather than genome-wide signals, and they do not explicitly test against an ILS-only case. Genome scans using summary statistics such as F_{ST} (Wright 1949) and d_{xy} (Nei and Li 1979) are common, though relative measures of divergence such as F_{ST} are confounded by natural selection when used for this task (Charlesworth 1998, Noor and Bennett 2009, Nachman and Payseur 2012, Cruickshank and Hahn 2014). There are multiple statistics based on minimum pairwise distances between multiple haplotypes in two species that avoid problems caused by selection (Joly et al. 2009, Geneva et al. 2015, Rosenzweig et al. 2016), and new machine learning methods combine multiple summary statistics into a single comparative framework that is powerful and robust (e.g. Schrider et al. 2018). However, these methods also usually require coalescent simulation under known demographic history to evaluate patterns of introgression, and this information is not always available.

None of the aforementioned limitations mean that genome-wide tests with one sample per species are not possible. Introgression between sister taxa—at least when it occurs in relatively discrete pulses—should result in the same multimodal distribution of coalescence times described above for non-sister taxa. This may be the most promising avenue for a genome-wide test of sister introgression when only one sample per species is available, since coalescence times for two species should follow an exponential distribution under ILS alone. Nevertheless, no methods have been developed to date that explicitly test for this pattern (*QuIBL* can only infer it for non-sister taxa). However, *PhyloNet*'s *InferNetwork_ML* method appears to be capable of reliably inferring introgression (including estimating the timing and rate) between sister taxa using gene trees with branch lengths using this signal (Yu et al. 2014) (Supplementary Figures 3,5) at least when nested within a tree containing more taxa. The *MSci* method in *BPP* (Flouri et al. 2020) can also evaluate models involving introgression between sister species. Despite this, the direction of introgression between sister taxa may not be inferable from only one sample per species.

Finally, while introgression between extant sister species is not detectable using gene tree frequencies, this may not necessarily be the case for introgression between ancestral sister lineages. Several studies have now shown that when introgression occurs between *P3* and the ancestor of *P1* and *P2* (event 2 in Figure 2), it becomes possible under specific conditions for both discordant gene tree topologies to become more common than the species tree topology, while remaining at equal frequencies (Solís-Lemus et al. 2016, Long and Kubatko 2018, Jiao and Yang 2020). It should be possible in principle to infer introgression using this pattern, but it requires sufficiently high rates of introgression to result in the anomalous trees, in addition to independent knowledge of the species tree topology.

Likelihood methods for detecting introgression

Perhaps the most powerful phylogenomic methods for inferring introgression are those that use model-based maximum likelihood or Bayesian inference. These methods can be constructed from a variety of different introgression models, can estimate a variety of different parameters, and can be applied to different types of data. Some methods infer introgression directly from a

multiple sequence alignment, while others use estimated gene trees; some are based on the multispecies network coalescent framework for modelling introgression, while others use the isolation-with-migration model; finally, some perform full likelihood calculations, while others estimate approximate likelihoods or pseudolikelihoods. Common to all of these approaches is the ability to widely search the space of possible introgression scenarios, making the best possible use (in principle) of available datasets to construct a phylogenetic network.

Likelihood methods for inferring introgression generally use one of two underlying models: either the multispecies network coalescent (MSNC) model (Meng and Kubatko 2009) or the isolation-with-migration (IM) model (Wakeley and Hey 1998, Nielsen and Wakeley 2001). The models are quite similar, differing mainly as to whether introgression occurs in discrete pulses (MSNC) or over a continuous time interval (IM). The models provide expectations for the probability and coalescence times of gene tree topologies under incomplete lineage sorting and introgression. These expectations—sometimes combined with models for sequence evolution along trees—allow maximum likelihood or Bayesian inference to be applied to either an inferred set of gene trees or to a set of sequence alignments. From these data, methods can infer the taxa involved in introgression, as well as the rate, timing, and direction of introgression.

Methods that use more data can provide more information, though this comes at a computational cost. Two methods implemented in *PhyloNet*, *InferNetwork_ML* (Yu et al. 2014) and *MCMC_GT* (Wen et al. 2016), can use gene trees without branch lengths, while *InferNetwork_ML* can also use trees with branch lengths. If branch lengths are not provided, only introgression between non-sister lineages can be identified (as with summary statistics such as *D*), with accurate estimates of the rate and potentially the direction of introgression. With branch lengths, the timing of introgression can also be accurately estimated, along with the identification of introgression between sister lineages. Using full sequences from each locus rather than gene trees can provide still more information, although maximum likelihood inference is only possible in the simplest scenarios (e.g. Lohse and Frantz 2014, Dalquen et al. 2017). Instead, most methods that take sequence data as input use Bayesian approaches for inference. These methods include the MSNC-based *MCMC_SEQ* (Wen and Nakhleh 2018) and *MCMC_BiMarkers* (Zhu et al. 2018) methods in *PhyloNet*, the *SpeciesNetwork* (Zhang et al. 2018) method in *BEAST2*, and the *MSci* method in *BPP* (Flouri et al. 2020). Examples of IM-based Bayesian methods include *IMa3* (Hey et al. 2018) and *G-PhoCS* (Gronau et al. 2011).

A major disadvantage of full maximum likelihood and Bayesian methods for inferring introgression is that the computational performance of these approaches tends to scale poorly to larger datasets. For example, the *InferNetwork_ML* method can only be practically applied to datasets of up to 10 species (Hejase and Liu 2016). Bayesian approaches scale especially poorly, and are limited to datasets of dozens to hundreds of loci (Flouri et al. 2020). Some methods have addressed this problem by estimating approximate likelihoods or pseudolikelihoods. The *InferNetwork_MPL* (Yu and Nakhleh 2015) method in *PhyloNet* and *SNaQ* (Solís-Lemus and Ané 2016) both maximize the pseudolikelihood of a set of gene tree topologies (in *SNaQ* the gene trees are first used to calculate gene concordance factors). By using pseudolikelihoods, these methods can be applied to larger datasets with more than ten species and thousands of loci

(Hejase and Liu 2016, Solís-Lemus and Ané 2016). However, in some regions of parameter space, the phylogenetic network is unidentifiable with these methods; that is, many different combinations of network parameters could be equally consistent with the observed data. These pseudolikelihood methods are also not ideal for use with information criteria, which makes it challenging to evaluate the fit of different inferred networks (see section on *"Inferring the number of introgression events*").

The richness of parameters estimated by likelihood methods can also be a double-edged sword, as these inferences are only possible with relatively strong assumptions. In addition to assumptions about no recombination within loci and free recombination between loci, all methods assume that sequences are evolving neutrally. While many methods make assumptions about neutrality, those that detect introgression using only gene tree topologies are quite robust to this assumption (Przeworski et al. 1999, Williamson and Orive 2002, Vanderpool et al. 2020). By contrast, the effect of various forms of selection is to cause changes in the distribution of gene tree branch lengths (Adams et al. 2018), a change that can be interpreted as introgression by full likelihood methods. This is especially true for inferences of introgression between sister lineages, where information on gene tree topologies is often not useful in distinguishing between these two scenarios (Ewing and Jensen 2016; Roux et al. 2016). Since interpreting likelihood methods can be difficult under such circumstances, we recommend complementing these analyses with other approaches that are formulated to be more robust to common model violations. Despite these limitations, likelihood methods for inferring introgression can have many advantages in terms of the power and richness of inference when compared to simpler approaches.

Challenges for inferring introgression

Dealing with phylogenetic uncertainty in introgression analyses

Most methods for inferring introgression require that the species phylogeny is known or can be inferred accurately. More precisely, they require a model of the possible histories of coalescence of samples in the absence of introgression, against which introgression hypotheses can be tested. However, for both technical and biological reasons, a single phylogeny often cannot be inferred accurately and/or with a high confidence. If the wrong species tree is chosen, then introgression may be erroneously inferred. In the case where certain regions of the phylogeny are poorly resolved, one approach is to permute only the poorly resolved regions in different introgression analyses, leaving the more confidently resolved "backbone" constant (Beckman et al. 2018, Pease 2018). Alternatively, it may be that the wrong species phylogeny is inferred with high confidence; in this case, careful examination of local genealogical patterns and coalescence times can uncover which histories correspond to speciation vs. introgression (Fontaine et al. 2015, Forsythe et al. 2020). Finally, likelihood methods should be less vulnerable to uncertainty, since the phylogeny and introgression events are typically co-estimated. However, computational and visual representations of these results can often be uninformative or misleading with regard to the true species branching order (see section below entitled "*Distinguishing among models of introgression*")

Evaluating introgression from unsampled ghost lineages

As we briefly mentioned above, there is always the possibility that the species being studied may have exchanged genes with unsampled "ghost" lineages. These lineages may be unsampled because appropriate specimens were not available for sequencing, because they are currently extinct, or simply because they are unknown taxa. Regardless of their origin, introgression from a distant ghost lineage into a sampled lineage can generate gene tree asymmetry in a rooted triplet. In the scenario considered here (Figure 5a), the ghost lineage is the donor of introgressed alleles into species *P1a*. As a result, at some introgressed loci *P2* and *P3* will appear to be sister lineages (Figure 5b), possibly resulting in an inference of introgression.

Our simulation study (Supplementary Figures 2 and 3), in addition to recent work from Tricou et al. (2021), demonstrates that introgression between a ghost lineage and a sampled taxon can result in significant tests for introgression, using both summary statistic and likelihood approaches. While introgression has indeed occurred, the problem is that the timing, direction, and identity of lineages involved in introgression may all be inferred incorrectly. As with results from sampled taxa, significant results are most likely to occur when the ghost taxon is not sister to the species it is exchanging genes with and when the ghost taxon is the donor of introgressed alleles rather than the recipient (Supplementary Figure 3).

There are a number of approaches researchers can take to detect the presence of ghost introgression. If multiple ingroup lineages are available for testing—but only one of them has been the recipient of introgression—switching the species used in the quartet being tested can reveal ghost introgression. Imagine we have two lineages available to serve as species *P1*: *P1a* and *P1b* (Figure 5a). *P1a* is the recipient of introgression from an unsampled lineage, *X,* which is more distant than *P3*. If species *P1a* is sampled, we may incorrectly infer introgression between *P2* and *P3* (Figure 5b). In contrast, *P1b* is uninvolved in ghost introgression; if the quartet (((*P1b*,*P2*),*P3*),*O*) is tested for introgression, the result should no longer be significant (Figure 5b). Such a result would be consistent with ghost introgression into *P1a*. If both quartets are significant, this would rule out ghost introgression into *P1a* alone, but could still be explained by ghost introgression into the ancestor of *P1a* and *P1b*.

Given an excess of gene trees with *P2* and *P3* sister to one another, another sign of ghost introgression is that the genetic distance between *P2* and *P3* at discordant loci will not be reduced relative to concordant loci, as would occur if they were truly exchanging alleles (Figure 5c). Although the *D3* statistic is still significant under ghost introgression (Supplementary Figure 3), this is because *P3* is also being compared to *P1*. A simple comparison of the distance between *P2* and *P3* at concordant and discordant loci should reveal if there is any signal of ghost introgression. Conversely, the presence of exceptionally divergent haplotypes in *P1* that are unlikely to have originated from known extant species are also consistent with ghost introgression (Figure 5c). In fact, most known cases of putative ghost introgression have been identified this way (i.e. Ai et al. 2015, Kuhlwilm et al. 2019, Zhang et al. 2019). Finally, as noted by Ottenburghs (2020), recent advances in model-based demographic inference may make it

possible to explicitly evaluate ghost introgression scenarios against scenarios involving gene flow between sampled taxa. The vast array of possible ghost introgression scenarios may make model selection difficult, but plausible scenarios can potentially be identified using the approaches described above.

Distinguishing among models of introgression

Introgression events are often depicted using a phylogenetic network. In these representations, a reticulation edge connects two lineages in the tree that have exchanged genes. However, the placement and orientation of these reticulations can imply specific information about the timing, direction, and species involved in introgression. While methods for inferring introgression are developed under a specific introgression model, many of them are agnostic to the true underlying model when applied to empirical data. More importantly, many methods that construct phylogenetic networks will produce the same network from data generated under very different underlying models (Huson and Bryant 2005). In this section we highlight the challenges associated with interpreting the results of introgression tests in the context of the underlying model of introgression

Two important models to consider are introgression that occurs between already-existing lineages and introgression that results in the formation of a new lineage. Figure 6A depicts the former scenario, which corresponds to the introgression scenarios considered in the paper thus far. In such cases, a single horizontal reticulation edge is typically used to connect the two taxa involved. This does not naturally convey any information about the direction of introgression, unless the donor and recipient lineages are explicitly identified (e.g. with an arrowhead). By contrast, methods that assume the formation of an admixed population (e.g., Bertorelle and Excoffier 1998, Wang 2003) or hybrid species (e.g., Meng and Kubatko 2009) often use the visualization shown in Figure 6B, where reticulations connect each parent lineage to the newly formed lineage. This representation implies a directionality of introgression without any additional labelling: from the two parent lineages into the newly formed lineage. In both cases, a horizontal reticulation edge can be used to denote the instantaneous exchange of alleles between the involved lineages. Alternatively, Figure 6C shows an example using non-horizontal branches, which may imply a period of branching off and independent evolution from the parent species before the hybrid lineage is formed (e.g., Patterson et al 2012, Yu et al. 2014, Zhang et al. 2018). An alternative interpretation of this representation is that it shows "standard" introgression involving a now extinct species, in which case the extinct lineage was the donor in the introgression scenario. In this case there really was a period of independent evolution, but it occurred along a lineage that was not sampled. In all three cases, the placement of the reticulation edge conveys information about the timing of introgression and/or lineage formation.

It important to consider how the methods for detecting introgression discussed here relate to the underlying introgression scenarios, and how this may affect our interpretation of results. Many tests for introgression are agnostic to the particulars of the underlying introgression scenario, and will therefore be significant under different models. For example, the *D*-statistic can detect introgression between non-sister taxa regardless of the direction of gene flow (Martin et al. 2015, Supplementary Figure 3), or whether introgression results in the formation of a new lineage

(Kong and Kubatko 2021). Other methods enforce a particular model of introgression, even though it may not reflect the underlying data. For example, *HyDe* (Blischak et al. 2018) is less accurate when estimating the admixture proportion if its hybrid speciation assumption is violated (Kong and Kubatko 2021), while other tests explicitly require the labelling of a putative admixed population under a lineage-formation scenario (Peter 2016). Some statistical methods can explicitly distinguish among these scenarios. The D_1 statistic (Hibbins and Hahn 2019) tests whether gene tree branch lengths are more consistent with hybrid speciation (Figure 6B) or post-speciation introgression (Figure 6A). The multispecies network coalescent implementation in *BPP* (Flouri et al. 2020) may also be able to differentiate among a variety of possible introgression scenarios.

One additional obstacle to distinguishing among models of introgression is a consequence of the information required by machine-readable formats for representing phylogenetic networks. In general, methods return inferred phylogenetic networks in the Extended Newick format (Cardona et al. 2008), which requires the specification of a bifurcating "parent" node that occurs closer to the root than the "hybrid" node, which has two incoming lineages. While it is possible for the hybrid node in this format to represent a lateral gene transfer event that does not have a parent closer to the root (Cardona et al. 2008), this format is often not used to represent introgression (though it could be).

Visualizing these results often complicates their interpretation even further. To highlight this, we inferred networks using *PhyloNet*'s *InferNetwork_ML* method (Yu et al. 2014) for simulated *P3* → *P1* and *P1* → *P3* introgression after speciation (see Supplementary Figure 2), and plotted the results using three popular tools (Figure 7): *Dendroscope* (Huson and Scornavacca 2012), *IcyTree* (Vaughan 2017), and *PhyloPlots*, which is part of the Julia package *PhyloNetworks* (Solís-Lemus et al. 2017). All three methods handle the placement of parent and daughter nodes differently. *Dendroscope* visualizes the two incoming lineages to the hybrid node with blue reticulations, which can erroneously imply a lineage-formation or hybrid speciation scenario with *P2* involved in hybridization when introgression is *P3* → *P1* (Figure 7A). As a consequence of the parent/hybrid node structure, all three methods use non-horizontal reticulations (Figure 7A-F), which may imply periods of independent evolution in the donor population prior to introgression, even under an instantaneous "pulse" scenario. The general use of reticulations to connect parent and daughter nodes also heavily implies a discrete-time event or series of discrete-time events, rather than a continuous window of gene flow as conceptualized in the isolation-with-migration model. While none of the output networks contained branch lengths, the arbitrary location of placement of the reticulations could imply an inferred time of introgression. We should stress that *PhyloNet*'s *InferNetwork_ML* method was accurate in its inferences about the presence and direction of introgression (Supplementary Figure 3)—it is only the visualization that is misleading.

The visualization of introgression results is especially difficult when information on the timing and direction of gene flow cannot be inferred. The software *admixturegraph* (Leppälä et al. 2017) plots a network representation solely from the results of a series of *D* tests. We applied this visualization to simulated *P3* → *P1* and *P1* → *P3* introgression (Supplementary Figure 2). The

resulting plots shown in Figures 7G and 7H imply that *P1* formed from hybridization after periods of independent evolution in *P2* and *P3*. However, none of these processes are knowable from a *D*-statistic result (because the direction of introgression cannot be inferred), and this is not the scenario that produced the data. In general, special care should be taken when visualizing the results of *D*-statistics and related test statistics on a phylogeny, since they only provide information on the presence/absence of introgression, and not the direction of introgression.

Clearly differentiating among different possible models of introgression remains challenging. Care should be taken not to over-interpret the results of methods that are model-agnostic, or that rely on a particular model of introgression rather than inferring it from data. This is especially true when interpreting results from common machine-readable visualizations. If possible, hand-drawn "tube tree" representations (e.g. Figure 4) may be more effective in accurately conveying the information available. If automated plotting software is being used, it appears that the visualizations produced by *PhyloPlots* (Figure 7E-F) are most faithful to the true model of introgression.

Inferring the number of introgression events

A major challenge that remains in the inference of introgression is how to assess the fit of different numbers of introgression events inferred on the same tree. The mostly widely used methods are formulated to test for the presence of introgression versus no introgression, but provide no rigorous way to evaluate the number of distinct introgression events. One approach is to perform many quartet-based tests, and then to infer the most parsimonious set of introgression events by collapsing sets of positive tests that share the same ancestral populations (Pease et al. 2016, Suvorov et al. 2021). However, this approach is highly conservative, as it can collapse cases where there truly are multiple instances of post-speciation introgression within a clade. Additionally, it requires large datasets and the piecing together of many quartets, which makes it impractical in many cases. Nonetheless, such approaches can be used to generate a conservative estimate for the minimum number of introgression events.

Even with likelihood methods, estimating the number of introgression events is not a solved problem. One issue is that adding additional parameters to the likelihood model always improves the likelihood score. This makes it necessary to penalize model complexity when comparing estimated likelihoods. Unfortunately, the information measures that are classically used to perform model selection, such as AIC and BIC, do not adequately scale with the increased complexity of adding a new reticulation to a phylogenetic network. This is because adding a new reticulation does not just add a single new model parameter—it adds a whole new space of possible networks, with different taxa involved in introgression, at different times, and in different directions (Blair and Ané 2020). AIC and BIC penalize the increased complexity of model parameters, but not the increased complexity of models within a set of parameters. The problem is greater for methods based on pseudolikelihood such as *SNaQ*, because these information measures are not intended for pseudolikelihood estimates. Bayesian approaches such as those implemented in *PhyloNet* (Wen and Nakhleh 2018) and *SpeciesNetwork* can incorporate appropriate penalties for model complexity, but unfortunately scale poorly to larger datasets and larger numbers of reticulations (Elworth et al. 2019).

While no methods currently exist that can both explicitly penalize model complexity and scale to large datasets, there are several alternate approaches available for assessing the fit of phylogenetic networks. One simple, empirical approach is to use a slope heuristic where networks are inferred across different numbers of reticulations, and the best network is taken as the least complex one after which the likelihood score appears to stop improving. This is the method recommended for use with *SNaQ* (Solís-Lemus and Ané 2016). *PhyloNet* has methods that can evaluate the fit of a network using *k*-fold cross-validation or parametric bootstrapping (Yu et al. 2014), which can both address this problem. Finally, a promising approach from Cai and Ané (2020) involves using the multispecies network coalescent to calculate the quartet concordance factors expected from an estimated network. A goodness-of-fit function is then used to evaluate the fit of these expected concordance factors to those observed in the data. This is similar to the method implemented in *admixturegraph* (Leppälä et al. 2017) for use with *D* statistics.

Conclusions

In conclusion, several methodological and technical challenges remain in the inference of introgression, including: more accurate estimation of the rate, timing, and direction of introgression; detection of introgression between sister taxa; spurious results generated by unsampled lineages; inference of the number of introgression events in a clade; and accurate automated visualization of phylogenetic networks. Despite these challenges, currently available approaches have remarkable power to detect and characterize introgression under a wide variety of conditions, especially when used in a complementary fashion. Overall, these methods will continue to reveal the nature and influence of introgression throughout the natural world.

Acknowledgements

We thank Leonie Moyle, Rafael Guerrero, Claudia Solís-Lemus, Cécile Ané, Mia Miyagi, and Luay Nakhleh for helpful comments and discussion. Michael Turelli and three reviewers all also provided constructive criticisms. This work was supported by National Science Foundation grant DEB-1936187.

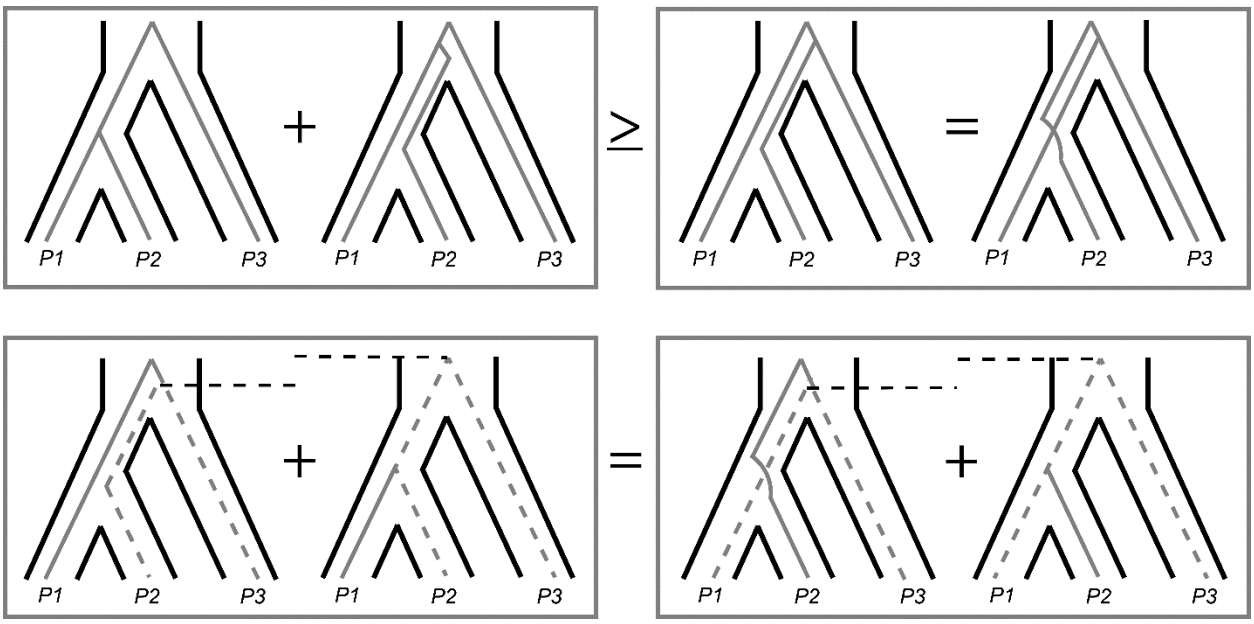


Figure 1: Expected gene tree topologies and coalescence times under ILS only. For a rooted triplet, four topologies are possible (top row): two concordant with the species tree, which can result either from lineage sorting or ILS (top left), and two that are discordant with the species tree and arise from ILS only (top right). The two concordant trees must be at least as frequent as the two discordant trees, which are equally frequent to each other. For non-sister pairs of taxa—either *P2-P3* (bottom left) or *P1-P3* (bottom right)—coalescence is expected to occur at one of two times, depending on whether they coalesce first or second in a gene tree (grey dotted lines). These expected times are symmetrical across gene trees, and so pairwise divergences between the non-sister lineages are expected to be equal when averaged across loci.

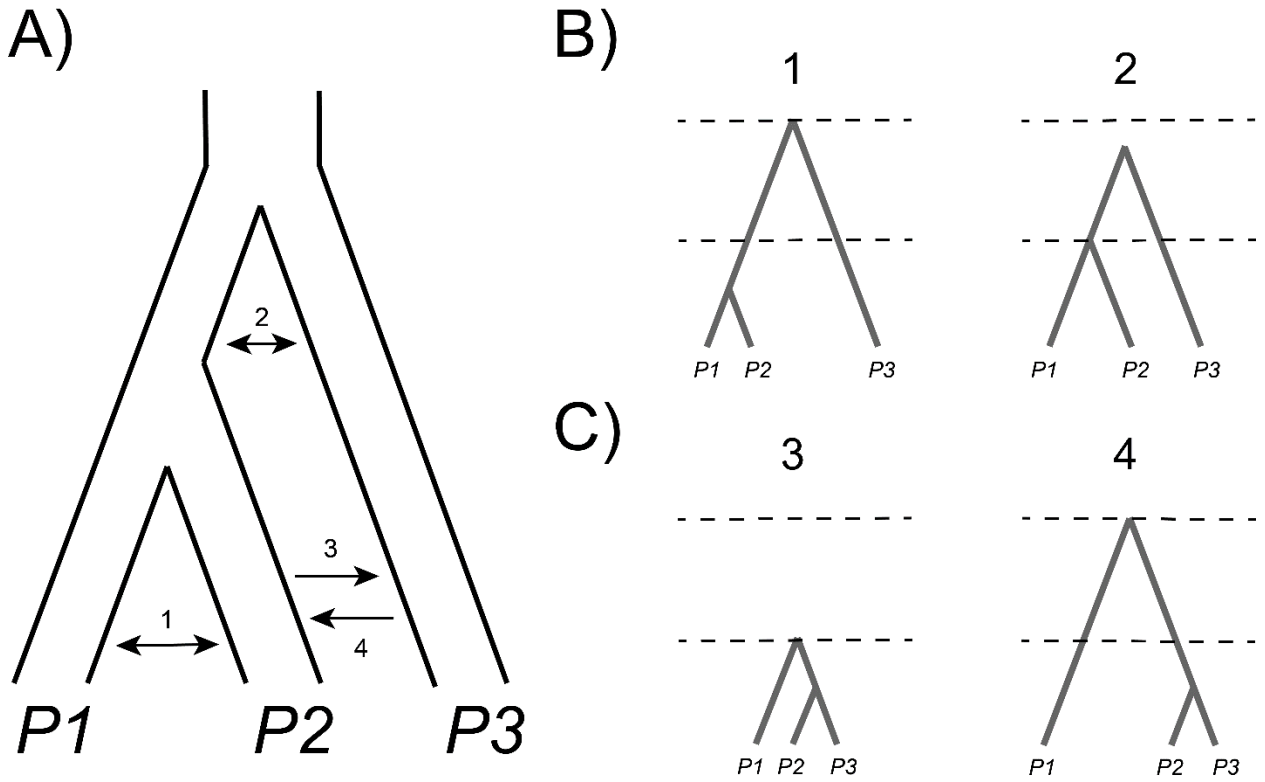
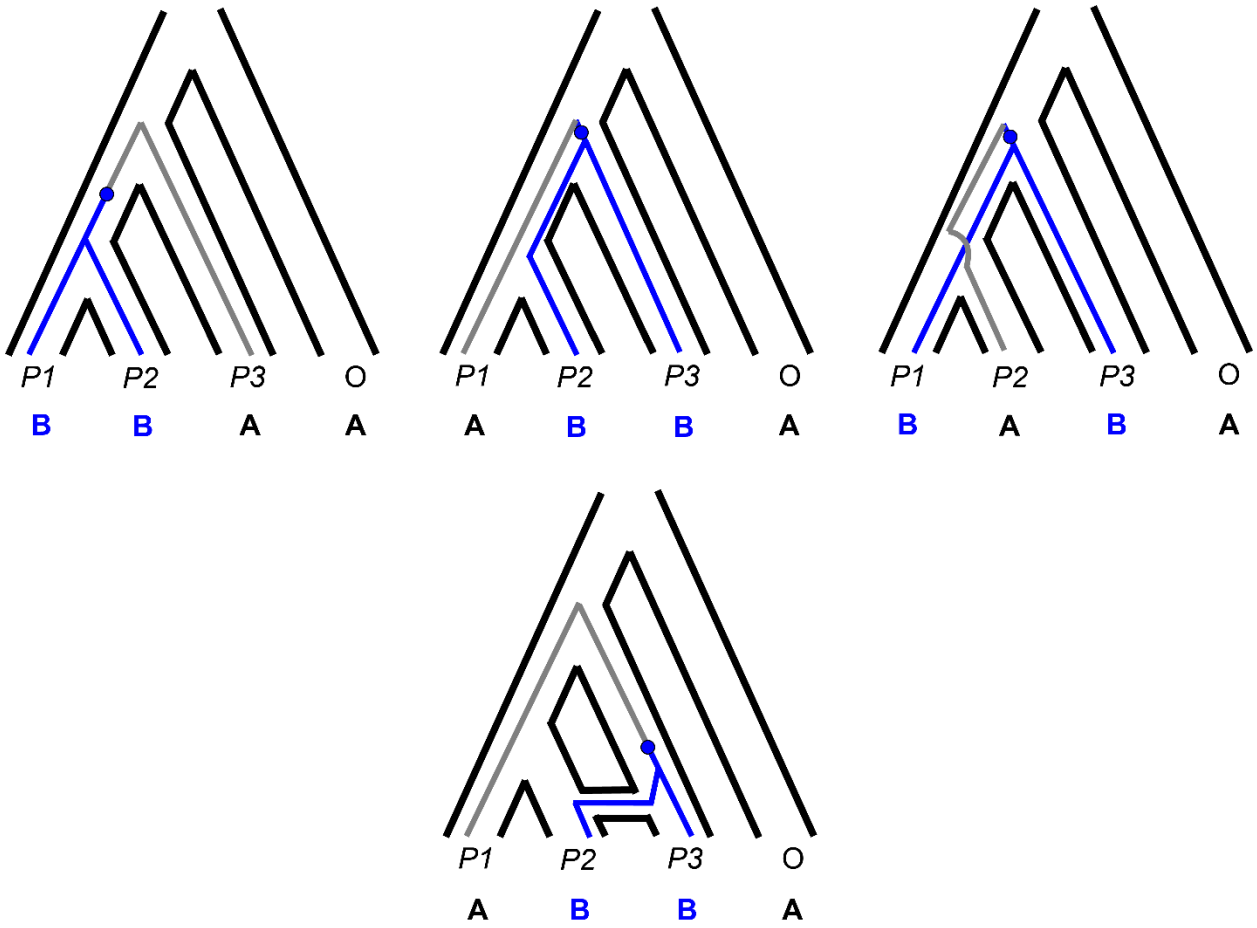


Figure 2: An overview of detectable introgression scenarios for a rooted triplet, and their effects on gene tree frequencies and branch lengths. A) The species tree relating three lineages. Introgression can occur between extant (1) or ancestral (2) sister lineages, or between non-sister taxa, with *P3* as either the recipient (3) or the donor (4). B) Gene trees at introgressed loci for introgression between sister lineages. Introgression between sister taxa reduces divergence between the involved taxa but does not generate discordant gene trees (events 1 and 2). In both trees the expected time to coalescence for pairs of lineages in the absence of introgression is denoted with dashed horizontal lines. C) Gene trees at introgressed loci for introgression between non-sister lineages. When *P3* is the recipient of introgression (event 3), discordant gene trees are generated uniting *P2* and *P3*. In addition, divergence is reduced between both *P2* and *P3* and between *P1* and *P3*. When *P3* is the donor of introgression (event 4) discordant gene trees are again generated uniting *P2* and *P3*. In this case divergence is reduced only between *P2* and *P3*, while divergence is increased between *P1* and *P2*. In both trees the expected time to coalescence for pairs of lineages in the absence of introgression is denoted with dashed horizontal lines.



870

871 *Figure 3*: Biallelic site patterns are informative of underlying gene tree topologies. With the
872 exception of low levels of homoplasy, such patterns can only arise from mutations (blue) on
873 internal branches of the local genealogy. The occurrence of the incongruent site patterns
874 "ABBA" (top middle) and "BABA" (top right) are therefore expected to reflect the frequency of
875 discordant gene tree topologies. With introgression between a specific non-sister species pair,
876 one incongruent pattern (bottom) can increase in frequency over the other due to the underlying
877 asymmetry in gene tree frequencies.

878

879

880

881

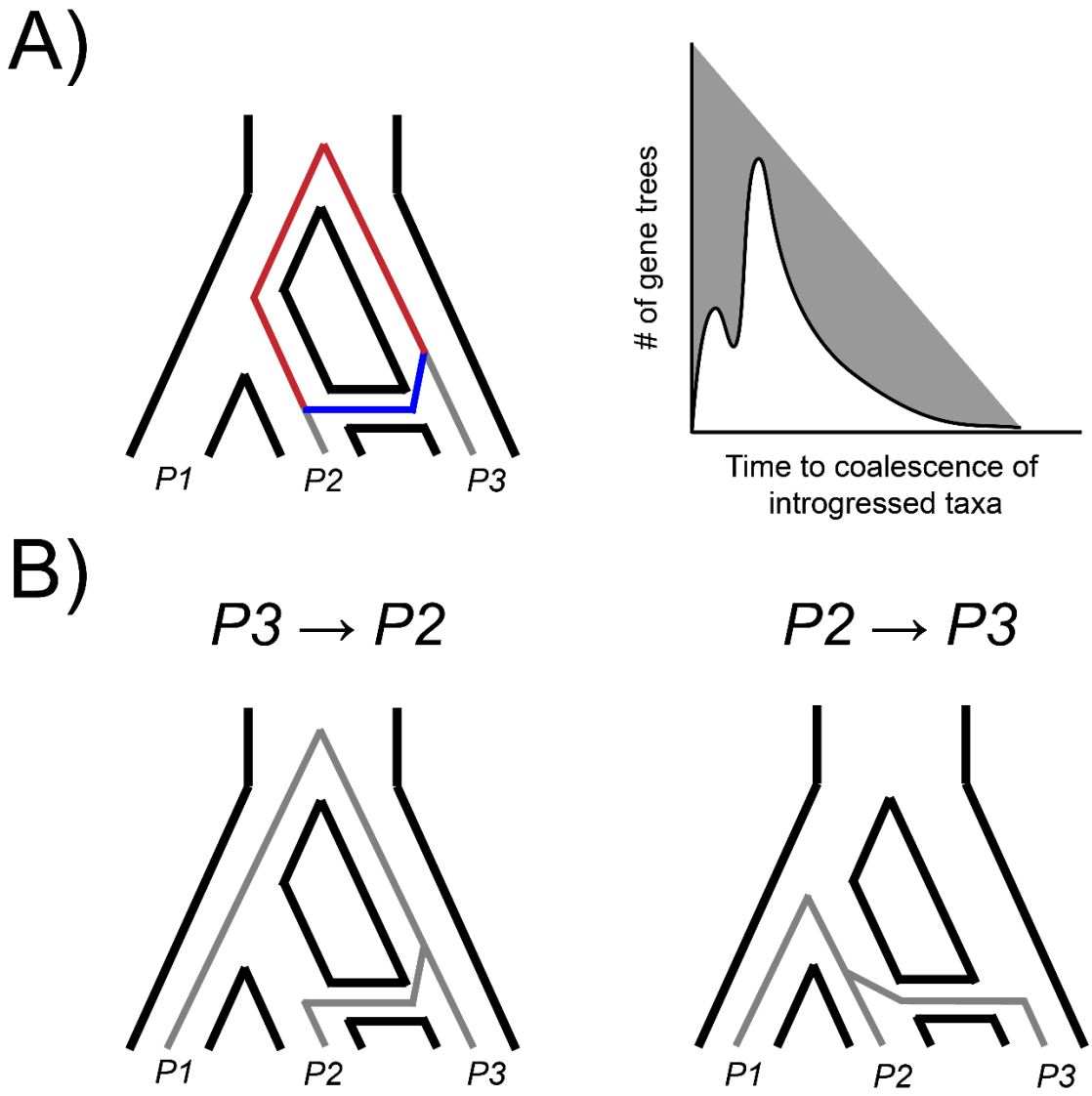
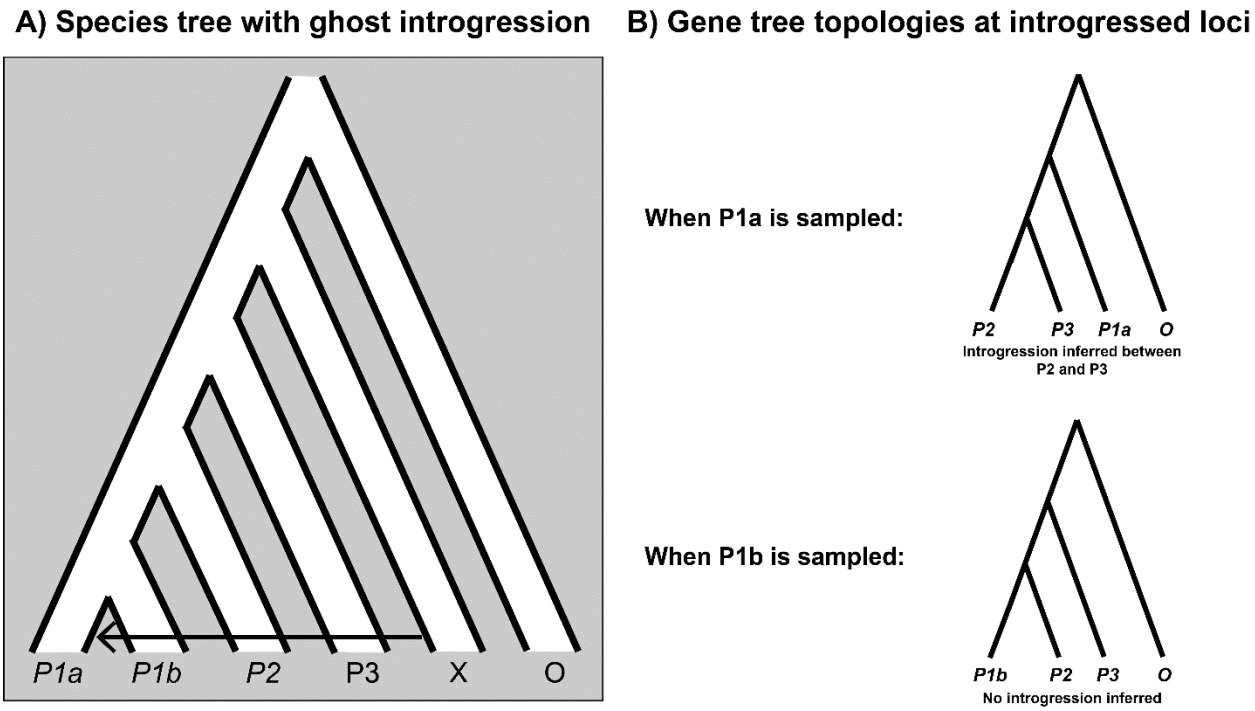


Figure 4: Coalescence times provide information on the timing, direction, and presence of introgression. A) Post-speciation introgression between *P2* and *P3* allows them to coalesce more quickly at introgressed loci (blue). This reduces their whole-genome divergence relative to *P1* and *P3*, an asymmetry that can be used to test for introgression. Since coalescence can now occur at one of two times, after introgression (blue) or after speciation (red), it also results in a bimodal distribution of coalescence times across loci (right figure). The more recent peak of this distribution can be used to estimate the timing of introgression. B) The direction of introgression between *P2* and *P3* affects the time to coalesce of *P1* and *P3* at introgressed loci. *P2* → *P3* introgression allows *P1* and *P3* to coalesce more quickly (right), reducing their divergence at introgressed loci.



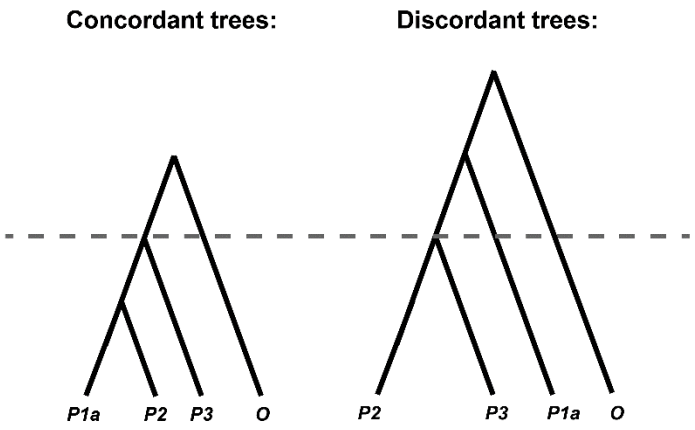


Figure 5: Understanding and detecting ghost introgression. A) A scenario of ghost introgression from an unsampled outgroup lineage, *X*, into *P1a*. B) When ghost introgression has occurred and a quartet including *P1a* is sampled, introgression may be erroneously inferred between *P2* and *P3.* This occurs because at some introgressed loci *P1a* will be more distantly related to both *P2* and *P3*, leading to an excess of discordant trees with *P2* and *P3* sister to one another (top). If instead a quartet including *P1b* is sampled, there should no longer be an excess of discordant trees (bottom). C) Ghost introgression should also be detectable via a change (or a lack of change) in branch lengths. True introgression between *P2* and *P3* should cause them to be more similar; i.e. shorter branch lengths separating them in discordant trees. In contrast, ghost introgression will not make them more closely related in discordant trees than in concordant trees on average. Similarly, the distance between *P1a* and all ingroup lineages will be higher when it is the recipient of ghost introgression from an outgroup.

909

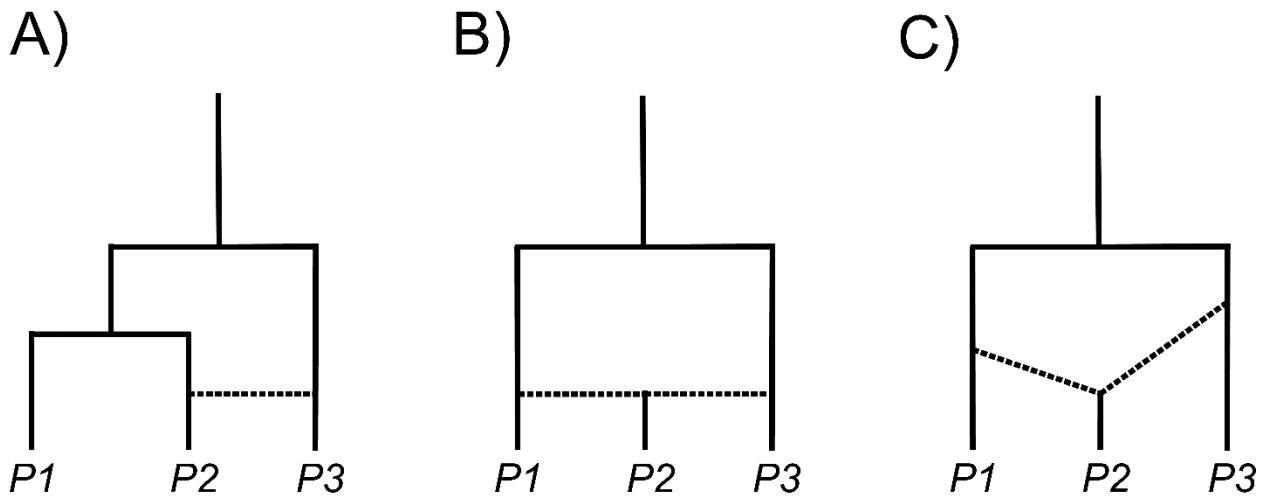


Figure 6: Conceptualizing different models of introgression. A) Introgression between extant lineages. B) and C) Introgression that results in the formation of a new lineage, differing only with respect to whether there appears to be a period of independent evolution before lineage formation.

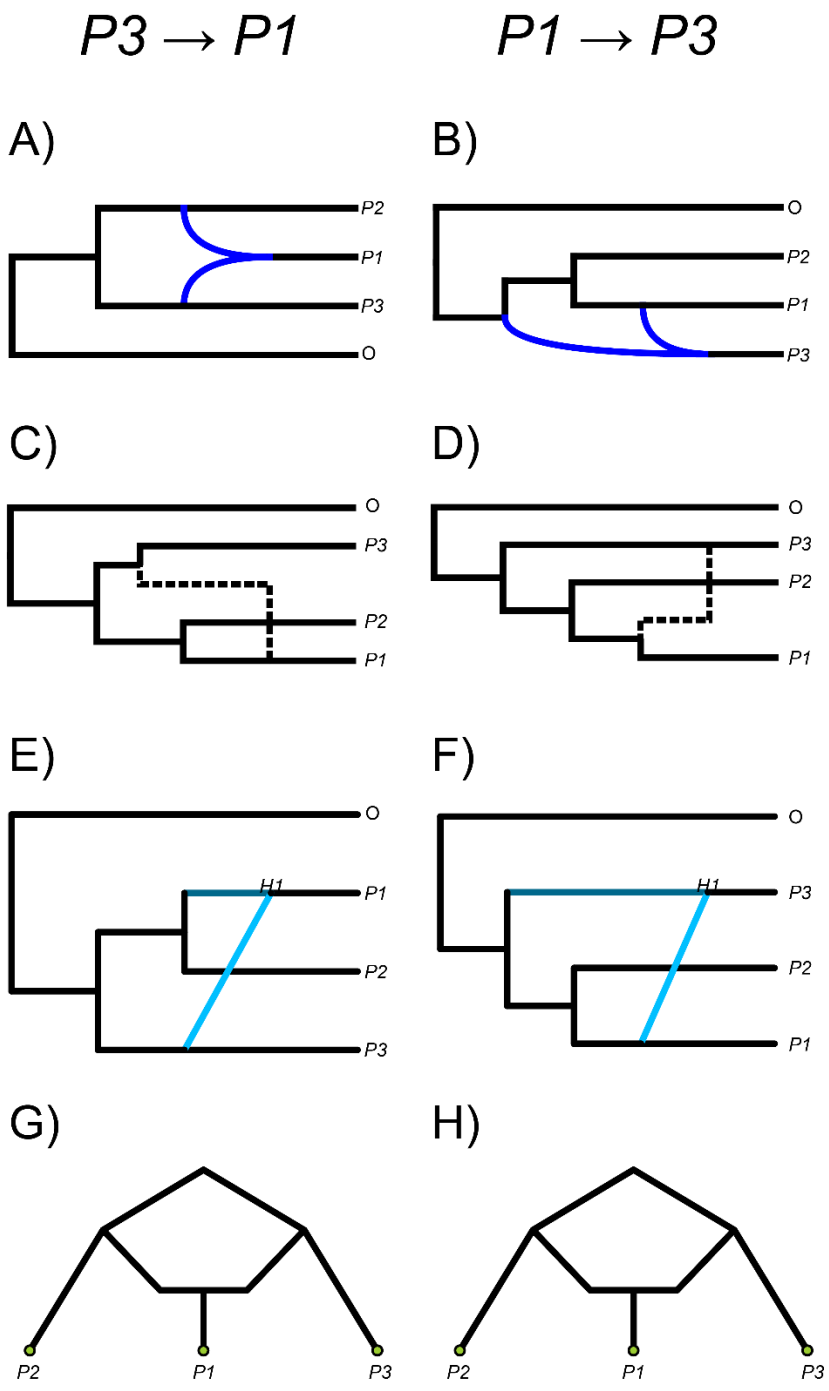


Figure 7: Different visualizations of the same underlying phylogenetic networks. The left column comes from a network representing *P3 → P1* introgression, while the right column comes from a network representing *P1 → P3* introgression. The rows, from top to bottom, show visualizations from A) and B) *Dendroscope*, C) and D) *IcyTree*, E) and F) *PhyloPlots*, and G) and H) *admixturegraph*.

References

Adams, R. H., D. R. Schield, D. C. Card, and T. A. Castoe, 2018 Assessing the impacts of positive selection on coalescent-based species tree estimation and species delimitation. Systematic Biology 67**:** 1076-1090.

Ai, H., X. Fang, B. Yang, Z. Huang, H. Chen *et al.*, 2015 Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nature Genetics 47**:** 217-225.

Akaike, H., 1974 A new look at the statistical model identification. IEEE Transactions on Automatic Control 19**:** 716-723.

Baum, D. A., 2007 Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56**:** 417-426.

Beckman, J. B., P. M. Benham, Z. A. Cheviron, and C. C. Witt, 2018 Detecting introgression despite phylogenetic uncertainty: The case of the South American siskins. Molecular Ecology 27**:** 4350-4367.

Bertorelle, G., and L. Excoffier, 1998 Inferring admixture proportions from molecular data. Molecular Biology and Evolution 15**:** 1298-1311.

Blischak, P. D., J. Chifman, A. D. Wolfe and L. S. Kubatko, 2018 HyDe: A Python package for genome-scale hybridization detection. Systematic Biology 67**:** 821-829.

Brandvain, Y., A. M. Kenney, L. Flagel, G. Coop and A. L. Sweigart, 2014 Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. PLoS Genetics 10**:** e1004410.

Burnham, K. P., and D. R. Anderson, 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, New York.

Cai, R., and C. Ané, 2020 Assessing the fit of the multi-species network coalescent to multi-locus data. Bioinformatics. 37**:** 634-641.

Cardona, G., F. Rossello and G. Valiente, 2008 Extended Newick: it is time for a standard representation of phylogenetic networks. BMC Bioinformatics 9**:** 532.

Charlesworth, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution 15**:** 538-543.

Copetti, D., A. Burquez, E. Bustamante, J. L. M. Charboneau, K. L. Childs *et al.*, 2017 Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. Proceedings of the National Academy of Sciences of the United States of America 114**:** 12003-12008.

961 Cruickshank, T. E., and M. W. Hahn, 2014 Reanalysis suggests that genomic islands of
962 speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:
963 3133-3157.

964 Dagilis, A. J., D. Peede, J. M. Coughlan, G. I. Jofre, E. R. R. D'Agostino *et al.*, 2021 15 years of
965 introgression studies: quantifying gene flow across eukaryotes. BioRxiv.

966 Dalquen, D. A., T. Zhu and Z. Yang, 2017 Maximum likelihood implementation of an isolation-
967 with-migration model for three species. Systematic Biology 66: 379-398.

968 Degnan, J. H., 2018 Modeling hybridization under the network multispecies coalescent.
969 Systematic Biology 67: 786-799.

970 Dowling T.E., S. C. L., 1997 The role of hybridization and introgression in the diversification of
971 animals. Annual Review of Ecology, Evolution, and Systematics 28: 593-619.

972 Durand, E. Y., N. Patterson, D. Reich and M. Slatkin, 2011 Testing for ancient admixture
973 between closely related populations. Molecular Biology and Evolution 28: 2239-2252.

974 Eaton, D. A., and R. H. Ree, 2013 Inferring phylogeny and introgression using RADseq data: an
975 example from flowering plants (Pedicularis: Orobanchaceae). Systematic Biology 62:
976 689-706.

977 Edelman, N. B., P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey *et al.*, 2019 Genomic
978 architecture and introgression shape a butterfly radiation. Science 366: 594-599.

979 Ellstrand N.C., M. P., Rong J., Bartsch D., Ghosh A., de Jong T.J., Haccou P., Lu B.R., Snow
980 A.A., Stewart C.N., Strasburg J.L., van Tienderen P.H., Vrieling K., Hooftman D. , 2013
981 Introgression of crop alleles into wild or weedy populations. Annual Review of Ecology,
982 Evolution, and Systematics 44: 325-345.

983 Elworth, R. A. L., H. A. Ogilvie, J. Zhu and L. Nakhleh, 2019 Advances in computational
984 methods for phylogenetic networks in the presence of hybridization, pp. 317-360 in
985 *Bioinformatics and Phylogenetics*, edited by T. Warnow. Springer, Cham.

986 Ewing, G. B., and J. D. Jensen, 2016 The consequences of not accounting for background
987 selection in demographic inference. Molecular Ecology 25: 135-141.

988 Felsenstein, J., 2004 *Inferring phylogenies*. Sinauer Associates, Sunderland, MA.

989 Flouri, T., X. Jiao, B. Rannala and Z. Yang, 2020 A Bayesian implementation of the
990 multispecies coalescent model with introgression for phylogenomic analysis. Molecular
991 Biology and Evolution 37: 1211-1223.

992 Folk, R. A., P. S. Soltis, D. E. Soltis and R. Guralnick, 2018 New prospects in the detection and
993 comparative analysis of hybridization in the tree of life. American Journal of Botany 105:

364-375.

Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey *et al.*, 2015 Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347**:** 1258524.

Forsythe, E. S., A. D. L. Nelson and M. A. Beilstein, 2020 Biased gene retention in the face of introgression obscures species relationships. Genome Biology and Evolution 12**:** 1646-1663.

Forsythe, E. S., D. B. Sloan and M. A. Beilstein, 2020 Divergence-based introgression polarization. Genome Biology and Evolution 12**:** 463-478.

Fuller, Z. L., C. J. Leonard, R. E. Young, S. W. Schaeffer and N. Phadnis, 2018 Ancestral polymorphisms explain the role of chromosomal inversions in speciation. PLoS Genetics 14**:** e1007526.

Geneva, A. J., C. A. Muirhead, S. B. Kingan and D. Garrigan, 2015 A new method to scan genomes for introgression in a secondary contact model. PLoS One 10**:** e0118621.

Gillespie, J. H., and C. H. Langley, 1979 Are evolutionary rates really variable? Journal of Molecular Evolution 13**:** 27-34.

Grau-Bove, X., S. Tomlinson, A. O. O'Reilly, N. J. Harding, A. Miles *et al.*, 2020 Evolution of the insecticide target Rdl in African Anopheles is driven by interspecific and interkaryotypic introgression. Molecular Biology and Evolution 37**:** 2900-2917.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel *et al.*, 2010 A draft sequence of the Neandertal genome. Science 328**:** 710-722.

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko and A. Siepel, 2011 Bayesian inference of ancient human demography from individual genome sequences. Nature Genetics 43**:** 1031-1034.

Hahn, M. W., 2018 *Molecular population genetics* Oxford University Press, Sunderland, MA.

Hahn, M. W., and M. S. Hibbins, 2019 A three-sample test for introgression. Molecular Biology and Evolution 36**:** 2878-2882.

Hamlin, J. A. P., M. S. Hibbins and L. C. Moyle, 2020 Assessing biological factors affecting postspeciation introgression. Evolution Letters 4**:** 137-154.

Harrison, R. G., and E. L. Larson, 2014 Hybridization, introgression, and the nature of species boundaries. Journal of Heredity 105 Suppl 1**:** 795-809.

He, C., D. Liang and P. Zhang, 2020 Asymmetric distribution of gene trees can arise under

purifying selection if differences in population size exist. Molecular Biology and Evolution 37**:** 881-892.

Hedrick, P. W., 2013 Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology 22**:** 4606-4618.

Heiser, C. B., 1949 Natural hybridization with particular reference to introgression. Journal of Heredity 15**:** 795-809.

Heiser, C. B., 1973 Introgression reexamined. Botanical Review 39**:** 347-366.

Hejase, H. A., and K. J. Liu, 2016 A scalability study of phylogenetic network inference methods using empirical datasets and simulations involving a single reticulation. BMC Bioinformatics 17**:** 422.

Hey, J., Y. Chung, A. Sethuraman, J. Lachance, S. Tishkoff *et al.*, 2018 Phylogeny estimation by integration over isolation with migration models. Molecular Biology and Evolution 35**:** 2805-2818.

Hey, J., and R. Nielsen, 2007 Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proceedings of the National Academy of Sciences of the United States of America 104**:** 2785-2790.

Hibbins, M. S., and M. W. Hahn, 2019 The timing and direction of introgression under the multispecies network coalescent. Genetics 211**:** 1059-1073.

Hudson, R. R., 1983 Testing the constant-rate neutral allele model with protein sequence data. Evolution 37**:** 203-217.

Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18**:** 337-338.

Huerta-Sanchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter *et al.*, 2014 Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512**:** 194-197.

Huson, D. H., and D. Bryant, 2006 Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23**:** 254-267.

Huson, D. H., T. Klöpper, P. J. Lockhart and M. A. Steel, 2005 Reconstruction of reticulate networks from gene trees, pp. 233-249 in *The 9th Annual International Conference Research in Computational Molecular Biology*, edited by S. Miyano, J. Mesirov, S. Kasif, S. Istrail, P. A. Pevzner *et al.* Springer, Berlin.

Huson, D. H., and C. Scornavacca, 2012 Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Systematic Biology 61**:** 1061-1067.

Jiao, X., and Z. Yang, 2020 Defining species when there is gene flow. Systematic Biology. 70: 108-119.

Joly, S., P. A. McLenachan and P. J. Lockhart, 2009 A statistical approach for distinguishing hybridization and incomplete lineage sorting. The American Naturalist 174: E54-70.

Kingman, J. F. C., 1982 The coalescent. Stochastic Processes and their Applications 13: 235-248.

Knowles, L. L., H. Huang, J. Sukumaran and S. A. Smith, 2018 A matter of phylogenetic scale: distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. American Journal of Botany 105: 376-384.

Kong, S., and L. S. Kubatko, 2021 Comparative performance of popular methods for hybrid detection using genomic data. Systematic Biology: syaa092.

Kronforst, M. R., M. E. Hansen, N. G. Crawford, J. R. Gallant, W. Zhang *et al.*, 2013 Hybridization reveals the evolving genomic architecture of speciation. Cell Reports 5: 666-677.

Kubatko, L. S., and J. Chifman, 2019 An invariants-based method for efficient identification of hybrid species from large-scale genomic data. BMC Evolutionary Biology 19: 112.

Kuhlwilm, M., S. Han, V. C. Sousa, L. Excoffier and T. Marques-Bonet, 2019 Ancient admixture from an extinct ape lineage into bonobos. Nature Ecology and Evolution 3: 957-965.

Leppala, K., S. V. Nielsen and T. Mailund, 2017 admixturegraph: an R package for admixture graph manipulation and fitting. Bioinformatics 33: 1738-1740.

Lohse, K., and L. A. Frantz, 2014 Neandertal admixture in Eurasia confirmed by maximum-likelihood analysis of three genomes. Genetics 196: 1241-1251.

Long, C., and L. Kubatko, 2018 The effect of gene flow on coalescent-based species-tree inference. Systematic Biology 67: 770-785.

Mallet, J., N. Besansky and M. W. Hahn, 2016 How reticulated are species? BioEssays 38: 140-149.

Martin, S. H., and W. Amos, 2020 Signatures of introgression across the allele frequency spectrum. Molecular Biology and Evolution. 38:716-726.

Martin, S. H., J. W. Davey and C. D. Jiggins, 2015 Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Molecular Biology and Evolution 32: 244-257.

Mendes, F. K., and M. W. Hahn, 2018 Why concatenation fails near the anomaly zone. Systematic Biology 67**:** 158-169.

Meng, C., and L. S. Kubatko, 2009 Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theoretical Population Biology 75**:** 35-45.

Nachman, M. W., and B. A. Payseur, 2012 Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal Society B Biological Sciences 367**:** 409-421.

Nei, M., and W. H. Li, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America 76**:** 5269-5273.

Nielsen, R., and J. Wakeley, 2001 Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158**:** 885-896.

Noor, M. A., and S. M. Bennett, 2009 Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb) 103**:** 439-444.

Novikova, P. Y., N. Hohmann, V. Nizhynska, T. Tsuchimatsu, J. Ali *et al.*, 2016 Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nature Genetics 48**:** 1077-1082.

Ottenburghs, J., 2020 Ghost introgression: spooky gene flow in the distant past. BioEssays 42**:** e2000012.

Ottenburghs J., K. R. H. S., van Hooft P., van Wieren S.E., Ydenberg R.C., Prins H.H.T., 2017 Avian introgression in the genomic era. Avian Research 8**:** 30.

Pamilo, P., and M. Nei, 1988 Relationships between gene trees and species trees. Molecular Biology and Evolution 5**:** 568-583.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland *et al.*, 2012 Ancient admixture in human history. Genetics 192**:** 1065-1093.

Pease, J. B., 2018 Why phylogenomic uncertainty enhances introgression analyses. Molecular Ecology 27**:** 4347-4349.

Pease, J. B., D. C. Haak, M. W. Hahn and L. C. Moyle, 2016 Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biology 14**:** e1002379.

Pease, J. B., and M. W. Hahn, 2015 Detection and polarization of introgression in a five-taxon phylogeny. Systematic Biology 64**:** 651-662.

1124 Peter, B. M., 2016 Admixture, population structure, and *F*-statistics. Genetics 202**:** 1485-1501.

1125 Pollard, D. A., V. N. Iyer, A. M. Moses and M. B. Eisen, 2006 Widespread discordance of gene
1126 trees with species tree in Drosophila: evidence for incomplete lineage sorting. PLoS
1127 Genetics 2**:** e173.

1128 Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using
1129 multilocus genotype data. Genetics 155**:** 945-959.

1130 Przeworski, M., B. Charlesworth and J. D. Wall, 1999 Genealogies and weak purifying selection.
1131 Molecular Biology and Evolution 16**:** 246-252.

1132 Racimo, F., S. Sankararaman, R. Nielsen and E. Huerta-Sanchez, 2015 Evidence for archaic
1133 adaptive introgression in humans. Nature Reviews Genetics 16**:** 359-371.

1134 Rambaut, A., and N. C. Grassly, 1997 Seq-Gen: an application for the Monte Carlo simulation of
1135 DNA sequence evolution along phylogenetic trees. Computer Applications in the
1136 Biosciences 13**:** 235-238.

1137 Rieseberg L.H., W. J. F., 1993 Introgression and its consequences in plants, pp. 70-109 in *Hybrid*
1138 *Zones and the Evolutionary Process*. Oxford University Press.

1139 Rosenzweig, B. K., J. B. Pease, N. J. Besansky and M. W. Hahn, 2016 Powerful methods for
1140 detecting introgressed regions from population genomic data. Molecular Ecology 25**:**
1141 2387-2397.

1142 Roux, C., C. Fraisse, J. Romiguier, Y. Anciaux, N. Galtier *et al.*, 2016 Shedding light on the grey
1143 zone of speciation along a continuum of genomic divergence. PLoS Biology 14**:**
1144 e2000234.

1145 Schrider, D. R., J. Ayroles, D. R. Matute and A. D. Kern, 2018 Supervised machine learning
1146 reveals introgressed loci in the genomes of *Drosophila simulans* and *D. sechellia*. PLoS
1147 Genetics 14**:** e1007341.

1148 Schumer, M., G. G. Rosenthal and P. Andolfatto, 2014 How common is homoploid hybrid
1149 speciation? Evolution 68**:** 1553-1560.

1150 Schwarz, G., 1978 Estimating the dimension of a model. The Annals of Statistics 6**:** 461-464.

1151 Sethuraman, A., V. Sousa and J. Hey, 2019 Model-based assessments of differential
1152 introgression and linked natural selection during divergence and speciation. BioRxiv.

1153 Slatkin, M., and J. L. Pollack, 2008 Subdivision in an ancestral species creates asymmetry in
1154 gene trees. Molecular Biology and Evolution 25**:** 2241-2246.

1155 Solís-Lemus, C., and C. Ané, 2016 Inferring phylogenetic networks with maximum

pseudolikelihood under incomplete lineage sorting. PLoS Genetics 12**:** e1005896.

Solís-Lemus, C., P. Bastide and C. Ané, 2017 PhyloNetworks: A package for phylogenetic networks. Molecular Biology and Evolution 34**:** 3292-3298.

Solís-Lemus, C., M. Yang and C. Ané, 2016 Inconsistency of species tree methods under gene flow. Systematic Biology 65**:** 843-851.

Suarez-Gonzalez, A., C. Lexer and Q. C. B. Cronk, 2018 Adaptive introgression: a plant perspective. Biology Letters 14**:** 20170688

Suvorov, A., B. Y. Kim, J. Wang, E. E. Armstrong, D. Peede *et al.*, 2021 Widespread introgression across a phylogeny of 155 Drosophila genomes. BioRxiv.

Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics 105**:** 437-460.

Taylor, S. A., and E. L. Larson, 2019 Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nature Ecology and Evolution 3**:** 170-177.

Than, C., D. Ruths and L. Nakhleh, 2008 PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9**:** 322.

Tricou, T., E. Tannier and D. M. de Vienne, 2021 Ghost lineages deceive introgression tests and call for a new null hypothesis. BioRxiv.

Vanderpool, D., B. Q. Minh, R. Lanfear, D. Hughes, S. Murali *et al.*, 2020 Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. PLoS Biology 18**:** e3000954.

Vaughan, T. G., 2017 IcyTree: rapid browser-based visualization for phylogenetic trees and networks. Bioinformatics 33**:** 2392-2394.

Wakeley, J., and J. Hey, 1997 Estimating ancestral population parameters. Genetics 145**:** 847-855.

Wakeley, J., and J. Hey, 1998 Testing speciation models with DNA sequence data in *Molecular Approaches to Ecology and Evolution*, edited by R. DeSalle and B. Schierwater. Birkhäuser, Basel.

Wang, J., 2003 Maximum-likelihood estimation of admixture proportions from genetic data. Genetics 164**:** 747-765.

Wen, D., and L. Nakhleh, 2018 Coestimating reticulate phylogenies and gene trees from multilocus sequence data. Systematic Biology 67**:** 439-457.

Wen, D., Y. Yu and L. Nakhleh, 2016 Bayesian inference of reticulate phylogenies under the multispecies network coalescent. PLoS Genetics 12**:** e1006006.

Wen, D., Y. Yu, J. Zhu and L. Nakhleh, 2018 Inferring phylogenetic networks using PhyloNet. Systematic Biology 67**:** 735-740.

Williamson, S., and M. E. Orive, 2002 The genealogy of a sequence subject to purifying selection at multiple sites. Molecular Biology and Evolution 19**:** 1376-1384.

Wright, S., 1931 Evolution in Mendelian populations. Genetics 16**:** 97-159.

Wu, D. D., X. D. Ding, S. Wang, J. M. Wojcik, Y. Zhang *et al.*, 2018 Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. Nature Ecology and Evolution 2**:** 1139-1145.

Wu, M., J. L. Kostyun, M. W. Hahn and L. C. Moyle, 2018 Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. Molecular Ecology 27**:** 3301-3316.

Yu, Y., J. H. Degnan and L. Nakhleh, 2012 The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. PLoS Genetics 8**:** e1002660.

Yu, Y., J. Dong, K. J. Liu and L. Nakhleh, 2014 Maximum likelihood inference of reticulate evolutionary histories. Proceedings of the National Academy of Sciences of the United States of America 111**:** 16448-16453.

Yu, Y., and L. Nakhleh, 2015 A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics 16 Suppl 10**:** S10.

Zhang, C., H. A. Ogilvie, A. J. Drummond and T. Stadler, 2018 Bayesian inference of species networks from multilocus sequence data. Molecular Biology and Evolution 35**:** 504-517.

Zhang, D., L. Tang, Y. Cheng, Y. Hao, Y. Xiong *et al.*, 2019 'Ghost introgression' as a cause of deep mitochondrial divergence in a bird species complex. Molecular Biology and Evolution 36**:** 2375-2386.

Zhang, W., K. K. Dasmahapatra, J. Mallet, G. R. Moreira and M. R. Kronforst, 2016 Genome-wide introgression among distantly related *Heliconius* butterfly species. Genome Biology 17**:** 25.

Zheng, Y., and A. Janke, 2018 Gene flow analysis method, the D-statistic, is robust in a wide parameter space. BMC Bioinformatics 19**:** 10.

Zhu, J., D. Wen, Y. Yu, H. M. Meudt and L. Nakhleh, 2018 Bayesian inference of phylogenetic networks from bi-allelic genetic markers. PLoS Computational Biology 14**:** e1005932.

1220

Supplementary Materials and Methods for Hibbins & Hahn 2021

Simulation study under different introgression scenarios

To illuminate many of the patterns and approaches discussed in this review, we conducted a small simulation study. We used the five introgression scenarios shown in Figure 2, as well as one scenario with only ILS and several additional scenarios involving ghost introgression (Supplementary Figure 2). Introgression was simulated in *ms* by specifying an instantaneous population split and join event; this is equivalent to simulating under the multispecies network coalescent framework (Hibbins and Hahn 2019). For each set of conditions, we performed 100 replicate simulations each consisting of 3000 gene trees with branch lengths. We evaluated the performance of three different test statistics designed to capture slightly different information about introgression: D, D_3, and Δ. In addition, we applied the InferNetwork_ML method (Yu et al. 2014) in *PhyloNet*, which infers a phylogenetic network using maximum-likelihood. For the three test statistics, we evaluated significance by bootstrap-resampling the gene trees in each dataset to estimate the sampling variance. The z-score obtained from bootstrap-resampling was used to estimate a two-tailed p-value. The method we use in *PhyloNet* evaluates the fit of a phylogenetic network internally (Yu et al. 2012) using a combination of the model selection measures AIC (Akaike 1974), AICc (Burnham and Anderson 2002), and BIC (Schwarz 1978). For our purposes, a positive result was taken as any result where *PhyloNet* selected a network over a strictly bifurcating tree. See Supplementary Table 1 for the simulation parameters used for each condition.

The power of each method to detect introgression under each scenario is shown in Supplementary Figure 3. All four methods yielded low false positive rates in the presence of high ILS but no introgression, confirming that they are effective tests against an ILS-only null hypothesis. For non-sister taxa, *PhyloNet* was always capable of identifying introgression, while the power of the other methods was strongly affected by the direction of introgression. A reduction of power for *P1* → *P3* introgression is consistent with the effect of direction on gene tree branch lengths described above, but the magnitude of the reduction is somewhat surprising: there is almost three times as much power to detect *P3* → *P1* introgression. Of the four methods, only *PhyloNet* appears capable of reliably inferring introgression between sister lineages, again consistent with expectations.

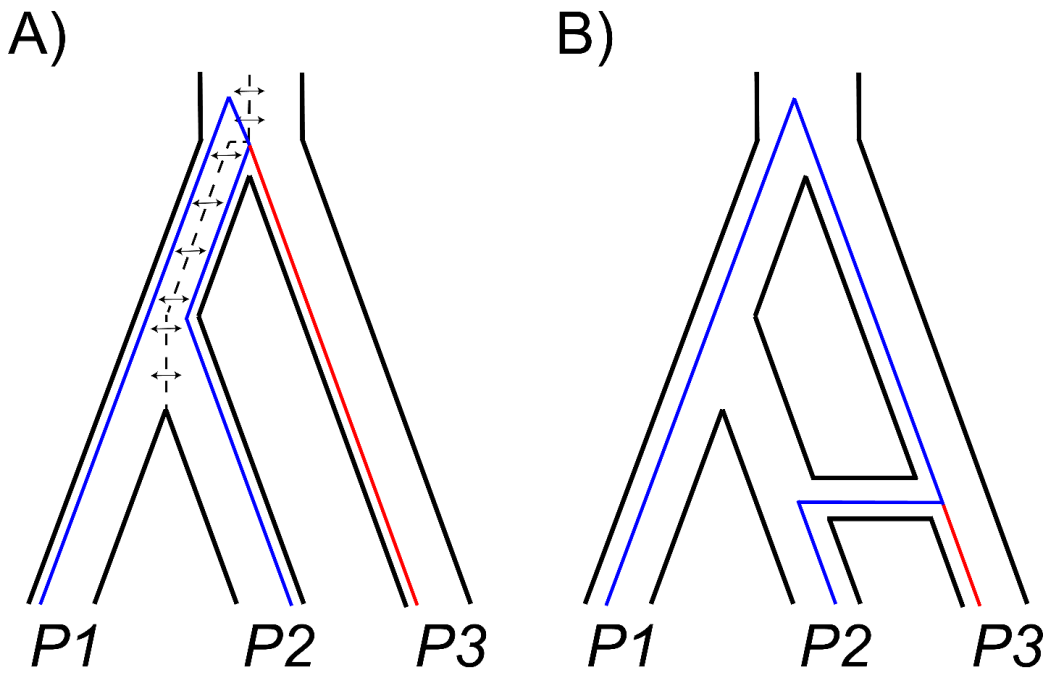
The D and Δ statistics, as well as *PhyloNet*, did not give significant results when introgression occurred between *P1* and an unsampled ingroup lineage. The D_3 statistic, interestingly, does appear to be sensitive to this scenario, at least when the ghost population is the donor. This suggests that patterns of pairwise divergence may be especially useful for detecting introgression with unsampled populations. When introgression occurs between *P1* and an outgroup ghost lineage, there is no effect when the ghost is the recipient, while all four methods are strongly affected when the ghost is the donor. These observations are consistent with expectations for ghost populations, highlighting the importance of careful interpretation of the potential taxa involved in a positive result. In this case, all methods appear to suggest introgression between *P2* and *P3*, even though neither of these lineages was involved in the introgression. This occurs

because introgression from outside the rooted triple draws *P1* to the outside as well, leaving *P3* more closely related to *P2.*

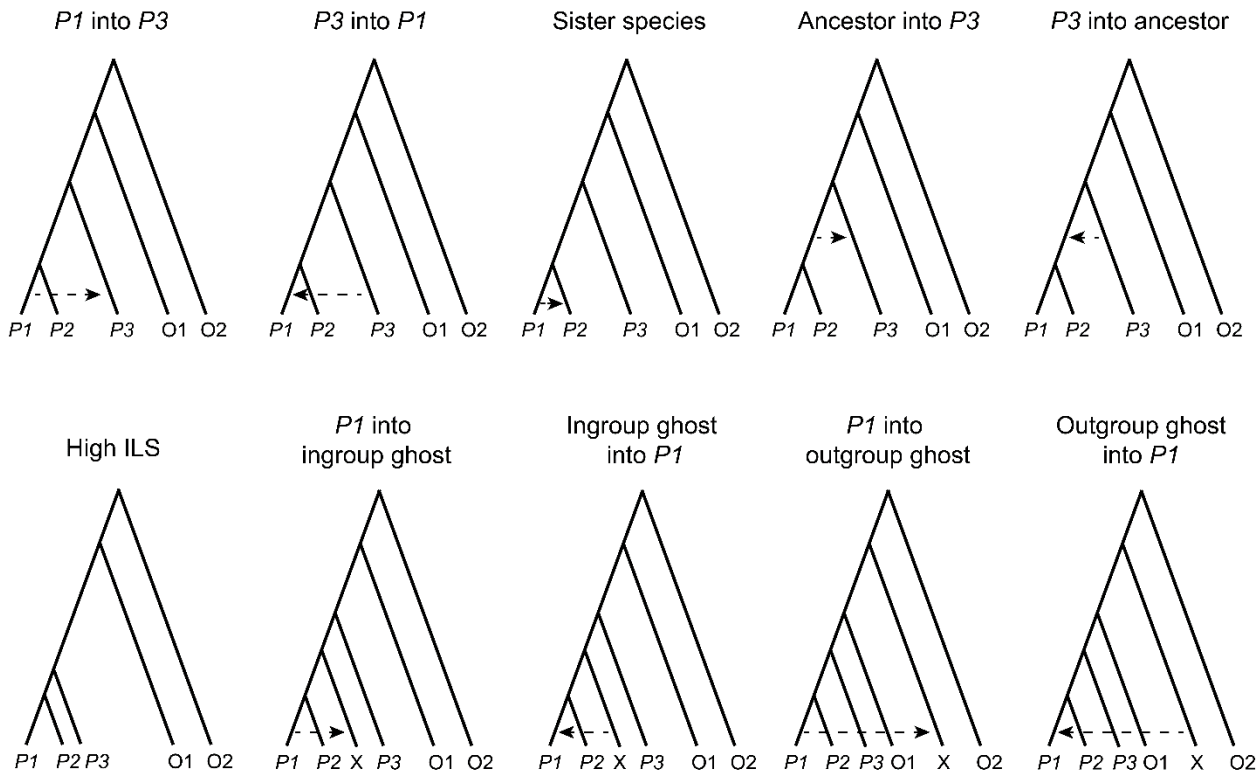
In addition to testing for the presence of introgression, we evaluated the ability of *PhyloNet* to infer the direction of introgression, and of several methods to infer the rate of introgression. We evaluated the ability of *PhyloNet* to correctly identify the taxa involved, the donor and recipient lineages, and the rate of introgression. For the two conditions involving introgression between non-sister taxa, we additionally estimated the rate of introgression using the D_p statistic and an analogous version of the Δ statistic where the count of the concordant tree topology was added to the denominator; we refer to this statistic as Δ_p.

We found that *PhyloNet* was highly accurate at identifying the taxa and direction for *P1* → *P3* introgression (Supplementary Figure 3). However, somewhat surprisingly, it often failed to identify the taxa involved when introgression was *P3* → *P1* (although it always correctly identified that introgression had occurred somewhere). While it is more difficult to detect introgression in the *P1* → *P3* direction, once it is detected it appears that the additional signal in gene tree branch lengths makes it easier for *PhyloNet* to infer the direction. For sister lineages, *PhyloNet* always correctly identified the taxa, but cannot accurately infer the direction. However, *PhyloNet* must always specify the direction of introgression, and its behavior differs between scenarios. For introgression between extant sister species, the direction of introgression appears to be assigned randomly, while for ancestral sister species introgression is always inferred to be in one direction. For the rate of introgression, *PhyloNet* appears to slightly overestimate the true rate under all scenarios in which it correctly identified introgression (Supplementary Figure 4). By contrast, D_p and Δ_p tend to slightly underestimate the rate of introgression between non-sister taxa (Supplementary Figure 4).

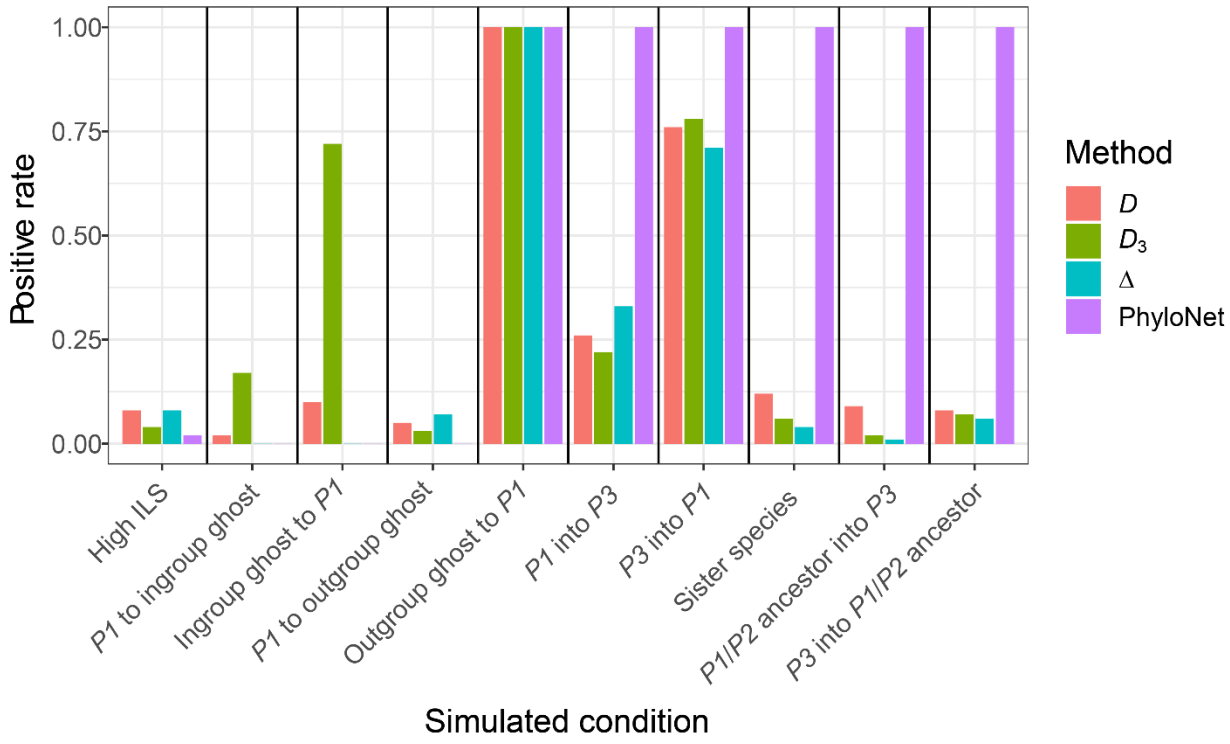
Supplementary Figures and Tables



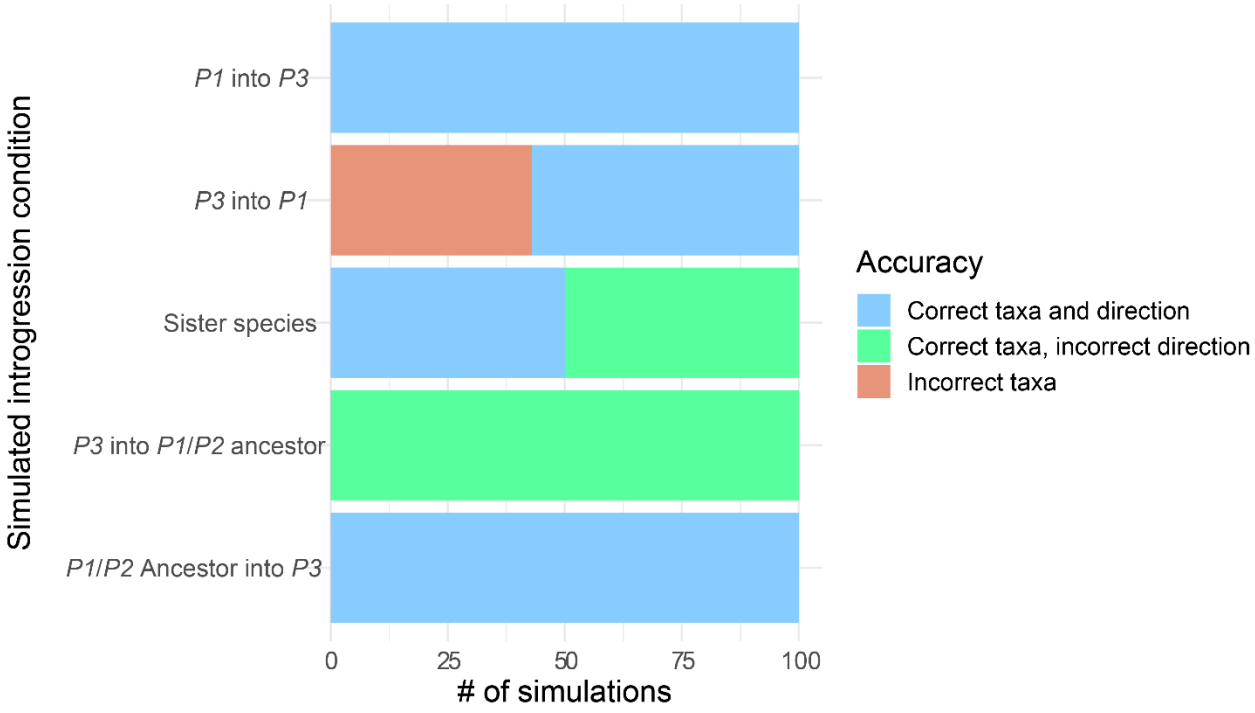
Supplementary Figure 1: Distinguishing ancestral population structure (A) from introgression (B). Persistent structure in the ancestral population of a quartet, which may or may not continue after the first speciation event, can result in the same asymmetries in gene tree topologies and divergence times that are expected from introgression between non-sister taxa. These two scenarios are distinguishable by studying the distribution of branch lengths, in particular the length of the tip branch leading to *P3* (red).



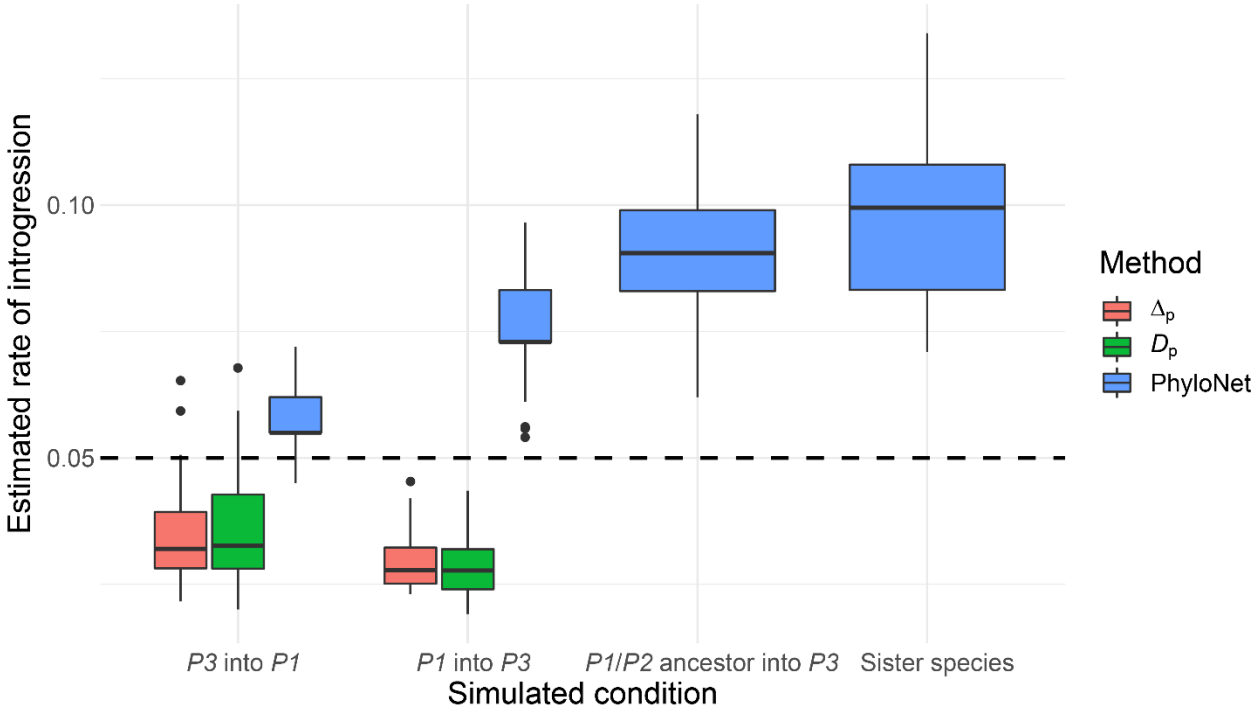
Supplementary Figure 2: A visual overview of the ten different conditions used in our simulation study. Branch lengths are not to scale.



Supplementary Figure 3: Power (y-axis) of four different methods (color legend) to infer the presence of introgression across ten different simulation conditions (x-axis). Power is measured as the proportion of tests that are significant; for the "High ILS" condition it therefore represents the false positive rate.



Supplementary Figure 4: The power of *PhyloNet* to identify the taxa involved and direction of introgression across five simulation conditions.



Supplementary Figure 5: Accuracy of three methods (color legend) for estimating the rate of introgression (y-axis) across four simulation conditions (x-axis). The horizontal dashed line shows the true simulated rate of introgression.

Condition	P1/P2_split	P1P2/P3_split	P1P2P3/O1_split	O1/O2_split	intro_timing	intro_rate	ghostpop_split	theta
P1 into P3	0.6	1.2	8	20	0.3	0.05	N/A	0.005
P3 into P1	0.6	1.2	8	20	0.3	0.05	N/A	0.005
Sister species	0.6	1.2	8	20	0.3	0.05	N/A	0.005
Ancestor into P3	0.6	1.2	8	20	0.9	0.05	N/A	0.005
P3 into ancestor	0.6	1.2	8	20	0.9	0.05	N/A	0.005
High ILS	0.6	0.62	8	20	N/A	0.05	N/A	0.005
P1 into ingroup ghost	0.6	8	20	30	0.3	0.05	1.2	0.005
Ingroup ghost into P1	0.6	8	20	30	0.3	0.05	1.2	0.005
P1 into outgroup ghost	0.6	1.2	8	30	0.3	0.05	20	0.005
Outgroup ghost into P1	0.6	1.2	8	30	0.3	0.05	20	0.005

Supplementary Table 1: Parameters used for introgression simulation conditions in *ms*. Split times and theta are in units of *2N* generations.