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, using a small simulation study to highlight the different signals of introgression and the power of each method to detect them. We conclude with a discussion of how to visualize and interpret the results of introgression analyses.

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). However, until genome sequencing became widely available to biologists, it was difficult to quantify patterns of introgression effectively and reliably. As a result, introgression's role in evolution was under-appreciated, especially in animal systems. 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), comparatively little synthesis has been provided on the accompanying growth in methods used to detect and characterize introgression.

The information that can be gleaned from genomic data about introgression depends on both the number of sampled species and the number of sampled individuals. Methods with only two species or populations depend on sampling multiple individuals within at least one of them. Patterns of nucleotide variation among individuals and across loci can then be used to make inferences about introgression (e.g. Wakeley and Hey 1997, Nielsen and Wakeley 2001, Joly et al. 2009, Lohse and Frantz 2014, Rosenzweig et al. 2016, Schrider et al. 2018). Because less information is available about phylogenetic relationships, these methods often rely on the assumption that all sequenced loci are evolving neutrally or that all loci have the same rate of nucleotide substitution (or both). For these reasons such methods are more prone to model violations, such as the heterogeneous effects of background selection across loci (Roux et al. 2016). Despite these limitations, population-genetic methods are one of the few approaches that can infer gene flow between pairs of sister taxa (see Hahn 2018 for more details).

When there is data for a rooted triplet of species—or an unrooted quartet—it becomes possible to construct more powerful tests for introgression using genome-scale datasets. Importantly, this can be done using only a single sample per species and without assumptions about neutrality. The robustness to non-neutral processes 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). This class of "phylogenomic" methods is largely based on one sample per species, but also includes methods based on multiple samples. One-sample 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). When multiple individuals are sampled from very closely related species or populations, additional power may be gained by measuring the

deviation from covariances in allele frequency expected under strictly treelike evolution (Reich et al. 2009, Patterson et al. 2012, Pickrell and Pritchard 2012, Peter 2016).

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, and the methods that are designed to detect these signals. We discuss common misuses and misinterpretations of these methods, as well as providing recommendations for best-use practices. Finally, we present results from a small simulation study conducted across different introgression scenarios to highlight the advantages and limitations of currently available methods. Based on these results, we identify areas for future theoretical and methodological advancement.

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, 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 with each other before reaching an 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 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. In addition, the mating requirement means that the hybridizing species are closely related enough such that the tree topologies produced by introgression will likely be the same as those produced by ILS. 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, 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*. Loci with an introgressed history therefore have a reduced probability of ILS because of the increased time for them 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; Solis-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 sample 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). Although 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.

Finally, event 5 depicts an introgression event involving an unsampled ("ghost") lineage. For many of the signals of introgression discussed here, the sampled lineages included in a study may not be the ones that actually hybridized. Whether species go unsampled because they are extinct or simply unavailable, non-sister ghost lineages that act as donors in introgression events will often generate detectable patterns of gene flow. These patterns can result in misleading inferences about the lineages involved in gene flow and the direction of gene flow, and should therefore always be kept in mind; we include introgression from ghost lineages in our simulation study below to demonstrate some of these effects. Ghost lineages that are either the recipients of gene flow or are sister to sampled taxa are much less likely to leave any signal of introgression. For similar reasons, the sister lineages shown in Figure 2 do not need to be one another's most closely related species in nature; what is important is whether they are sister (or non-sister) among sampled species.

Detecting introgression using gene tree frequencies

The D *statistic*

Perhaps the most 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. This test was originally formulated to test for evidence of gene flow between

Neanderthals and archaic humans (Green et al. 2010, Durand et al. 2011), and is based on the effect of introgression between non-sister taxa on gene tree frequencies.

The 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 phylogenetically 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 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. 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 phylogenetically 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.

Finally, the *D*-statistic does not provide any information about introgression other than its presence or absence. While its value does increase with the rate of introgression, 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 *P2* and *P3* could involve the $P3 \rightarrow P2$ direction, the $P2 \rightarrow P3$ direction, or some combination of the two. 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. Methods for inferring introgression can often be used to estimate its "rate," which 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 former definition.

The degree of asymmetry in discordant gene tree topologies contains information about the proportion of introgressed loci across the genome. However, simply using the *D* statistic does not provide an unbiased estimation of the rate (Martin et al. 2015, Hamlin et al. 2020). A recently

proposed extension of D called D_p makes one simple addition that improves the estimate of the proportion of introgressed loci. The statistic 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 phylogenetically 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—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).

In an alternative approach, the observed value of an introgression test statistic is compared to the value that would be expected under a scenario where the entire genome was introgressed. The F_4-ratio or α statistic (Green et al. 2010, Patterson et al. 2012, Peter 2016) makes this comparison by taking the ratio of two F_4 statistics (a genome-wide test for introgression based on allele frequencies). The α statistic requires data from five samples and assumes an admixed population with two parent populations. *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 these methods are somewhat restrictive and are likely not 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).

The f_d statistic of Martin et al. (2015) also takes the ratio of two D-statistics. However, by making the assumption that allele frequencies would be completely homogenized in a complete introgression scenario, f_d can be applied to quartets rather than requiring an additional sample. Like D_p, f_d is sensitive to the direction of introgression because it estimates the proportion of the genome that came from the donor population during introgression. The f_d statistic somewhat overcomes this issue by assuming that the population with the higher derived allele frequency is the donor at each site. Nonetheless, f_d tends to underestimate the proportion of introgressed loci when *P2* is the donor population.

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) showed that information about the rate, direction, and timing of introgression in a quartet becomes available using site patterns if multiple individuals are sampled per lineage. Their approach, called the "*D* frequency spectrum" or D_{FS} for short, 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 of introgression. If one of the sister taxa is the recipient, then the spectrum is left-skewed, as the signal of introgression will be enriched among low-frequency alleles. In contrast, if a sister lineage is the donor and the non-sister lineage is the recipient, the spectrum is expected to be flat, because the frequency spectrum of the non-sister lineage is not used to construct the D_{FS}. The degree of left-skewness is affected by the timing of introgression, while the rate of introgression affects the magnitude of the *D*-statistic across bins. 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 it becomes more likely that ABBA and BABA site patterns can accumulate due to convergent substitutions. For this reason, site patterns are generally not 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, 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. 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. Phylogenetic network methods can handle several different data types (which will be discussed in subsequent sections), but some of them can make inferences using only gene tree topologies as input. *PhyloNet* (Than et al. 2008, Wen et al. 2018) infers networks directly from gene tree topologies using either maximum likelihood or maximum pseudo-likelihood. Similarly, *SNaQ* (Solis-Lemus & Ane 2016) estimates a network with reticulation edges via maximum pseudo-likelihood using quartet concordance factors (Baum 2007)—essentially just the counts of the three possible unrooted tree topologies. The principles outlined in previous sections apply equally well to these methods, showing how phylogenetic network approaches can detect, polarize, and estimate the rate of introgression events in a tree with a minimum of five taxa. Additionally, they can be applied to species trees with many more than five taxa, making use of all available information available. We will discuss phylogenetic network methods in more detail later, in the section entitled "*Application and interpretation of methods for inferring 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 sample from three ingroup species. As with *D*, D_3 can only detect introgression between non-sister lineages.

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. 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). *QuIBL* takes gene trees with inferred branch lengths as input, using maximum-likelihood to infer whether one distribution (ILS-only) or two distributions (ILS + introgression) is a better fit. If two distributions is a better fit, then introgression between non-sister species is inferred. QuIBL may also be able to infer the timing and rate of introgression using information contained within these distributions.

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* (Than et al. 2008, Wen et al. 2018) is able to infer phylogenetic networks with reticulation edges (i.e. discrete introgression events) from gene trees with branch lengths using maximum likelihood. Based on the previously discussed patterns, this method should be capable of accurately estimating the presence, timing, direction, and rate of introgression by making use of all available information on the distribution of coalescence times. It can also estimate multiple independent events on the same species tree, and on trees with more than five taxa. As we discuss in a following section, it is also capable of detecting the signals of introgression between sister species.

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*, 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 sample 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 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 (Wen and Nakhleh 2018), at least when nested within a tree containing more taxa. 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.

Application and interpretation of methods for inferring introgression

Evaluating the power to detect and characterize introgression

To illuminate many of the patterns and approaches discussed in this review, we conducted a small simulation study using *ms* (Hudson 2002) and *Seq-Gen* (Rambaut and Grassly 1997). 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). For each set of conditions, we performed 100 replicate simulations each consisting of 3000 gene trees with branch lengths. We simulated 1kb per locus using *Seq-Gen* with $\theta = 0.005$ per *2N* generations. 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 Figure 6. 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 $P1 \rightarrow P3$ 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 $P3 \rightarrow P1$ 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 (see below for more explanation), 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).

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, it 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 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 pseudo-likelihood such as *SNaQ*, because these information measures are not intended for pseudo-likelihood 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 (see next section).

Visualizing and interpreting phylogenetic networks

When visualizing inferred phylogenetic networks, reticulations represent the histories of loci that have introgressed. Visually, the relative placement, orientation, and length of these reticulations imply specific information about the timing and direction of introgression, as well as the identity of the species involved. However, not all phylogenetic networks are constructed from the same underlying models, and therefore they may not always convey the same information (Huson and Bryant, 2005). As a result, choices for network visualization that are primarily stylistic can unintentionally imply specific introgression processes. In this section we discuss these different visualizations and how to interpret them.

One important distinction when visualizing networks is the contrast between introgression that occurs among extant lineages and introgression that results in the formation of a new lineage. Supplementary Figure 5A depicts the former scenario, which corresponds to the introgression scenarios considered in this paper thus far. In such cases, a single horizontal reticulation is

typically used to connect the two taxa involved. Such visualizations do not naturally convey any information about the direction of introgression. 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 Supplementary Figure 5B, where reticulations connect each parent lineage to the newly formed lineage. This representation implies a directionality of introgression: from the two parent lineages into the newly formed lineage. In both cases, a horizontal reticulation edge is used to denote the instantaneous exchange of alleles between the involved lineages. Supplementary Figure 5C 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). Alternatively, this could represent "standard" introgression involving a now-extinct species, in which case the extinct lineage was the donor in the introgression scenario. In all three cases, the placement of the reticulation edge conveys information about the timing of introgression and/or lineage formation.

The key take-away from this last representation is that non-horizontal reticulation edges often imply directionality, with the introgressed alleles travelling toward the tips. Unfortunately, many automated methods for visualizing species networks do not allow strictly horizontal edges—instead, all reticulations must have a bifurcating "parent" node that occurs closer to the root than the "daughter" node, which has two incoming lineages. This was the behavior observed in *PhyloNet* in the previous section. To highlight how different network visualizations can potentially be (mis-)interpreted, with particular emphasis on the direction of introgression, we plotted the same inferred networks using four popular tools (Figure 5): *Dendroscope* (Huson and Scornavacca 2012), *IcyTree* (Vaughan 2017), *PhyloPlots*, which is part of the Julia package *PhyloNetworks* (Solís-Lemus et al. 2017), and *admixturegraph* (Leppälä et al. 2017). The networks were inferred using *PhyloNet* on simulated gene trees from the two non-sister introgression scenarios (i.e. both *P1* → *P3* and *P3* → *P1*) discussed in the previous section. For *admixturegraph*, we simply plotted the outcome of applying the *D*-statistic to the data under both scenarios.

In *Dendroscope*'s visualizations (Figure 5A, 5B), the position of the hybrid node (the node with two parents) is made clear, but it is not clear which parent corresponds to a history of introgression vs. the species tree history, since both are represented using a curved blue line (and therefore both resemble reticulation edges). As a result, the direction of introgression confounds accurate representation of the underlying introgression scenario. For the *P3* → *P1* direction (Figure 5A), the visualization strongly implies that *P1* is a hybrid species that formed from hybridization between *P2* and *P3*. While this representation accurately conveys the fact that *P1* is the recipient of introgressed alleles, it unfortunately suggests that *P2* was involved in hybridization when it was not. The *P1* → *P3* visualization (Figure 5B) is easier to interpret, because one of the blue edges cannot possibly represent an introgression history. Additionally, the curvature of the blue edges suggests a non-horizontal reticulation, which may imply a period of independent evolution. However, in this case it is purely stylistic as the network does not contain any branch lengths. Finally, it is important to note that all the visualizations we discuss do not take branch lengths into account, so the placement of the reticulation edges within

branches of the species tree are arbitrary and do not convey information about the timing of introgression. *IcyTree* and *PhyloPlots* are capable of plotting networks with branch lengths, in which case the timing of introgression within lineages can be displayed. Since our primary concern is with the direction of introgression, we have not shown these visualizations.

IcyTree uses a different style of visualization (Figure 5C, D). A dashed line represents the reticulation edge, which branches off from the donor population and enters the recipient population. This allows the introgression and species tree histories to be more visually distinct, while still depicting the direction of introgression. However, it implies that a lineage branched off from the donor and underwent a period of independent evolution before entering the recipient, which did not happen in either case. As the network is plotted without branch lengths, the point at which the reticulation leaves the donor branch is arbitrary. *PhyloPlots* (Figure 5E, F) visually distinguishes the reticulation edge (light blue) from the species tree history (black) while explicitly labelling the hybrid node. The reticulation edge is not horizontal, erroneously implying some period of independent evolution, though it does effectively convey the direction of introgression. The distinct coloration of each history, in combination with labelling of the hybrid node, means that the direction of introgression can be easily visualized. Finally, *admixturegraph* (Figure 5G, 5H) plots the network solely from the results of a series of *D* tests. This means that no inference of directionality is possible. As with *Dendroscope*, this method plots phylogenetic networks as admixture graphs, which have the same issues with implied directionality and hybrid speciation. In our case, this results in *P1* being the implied recipient of introgression regardless of the true direction.

The message we hope to convey from this discussion is that it is very difficult to simultaneously visualize the direction of introgression and to preserve the underlying model of hybridization. This is especially challenging for cases when network visualization needs to be automated, because the standard computational representation of phylogenetic networks, the Extended Newick format (Cardona et al. 2008), requires labeling of parent and daughter nodes, and therefore implies directionality any time a hybrid node is inferred. “Tube tree” representations like the ones we use for figures in this paper (e.g. Figure 4) can be effective for individual cases, but to our knowledge no automated approaches exist as of yet that can accurately convey all the necessary information. In general, care should be taken not to over-interpret phylogenetic network visualizations.

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. This work was supported by National Science Foundation grant DEB-1936187.

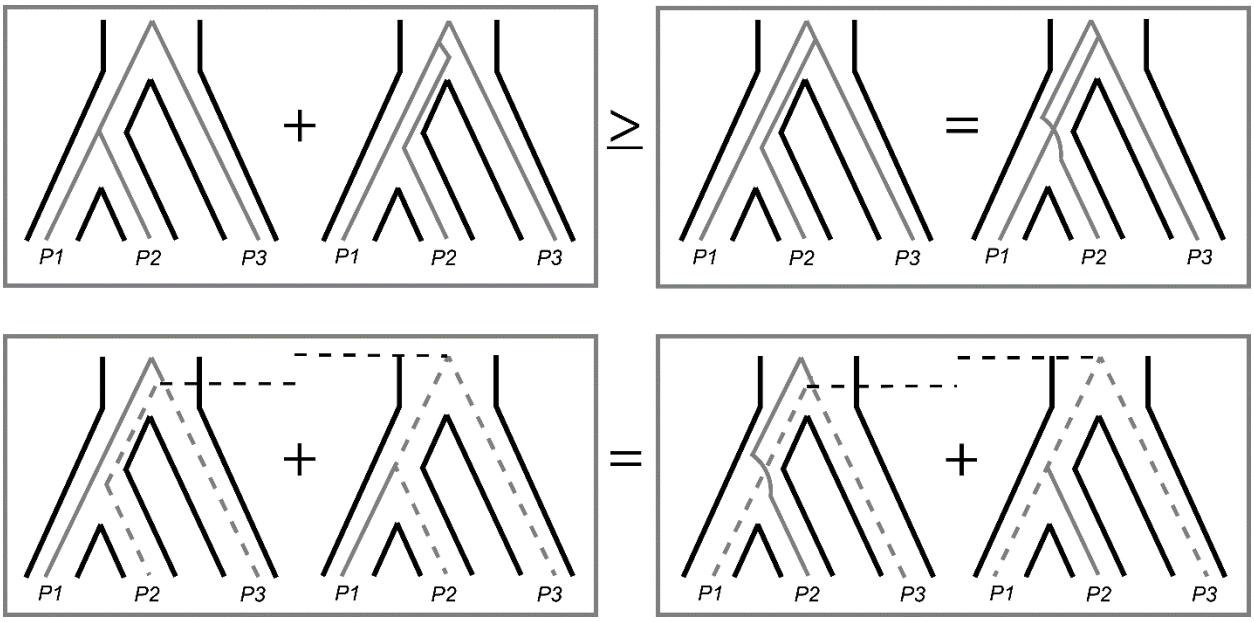


Figure 1: Expected gene tree topologies and coalescence times under ILS only. For a rooted triplet, four topologies are possible (top row): two concordant with the species tree, which can result either from lineage sorting or ILS (top left), and two that are discordant with the species tree and arise 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.

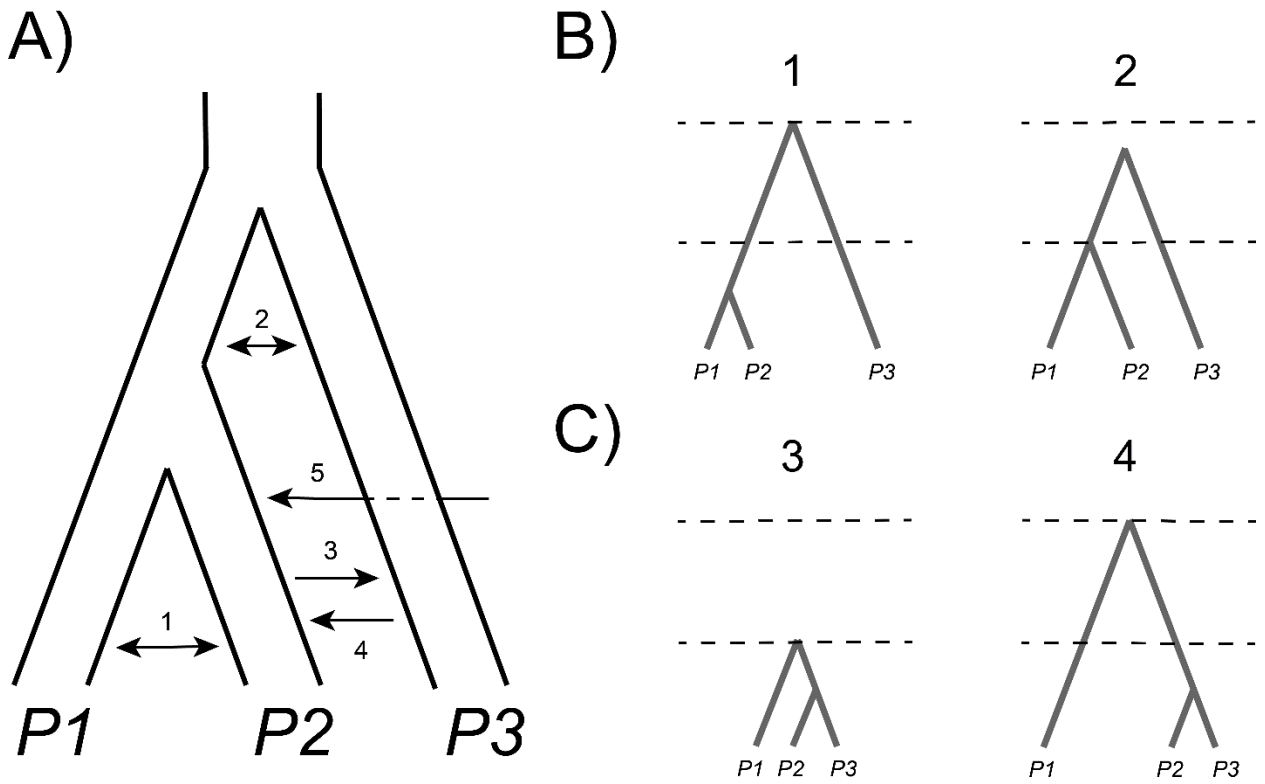


Figure 2: An overview of detectable introgression scenarios for a rooted triplet, and their effects on gene tree frequencies and branch lengths. A. The species tree relating three lineages. Introgression can occur between extant (1) or ancestral (2) sister lineages, or between non-sister taxa, with *P3* as either the recipient (3) or the donor (4). One of the sampled taxa may also be the recipient of introgression from an unsampled taxon (5). B. Gene trees 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 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. No example gene tree is shown for introgression from a ghost lineage outside the triplet (event 5). The expectation is that these events will generate topologies with *P2* pulled outside the clade, sister to the two unintrogressed lineages.

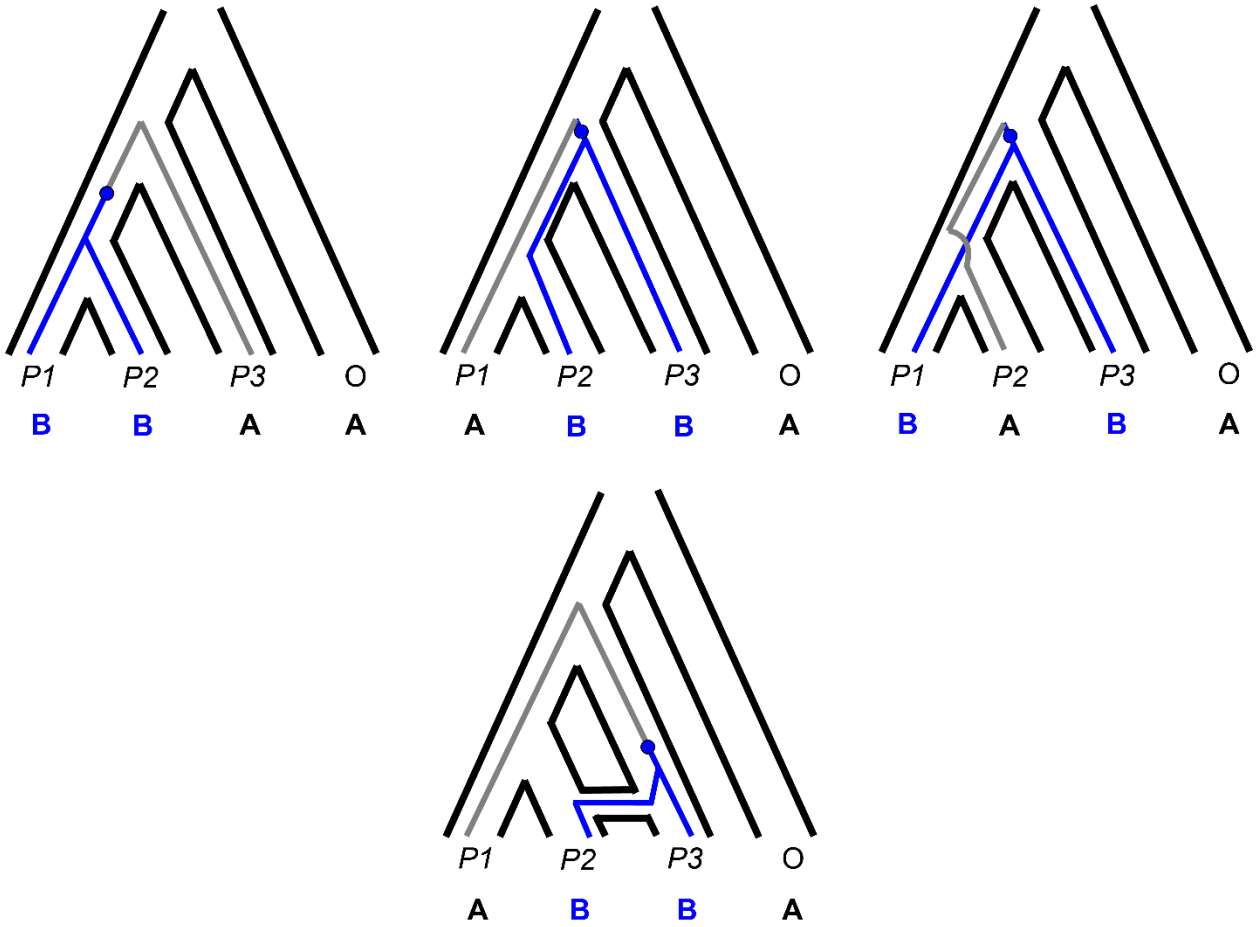


Figure 3: Biallelic site patterns are informative of underlying gene tree topologies. With the exception of low levels of homoplasy, such patterns can only arise from mutations (blue) on internal branches of the local genealogy. The occurrence of the incongruent site patterns "ABBA" (top middle) and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

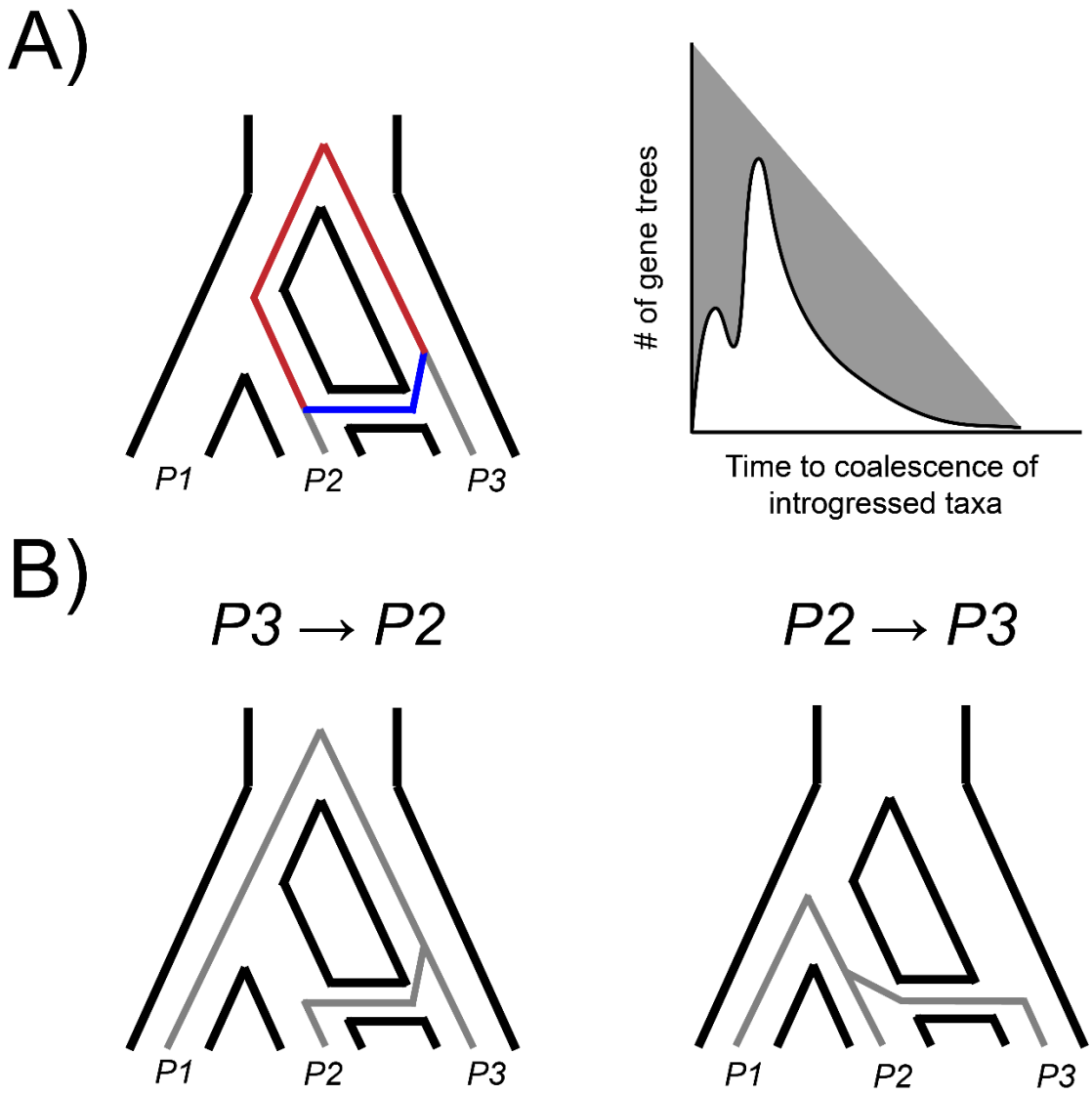


Figure 4: Coalescence times provide information on the timing, direction, and presence of introgression. A) Post-speciation introgression between *P2* and *P3* allows them to coalesce more quickly at introgressed loci (blue). This reduces their whole-genome divergence relative to *P1* and *P3*, an asymmetry 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.

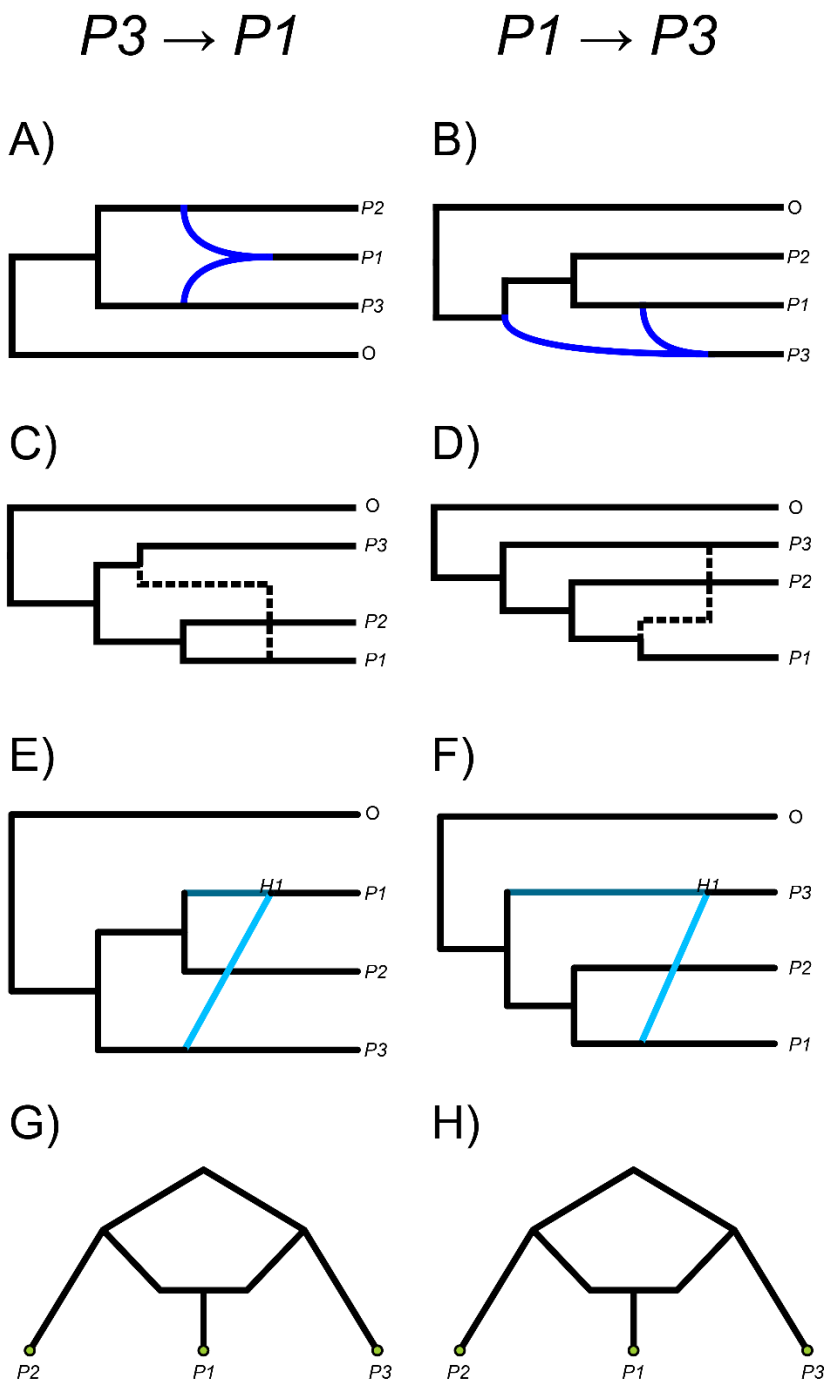


Figure 5: Different visualizations of the same underlying phylogenetic networks. The left column comes from a network representing *P3* → *P1* introgression, while the right column comes from a network representing *P1* → *P3* introgression. The rows, from top to bottom, show visualizations from *Dendroscope* (A, B), *IcyTree* (C, D), *PhyloPlots* (E, F), and *admixturegraph* (G, H), respectively.

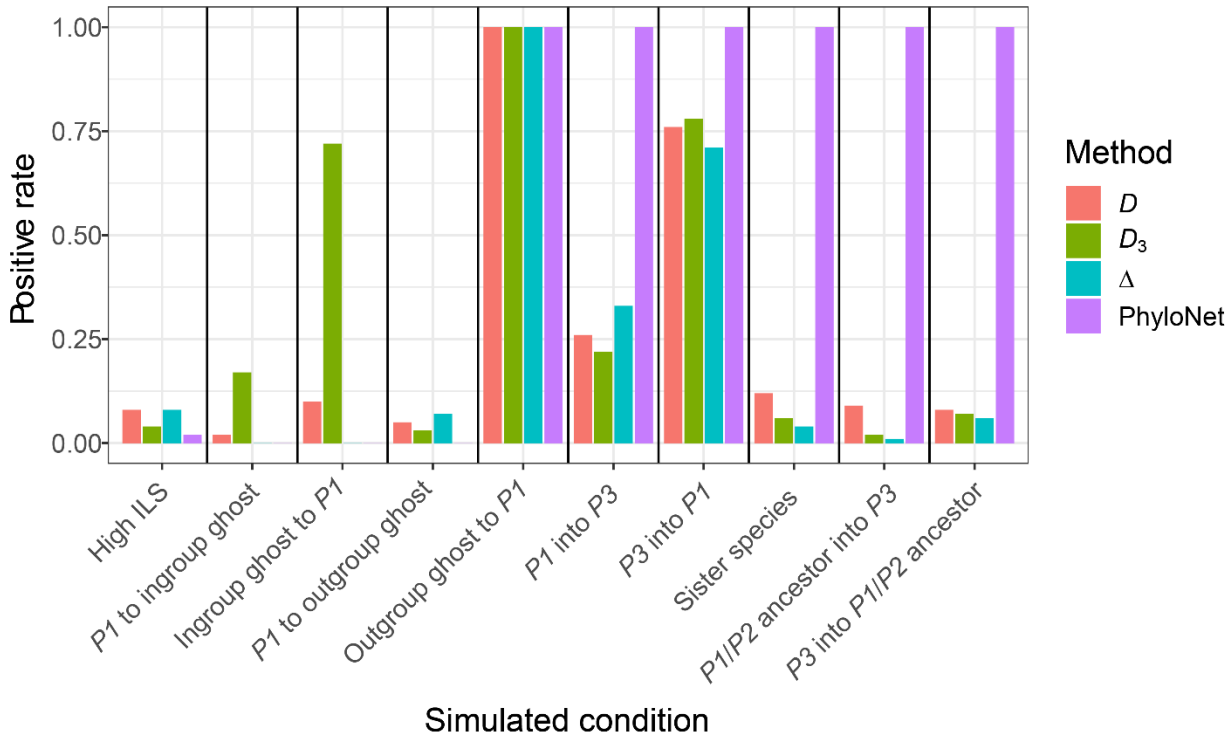
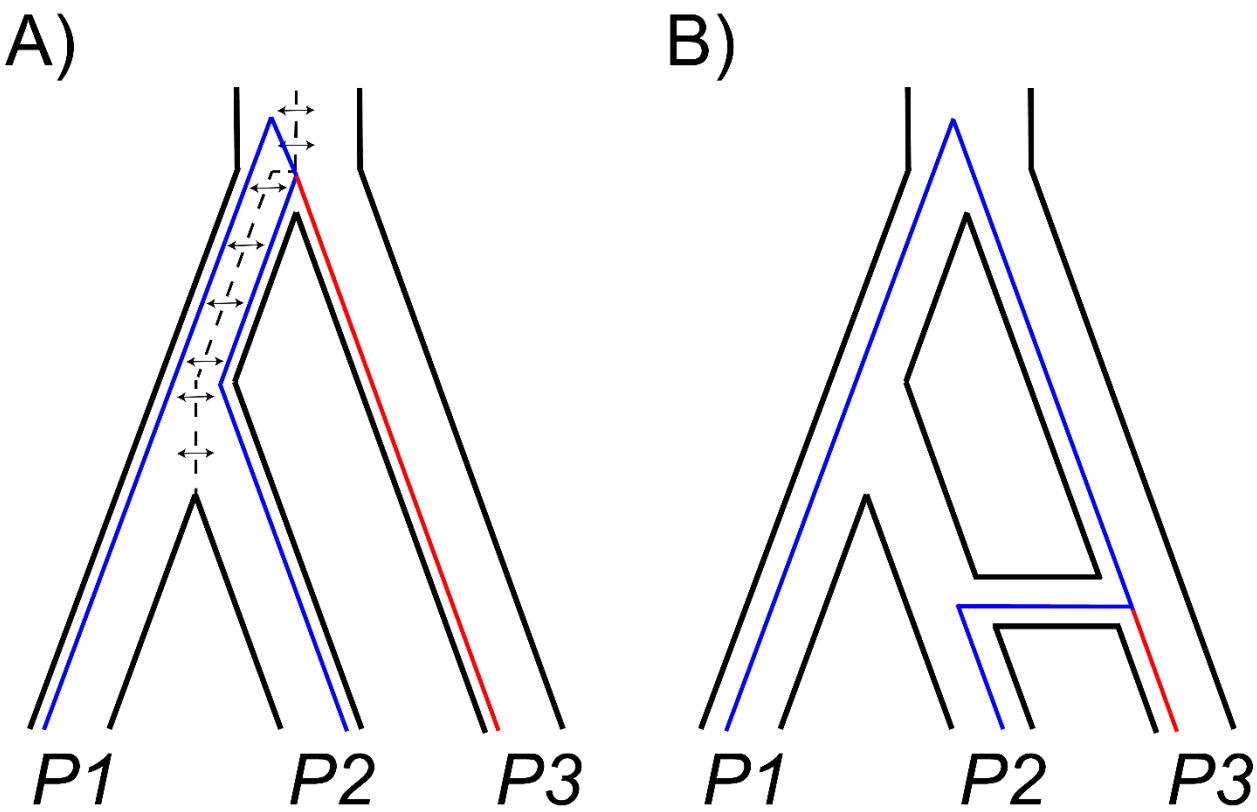
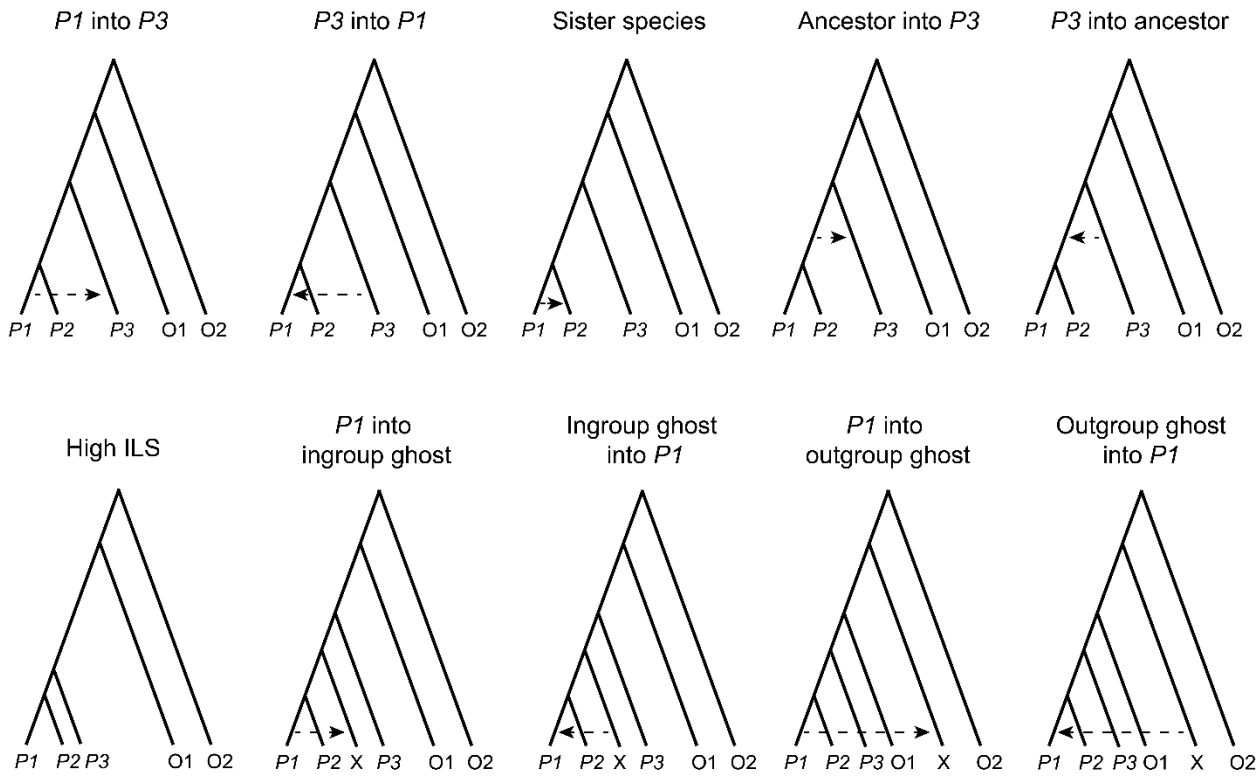


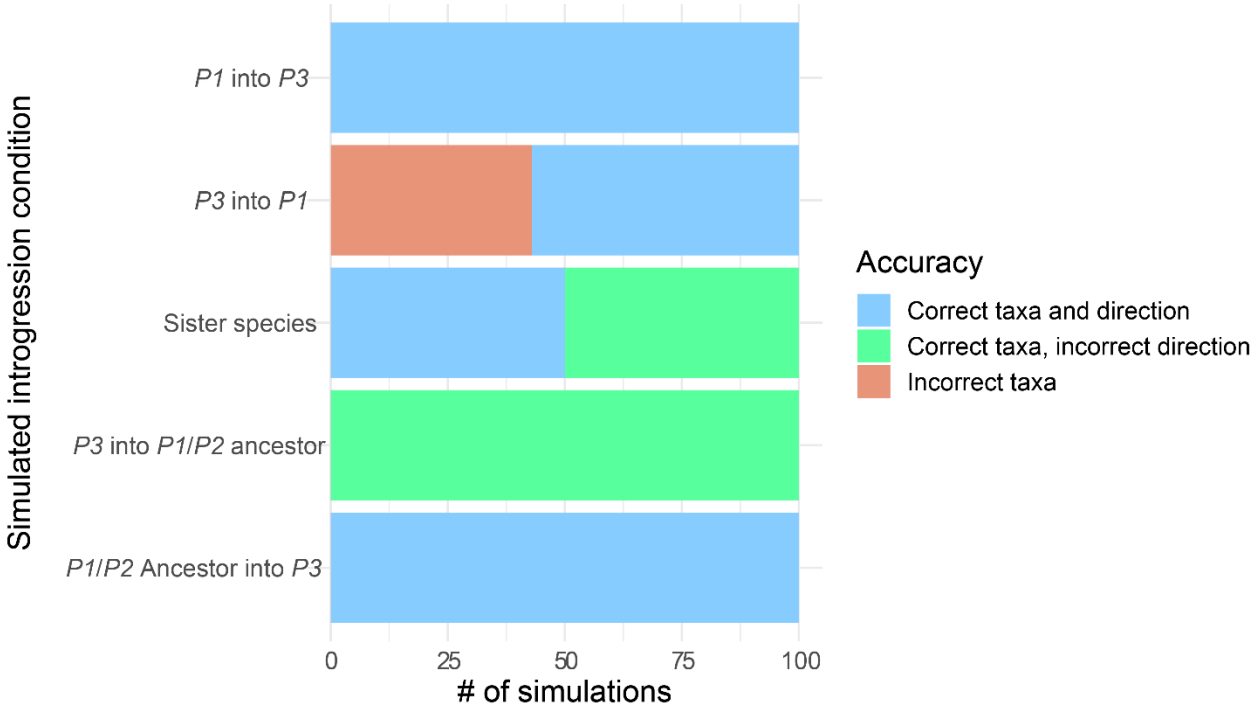
Figure 6: Power (y-axis) of four different methods (color legend) to infer the presence of introgression across ten different simulation conditions (x-axis). Power is measured as the proportion of tests that are significant; for the "High ILS" condition it therefore represents the false positive rate.



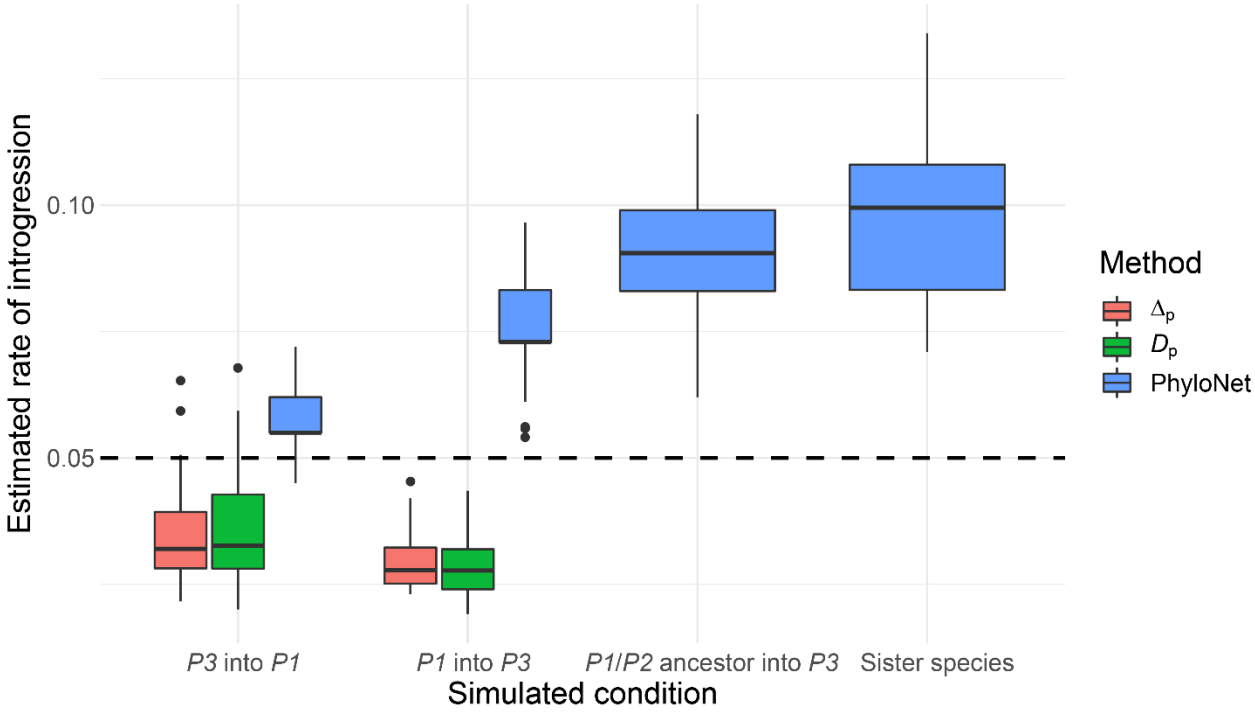
Supplementary Figure 1: Distinguishing ancestral population structure (A) from introgression (B). Persistent structure in the ancestral population of a quartet, which may or may not continue after the first speciation event, can result in the same asymmetries in gene tree topologies and divergence times 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).



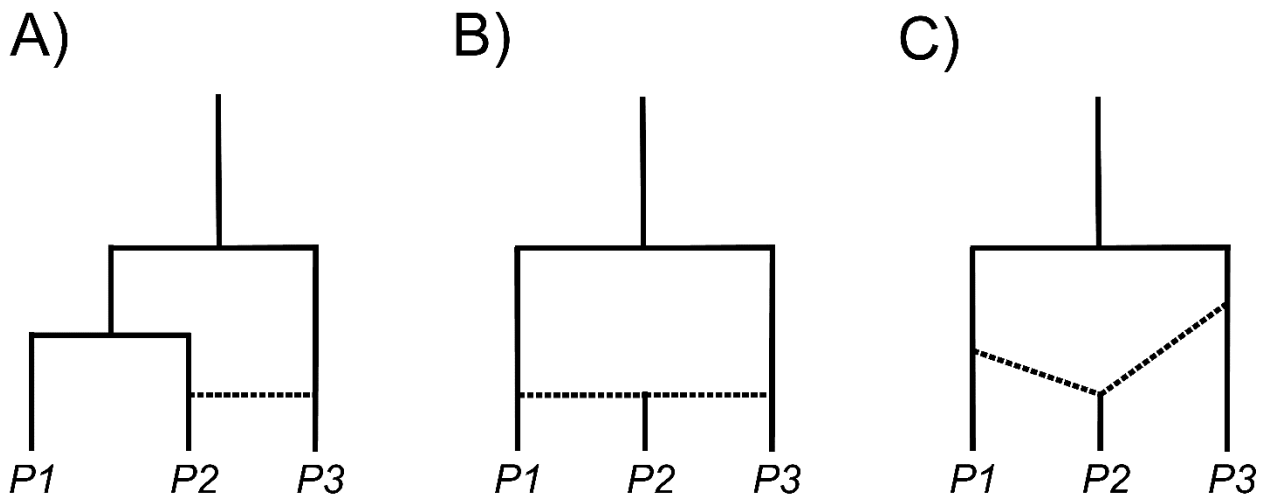
Supplementary Figure 2: A visual overview of the ten different conditions used in our simulation study. Branch lengths are not to scale.



Supplementary Figure 3: The power of *PhyloNet* to identify the taxa involved and direction of introgression across five simulation conditions.



Supplementary Figure 4: 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.



Supplementary Figure 5: Network representations of introgression between extant lineages (A) vs. introgression that results in the formation of a new lineage (B, C).

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.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716-723. doi:10.1109/TAC.1974.1100705

Baum, D. A. (2007). Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* 56(12), 417-426. doi:10.1002/tax.562013

Bertorelle, G., & Excoffier, L. (1998). Inferring admixture proportions from molecular data. *Molecular Biology and Evolution,* 15(10), 1298-1311. doi:10.1093/oxfordjournals.molbev.a025858

Blischak, P. D., Chifman, J., Wolfe, A. D., & Kubatko, L. S. (2018). HyDe: a Python package for genome-scale hybridization detection. *Systematic Biology,* 67(5), 821-829. doi:10.1093/sysbio/syy023

Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genetics,* 10(6), e1004410. doi:10.1371/journal.pgen.1004410

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2 ed.). New York: Springer-Verlag.

Cai, R., & Ané, C. (2020). Assessing the fit of the multi-species network coalescent to multi-locus data. *Bioinformatics*, btaa863. doi:10.1093/bioinformatics/btaa863

Cardona, G., Rossello, F., & Valiente, G. (2008). Extended Newick: it is time for a standard representation of phylogenetic networks. *BMC Bioinformatics,* 9, 532. doi:10.1186/1471-2105-9-532

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution,* 15(5), 538-543. doi:10.1093/oxfordjournals.molbev.a025953

Copetti, D., Burquez, A., Bustamante, E., Charboneau, J. L. M., Childs, K. L., *et al*. (2017). Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Science of the United States of America,* 114(45), 12003-12008. doi:10.1073/pnas.1706367114

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology,* 23(13), 3133-3157. doi:10.1111/mec.12796

Degnan, J. H. (2018). Modeling hybridization under the network multispecies coalescent. *Systematic Biology,* 67(5), 786-799. doi:10.1093/sysbio/syy040

Dowling, T.E., Secor, C. L. (1997). The role of hybridization and introgression in the

diversification of animals. *Annual Review of Ecology, Evolution, and Systematics,* 28, 593-619. doi:10.1146/annurev.ecolsys.28.1.593

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution,* 28(8), 2239-2252. doi:10.1093/molbev/msr048

Eaton, D. A., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology,* 62(5), 689-706. doi:10.1093/sysbio/syt032

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., *et al*. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science,* 366(6465), 594-599. doi:10.1126/science.aaw2090

Ellstrand N.C., M. P., Rong J., Bartsch D., Ghosh A., de Jong T.J., *et al*. (2013). Introgression of crop alleles into wild or weedy populations. *Annual Review of Ecology, Evolution, and Systematics,* 44, 325-345. doi:doi.org/10.1146/annurev-ecolsys-110512-135840

Elworth, R. A. L., Ogilvie, H. A., Zhu, J., & Nakhleh, L. (2019). Advances in computational methods for phylogenetic networks in the presence of hybridization. In T. Warnow (Ed.), *Bioinformatics and Phylogenetics* (Vol. 29, pp. 317-360). Cham: Springer.

Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.

Folk, R. A., Soltis, P. S., Soltis, D. E., & Guralnick, R. (2018). New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany,* 105(3), 364-375. doi:10.1002/ajb2.1018

Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E *et al*. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science,* 347(6217), 1258524. doi:10.1126/science.1258524

Forsythe, E. S., Nelson, A. D. L., & Beilstein, M. A. (2020). Biased gene retention in the face of introgression obscures species relationships. *Genome Biology and Evolution,* 12(9), 1646-1663. doi:10.1093/gbe/evaa149

Forsythe, E. S., Sloan, D. B., & Beilstein, M. A. (2020). Divergence-based introgression polarization. *Genome Biology and Evolution,* 12(4), 463-478. doi:10.1093/gbe/evaa053

Fuller, Z. L., Leonard, C. J., Young, R. E., Schaeffer, S. W., & Phadnis, N. (2018). Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLoS Genetics,* 14(7), e1007526. doi:10.1371/journal.pgen.1007526

Geneva, A. J., Muirhead, C. A., Kingan, S. B., & Garrigan, D. (2015). A new method to scan genomes for introgression in a secondary contact model. *PLoS One,* 10(4), e0118621.

doi:10.1371/journal.pone.0118621

Gillespie, J. H., & Langley, C. H. (1979). Are evolutionary rates really variable? *Journal of Molecular Evolution,* 13(1), 27-34. doi:10.1007/BF01732751

Grau-Bové, X., Tomlinson, S., O'Reilly, A. O., Harding, N. J., Miles, A., *et al*. (2020). Evolution of the insecticide target *Rdl* in African *Anopheles* is driven by interspecific and interkaryotypic introgression. *Molecular Biology and Evolution,* 37(10), 2900-2917. doi:10.1093/molbev/msaa128

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., *et al.* (2010). A draft sequence of the Neandertal genome. *Science,* 328(5979), 710-722. doi:10.1126/science.1188021

Hahn, M. W. (2018). Molecular population genetics Sunderland, MA: Oxford University Press.

Hahn, M. W., & Hibbins, M. S. (2019). A three-sample test for introgression. *Molecular Biology and Evolution,* 36(12), 2878-2882. doi:10.1093/molbev/msz178

Hamlin, J. A. P., Hibbins, M. S., & Moyle, L. C. (2020). Assessing biological factors affecting postspeciation introgression. *Evolution Letters,* 4(2), 137-154. doi:10.1002/evl3.159

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity,* 105(1), 795-809. doi:10.1093/jhered/esu033

He, C., Liang, D., & Zhang, P. (2020). Asymmetric distribution of gene trees can arise under purifying selection if differences in population size exist. *Molecular Biology and Evolution,* 37(3), 881-892. doi:10.1093/molbev/msz232

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(18), 4606-4618. doi:10.1111/mec.12415

Heiser, C. B. (1949). Natural hybridization with particular reference to introgression. *Journal of Heredity,* 15(10), 795-809.

Heiser, C. B. (1973). Introgression reexamined. *Botanical Review,* 39(4), 347-366.

Hibbins, M. S., & Hahn, M. W. (2019). The timing and direction of introgression under the multispecies network coalescent. *Genetics,* 211(3), 1059-1073. doi:10.1534/genetics.118.301831

Hudson, R. R. (1983). Testing the constant-rate neutral allele model with protein sequence data. *Evolution,* 37(1), 203-217. doi:10.1111/j.1558-5646.1983.tb05528.x

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics,* 18(2), 337-338. doi:10.1093/bioinformatics/18.2.337

Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., *et al.* (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature,* 512(7513), 194-197. doi:10.1038/nature13408

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution,* 23(2), 254-267. doi:10.1093/molbev/msj030

Huson, D. H., Klöpper, T., Lockhart, P. J., & Steel, M. A. (2005). *Reconstruction of reticulate networks from gene trees*. Paper presented at the The 9th Annual International Conference Research in Computational Molecular Biology, Berlin.

Huson, D. H., & Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology,* 61(6), 1061-1067. doi:10.1093/sysbio/sys062

Jiao, X., & Yang, Z. (2020). Defining species when there is gene flow. *Systematic Biology*, 70(1), 108-119. doi:10.1093/sysbio/syaa052

Joly, S., McLenachan, P. A., & Lockhart, P. J. (2009). A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist,* 174(2), E54-70. doi:10.1086/600082

Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications,* 13(3), 235-248.

Knowles, L. L., Huang, H., Sukumaran, J., & Smith, S. A. (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(3), 376-384. doi:10.1002/ajb2.1064

Kong, S., & Kubatko, L. S. (2021). Comparative performance of popular methods for hybrid detection using genomic data. *Systematic Biology*, syaa092. doi:10.1093/sysbio/syaa092

Kronforst, M. R., Hansen, M. E., Crawford, N. G., Gallant, J. R., Zhang, W., *et al*. (2013). Hybridization reveals the evolving genomic architecture of speciation. *Cell Reports,* 5(3), 666-677. doi:10.1016/j.celrep.2013.09.042

Kubatko, L. S., & Chifman, J. (2019). An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *BMC Evolutionary Biology,* 19(1), 112. doi:10.1186/s12862-019-1439-7

Leppälä, K., Nielsen, S. V., & Mailund, T. (2017). admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics,* 33(11), 1738-1740. doi:10.1093/bioinformatics/btx048

Lohse, K., & Frantz, L. A. (2014). Neandertal admixture in Eurasia confirmed by maximum-likelihood analysis of three genomes. *Genetics,* 196(4), 1241-1251. doi:10.1534/genetics.114.162396

Long, C., & Kubatko, L. (2018). The effect of gene flow on coalescent-based species-tree inference. *Systematic Biology,* 67(5), 770-785. doi:10.1093/sysbio/syy020

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *Bioessays,* 38(2), 140-149. doi:10.1002/bies.201500149

Martin, S. H., & Amos, W. (2020). Signatures of introgression across the allele frequency spectrum. *Molecular Biology and Evolution*, 38(2), 716-726. doi:10.1093/molbev/msaa239

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution,* 32(1), 244-257. doi:10.1093/molbev/msu269

Mendes, F. K., & Hahn, M. W. (2018). Why concatenation fails near the anomaly zone. *Systematic Biology,* 67(1), 158-169. doi:10.1093/sysbio/syx063

Meng, C., & Kubatko, L. S. (2009). Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theoretical Population Biology,* 75(1), 35-45. doi:10.1016/j.tpb.2008.10.004

Nachman, M. W., & Payseur, B. A. (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(1587), 409-421. doi:10.1098/rstb.2011.0249

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science of the United States of America,* 76(10), 5269-5273. doi:10.1073/pnas.76.10.5269

Nielsen, R., & Wakeley, J. (2001). Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics,* 158(2), 885-896.

Noor, M. A., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* 103(6), 439-444. doi:10.1038/hdy.2009.151

Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., *et al*. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics, 48*(9), 1077-1082. doi:10.1038/ng.3617

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. doi:10.1186/s40657-017-0088-z

Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution,* 5(5), 568-583. doi:10.1093/oxfordjournals.molbev.a040517

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., *et al*. (2012). Ancient admixture in human history. *Genetics,* 192(3), 1065-1093. doi:10.1534/genetics.112.145037

Pease, J. B., Haak, D. C., Hahn, M. W., & Moyle, L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology,* 14(2), e1002379. doi:10.1371/journal.pbio.1002379

Pease, J. B., & Hahn, M. W. (2015). Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology,* 64(4), 651-662. doi:10.1093/sysbio/syv023

Peter, B. M. (2016). Admixture, population structure, and *F*-statistics. *Genetics,* 202(4), 1485-1501. doi:10.1534/genetics.115.183913

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics,* 8(11), e1002967. doi:10.1371/journal.pgen.1002967

Pollard, D. A., Iyer, V. N., Moses, A. M., & Eisen, M. B. (2006). Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. *PLoS Genetics,* 2(10), e173. doi:10.1371/journal.pgen.0020173

Przeworski, M., Charlesworth, B., & Wall, J. D. (1999). Genealogies and weak purifying selection. *Molecular Biology and Evolution,* 16(2), 246-252. doi:10.1093/oxfordjournals.molbev.a026106

Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sanchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics, 16*(6), 359-371. doi:10.1038/nrg3936

Rambaut, A., & Grassly, N. C. (1997). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics*, 13(3), 235-238. doi:10.1093/bioinformatics/13.3.235

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature,* 461(7263), 489-494. doi:10.1038/nature08365

Rieseberg L.H., Wendel. J. F. (1993). Introgression and its consequences in plants. In Hybrid Zones and the Evolutionary Process (pp. 70-109): Oxford University Press.

Rosenzweig, B. K., Pease, J. B., Besansky, N. J., & Hahn, M. W. (2016). Powerful methods for detecting introgressed regions from population genomic data. *Molecular Ecology,* 25(11), 2387-2397. doi:10.1111/mec.13610

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology, 14*(12), e2000234. doi:10.1371/journal.pbio.2000234

Schrider, D. R., Ayroles, J., Matute, D. R., & Kern, A. D. (2018). Supervised machine learning reveals introgressed loci in the genomes of *Drosophila simulans* and *D. sechellia*. *PLoS Genetics,* 14(4), e1007341. doi:10.1371/journal.pgen.1007341

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2014). How common is homoploid hybrid speciation? *Evolution, 68*(6), 1553-1560. doi:10.1111/evo.12399

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics,* 6(2), 461-464.

Sethuraman, A., Sousa, V., & Hey, J. (2019). Model-based assessments of differential introgression and linked natural selection during divergence and speciation. *BioRxiv*. doi:10.1101/786038

Slatkin, M., & Pollack, J. L. (2008). Subdivision in an ancestral species creates asymmetry in gene trees. *Molecular Biology and Evolution,* 25(10), 2241-2246. doi:10.1093/molbev/msn172

Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics,* 12(3), e1005896. doi:10.1371/journal.pgen.1005896

Solís-Lemus, C., Bastide, P., & Ané, C. (2017). PhyloNetworks: A package for phylogenetic networks. *Molecular Biology and Evolution,* 34(12), 3292-3298. doi:10.1093/molbev/msx235

Solís-Lemus, C., Yang, M., & Ané, C. (2016). Inconsistency of species tree methods under gene flow. *Systematic Biology,* 65(5), 843-851. doi:10.1093/sysbio/syw030

Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. B. (2018). Adaptive introgression: a plant perspective. *Biology Letters,* 14(3). doi:10.1098/rsbl.2017.0688

Suvorov, A., Kim, B. Y., Wang, J., Armstrong, E. E., Peede, D., *et al.* (2021). Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *BioRxiv*. doi:10.1101/2020.12.14.422758

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics,* 105(2), 437-460.

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution,* 3(2), 170-177. doi:10.1038/s41559-018-0777-y

Than, C., Ruths, D., & Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics,* 9, 322. doi:10.1186/1471-2105-9-322

Vanderpool, D., Minh, B. Q., Lanfear, R., Hughes, D., Murali, *et al.* (2020). Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biology*, 18(12), e3000954. doi:10.1317/journal.pbio.3000954

Vaughan, T. G. (2017). IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics,* 33(15), 2392-2394. doi:10.1093/bioinformatics/btx155

Wakeley, J., & Hey, J. (1997). Estimating ancestral population parameters. *Genetics,* 145(3), 847-855.

Wakeley, J., & Hey, J. (1998). Testing speciation models with DNA sequence data. In R. DeSalle & B. Schierwater (Eds.), *Molecular Approaches to Ecology and Evolution*: Birkhäuser, Basel.

Wang, J. (2003). Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics,* 164(2), 747-765.

Wen, D., & Nakhleh, L. (2018). Coestimating reticulate phylogenies and gene trees from multilocus sequence data. *Systematic Biology,* 67(3), 439-457. doi:10.1093/sysbio/syx085

Wen, D., Yu, Y., Zhu, J., & Nakhleh, L. (2018). Inferring phylogenetic networks using PhyloNet. *Systematic Biology, 67*(4), 735-740. doi:10.1093/sysbio/syy015

Williamson, S., & Orive, M. E. (2002). The genealogy of a sequence subject to purifying selection at multiple sites. *Molecular Biology and Evolution,* 19(8), 1376-1384. doi:10.1093/oxfordjournals.molbev.a004199

Wright, S. (1931). Evolution in Mendelian Populations. *Genetics,* 16(2), 97-159.

Wu, D. D., Ding, X. D., Wang, S., Wojcik, J. M., Zhang, *et al.* (2018a). Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nature Ecology and Evolution,* 2(7), 1139-1145. doi:10.1038/s41559-018-0562-y

Wu, M., Kostyun, J. L., Hahn, M. W., & Moyle, L. C. (2018b). Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Molecular Ecology*, 27(16), 3301-3316. doi:10.1111/mec.14780

Yu, Y., Degnan, J. H., & Nakhleh, L. (2012). The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genetics,* 8(4), e1002660. doi:10.1371/journal.pgen.1002660

Yu, Y., Dong, J., Liu, K. J., & Nakhleh, L. (2014). Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Science of the United States of America,* 111(46), 16448-16453. doi:10.1073/pnas.1407950111

Zhang, C., Ogilvie, H. A., Drummond, A. J., & Stadler, T. (2018). Bayesian inference of species networks from multilocus sequence data. *Molecular Biology and Evolution,* 35(2), 504-517. doi:10.1093/molbev/msx307

Zhang, W., Dasmahapatra, K. K., Mallet, J., Moreira, G. R., & Kronforst, M. R. (2016). Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biology,* 17, 25. doi:10.1186/s13059-016-0889-0

Zheng, Y., & Janke, A. (2018). Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics,* 19(1), 10. doi:10.1186/s12859-017-2002-4