Distinguishing Introgression and Incomplete Lineage Sorting in Evolutionary Genetics using D-statistic and FST

Rachael Chew
Minerva University
rachaelchewyuen@gmail.com

Abstract

Incomplete lineage sorting (ILS) and introgression are two major sources of gene tree discordance in evolutionary genomics, yet distinguishing between them remains a central challenge. This study systematically evaluates the performance of two key population genetic statistics—the D-statistic and fixation index (FST)—in differentiating these processes under a range of realistic evolutionary scenarios. Using coalescent simulations on a four-taxon phylogeny, we vary effective population sizes, divergence times, migration rates, timing, and directionality of gene flow to generate pure ILS, pure introgression, and mixed conditions. Our results reveal that D-statistics are highly sensitive to directional gene flow but exhibit saturation at high migration rates, while FST remains robust to weak introgression but declines with increased homogenization.

Crucially, under intermediate divergence and overlapping ILS-introgression regimes, neither metric alone provides reliable inference. However, their joint distribution captures the nonlinear and complementary behavior of each, substantially improving power to distinguish evolutionary histories. We propose a two-metric inference framework and identify parameter regimes where signal detection is most effective. These findings provide theoretical and practical insight for interpreting gene tree discordance in empirical systems, particularly those involving rapid radiations or ancient introgression events.

1 Introduction

In evolutionary genetics, researchers often face the challenge of explaining gene tree discordances – cases where the genealogical relationships among alleles do not mirror the species tree. Two biological processes can cause such discordance: introgression and incomplete lineage sorting (ILS) (Degnan & Rosenberg, 2009). Understanding the fundamental differences between these processes is critical for accurate evolutionary inference, yet distinguishing between them remains one of the major challenges in phylogenomics.

1.1 Incomplete Lineage Sorting

Incomplete lineage sorting (ILS) occurs when ancestral genetic polymorphisms persist through successive speciation events, resulting in gene trees that do not match the species tree (Figure 1).

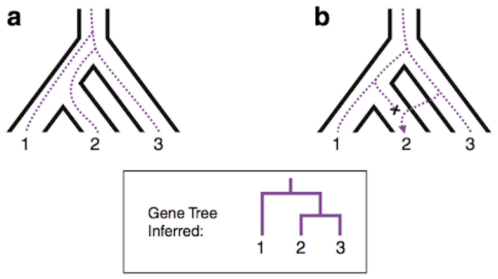

Figure 1: (a) ILS with no introgression showing how gene trees can differ from species trees. Despite species 1 and 2 being true sister species, the inferred gene tree shows a different relationship. (b) Introgression event (indicated by X) from species 3 to species 2, producing a similar discordant gene tree pattern. From Edelman et al. (2019).

This phenomenon emerges directly from population genetic processes: when an ancestral population contains multiple allelic variants at a locus and subsequently splits into daughter species, these variants may be sorted randomly into descendant lineages rather than strictly following the species branching pattern.

ILS is particularly common when speciation events occur in rapid succession or when ancestral population sizes are large, as these conditions increase the likelihood that lineages will not co-

alesce within their respective species boundaries (Degnan & Rosenberg, 2009). The probability of ILS is mathematically related to the ratio of population size (Ne) to the time between speciation events – larger populations and shorter intervals between speciations both increase ILS prevalence.

For example, in a standard three-species model where species A and B are most closely related, an ancestral polymorphism may persist through their common ancestor's speciation, leading to a scenario where some genes in species B share a more recent common ancestor with genes in species C than with those in species A. This creates gene tree discordance purely through stochastic lineage sorting, without any post-divergence gene flow.

The evolutionary implications of ILS are significant: it reflects historical demographic processes and speciation tempos rather than hybridization or adaptation. ILS is expected to affect the genome broadly, with discordant patterns distributed randomly across chromosomes rather than concentrated in specific regions (barring the effects of selection or recombination rate variation). While ILS complicates phylogenetic inference, it also contains valuable information about ancestral population sizes and the timing of speciation events (Pamilo & Nei, 1988).

1.2 Introgression

Introgression (or introgressive hybridization) involves the transfer of genetic material between divergent species through hybridization followed by backcrossing (Figure 2C). Unlike ILS, which reflects the persistence of ancestral variation, introgression introduces new genetic material across species boundaries after divergence has occurred (Mallet, 2005). This process results in the permanent incorporation of foreign alleles into a species' gene pool, creating regions of the genome where the evolutionary history differs from the species history due to post-divergence gene flow rather than ancestral polymorphism.

Once considered rare, introgression is now recognized as widespread and consequential across the tree of life. High-throughput sequencing has revealed evidence of gene flow between divergent species in diverse taxa, from plants (e.g., sunflowers and oaks) to animals (e.g., butterflies, finches, and hominins). Hybridization followed by backcrossing can introduce genetic material from one lineage into another, leaving patches of the genome with a history of interspecific exchange (Mallet et al., 2016).

Introgression can have profound evolutionary consequences. It can transfer adaptive traits between species, increasing genetic variation and potentially accelerating adaptation to new environments (Arnold & Martin, 2009). It can blur species boundaries in hybrid zones, potentially leading to speciation reversal or reinforcement (Servedio & Noor, 2003). In some cases, introgression can even give rise to new hybrid lineages with novel combinations of traits, thereby playing an important role in adaptive evolution (Hedrick, 2013). This process challenges the traditional view of species as strictly separated lineages and suggests that the tree of life is often more web-like than strictly bifurcating.

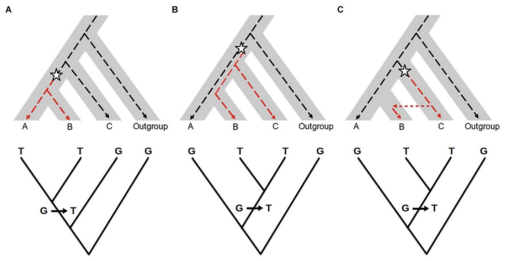

Figure 2: Conceptual illustration of gene tree discordance. (A) Gene tree is congruent with species tree. (B) ILS scenario - ancestral polymorphism persists through speciation, leading to discordant gene trees where B and C share an allele by chance. (C) Introgression scenario - gene flow from C to B transfers an allele, creating a similar discordant pattern. Below each scenario are the resulting nucleotides at this locus. From Burgarella et al. (2019).

1.3 Distinguishing Between ILS and Introgression

As illustrated in Figure 2, both processes can produce identical patterns of gene tree discordance despite fundamentally different mechanisms. In both scenarios, species B and C may share derived alleles (such as the T allele shown in the figure) despite not being sister species in the true species tree. However, with ILS, this pattern results from the random sorting of ancestral polymorphisms that existed before speciation, while with introgression, it results from post-divergence gene flow between species.

Figure 3 further illustrates how both ILS and introgression can produce discordant gene trees. In a three-species tree where A and B are true sister species, both processes can lead to alternative topologies where either B and C or A and C appear as sister species. With ILS, this results from incomplete coalescence within the ancestral population, while with introgression, it results from genetic exchange between non-sister species.

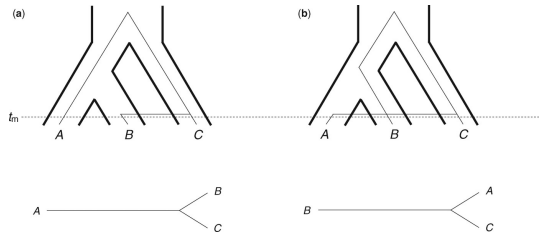

Figure 3: Two discordance arrangements (BC and AC) are expected in equal frequencies when there is no gene flow. With introgression, one pattern becomes more common and results in a decreased genetic distance between some species. From: Hahn & Hibbins (2019) Molecular Biology and Evolution

This similarity in observed patterns despite different underlying processes presents a significant challenge for evolutionary biologists. Misidentifying one process as the other can lead to incorrect evolutionary inferences – for instance, inferring a phantom hybridization event when discordant gene trees were actually due to ancestral polymorphism, or overlooking real introgressive events critical to adaptive evolution.

The key differences between ILS and introgression lie in:

1. **Timing**: ILS involves ancestral polymorphisms that predate speciation, while introgression involves genetic exchange after species divergence.

2. **Genomic distribution**: ILS tends to affect the genome broadly with random distribution, while introgression often creates blocks of shared variation clustered in specific regions.

3. **Allele frequency patterns**: Introgressed alleles, especially from recent gene flow, often show skewed frequency distributions, whereas ILS-derived shared alleles may have more balanced frequencies.

4. **Branch length patterns**: Gene trees affected by introgression often show distinctive branch length patterns that differ from those expected under pure ILS.

5. **Population differentiation effects**: Introgression tends to reduce genetic differentiation between the species involved, while ILS alone does not systematically reduce differentiation.

Accurately distinguishing introgression from ILS is essential for understanding speciation and adaptation. Introgression provides evidence of historical hybridization and potential adaptive exchange, while ILS informs us about ancestral population dynamics and speciation rates. The genomic architecture of these processes—which regions are affected and how—can reveal the mechanisms underlying species formation and maintenance.

2 Focus of the Present Study

This study addresses the challenge of distinguishing introgression from ILS by focusing on two genomic signals: the D-statistic (also known as the ABBA–BABA test) and the fixation index (FST).

The D-statistic detects asymmetrical allele sharing among taxa, beyond what neutral ILS would produce, providing a statistical test for historical gene flow (Green et al., 2010; Durand et al., 2011). As illustrated in Figure 4, this approach can effectively identify introgression in real biological systems, such as gene flow between Cackling Goose (*Branta hutchinsii*) and Canada Goose (*B. canadensis*) with a significant positive D-statistic (D = 0.087, Z = 4.32).

FST, also known as Wright's fixation index, measures genetic differentiation between populations; introgression is expected to reduce interspecific differentiation by introducing shared alleles, whereas purely ILS-driven discordance may not systematically lower between-species FST (Wright, 1949). When used together, these complementary metrics can provide more robust evidence for distinguishing between these evolutionary processes.

By analyzing how these statistics behave under various simulated evolutionary scenarios – varying population sizes, divergence times, and gene flow patterns – we aim to establish diagnostic signatures that reliably differentiate introgression from ILS. A brief mention is warranted that branch length

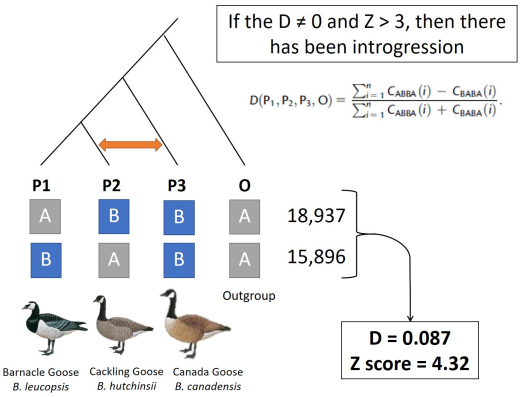

Figure 4: The D-statistic test applied to goose species. The positive D-statistic indicates an excess of ABBA-patterns (18,937) compared to BABA patterns (15,896) in the genomes of these geese, suggesting introgression between Cackling Goose (Branta hutchinsii) and Canada Goose (B. canadensis). Based on Ottenburghs et al. (2017) BMC Evolutionary Biology.

patterns in gene trees have also been explored as a signal (e.g., through coalescent-based methods and tools like QuIBL). However, in this work the emphasis shifts to allele frequency and population differentiation measures as the primary focus for teasing apart the two processes.

We hypothesize that a combined examination of D-statistics and FST across many loci will more reliably discriminate introgression from ILS "noise" than topology alone, given that each metric captures different aspects of the underlying history. This approach is particularly valuable for researchers studying rapid radiations or recently diverged species where both ILS and potential introgression are expected to occur.

In the following sections, we provide a detailed literature review on current methodological approaches to distinguishing introgression from ILS – including their theoretical basis, applications, and limitations – to underscore the importance of this research question in evolutionary genetics and how addressing it will advance the field. We then describe our simulation framework and analyses, present key results from our parameter explorations, and discuss the implications of our findings for empirical research in evolutionary genetics.

3 Literature Review

3.1 Evolutionary Genetics and Phylogenetics – Gene Trees vs. Species Trees

Evolutionary genetics and phylogenetics are grounded in analyzing genetic variation to infer evolutionary relationships. A fundamental concept is the distinction between a gene tree (the genealogy of alleles at a locus) and the species tree (the actual branching history of species). Gene trees do not always align with the species tree. This phenomenon, known as gene tree discordance (Figure 1), has significant implications: it complicates the reconstruction of true species histories and can obscure our understanding of speciation and gene flow. Early theoretical work (e.g., coalescent theory developed by Sewall Wright and later formalized by Kingman) explained how gene lineages coalesce back to common ancestors and laid the groundwork for understanding ILS.

The coalescent process describes how genetic lineages merge as we trace them backward in time, eventually reaching their most recent common ancestor. This framework provides a mathematical basis for understanding how gene trees can vary due to stochastic processes, even within a fixed species tree (Kingman, 1982; Hudson, 1983). The probability that lineages from different species coalesce in their common ancestral population depends on both the population size and the time between speciation events. When effective population sizes are large or when speciation events occur in rapid succession, there is a higher probability that ancestral polymorphisms will persist through speciation events, leading to ILS.

The disconnect between gene trees and species trees was formally described by Maddison (1997) as the "gene tree/species tree problem." This recognition has led to the development of numerous methods that explicitly account for ILS when inferring species trees, such as the multispecies coalescent model (Rannala & Yang, 2003; Liu et al., 2009). However, these methods typically assume that discordance arises solely from ILS and do not account for introgression, which can produce similar patterns of gene tree heterogeneity.

3.2 D-statistic and FST as Diagnostic Tools

3.2.1 D-statistic

The D-statistic, also known as the ABBA-BABA test, provides a powerful method for detecting introgression by comparing patterns of shared de-

rived alleles among taxa. Developed by Green et al. (2010) and formalized by Durand et al. (2011), the D-statistic was initially used to detect admixture between Neanderthals and modern humans but has since become a standard tool in evolutionary genomics.

The D-statistic is calculated using a four-taxon phylogeny, typically represented as $(((P_1, P_2), P_3), O)$, where P_1 and P_2 are sister taxa, P_3 is an outgroup to $P_1 + P_2$, and O is an outgroup to all three ingroup taxa. For each biallelic site in the genome, the ancestral state (A) and derived state (B) are determined by comparison to the outgroup O. The D-statistic focuses on two specific site patterns:

- **ABBA sites**: where P_1 has the ancestral allele (A), while P_2 and P_3 share the derived allele (B), and O has the ancestral allele (A).
- **BABA sites**: where P_1 and P_3 share the derived allele (B), while P_2 has the ancestral allele (A), and O has the ancestral allele (A).

The D-statistic is then calculated as:

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

The mathematical foundation of the D-statistic lies in coalescent theory. Under a null model of no introgression and strict bifurcating speciation, ABBA and BABA patterns should result solely from ILS, and their expected frequencies should be equal (Patterson et al., 2012). This equality occurs because, under pure ILS with no directional gene flow, the probability of a coalescence event between lineages from P2 and P3 should equal the probability of a coalescence event between lineages from P1 and P3 before these lineages coalesce with the common ancestor of P1 and P2.

When D = 0, the observed ABBA and BABA patterns occur with equal frequency, consistent with pure ILS. A significant deviation from zero indicates an excess of one pattern, suggesting directional gene flow. Specifically:

- $D > 0$ (excess of ABBA) suggests gene flow between P_2 and P_3
- $D < 0$ (excess of BABA) suggests gene flow between P_1 and P_3

Statistical significance is typically assessed through a block jackknife procedure to obtain standard errors and Z-scores (Green et al., 2010). A common threshold is $|Z| > 3$, corresponding to a p-value of approximately 0.001, although more conservative thresholds (e.g., $|Z| > 4$) are sometimes used for genome-wide analyses (Martin et al., 2015).

The D-statistic has been implemented in various software packages, including ADMIXTOOLS (Patterson et al., 2012), Dsuite (Malinsky et al., 2021), and custom scripts in languages like Python and R. In genomic studies, the D-statistic is often calculated across sliding windows to identify regions with particularly strong signals of introgression.

The mathematical properties of the D-statistic make it particularly suitable for distinguishing introgression from ILS for several reasons:

1. **Direct test against an ILS null model**: By comparing the frequencies of ABBA and BABA patterns, the D-statistic provides an explicit test against the expectations under pure ILS.
2. **Robustness to mutation rate variation**: Because the D-statistic compares patterns within the same genomic regions, it is relatively robust to variation in mutation rates across the genome.
3. **Ability to detect ancient admixture**: The D-statistic can detect signals of introgression even when the admixture occurred in the distant past, as demonstrated by its successful application to ancient hominins (Green et al., 2010; Reich et al., 2010).
4. **Scalability to whole genomes**: The D-statistic can be efficiently calculated from genome-wide data, making it suitable for large-scale genomic studies.

Extension of the basic D-statistic include the f-ratio:

$$f_4 = D \times \frac{\text{ABBA} + \text{BABA}}{2 \times \text{Total}}$$

which estimates the proportion of the genome derived from introgression (Patterson et al., 2012), and DFOIL, which uses five-taxon comparisons to infer the direction of gene flow (Pease & Hahn, 2015). These extensions address some limitations of the standard D-statistic, providing more detailed insights into introgression patterns.

A notable innovation is the three-sample test (D3) proposed by Hahn and Hibbins (2019), which

eliminates the need for an outgroup to determine ancestral states:

$$D_3 = \frac{d_{B-C} - d_{A-C}}{d_{B-C} + d_{A-C} - 2d_{A-B}}$$

where d_{X-Y} represents the genetic distance between species X and Y. This approach is valuable for systems where appropriate outgroups are unavailable, as it relies solely on pairwise distances between the three ingroup taxa.

Despite these strengths, the D-statistic has important limitations. It does not directly indicate the timing or direction of gene flow without additional information or extensions. The standard D-statistic also assumes no ancestral population structure beyond the specified tree, and violation of this assumption can produce false positives. Eriksson and Manica (2012) demonstrated that deep structure in an ancestral population could cause non-random sharing of alleles between non-sister taxa even without hybridization.

To address these limitations, researchers have developed methods to distinguish ancestral structure from introgression by examining the allele frequency spectrum of shared variants. Yang et al. (2012) showed that the frequency distribution of Neanderthal alleles in modern humans (skewed toward rare variants) better supported recent introgression than ancient population structure. More recently, the D*FS statistic (Martin et al., 2015) partitions the D-statistic by allele frequency class to provide insights into introgression timing.

3.2.2 FST

While the D-statistic provides a direct test for introgression, Wright's fixation index (FST) offers a complementary perspective by measuring the degree of genetic differentiation between populations. Originally developed by Sewall Wright (1943, 1949), FST quantifies the proportion of genetic variance in a metapopulation that is explained by population structure.

The mathematical definition of F_{ST} is:

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Where:

- H_T is the expected heterozygosity in the total (pooled) population
- H_S is the average expected heterozygosity within subpopulations

For biallelic loci, this can be expressed as:

$$F_{ST} = \frac{\sigma_p^2}{\bar{p}(1 - \bar{p})}$$

Where:

- σ_p^2 is the variance of allele frequencies across populations
- $\bar{p}$ is the average allele frequency across all populations

Several estimators have been developed for calculating F_{ST} from genetic data, including the Weir and Cockerham (1984) estimator, which accounts for sampling variance and is widely used in population genetic studies:

$$\theta = \frac{a}{a + b + c}$$

Where:

- a represents the variance component due to differences between populations
- b represents the variance component due to differences between individuals within populations
- c represents the variance component due to differences between gametes within individuals

FST values range from 0 to 1, with values close to 0 indicating little genetic differentiation between populations and values close to 1 indicating substantial differentiation. In the context of speciation, FST provides a measure of the progress towards complete reproductive isolation (Charlesworth, 1998).

The relationship between FST and introgression stems from population genetic theory: gene flow between populations tends to homogenize allele frequencies, reducing genetic differentiation. Consequently, ongoing or recent introgression is expected to decrease FST relative to what would be observed under complete isolation. In contrast, ILS alone does not systematically reduce FST between populations; it causes random gene tree discordances but does not create directional gene flow that would homogenize allele frequencies (Cruickshank & Hahn, 2014).

FST has been implemented in numerous population genetic software packages, including Arlequin

(Excoffier & Lischer, 2010), GENEPOP (Rousset, 2008), and vcftools (Danecek et al., 2011). In genomic studies, FST is often calculated in sliding windows across the genome to identify regions of unusually high or low differentiation, potentially indicating selection or introgression, respectively.

The mathematical properties of FST make it valuable for distinguishing introgression from ILS in several ways:

1. **Sensitivity to recent gene flow**: FST is particularly sensitive to recent or ongoing gene flow, which rapidly reduces genetic differentiation between populations.

2. **Genomic heterogeneity**: FST can identify specific genomic regions with reduced differentiation, potentially indicating localized introgression.

3. **Quantitative measure**: Unlike the D-statistic, which primarily indicates presence/absence of introgression, FST provides a quantitative measure of differentiation that can be compared across populations and genomic regions.

4. **Integration with other statistics**: FST can be effectively combined with other statistics like d_{XY} (absolute divergence) to distinguish between reduced differentiation due to gene flow versus shared ancestral variation (Martin et al., 2015).

5. **Direct connection to gene flow in population genetic models**: Under island models of migration, F_{ST} is inversely related to the number of migrants per generation (N_m), with

$$F_{ST} \approx \frac{1}{4N_m + 1}$$

for neutral loci at equilibrium (Wright, 1931).

However, interpretation of FST patterns comes with caveats. FST is a relative measure of differentiation and can be confounded by factors like mutation rate and natural selection. For instance, a genomic "island" of high differentiation might not necessarily indicate a barrier to gene flow; it could result from reduced genetic diversity in one lineage due to selective sweeps or background selection, inflating the between-group FST (Charlesworth, 1998; Cruickshank & Hahn, 2014).

Similarly, regions of low FST might arise from balancing selection maintaining similar alleles in both populations, rather than admixture. Noor and Bennett (2009) and Nachman and Payseur (2012) highlighted that selection on linked sites can heterogeneously shape FST across the genome, complicating the inference of introgression from differentiation alone. To mitigate this, researchers often combine evidence: if a region shows low FST and a significant D-statistic signal, it strengthens the case for introgression rather than ILS or selection alone.

3.2.3 Complementarity of D-statistic and FST for Distinguishing Introgression from ILS

The D-statistic and FST provide complementary information for distinguishing introgression from ILS, with different strengths and sensitivities. The D-statistic offers a direct statistical test for introgression based on asymmetry in shared derived alleles, while FST quantifies the overall genetic differentiation between populations, which is expected to be reduced by gene flow but not by ILS alone.

These metrics complement each other in several key ways:

1. **Different aspects of the data**: The D-statistic focuses specifically on patterns of shared derived alleles, while FST examines overall allele frequency differentiation. This dual perspective provides a more comprehensive view of potential introgression signals.

2. **Robustness to different confounding factors**: The D-statistic is robust to mutation rate variation but can be affected by ancestral population structure. FST is less affected by ancient population structure but can be influenced by selection and demographic history. Using both metrics helps mitigate these confounding factors.

3. **Time sensitivity**: The D-statistic can detect signals of ancient introgression, while FST is more sensitive to recent gene flow. This difference in temporal sensitivity allows for a more nuanced understanding of the timing of potential introgression events.

4. **Spatial resolution**: FST can be calculated in smaller genomic windows than the D-statistic

(which requires sufficient polymorphic sites for statistical power), potentially providing higher spatial resolution of introgression patterns.

5. **Quantitative versus qualitative information**: The D-statistic primarily indicates the presence and direction of introgression, while FST provides a quantitative measure of differentiation that can be compared across the genome and between different population pairs.

Martin et al. (2015) demonstrated the value of combining these approaches in their study of Heliconius butterflies, where the integration of D-statistics with FST and other differentiation measures allowed for more robust identification of introgressed regions and helped distinguish adaptive introgression from background noise.

For these reasons, our study employs both the D-statistic and FST to develop a more comprehensive framework for distinguishing introgression from ILS. By examining how these metrics respond to varying demographic parameters, we aim to identify diagnostic signatures that reliably differentiate these processes across a range of evolutionary scenarios.

3.3 Other Methods for Distinguishing Introgression from ILS

While our study focuses on the D-statistic and FST, several other methods have been developed to distinguish introgression from ILS. Phylogenetic networks (e.g., NeighborNet, PhyloNet) offer a visualization of non-tree-like evolution, explicitly showing reticulation events as network connections (Huson & Bryant, 2006). These approaches can effectively represent the complex evolutionary relationships that arise from both ILS and introgression, but they may not always clearly distinguish between these processes.

Coalescent-based frameworks, like BEAST or MSC (multi-species coalescent) methods, incorporate ILS by modeling gene tree distributions under a species tree with ILS and can test for introgression by comparing support for network vs. tree models (Edwards et al., 2016). These approaches are statistically powerful but computationally intensive, and they may struggle as the number of taxa or loci grows (Heled & Drummond, 2010).

Of particular note, recent methods explicitly leverage branch length information in gene trees to infer introgression. For example, QuIBL (Quantifying Introgression via Branch Lengths) analyzes the distribution of internal branch lengths in gene genealogies (Edelman et al., 2019). The rationale is that under pure ILS, the internal branch lengths (reflecting coalescent times) fit an exponential distribution, whereas introgression adds a secondary peak of longer branches corresponding to more recent coalescent events introduced by gene flow. This approach provides another dimension of evidence but requires robust gene tree inference and can be sensitive to model assumptions.

Machine learning approaches represent an emerging frontier in distinguishing introgression from ILS. Methods like DILS (Detecting Introgression in a Likelihood framework with Simulation) use supervised machine learning trained on simulated datasets to classify empirical data based on multiple summary statistics (Aeschbacher et al., 2017). Similarly, Schrider et al. (2018) developed a machine learning classifier that integrates multiple population genetic statistics to identify introgressed regions more accurately than any single metric alone. These approaches are promising but require extensive training data and may be sensitive to the specific evolutionary scenarios used for training.

Allele frequency spectrum-based methods, such as those implemented in ai (Gutenkunst et al., 2009) and fastsimcoal2 (Excoffier et al., 2013), can fit demographic models that include both ILS and introgression to observed data. These methods can estimate parameters like migration rates and divergence times, providing insights into the relative contributions of these processes to observed genetic patterns. However, they typically require simplifying assumptions about population history and may struggle with complex demographic scenarios.

3.4 Empirical Applications in Diverse Taxa

The application of these methods has revealed widespread introgression across the tree of life, challenging traditional views of species boundaries. In birds, a recent study by Singhal et al. (2021) used D-statistics to demonstrate that approximately 38% of suboscine bird trios showed significant evidence of introgression. This finding is likely an underestimate, as the D-statistic cannot detect gene flow between sister species and may miss introgression between more distantly related

species not included in the analysis.

In mammals, the discovery of introgression between Neanderthals and modern humans (Green et al., 2010) revolutionized our understanding of human evolution. Subsequent studies have identified introgression from Denisovans into modern human populations, particularly in Oceania (Reich et al., 2010), and complex patterns of gene flow among archaic and modern humans. These findings have transformed our view of human evolution from a simple tree-like model to a more reticulated network of genetic exchange.

In butterflies, Martin et al. (2013) used both D-statistics and FST to identify genomic regions involved in adaptive introgression of wing pattern mimicry genes in Heliconius. Their work demonstrated how introgression can facilitate adaptive evolution by transferring beneficial alleles between species. Similar patterns have been observed in Anopheles mosquitoes, where introgression has contributed to the spread of insecticide resistance alleles (Fontaine et al., 2015).

In plants, introgression appears to be even more widespread. Hybridization and introgression have played major roles in the evolution of many plant groups, including oaks (Quercus), poplars (Populus), and sunflowers (Helianthus). For example, Whitney et al. (2010) demonstrated that introgression has facilitated adaptation to extreme habitats in sunflowers, transferring drought tolerance alleles between species.

These empirical examples highlight the importance of accurately distinguishing introgression from ILS. In many cases, the adaptive significance of introgressed alleles would have been overlooked if gene tree discordance had been attributed solely to ILS. Conversely, failing to account for ILS could lead to overestimation of introgression rates or misidentification of introgressed regions.

3.5 Current Challenges and Future Directions

Despite methodological advances, several challenges remain in distinguishing introgression from ILS. One key issue is the confounding effect of ancestral population structure, which can produce patterns similar to introgression. Additionally, the timing of introgression events can be difficult to determine, especially for ancient events where the signal may be weakened by subsequent evolution. The genomic extent of introgression is also challenging to estimate accurately, as it may vary across the genome due to selection and recombination.

Future directions in this field include the development of integrated approaches that combine multiple lines of evidence, such as D-statistics, FST, and branch length patterns, to provide more robust inferences. Machine learning approaches that train on simulated data with known introgression and ILS scenarios show promise for classifying empirical patterns (e.g., Schrider et al., 2018). Additionally, as genomic data become more readily available, comparative analyses across multiple taxa can provide insights into the general patterns and processes of introgression.

Another promising direction is the integration of ancient DNA into analyses of introgression and ILS. Ancient samples can provide direct evidence of historical population structure and admixture events, helping to distinguish between alternative scenarios that might be indistinguishable from modern data alone (Slatkin & Racimo, 2016).

Finally, the development of methods that can handle complex demographic scenarios, including multiple episodes of gene flow, changing population sizes, and selection, represents an important frontier in this field. Such methods will be crucial for understanding the evolutionary history of groups with complex demographic histories, such as humans, great apes, and many domesticated species.

3.6 Summary of Progress and Gaps

Advances in genomic analysis – from D-statistics to phylogenetic networks to differentiation scans – have greatly improved our ability to detect introgression in the presence of ILS. Researchers have successfully uncovered many instances of reticulate evolution that traditional trees alone would miss. However, each method has limitations: D-statistics require specific taxon sampling and cannot pinpoint direction or location of gene flow without extensions; FST scans can flag unusual patterns but are subject to interpretative pitfalls; network and coalescent methods are computationally demanding and may not scale well. There is also the issue that introgression and ILS often co-occur – rapid radiations can involve bursts of divergence (promoting ILS) alongside intermittent hybridization. Disentangling these overlapping signals remains difficult.

Current research is moving toward combining

multiple lines of evidence and developing statistical frameworks that jointly model introgression and ILS. For example, integrated approaches use machine learning to classify genealogical patterns by training on simulations of introgression vs. ILS (e.g., Schrider et al., 2018), and frequency-spectrum methods like D*FS adds nuance to D-tests. These efforts underscore the importance of our research question: how can we more reliably distinguish introgression from incomplete lineage sorting in evolutionary genomics? By systematically evaluating how D-statistic signals and FST differentiation respond to varying demographic scenarios (population size, divergence time, gene flow rate), our study contributes to this goal. Distinguishing introgression from ILS with greater confidence will refine species tree reconstructions and improve our understanding of speciation with gene flow.

4 Methods

4.1 Study Design

This study employs computational simulations to systematically investigate the behavior of two population genetic statistics—the D-statistic and FST—under various evolutionary scenarios involving incomplete lineage sorting (ILS) and introgression. While empirical studies provide valuable insights into real evolutionary processes, controlled simulations offer several distinct advantages for addressing our specific research questions: they allow precise manipulation of evolutionary parameters, enable the generation of data under known scenarios (establishing ground truth), and facilitate the exploration of parameter spaces that might be rare or difficult to detect in natural systems.

Our research addresses three primary questions:

1. What are the characteristic distributions of D-statistic and FST values under pure ILS versus scenarios with introgression?

2. How do key parameters—effective population size, divergence time, migration rate, and timing—affect these distributions?

3. Can we establish diagnostic signatures that reliably distinguish ILS from introgression across different evolutionary conditions?

To answer these questions, we implemented a three-phase simulation approach:

1. **Baseline establishment**: We first simulated pure ILS scenarios (without any migration) and pure introgression scenarios (with defined migration between non-sister taxa) to establish baseline distributions for our statistics.

2. **Parameter exploration**: We then systematically varied evolutionary parameters to determine their effects on D-statistic and FST distributions.

3. **Integration analysis**: Finally, we examined more complex scenarios with varying combinations of ILS and introgression to understand their interaction and develop a robust diagnostic framework.

4.2 Simulation Framework

All simulations were performed using the `msprime` library (Kelleher et al., 2016), which implements a coalescent simulator that efficiently generates tree sequence data. The tree sequence format allows for the calculation of various population genetic statistics while maintaining the full ancestral history of simulated samples.

4.2.1 Simulation Model

We simulated a three-species model consisting of populations A, B, and C, with an outgroup O, following the topology $(((A, B), C), O)$. In our notation, A corresponds to P_1, B to P_2, and C to P_3 in the standard D-statistic calculation. This topology represents a scenario in which A and B are sister species that diverged more recently from each other than either did from C.

For each simulation, we sampled 20 haploid individuals from each of the four populations (A, B, C, and outgroup O). We simulated a 1 megabase (Mb) genomic region under a uniform mutation rate of 1×10^{-8} mutations per site per generation and a uniform recombination rate of 1×10^{-8} recombinations per base pair per generation. These parameters were chosen to reflect realistic levels of genetic diversity while maintaining computational efficiency.

4.2.2 Parameter Space

We explored a parameter space defined by five key evolutionary variables:

4.3 Simulation Scenarios

Our simulation study explores three main scenarios, each designed to isolate and understand specific aspects of gene tree discordance:

Table 1: **Primary Simulation Parameters**

Parameter	Values	Significance in Evolutionary Context
Effective population size (N_e)	Small (10,000), Large (100,000)	Controls the strength of genetic drift and the rate of lineage coalescence. Larger N_e increases the probability of maintaining ancestral polymorphisms across speciation events, increasing ILS. Smaller N_e leads to faster lineage sorting.
Divergence times	**Recent**: AB = 25,000, ABC = 50,000 **Intermediate**: AB = 100,000, ABC = 200,000 **Ancient**: AB = 400,000, ABC = 800,000	Determines the time between speciation events. Shorter intervals increase ILS, while ancient divergence allows lineages more time to coalesce, reducing ILS. Timing also impacts introgression detectability.
Migration rate (m)	None (0.0), Low (10^{-6}), Moderate (10^{-4}), High (10^{-2}), Very high (0.1), Super high (0.3)	Governs the intensity of gene flow. Higher m increases introgression signals by elevating foreign allele proportions. The product $N_e m$ reflects the number of migrants per generation, key in population genetics.
Migration timing	Recent (10,000), Intermediate (50,000), Ancient (200,000)	Influences the persistence of introgression signals. Recent events yield stronger signals. Ancient events allow recombination and selection to erode signal strength.
Migration mode	Episodic (pulse event), Continuous (ongoing gene flow)	Models different hybridization scenarios. Episodic introgression reflects discrete events; continuous introgression reflects long-term gene flow across semi-permeable boundaries.

1. **Pure ILS Scenario**: Simulations with no migration between populations, varying Ne and divergence times to establish baseline D-statistic and FST distributions when gene tree discordance arises solely from incomplete lineage sorting. This serves as our null model for subsequent comparisons.

2. **Pure Introgression Scenario**: Simulations with controlled migration between non-sister taxa (B and C), systematically varying migration rates, timing, and modes to characterize the distinctive signatures of introgression in isolation from other confounding factors.

3. **Mixed Scenarios**: Simulations combining varying levels of both ILS and introgression to explore their interaction and identify parameter regimes where the signals can be reliably distinguished. This expanded parameter grid helps establish boundaries of statistical power for detection methods.

These scenarios progressively build in complexity, allowing us to first establish clear baseline expectations for each process in isolation before examining their interactions. In Section 5, we will provide detailed analyses of the results from each scenario, demonstrating how D-statistic and FST respond across the parameter space.

4.4 Statistical Analysis

4.4.1 D-statistic

The D-statistic (also known as the ABBA-BABA test) provides a formal test for gene flow by comparing the frequencies of discordant site patterns across the genomes of four taxa. Using the standard notation $(((P_1, P_2), P_3), O)$, where P_1 and P_2 are sister taxa, P_3 is an outgroup to P_1 + P_2, and O is an outgroup to all three ingroup taxa, the D-statistic compares two specific site patterns:

- **ABBA sites**: Where P_1 has the ancestral allele (A), while P_2 and P_3 share the derived allele (B)
- **BABA sites**: Where P_1 and P_3 share the derived allele (B), while P_2 has the ancestral allele (A)

The D-statistic is calculated as:

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

Under a null model of no introgression (only ILS), the expected value of D is zero, as ABBA and BABA patterns should occur with equal frequency. Significant deviations from zero suggest directional gene flow, with positive values indicating gene flow between P2 and P3, and negative values indicating gene flow between P1 and P3.

In our implementation, we calculate the D-statistic directly from the simulated tree sequences, using allele frequencies in each population to identify and count ABBA and BABA patterns.

4.4.2 FST

Wright's fixation index (FST) measures the proportion of genetic variance explained by population structure. In our study, we calculate pairwise FST values between all population pairs (A-B, B-C) using the standard Weir and Cockerham (1984) estimator.

FST ranges from 0 to 1, with values close to 0 indicating little genetic differentiation between populations and values close to 1 indicating substantial differentiation. Gene flow between populations is expected to reduce FST by homogenizing allele frequencies, making FST between introgressing populations lower than expected under a model of complete isolation.

We implemented the FST calculation directly using the tree sequence API, which provides an efficient way to compute this statistic across the simulated genomic regions.

4.4.3 Sliding Window Analysis

To examine local variation in introgression signals across the genome, we implemented a sliding window approach for both the D-statistic and FST. Windows of 50 kb were analyzed across the simulated 1 Mb region, providing insight into the heterogeneity of signals that might be expected in empirical data due to the stochastic nature of recombination and introgression.

Sliding window analyses help identify regions with particularly strong signals of introgression and provide a more detailed view of how these signals vary across the genome—an important consideration when analyzing empirical data where gene flow may affect only portions of the genome.

4.4.4 Statistical Power Analysis

To quantify the ability of D-statistic and FST to detect introgression across different parameter settings, we implemented a power analysis approach. For each parameter combination, we calculated:

1. The proportion of simulations where the D-statistic exceeded a critical threshold ($|D| > 0.2$), representing the power to detect introgression using this statistic.

2. The proportion of simulations where FST between introgressing populations FST_BC was below a critical threshold (defined as the 5th percentile of the null distribution from simulations with no migration), representing the power to detect introgression as abnormally low differentiation.

These power calculations were performed across the range of migration rates to generate power curves, allowing us to determine the minimum migration rate required for reliable detection of introgression using each statistic. Additionally, we examined how power varies with divergence time, effective population size, and other parameters to identify conditions under which each statistic performs optimally.

4.4.5 Joint Distribution Analysis

To explore the complementary nature of D-statistic and FST in distinguishing introgression from ILS, we analyzed their joint distribution across different simulation scenarios. Using kernel density estimation (KDE), we visualized the bivariate distribution of these statistics under varying migration rates, allowing us to identify regions in the D vs. FST parameter space that are characteristic of pure ILS versus different intensities of introgression.

This joint analysis provides a more powerful approach to distinguishing evolutionary processes, as it leverages the complementary sensitivities of these statistics to different aspects of introgression and ILS.

4.4.6 Data Visualization

For each simulation scenario, we compiled the calculated D-statistics and FST values into pandas DataFrames, storing all relevant parameters for subsequent analysis. We used kernel density estimation (KDE) to visualize the distribution of D-statistic and FST values across different parameter combinations, providing a visual representation of how these statistics respond to varying evolutionary conditions.

For power analysis, we created plots showing the detection rate (power) as a function of migration rate for both D-statistic and FST, allowing direct comparison of their relative sensitivity across the parameter space.

This comprehensive statistical approach allows us to thoroughly characterize the behavior of D-statistic and FST under various evolutionary scenarios, establishing a robust framework for distinguishing ILS from introgression in empirical studies.

5 Results

5.1 Scenario 1: Pure ILS

5.1.1 Simulation Setup and Rationale

This scenario evaluates the behavior of two summary statistics—D-statistic and FST—under conditions of pure incomplete lineage sorting (ILS), where there is no post-divergence gene flow. It establishes the null model for our study, simulating genealogical discordance that arises strictly from the stochastic sorting of ancestral alleles (Degnan Rosenberg, 2009). We simulated a four-taxon topology $(((A, B), C), O)$, with:

- **A and B** as sister taxa,
- **C** as the outgroup to A+B,
- **O** as an external outgroup for ancestral allele polarization.

We varied two evolutionary parameters:
Simulation details:

- 20 haploid samples per population
- 1 Mb genomic region
- Mutation and recombination rate: 1×10^{-8} per site per generation
- 30 replicates per parameter combination

Table 2: Key simulation parameters and rationale. Parameters were chosen to isolate the effects of effective population size and divergence time on the prevalence of incomplete lineage sorting (ILS).

Parameter	Values and Rationale
Effective Population Size (N_e)	**10,000 (small), 100,000 (large)**. Larger N_e preserves ancestral polymorphisms for longer, increasing the likelihood of ILS. Smaller N_e accelerates lineage sorting via genetic drift (PamiloNei, 1988)
Divergence Times (generations ago)	**Recent:** AB = 25k, ABC = 50k; **Intermediate:** AB = 100k, ABC = 200k; **Ancient:** AB = 400k, ABC = 800k. Shorter intervals increase ILS by limiting the time for lineage coalescence; longer intervals reduce ILS probability (Maddison, 1997).

This scenario sets the **baseline expectation** for D-statistic and FST distributions under ILS alone—providing the foundation for all comparative inference in subsequent introgression scenarios.

5.1.2 D-statistic Under Pure ILS

We first examined the distributions of D-statistic values under pure incomplete lineage sorting. In the absence of gene flow, ABBA and BABA patterns are expected to occur at equal frequencies, resulting in an expected D-statistic of approximately zero ($D \approx 0$) (Durand et al., 2011).

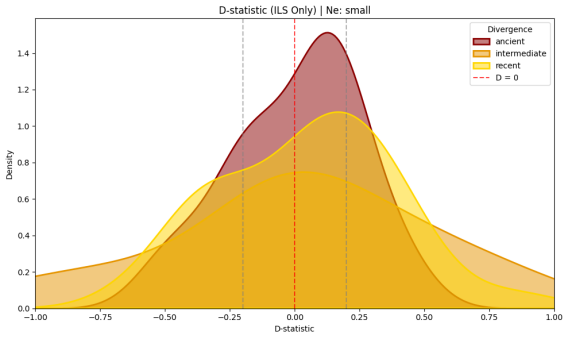

Figure 5: Distribution of D-statistic values under pure ILS with small effective population size ($N_e = 10{,}000$) across three divergence time scenarios. The red dashed line indicates $D = 0$, the theoretical expectation under ILS. Gray dashed lines show typical detection thresholds at $D = \pm 0.25$, used to assess whether introgression signals exceed the ILS null range. All divergence scenarios show wide distributions centered near zero, with the recent divergence scenario displaying the broadest spread and most pronounced tail density.

Across all simulations, the D-statistic distributions were centered near zero, as expected under neutral conditions without introgression. However, the variance and tail behavior of the distributions were strongly influenced by both effective population size and divergence time. Under

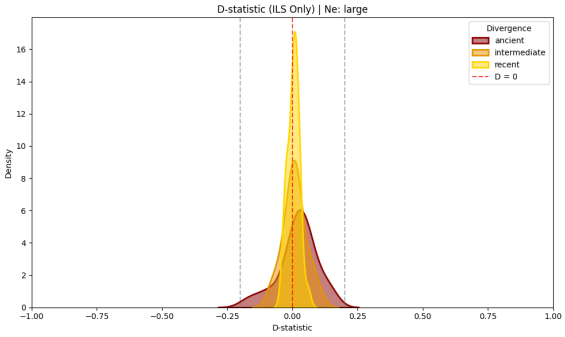

Figure 6: Distribution of D-statistic values under pure ILS with large effective population size (N_e = 100,000). The red dashed line indicates $D = 0$, while gray dashed lines denote significance thresholds at ±0.25. All three divergence scenarios produce sharply peaked distributions well within the neutral ILS range. The recent divergence distribution forms a steep peak at $D = 0$ with a density of approximately 17, compared to a broader, flatter peak near 6 for ancient divergence.

small Ne conditions (Figure 5), the distributions were notably wide, particularly for recent divergence times. The recent divergence curve showed extensive dispersion, with some values exceeding $|D| > 0.5$, indicating that extreme D values can arise under pure ILS when lineage sorting is incomplete. This broad variance reflects the combined effects of rapid speciation and strong genetic drift in small populations, which increase the likelihood of discordant gene trees through stochastic coalescent events.

In contrast, simulations with large Ne (Figure 6) produced D-statistic distributions that were narrowly concentrated around zero. The peak of the recent divergence distribution reached a density of approximately 17 at $D = 0$, while the ancient divergence scenario produced a broader, flatter peak around 6. Importantly, in this large Ne setting, all three divergence time scenarios fell entirely within the ±0.25 range, marked by the gray dashed lines. This range, often used as a practical significance threshold in empirical studies, serves here as a visual benchmark for distinguishing extreme deviations from neutral ILS expectations.

The difference between Figures 5 and 6 is particularly informative: in small Ne populations, D-statistic values often extend well beyond ±0.25, making it difficult to distinguish weak introgression from stochastic ILS. However, in large Ne settings, D values remain tightly constrained within the neutral range, improving the resolution to detect genuine introgression events in subsequent scenarios. Thus, when D-statistics exceed ±0.25 in a high-Ne context, such signals are less likely to be attributable to ILS alone and warrant further investigation.

These findings emphasize the importance of incorporating demographic context when interpreting D-statistic values. While the central expectation of $D = 0$ holds across all conditions, the width and extremity of the null distribution vary considerably with Ne and divergence time. In empirical datasets involving recent divergence or small Ne, extreme D values must be interpreted cautiously, as they may fall within the expected variance of pure ILS. Only deviations that exceed the ILS-induced variance—especially in large Ne scenarios—can be considered strong evidence for introgression (Hahn Hibbins, 2019).

5.1.3 FST Under Pure ILS

FST measures population differentiation. Under pure ILS (no gene flow), we expect FST to increase with divergence time and decrease with larger (N_e) (Wright, 1949). This pattern arises because longer divergence times allow more genetic drift to accumulate differences between populations, thereby increasing inter-population differentiation. Larger effective population sizes slow the rate of drift, leading to reduced divergence in allele frequencies over time (Charlesworth, 1998; Cruickshank & Hahn, 2014).

Divergence time was found to exert a strong influence on the location of F_{ST} distributions, with values consistently increasing as the time since population split increased. This rightward shift in the distributions reflects the accumulation of genetic differentiation over time in the absence of gene flow. For both small and large effective population sizes, the F_{ST} distributions were tightly constrained within characteristic ranges. In simulations with small N_e = 10,000, the F_{ST} values ranged from approximately 0.15 in the recent divergence scenario to around 0.85 in the ancient divergence scenario.

For large N_e = 100,000, the corresponding range was lower, from approximately 0.02 to 0.35. These results demonstrate that population size significantly affects the magnitude of F_{ST}, with larger N_e values leading to systematically lower F_{ST} across all divergence times. This pattern arises because larger populations maintain ancestral poly-

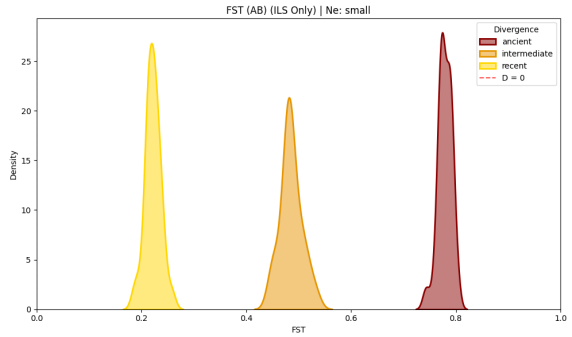

Figure 7: Distribution of F_{ST} between populations A and B under pure ILS with small effective population size (N_e = 10,000). Each curve represents a different divergence time: recent, intermediate, and ancient. As the time since speciation increases, the distributions shift rightward, with means around 0.21 (recent), 0.47 (intermediate), and 0.79 (ancient). This reflects greater population differentiation due to more time for genetic drift to act and for lineages to coalesce within species boundaries, reducing the extent of incomplete lineage sorting.

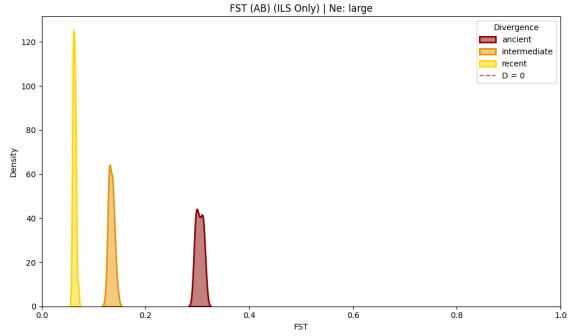

Figure 8: Distribution of F_{ST} between populations A and B under pure ILS with large effective population size (N_e = 100,000). All distributions are shifted leftward relative to the small N_e case, with lower mean values: 0.04 (recent), 0.13 (intermediate), and 0.30 (ancient). As with small N_e, increasing divergence time leads to greater differentiation. However, larger populations retain ancestral variation longer, resulting in slower allele frequency divergence and reduced F_{ST}. The sharply peaked distributions indicate low stochastic variance and reflect the predictability of drift-driven differentiation in large populations under ILS.

morphisms longer, reducing the rate at which allele frequencies diverge between populations. The observed relationship is consistent with theoretical expectations under the neutral model, where

$$F_{ST} \approx 1 - e^{-t/2N_e}$$

and larger N_e leads to slower accumulation of genetic differentiation over time (Wright, 1943).

In addition to differences in magnitude, the shape of the FST distributions also differed markedly between population sizes. FST distributions were highly peaked and narrow, especially under large Ne, indicating low stochastic variation across simulation replicates. This contrasts with the behavior of the D-statistic, which exhibited broader and more variable distributions due to its sensitivity to coalescent stochasticity. These findings underscore the utility of FST as a stable and robust measure of baseline genetic divergence in neutral evolutionary contexts. Importantly, in scenarios involving introgression, a reduction in FST between non-sister taxa—manifesting as a leftward shift relative to these null distributions—may serve as a diagnostic signal of gene flow rather than incomplete lineage sorting.

5.1.4 Establishing Null Distributions for Subsequent Scenarios

This ILS-only scenario provides foundational reference distributions for interpreting the D-statistic and FST under neutral demographic conditions. These null patterns serve as diagnostic baselines for distinguishing signals of introgression from the background noise generated by incomplete lineage sorting alone.

Table 3: Summary of null expectations under incomplete lineage sorting (ILS) and how deviations from these baselines inform diagnosis of introgression.

Statistic	Null Pattern (ILS-only)	Diagnostic Use
D-statistic	Centered at 0; variance depends on N_e and divergence time	Significant introgression must exceed this "ILS noise floor" (e.g., $\|D\| > 0.2$)
FST	Increases with divergence time; lower values for large N_e	Reduced F_{ST} between non-sister taxa can flag post-divergence gene flow

These baseline distributions are notably sensitive to underlying demographic parameters. The D-statistic is particularly influenced by variance and tail behavior, with the greatest spread observed in scenarios involving small effective population sizes or recent divergence events. Such conditions increase the prevalence of ILS and thereby elevate the probability of observing extreme D values even

in the absence of introgression. In contrast, FST is primarily shaped by differences in mean values and range shifts across parameter regimes, providing a more stable and robust measure of long-term genetic divergence between populations.

Together, these patterns underscore the importance of demographic context in the interpretation of both statistics. As we proceed to the next scenario, which introduces post-divergence introgression between non-sister taxa, these ILS-only distributions will be essential for identifying statistically meaningful deviations. By comparing observed D-statistic and FST values against their respective null distributions, we can diagnose whether gene flow has occurred and assess the demographic conditions under which introgression becomes reliably detectable.

5.1.5 Summary of Scenario 1: Key Takeaways

Table 4: Summary of D-statistic and FST behavior under pure ILS

Parameter	Effect on D-statistic	Effect on FST
Small N_e	Broad, noisy D-distributions; high tail density	High FST, due to faster fixation
Large N_e	Narrow D centered near 0; low tail density	Lower FST, due to slower drift
Recent divergence	High D-variance and wider tails	Lower FST (0.04–0.21)
Ancient divergence	Narrow D, low variance	High FST (0.30–0.79)

The results from Scenario 1 illustrate how the behavior of D-statistic and FST is shaped by variation in effective population size and divergence time under pure incomplete lineage sorting. Notably, the D-statistic is most sensitive to variance-inducing factors such as small Ne and recent speciation, resulting in broader distributions with extended tails. In contrast, larger Ne and deeper divergence reduce this variance, yielding narrower distributions centered around zero. FST, meanwhile, increases predictably with divergence time and decreases with larger Ne due to slower genetic drift and greater retention of ancestral variation.

These patterns reflect well-established theoretical expectations and reinforce the complementary nature of the two statistics: the D-statistic captures variation in allele sharing driven by recent evolutionary events, while FST provides a more stable indicator of long-term differentiation between populations. The table below summarizes the specific effects of demographic parameters on the two metrics, serving as a concise reference point for interpreting outcomes in subsequent scenarios involving introgression.

5.2 Scenario 2: Pure Introgression

5.2.1 Simulation Setup and Rationale

This scenario investigates the genomic consequences of *pure introgression*—a discrete, unidirectional gene flow event occurring post-speciation between non-sister taxa in the absence of incomplete lineage sorting (ILS) or other confounding demographic processes. The goal is to establish clear expectations for the behavior of D-statistic and F_{ST} metrics when introgression occurs in isolation, thereby providing critical contrast to the ILS-only null model in Scenario 1.

We modeled a standard four-taxon phylogeny $(((A, B), C), O)$, where A and B are sister taxa, C is a more distantly related population, and O serves as an outgroup to polarize derived versus ancestral alleles. In this topology, unidirectional gene flow was introduced from C to B—that is, between non-sister taxa—at one of three timepoints following divergence. Importantly, no gene flow was allowed between sister taxa (A and B), preserving the integrity of the species tree and eliminating confounding effects such as overlapping migration or bidirectional admixture.

Each simulation consisted of a single discrete pulse of migration (i.e., `continuous=False`) from $C \rightarrow B$, occurring at one of three divergence-informed times: recent (10,000 generations ago), intermediate (50,000), or ancient (200,000). This range was selected to span plausible post-speciation intervals over which introgression signals may persist or erode. Migration strength was also varied systematically from none ($m = 0$) to high ($m = 10^{-2}$) across biologically realistic levels, allowing us to test detection thresholds of D-statistic and F_{ST} under different intensities of gene flow.

To ensure statistical robustness, 50 replicate simulations were performed for each parameter combination. Each simulation generated a 1 Mb genomic region with 20 sampled haploid genomes per population and used a mutation and recombination rate of 1×10^{-8} per site per generation. For each replicate, we computed the D-statistic, which captures asymmetry in ABBA-BABA patterns (Durand et al., 2011), as well as F_{ST} between sister taxa (F_{ST}^{AB}) and between the introgressing lineages (F_{ST}^{BC}). These summary statistics were visualized using kernel density estimates and slid-

Table 5: Pure introgression simulation parameters. Parameters were chosen to minimize background ILS noise and systematically explore the effect of gene flow strength and timing on genomic statistics.

Parameter	Value(s)	Rationale
Effective population size (N_e)	100,000	Large N_e reduces stochastic variance from incomplete lineage sorting (ILS), allowing clearer detection of introgression signals.
Divergence times	AB = 800,000 ABC = 400,000	Deeper divergences decrease the probability of ancestral polymorphism persistence, minimizing ILS and sharpening introgression detection (Degnan & Rosenberg, 2009).
Migration rate (m)	None (0), Low (10^{-6}), Moderate (10^{-4}), High (10^{-2})	Varying m allows detection thresholds to be evaluated across a biologically plausible range of introgression intensities.
Migration timing (generations ago)	Recent (10,000), Intermediate (50,000), Ancient (200,000)	Temporal variation tests the hypothesis that older introgression events leave weaker genomic signals due to recombination and drift (Racimo et al., 2015).
Migration mode	Episodic (pulse)	A single, discrete introgression event isolates effects of hybridization from continuous gene flow dynamics.
Migration direction	$C \rightarrow B$	Directional introgression from a non-sister lineage ensures that introgression results in an excess of ABBA patterns, producing positive D values under the Durand test (Durand et al., 2011).

ing window analyses to capture both genome-wide trends and local heterogeneity.

Why this qualifies as "pure" introgression. By introducing gene flow only between non-sister taxa ($C \rightarrow B$), using a time-bound and directionally fixed migration event, and holding divergence history constant, this setup meets all theoretical criteria for simulating pure introgression (Hibbins & Hahn, 2022). No other population pairs exchange migrants, and each replicate represents an isolated, episodic gene flow event—free of complex demographic entanglements or ILS-driven discordance. As such, this scenario provides a textbook case for identifying how introgression uniquely shapes genomic summary statistics, thereby directly addressing the core research question of this study: how to distinguish introgression from incomplete lineage sorting using D-statistic and F_{ST}.

5.2.2 D-statistic Under Pure Introgression

This section analyzes the behavior of the D-statistic under conditions of pure introgression between non-sister taxa B and C, using the experimental setup described in Section 5.2.1. D-statistics were calculated across 50 replicates for each combination of three migration times (recent, intermediate, ancient) and four migration rates (none, low, moderate, high). The effective population size was held constant at $N_e = 100{,}000$ and divergence times were fixed at 800,000 generations (A vs B) and 400,000 generations (A/B vs C) to suppress ILS noise and isolate the signal of introgression.

The D-statistic measures asymmetry in allele sharing among taxa, where excess ABBA over BABA patterns (indicative of gene flow from C to B) leads to a positive D-value (Durrand, 2011). Under our setup, directional migration from C to B should result in systematically right-skewed D-statistic distributions, with the magnitude of the skew scaling with the strength and recency of migration.

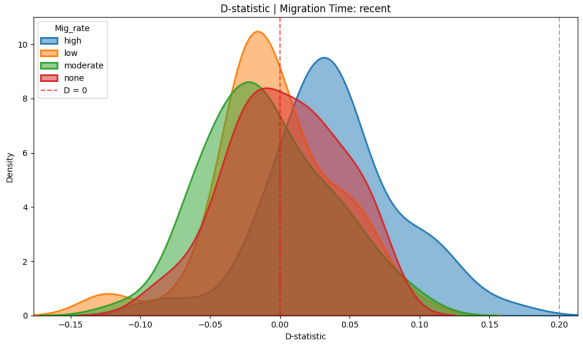

Figure 9: Distribution of D-statistic values under recent migration (10,000 generations ago). Increasing migration rates yield stronger rightward shifts: high migration (blue) is centered at $D \approx 0.08$, moderate (green) at $D \approx 0.03$, and low (orange) near $D \approx 0.005$. The control (red) remains centered at $D = 0$. The dashed vertical lines at $D = \pm 0.2$ represent typical empirical detection thresholds.

As shown in Figure 9, under recent introgression (10,000 generations), D-statistic distributions exhibit a monotonic shift toward positive values as migration rate increases. The high migration regime ($m = 10^{-2}$) produces a distribution with a mode at $D \approx 0.08$, and a substantial proportion of values exceed the empirical significance threshold of $|D| > 0.2$, indicating strong signal detection capability. In contrast, the no-migration control remains tightly centered at $D = 0$, with minimal variance, validating the specificity of D as a signal of gene flow.

With intermediate migration timing (50,000 generations ago), the D-statistic remains effective at detecting gene flow but shows attenuation in signal magnitude (Figure 10). The high migration condition now peaks at $D \approx 0.07$, reflecting a ~12.5% reduction from the recent case. This decline aligns with theoretical predictions that the detectability of introgression diminishes over time due to recombination breaking down introgressed

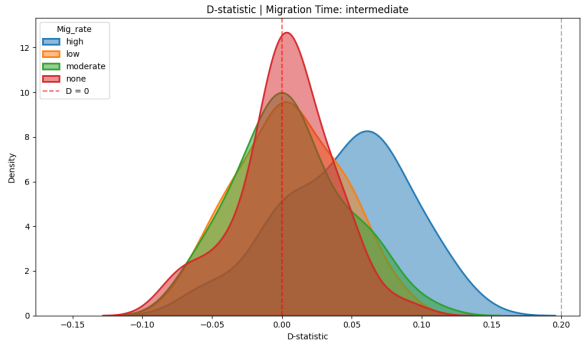

Figure 10: D-statistic distributions for intermediate introgression (50,000 generations ago). High migration (blue) yields a peak at $D \approx 0.07$, with lower rates showing marginal deflections. The control distribution remains tightly centered around 0, reinforcing the specificity of the D-statistic.

haplotypes and genetic drift eroding signal strength (Racimo, 2015).

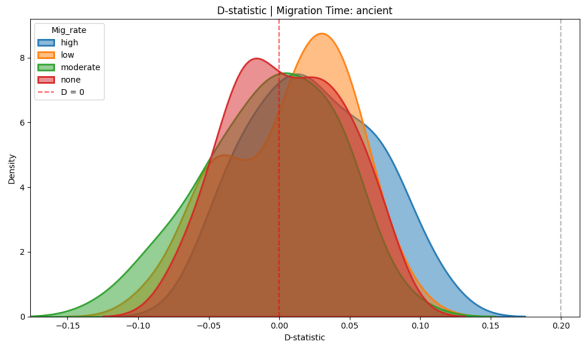

Figure 11: D-statistic distributions under ancient introgression (200,000 generations ago). The high migration curve (blue) still exhibits a rightward shift ($D \approx 0.05$), but there is substantial overlap with both moderate (green) and control (red) distributions, indicating signal degradation over time.

Under ancient migration conditions (200,000 generations ago), the D-statistic signal is further attenuated, with high migration values centered around $D \approx 0.05$ (Figure 11). This value remains distinguishable from the control but increasingly overlaps with moderate and low migration distributions. The signal-to-noise ratio declines considerably, illustrating the reduced power of the D-statistic to detect older introgression events, especially at low migration rates.

Distributional Behavior. Across all timepoints, the D-statistic distributions remain unimodal and right-skewed for high and moderate migration rates, with the variance increasing slightly with decreasing migration timing. Importantly, skewness in these distributions directly corresponds to the directionality and strength of gene flow, confirming the diagnostic specificity of the D-statistic in introgression studies.

Implications for Distinguishing ILS vs. Introgression. These results contrast sharply with the ILS-only scenario in Section 5.1, where D-statistic distributions were centered at 0 with heavy-tailed variance under small N_e. In contrast, the pure introgression scenario (with large N_e) yields narrower and more directional distributions whose means scale linearly with migration rate. This behavior strengthens the interpretive framework for distinguishing introgression from ILS: whereas ILS yields symmetric, wide-tailed distributions around $D = 0$, directional gene flow shifts the D-statistic distribution away from 0 in a migration-dependent manner.

Table 6: Summary of D-statistic mean values under varying migration rates and timings. Values represent approximate modal D-statistics.

Migration Rate	Recent (10k)	Intermediate (50k)	Ancient (200k)
None	0.00	0.00	0.00
Low (10^{-6})	~0.005	~0.002	~0.001
Moderate (10^{-4})	~0.03	~0.02	~0.015
High (10^{-2})	~0.08	~0.07	~0.05

Summary of D-statistic Behavior Under Introgression. In summary, the D-statistic emerges as a powerful and directionally sensitive indicator of introgression in this scenario. When demographic conditions such as N_e, divergence time, and gene flow direction are controlled, the D-statistic reliably reflects migration parameters. The scaling of signal strength with migration rate and time since introgression affirms theoretical expectations and provides practical thresholds for distinguishing historical introgression from background ILS.

5.2.3 FST Under Pure Introgression

Genetic differentiation, quantified using F_{ST}, provides complementary evidence to the D-statistic for detecting introgression. In this scenario, we evaluated how post-divergence gene flow between non-sister taxa B and C influences differentiation between both the introgressing populations ($F_{ST_{BC}}$) and the sister pair ($F_{ST_{AB}}$). All simulations were conducted under large effective population size (N_e = 100,000), with fixed divergence times (AB = 800k, ABC = 400k generations ago), thereby

minimizing the confounding effects of incomplete lineage sorting.

FST Between Sister Taxa (A and B). Figure 12 illustrates how increasing migration rates from C to B under *recent* introgression (10,000 generations ago) subtly reduce F_{ST} between A and B. The no-migration control (red) is centered near $F_{ST} \approx 0.442$, while high migration (blue) shows a modest leftward shift with a bimodal distribution and a primary peak around $F_{ST} \approx 0.428$. This reflects an approximate 3.2% reduction in genetic differentiation between the sister taxa.

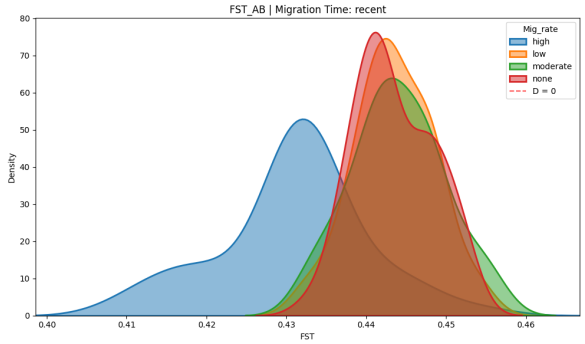

Figure 12: Density plots of $F_{ST_{AB}}$ under recent migration timing (10k generations ago). High migration (blue) causes a visible leftward shift in the distribution, indicating reduced divergence between A and B. The no-migration control (red) remains tightly centered around $F_{ST} \approx 0.442$.

This phenomenon, termed *divergence erosion* (Hibbins & Hahn, 2022), occurs because gene flow from C to B makes B genetically more similar to its non-sister lineage (C), thus indirectly reducing its differentiation from A due to shared ancestry. Importantly, this effect emerges even though A and C do not exchange migrants directly.

As migration timing becomes more ancient (Figures 13 and 14), the impact on $F_{ST_{AB}}$ diminishes. For intermediate migration (50k generations ago), the high migration curve is centered around $F_{ST} \approx 0.425$, while the no-migration distribution remains at $F_{ST} \approx 0.440$. In the ancient case (200k generations ago), all distributions converge near $F_{ST} \approx 0.44$, suggesting that the differentiating signal of introgression degrades over time due to recombination and drift (Racimo et al., 2015).

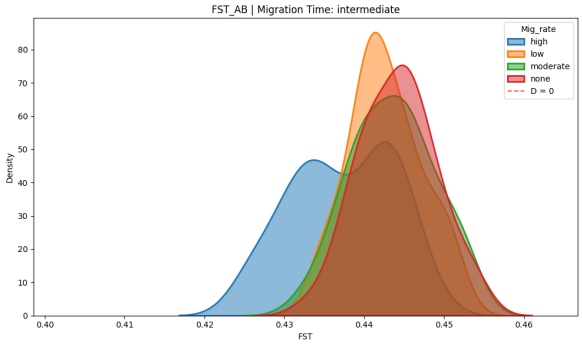

Figure 13: Density plots of $F_{ST_{AB}}$ under intermediate migration timing (50k generations ago). High migration (blue) still causes a leftward shift ($F_{ST} \approx 0.425$), though the magnitude of reduction is smaller than in the recent scenario.

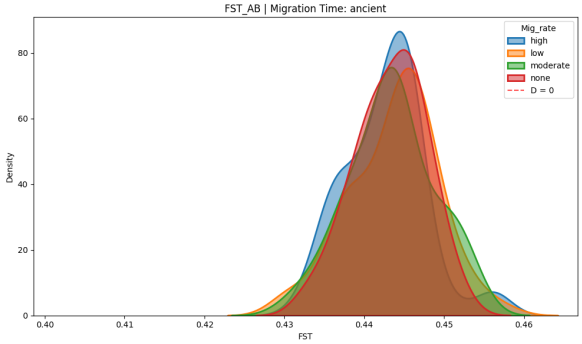

Figure 14: Density plots of $F_{ST_{AB}}$ under ancient migration timing (200k generations ago). All distributions converge, indicating minimal signal retention.

These findings affirm the hypothesis that the timing of gene flow governs the detectability of introgression. Older events leave subtler genomic imprints, making F_{ST} a less powerful statistic in such contexts, especially when analyzing sister taxa indirectly affected by migration.

FST Between Introgressing Taxa (B and C). To directly assess gene flow's homogenizing effect, we analyzed $F_{ST_{BC}}$—the genetic differentiation between B and C. Figure 15 shows the impact of varying migration *modes* (episodic vs. continuous) and directions (C→B vs. B→C), all under high migration ($m = 10^{-2}$) and recent timing.

While $F_{ST_{AB}}$ exhibits only mild changes across all configurations, $F_{ST_{BC}}$ responds dramatically to migration mode. Continuous migration leads to a sharp reduction ($F_{ST} \approx 0.03$), contrasting with episodic migration's much higher values ($F_{ST} \approx$

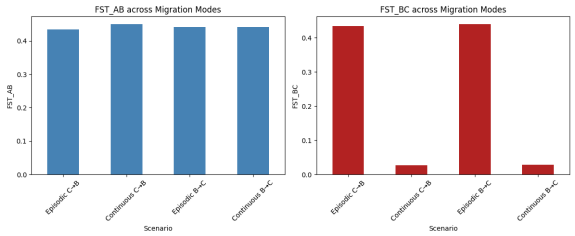

Figure 15: Bar plots comparing $F_{ST_{AB}}$ (left) and $F_{ST_{BC}}$ (right) across four migration scenarios under high recent introgression. Continuous migration drastically reduces $F_{ST_{BC}}$ (from $\approx$ 0.43 to $\approx$ 0.03), whereas $F_{ST_{AB}}$ remains relatively stable, confirming the localized effect of introgression.

0.43). This demonstrates the increased homogenization of allele frequencies across the genome under sustained gene flow, consistent with theoretical expectations (Cruickshank & Hahn, 2014).

Additionally, the direction of migration influences the outcome. Continuous C→B migration suppresses $F_{ST_{BC}}$ more effectively than B→C, revealing that the identity of the donor versus recipient population can leave asymmetric genomic signatures.

Table 7: Summary of F_{ST} behavior under pure introgression.

Parameter	Effect on $F_{ST_{AB}}$	Effect on $F_{ST_{BC}}$
Recent migration	3–4% decrease; left-shifted peak	Variable by mode; up to 93% reduction under continuous migration
Ancient migration	Minimal change; distributions converge	Slight reduction, minimal distinguishability from control
Migration mode: Episodic	Modest reduction in $F_{ST_{BC}}$	Retains signal at $\approx$ 0.43
Migration mode: Continuous	Sharp drop to $F_{ST} \approx 0.03$	Strong homogenization, clear introgression signature

Summary and Interpretation. Taken together, the F_{ST} analyses reveal two complementary insights: (1) gene flow between non-sister taxa can indirectly erode divergence between sister lineages, and (2) $F_{ST_{BC}}$ is highly responsive to the mode and direction of migration. These patterns directly address our core research question by showing that F_{ST} offers diagnostic value for distinguishing introgression from incomplete lineage sorting, particularly in high-recombination regions or under continuous gene flow.

5.2.4 Genomic Heterogeneity in Introgression Signals

To investigate the spatial consistency of introgression signals along the genome, we performed a sliding window analysis of F_{ST} between the two introgressing populations (B and C). The simulated scenario featured continuous high migration ($m = 10^{-2}$) from population C to B under a neutral evolutionary model. We sampled 50 windows of equal size across a 1 Mb genome and computed raw and smoothed $F_{ST_{BC}}$ values to assess regional variation.

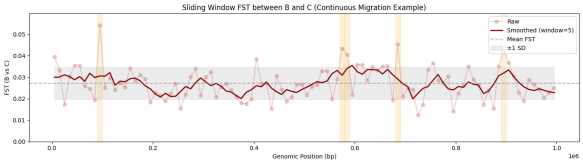

Figure 16: Sliding window analysis of F_{ST} between populations B and C under continuous high migration ($m = 10^{-2}$) in a neutral evolutionary model. Each pink dot represents raw F_{ST} calculated within a 20 kb window, while the solid dark red line shows the smoothed moving average (window = 5). The gray dashed line indicates the mean F_{ST} value (0.027) across all windows, with the shaded region representing ±1 standard deviation. Orange vertical bands highlight windows in the 95th percentile of genetic differentiation, revealing localized regions of elevated F_{ST} despite uniform migration rates across the genome. This heterogeneity emerges from stochastic coalescent processes and recombination alone, without selection, demonstrating how 'genomic islands' can arise purely from neutral processes—a critical consideration when interpreting empirical patterns of differentiation.

Despite uniform high migration across the entire genomic region, the sliding window results (Figure 16) reveal substantial heterogeneity in $F_{ST_{BC}}$ values, with raw window estimates ranging from 0.01 to over 0.05. The overall mean $F_{ST_{BC}}$ across the chromosome was 0.027, which is markedly lower than the no-migration baseline of $F_{ST} \approx 0.44$ observed in earlier sections, reinforcing the strong homogenizing effect of sustained gene flow.

The presence of peaks in the upper 5th percentile—highlighted in orange—indicates that even in the absence of selection, recombination and drift can create genomic regions that retain elevated differentiation. This result mirrors empirical findings where heterogeneous introgression landscapes have been linked to "genomic islands of differentiation"—regions that resist gene flow due to selection against foreign alleles or reduced recombination (Wolf & Ellegren, 2017).

Neutral Origins of Heterogeneity. Because our simulations were conducted under a strictly neutral model, without any form of selection, all observed variation in $F_{ST_{BC}}$ must be attributed to stochastic processes. One primary contributor is genetic drift, which can cause certain genomic windows—especially smaller ones—to retain elevated levels of divergence simply by chance. Additionally, even though recombination was modeled as uniform across the genome, the stochastic nature of genealogical coalescent events introduces localized heterogeneity in tree topologies and allele sharing patterns.

As Ravinet et al. (2017) emphasize, such variation can generate differentiation peaks that superficially resemble selection-driven "genomic islands." This reinforces the notion that elevated F_{ST} regions alone do not constitute conclusive evidence of resistance to introgression. Unless supported by complementary signals such as reduced nucleotide diversity, excess non-synonymous substitutions (dN/dS), or functional enrichment, apparent peaks may reflect neutral background processes rather than adaptive barriers to gene flow.

Implications for Detecting Introgression. The heterogeneity observed in sliding window F_{ST} has several methodological implications for studies seeking to identify introgression. First, computing average F_{ST} across the genome can obscure the presence of introgression if its effects are not uniformly distributed; focal signals may be diluted in the overall mean. Therefore, sliding window or outlier-based approaches are critical for detecting local patterns of resistance or facilitation of gene flow.

However, such methods must be interpreted within the context of expected neutral variation to avoid overattributing stochastic noise to biological causes. Importantly, integrating these spatially explicit F_{ST} measures with genome-wide D-statistic results enables a more nuanced perspective: while the D-statistic captures directional gene flow across the whole genome, windowed F_{ST} reveals where that signal may be regionally enhanced or suppressed. Together, they form a complementary toolkit for dissecting the genomic architecture of introgression.

Table 8: Summary of Genomic Heterogeneity in $F_{ST_{BC}}$ under High Continuous Introgression

Observation	Interpretation	Implication for Inference
Mean $F_{ST_{BC}} \approx 0.027$	Strong genome-wide homogenization	Signal of recent continuous introgression
5% of windows $F_{ST} > 0.045$	Residual divergence persists in specific windows	Possible outliers from drift; may mimic resistance
Smoothed variation along genome	Local fluctuations due to recombination	Importance of fine-scale spatial analysis

Conclusion. The heterogeneous $F_{ST_{BC}}$ landscape observed in this scenario—despite constant high gene flow—supports the growing consensus that introgression is rarely genome-wide in its effects. These findings validate the use of spatially explicit statistics in distinguishing introgression from incomplete lineage sorting and caution against overinterpretation of individual peaks without neutral expectations. Combined with the D-statistic, windowed F_{ST} analyses offer a powerful toolkit for parsing the complexity of hybridization in empirical data (Martin et al., 2013; Wolf & Ellegren, 2017).

5.2.5 Key Diagnostic Signatures of Pure Introgression

The pure introgression scenario described in this study allows us to isolate and identify several reliable diagnostic features that distinguish gene flow from incomplete lineage sorting (ILS). By varying only the timing and intensity of a single unidirectional migration event (C→B) in a neutral demographic setting, we were able to characterize the behavior of both the D-statistic and FST under conditions where introgression is the sole source of phylogenetic discordance. The following paragraphs synthesize these findings.

First, directional skew in the D-statistic is a clear and quantifiable signal of introgression. Across all migration timings, we observed consistent rightward shifts in the D-statistic distribution relative to the null expectation of zero, particularly for high migration rates. These shifts are aligned with the expected excess of ABBA patterns produced by gene flow from C into B, and their magnitude scales with the strength of migration. In contrast, ILS alone—as shown in Scenario 1—produced symmetrical D-statistic distributions centered at zero with wider tails due to stochastic lineage sorting.

Second, we observed a decay in signal strength with time since introgression, consistent with theoretical predictions (Durand et al., 2011; Martin et al., 2015). Older introgression events were harder to detect, both in terms of D-statistic shifts and re-

ductions in FST, due to the erosion of linkage patterns by recombination and drift. Despite this attenuation, strong gene flow (m = 1×10^2) remained detectable even 200,000 generations after the migration event—an important finding for researchers studying ancient hybridization.

Third, reduced genetic differentiation (FST) between introgressing populations served as a complementary signal of gene flow. Particularly in continuous migration scenarios, FST between B and C dropped dramatically (e.g., to 0.03), indicating near-complete homogenization of allele frequencies. This effect was not observed in ILS-only scenarios, where all FST values remained relatively high due to the absence of gene flow.

Notably, introgression also produced secondary effects on sister taxa: gene flow from C into B subtly but consistently reduced FST between A and B. This "divergence erosion" (Hibbins Hahn, 2022) arises because introgression alters B's genetic composition, causing it to drift away from its sister taxon A. Such indirect signatures may be particularly useful in empirical systems where direct donor-recipient comparisons are not available.

Finally, genomic heterogeneity in introgression signals emerged even in the absence of selection. Sliding window analyses of FST between B and C revealed considerable variance in differentiation levels along the genome. This heterogeneity, though arising from neutral processes like recombination and drift, mimicked patterns often attributed to selection—highlighting the need for caution when interpreting genomic islands of differentiation (Wolf Ellegren, 2017).

These key features are summarized in Table 9.

Table 9: Diagnostic features distinguishing introgression from incomplete lineage sorting (ILS) based on simulation outcomes.

Signature	Behavior under Introgression	Behavior under ILS (Scenario 1)
D-statistic Mean Shift	Directional skew (e.g., $D \approx 0.08$ at $m = 10^{-2}$, recent)	Symmetric around 0
D-statistic Variance	Narrower, more peaked distributions	Wider tails from lineage sorting
Temporal Signal Decay	Signal weakens with older events	No systematic decay pattern
FST (B vs C)	Reduced under gene flow (e.g., ≈ 0.03 under continuous migration)	High differentiation (≈ 0.44)
FST (A vs B)	Subtle reduction due to divergence erosion	Unaffected by non-sister gene flow
Genomic Heterogeneity	Windowed F_{ST} varies; peaks from drift	Similar heterogeneity possible, but not directional

In conclusion, this scenario reveals that introgression leaves a suite of interlocking genomic signals—from genome-wide D-statistic shifts to local FST reductions and heterogeneity—that can be used in concert to distinguish gene flow from ILS. The strongest diagnostic power is achieved by leveraging both global patterns (e.g., D-statistic skew, reduced FST) and local structure (e.g., sliding window FST peaks), particularly when interpreted within a well-calibrated demographic model.

5.3 Scenario 3: Mixed Introgression Scenarios

5.3.1 Simulation Design

To explore the interaction between incomplete lineage sorting (ILS) and introgression, we designed a simulation scenario that systematically varies both evolutionary processes. This mixed setup enables us to assess how introgression signatures behave in the presence of varying ILS backgrounds and to identify parameter regimes where these processes can be statistically distinguished—advancing the central question of how to disentangle gene flow from ancestral polymorphism.

We maintained the canonical four-taxon phylogeny $((A, B), C), O$, and implemented a comprehensive grid of parameter combinations that vary divergence times, population sizes, migration rates, timings, directions, and models. Table 10 summarizes the parameters used.

Simulation Set-up For each unique combination of the above parameters, we performed $n = 10$ replicate simulations, yielding a total of:

$$\begin{aligned}\text{Total simulations} = {} & 3 \text{ (divergence levels)} \\ & \times 3 \text{ (migration timings)} \\ & \times 5 \text{ (migration rates)} \\ & \times 3 \ (N_e \text{ values}) \\ & \times 2 \text{ (migration models)} \\ & \times 2 \text{ (directions)} \\ & \times 10 \text{ (replicates)} \\ = {} & \mathbf{5,400} \text{ simulations}\end{aligned}$$

Each replicate simulated a 1 Mb chromosome and produced full genealogical information using tree-sequence recording, resulting in over 5.4 Gb of structured simulation output for downstream metric analysis.

Table 10: Mixed Scenario Simulation Parameters

Parameter	Values	Rationale
Effective population size (N_e)	100,000, 200,000, 500,000	Varies the strength of ILS. Larger N_e retains deeper coalescent branches, enhancing signal overlap.
Divergence scenarios	Shallow: AB = 100,000; ABC = 200,000 Intermediate: AB = 400,000; ABC = 800,000 Ancient: AB = 800,000; ABC = 1,600,000	Provides a gradient of ILS backgrounds by changing time since shared ancestry.
Migration rates	0.0 (none), 10^{-6} (low), 10^{-4} (moderate), 10^{-2} (high), 0.1 (very high)	Covers the full spectrum from negligible to strong introgression.
Migration timings	10,000 (recent), 50,000 (intermediate), 200,000 (ancient)	Tests how temporal distance of gene flow affects signal decay.
Migration models	Episodic, Continuous	Episodic simplifies causal interpretation. Continuous adds realism and complexity.
Migration directions	$C \rightarrow B$, Symmetric ($C \leftrightarrow B$)	Allows comparison between asymmetric and bidirectional gene flow.

Rationale and Theoretical Justification. This design allows us to evaluate the D-statistic (Durand, 2011) and F_{ST} under diverse evolutionary pressures. Divergence times modulate ILS depth, while N_e controls retention of ancestral variation. Migration parameters shape the strength and persistence of gene flow signals. By varying each component, we probe how robust each metric is to confounding, and under what regimes one outperforms the other. Episodic and continuous models further allow contrasting interpretations of pulse vs. persistent introgression (Hibbins, 2022).

Goal Alignment. The mixed scenario reflects the complexity of natural systems, where ILS and introgression often co-occur. By generating this multidimensional landscape, we create a principled basis for identifying when each statistic is interpretable—and where both fail. The extensive grid enables us to assess metric behavior under overlapping ILS and introgression, a key challenge in evolutionary inference.

5.3.2 D-statistic Signatures Across Mixed Conditions

The D-statistic distributions across varying migration rates revealed distinctive patterns that differed markedly between shallow and intermediate divergence scenarios (Figures 17-18).

In the shallow divergence scenario (Figure 17), the D-statistic distributions showed a gradual progression from centered at zero (no migration) to increasingly positive values as migration rates increased. Notably, moderate and low migration rates produced only subtle rightward shifts, with substantial overlap with the no-migration distribution. Very high migration rates ($m = 0.1$) generated a distinct rightward shift (mean $D \approx 0.1$), while super-high rates ($m = 0.3$) produced a broader distribution centered around $D \approx 0.3$.

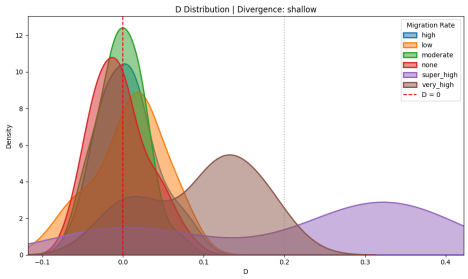

Figure 17: D-statistic distributions across migration rates under shallow divergence (ABC = 100,000, AB = 200,000 generations). Migration rates produce a systematic progression of D-value distributions from centered at zero (no migration, red) to strongly positive values with means of ≈ 0.1 (very_high, brown) and ≈ 0.3 (super_high, purple). The shallow divergence creates a background of substantial ILS, causing wider distribution spreads and more overlap between adjacent migration rates.

The distributions under shallow divergence exhibited considerable width, reflecting the substantial contribution of ILS in creating stochastic variation in allele sharing patterns. This width resulted in significant overlap between adjacent migration rate categories, particularly at the lower end of the spectrum. The overlap indicates that at this divergence depth, low to moderate introgression can be difficult to distinguish reliably from pure ILS based on D-statistics alone.

By contrast, in the intermediate divergence scenario (Figure 18), the D-statistic distributions showed markedly clearer separation between migration rate categories. The reduced ILS background (due to longer divergence times) resulted in narrower distributions with distinct modes for each migration rate: no migration centered tightly at $D = 0$, high migration at $D \approx 0.05$, very high at $D \approx 0.3$, and super high at $D \approx 0.6$. This pattern demonstrates how increased divergence time

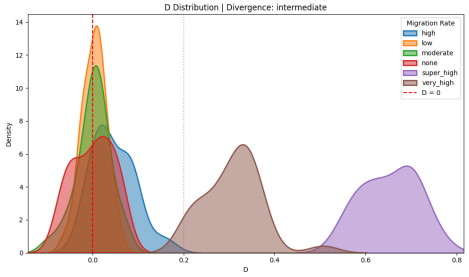

Figure 18: D-statistic distributions across migration rates under intermediate divergence (ABC = 400,000, AB = 800,000 generations). Migration effects are more clearly separated with distinct distribution modes for each rate: none (centered at $D = 0$), high ($D \approx 0.05$), very_high ($D \approx 0.3$), and super_high ($D \approx 0.6$). The reduced background ILS results in narrower distributions and less overlap compared to the shallow divergence scenario.

improves the resolution of introgression detection by reducing the confounding effect of ILS.

The transition from shallow to intermediate divergence showed a pronounced increase in maximum D-statistic values, with the super-high migration rate producing a distribution centered around $D \approx 0.6$ in the intermediate divergence scenario compared to $D \approx 0.3$ in the shallow divergence scenario. This pattern runs somewhat counter to expectations, as one might predict that older divergences would show weaker introgression signals. The observed pattern suggests that in the context of competing ILS and introgression, reducing the background ILS (through deeper divergence) can actually enhance the relative strength and detectability of introgression signals (Hibbins et al., 2020).

These results align with theoretical predictions by Durand et al. (2011) and expand upon them by demonstrating how the interplay between divergence time and migration rate affects the distribution of D-statistics in mixed scenarios. The clear progression of D-statistic distributions with increasing migration rates confirms the metric's sensitivity to introgression intensity, while the differences between divergence scenarios highlight the importance of accounting for background ILS when interpreting D-statistic values.

5.3.3 FST Patterns and Heterogeneity

FST values between the introgressing populations (B and C) provided complementary evidence for gene flow and exhibited distinctive patterns across migration rates and divergence scenarios (Figures 19–20).

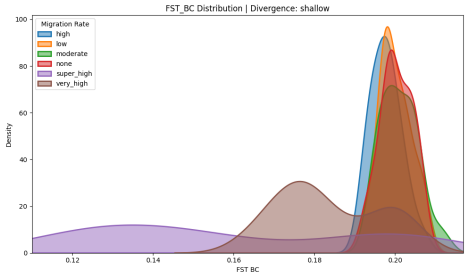

Figure 19: FST distributions between populations B and C across migration rates under shallow divergence. Distributions show a clear progression from high differentiation in the absence of gene flow (none, red, FST $\approx$ 0.20) to moderate differentiation with very_high migration (brown, FST $\approx$ 0.17), and near-complete homogenization with super_high migration (purple, broad distribution centered at FST $\approx$ 0.13). The super_high migration rate produces a distinctive flattened distribution spanning FST values from 0.11 to 0.18.

In the shallow divergence scenario (Figure 19), the baseline FST value between populations B and C without migration was approximately 0.20, reflecting their relatively recent common ancestry. As migration rates increased, FST distributions progressively shifted toward lower values, indicating reduced genetic differentiation due to gene flow. Very high migration ($m = 0.1$) produced a modest but detectable leftward shift to FST $\approx$ 0.17, while super high migration ($m = 0.3$) caused a more substantial reduction, creating a broad distribution centered at FST $\approx$ 0.13. This represents a 35% reduction in genetic differentiation relative to the no-migration baseline.

For the intermediate divergence scenario (Figure 20), the baseline FST was considerably higher (FST $\approx$ 0.45), consistent with the greater time for genetic divergence between populations. The effect of migration was more dramatic in this context, with very high migration producing a distinctive bimodal distribution with peaks at FST $\approx$ 0.37 and FST $\approx$ 0.45. This bimodality suggests the presence of genomic regions with differential susceptibility to introgression, even in our neutral simulation model. Super high migration caused a substantial reduction in FST to approximately 0.25, representing a 44% decrease in genetic differenti-

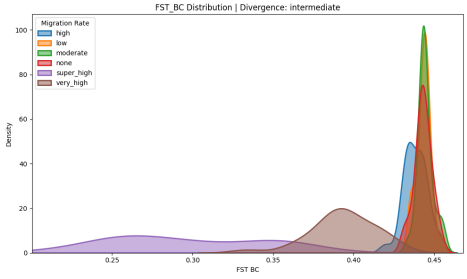

Figure 20: FST distributions between populations B and C across migration rates under intermediate divergence. The baseline differentiation without migration is substantially higher (FST $\approx$ 0.45) than in the shallow divergence scenario. Very_high migration produces a clear bimodal distribution with peaks at FST $\approx$ 0.37 and FST $\approx$ 0.45, while super_high migration dramatically reduces FST to a broad distribution centered around FST $\approx$ 0.25, representing a 44% reduction in genetic differentiation.

ation compared to the no-migration scenario.

The contrasting patterns between shallow and intermediate divergence scenarios highlight how the baseline level of differentiation affects the detectability of introgression via FST. In the shallow divergence case, the relatively low initial differentiation (FST $\approx$ 0.20) limits the magnitude of potential FST reduction due to gene flow. By contrast, the higher baseline differentiation in the intermediate divergence scenario (FST $\approx$ 0.45) provides more "dynamic range" for detecting introgression effects.

Notably, low and moderate migration rates had minimal impact on FST distributions in both divergence scenarios, indicating that FST may be less sensitive than the D-statistic for detecting weak introgression. This aligns with theoretical expectations, as FST reflects genome-wide patterns of differentiation that can be relatively insensitive to limited gene flow affecting only a portion of the genome (CruickshankHahn, 2014).

The super high migration rate produced qualitatively different FST distributions compared to other rates, with broader and more irregular shapes. This pattern suggests that extreme introgression can create complex genomic mosaics of ancestry, resulting in more variable patterns of genetic differentiation across the genome.

5.3.4 Joint Inference Using D and FST

To examine the complementary nature of D-statistic and F_{ST} in distinguishing ILS from introgression, we analyzed their joint behavior across a range of migration rates and effective population sizes (Figure 21).

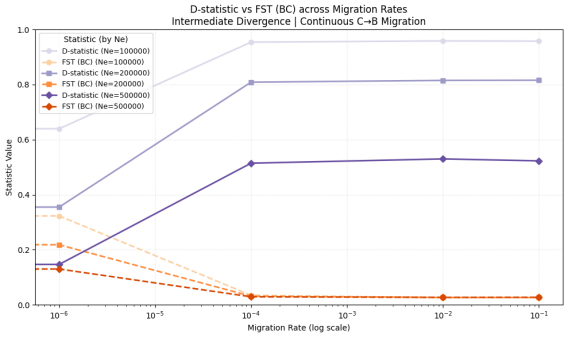

Figure 21: Joint behavior of D-statistic and $F_{ST_{BC}}$ across migration rates for different effective population sizes (100,000–500,000) under intermediate divergence and continuous $C \rightarrow B$ migration. D-statistics (solid lines) increase rapidly with migration rates until plateauing at approximately $m = 10^{-4}$, while F_{ST} values (dashed lines) show the opposite pattern, decreasing sharply as migration intensifies. Larger N_e values (purple) produce lower D-statistic plateaus ($\approx$ 0.5) compared to smaller N_e values (light gray, $\approx$ 0.95), demonstrating the moderating effect of ILS on introgression signals at larger population sizes.

Figure 21 demonstrates the opposing trends of D-statistic and $F_{ST_{BC}}$ as migration rates increase from very low (10^{-6}) to high (10^{-1}) values. For all three effective population sizes, D-statistics increased rapidly with migration rate until reaching a plateau at approximately $m = 10^{-4}$, while F_{ST} values showed the reverse pattern, decreasing sharply from their baseline values as migration intensified.

The plateau effect in D-statistics represents a saturation point beyond which further increases in migration rate do not produce proportional increases in the signal. This saturation occurs because the D-statistic measures relative rather than absolute introgression, and once a sufficient portion of the genome carries introgressed material, additional gene flow does not substantially alter the relative frequencies of ABBA versus BABA patterns.

Effective population size had a notable impact on both statistics. Larger N_e values (500,000, purple lines) resulted in lower D-statistic plateaus (approximately 0.5) compared to smaller N_e values

(100,000, light gray lines, approximately 0.95). This pattern reflects how larger population sizes maintain more ancestral variation and experience slower genetic drift, which moderates the genomic impact of introgression and increases the relative contribution of ILS to gene tree discordance.

The F_{ST} trends similarly varied with N_e, with larger population sizes showing higher initial F_{ST} values (consistent with slower differentiation in larger populations) but ultimately converging to similar low values (approximately 0.02–0.03) under high migration rates. This convergence indicates that intense gene flow eventually overcomes the effects of population size on genetic differentiation.

The crossover points between D-statistic and F_{ST} curves differed across N_e values. For N_e = 100,000, the crossover occurred at approximately $m = 10^{-6}$, while for N_e = 500,000, it shifted to approximately $m = 10^{-5}$. These crossover points potentially represent migration thresholds where introgression signals begin to dominate over ILS effects, with larger populations requiring higher migration rates to reach this transition.

This joint analysis demonstrates the power of combining D-statistic and F_{ST} for evolutionary inference. The D-statistic provides a sensitive detector of directional gene flow even at relatively low migration rates, while F_{ST} offers complementary information about overall genetic differentiation that becomes particularly informative at higher migration rates where D-statistics begin to saturate. Together, these metrics provide a more complete picture of the interplay between ILS and introgression than either would alone.

5.3.5 Genomic Heterogeneity in Mixed Scenarios

To examine how introgression signals vary across the genome in mixed scenarios, we conducted sliding window analyses of D-statistic and F_{ST} between populations B and C (Figures 22–23).

The D-statistic sliding window analysis (Figure 22) revealed extreme variability across the genome, with values ranging from −1.0 to +1.0. This dramatic heterogeneity occurred despite our simulation using uniform mutation, recombination, and migration rates across the genome. Approximately 65% of windows showed strongly positive D-values (> 0.2), consistent with the $C \rightarrow B$ introgression implemented in the simulation. However, a substantial number of windows also showed values close to or below zero, and some even reached the minimum possible value of −1.0.

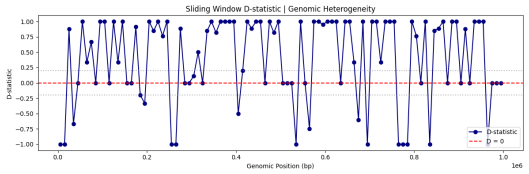

Figure 22: Sliding window analysis of D-statistic across a 1 Mb genomic region under intermediate divergence (ABC = 400,000, AB = 800,000 generations) with high migration ($m = 10^{-2}$). Window-specific D-values (blue line) exhibit extreme variation, ranging from -1.0 to +1.0, with approximately 65% of windows showing strongly positive values ($D > 0.2$) consistent with the implemented C→B introgression. The dramatic fluctuations between adjacent windows demonstrate the inherent stochasticity in genealogical histories, even under uniform simulation conditions. This pattern underscores why genome-wide averaging of the D-statistic is necessary for reliable inference, as individual genomic regions can produce highly discordant signals due to random coalescent events and local recombination histories.

This extreme window-to-window variation highlights the inherent stochasticity in genealogical histories across the genome, created by the interplay of recombination, ILS, and introgression. The pattern demonstrates why genome-wide averaging of the D-statistic is necessary for reliable inference, as any single genomic region can show extreme deviations due to stochastic processes alone.

The F_{ST} sliding window analysis (Figure 23) showed a similarly heterogeneous pattern, though with somewhat less extreme variation than the D-statistic. F_{ST} values between populations B and C fluctuated between approximately 0.15 and 0.50, with a mean of 0.37. The smoothed trend line (orange) revealed subtle regional variation in genetic differentiation, with some genomic regions consistently showing lower or higher F_{ST} values.

The contrast between these adjacent genomic regions with different levels of differentiation mirrors the "genomic islands" pattern often observed in empirical studies (Cruickshank & Hahn, 2014; Wolf & Ellegren, 2017). However, it is crucial to note that in our case, this pattern emerged from purely neutral processes without any selective pressures. This observation underscores a key point:

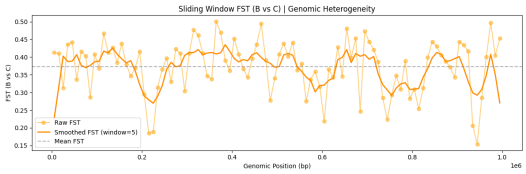

Figure 23: Sliding window analysis of F_{ST} between populations B and C across the same genomic region. Raw F_{ST} values (yellow dots) fluctuate between 0.15 and 0.50, with a mean of approximately 0.37 (gray dashed line). The smoothed trend (orange line) reveals regional variation in differentiation, with localized regions of both reduced and elevated F_{ST}, possibly corresponding to varying levels of gene flow influence across the genome.

heterogeneous differentiation across the genome can arise from stochastic genealogical processes alone and does not necessarily imply differential selection or "islands of speciation" (Noor & Bennett, 2009).

The joint consideration of D-statistic and F_{ST} variation across the genome provides a more nuanced view of how introgression signals manifest in genomic data. While both statistics showed considerable heterogeneity, they did not always covary perfectly. Some regions with highly positive D-values maintained moderate F_{ST} levels, while others showed both reduced F_{ST} and elevated D, potentially representing genomic hotspots of introgression.

These patterns align with empirical observations in studies of natural populations, where introgression is often found to affect the genome heterogeneously (Martin et al., 2013; Payseur & Rieseberg, 2016). Our simulations demonstrate that such heterogeneity can arise even under uniform migration without invoking selection or variation in recombination rates, though these factors would likely amplify the effect in real genomes.

5.3.6 Key Insights from Mixed Scenarios

The mixed introgression scenario provides a nuanced landscape in which both incomplete lineage sorting (ILS) and introgression are simultaneously active. By simulating a wide range of evolutionary conditions, we were able to evaluate the diagnostic power and interpretative limits of both the D-statistic and F_{ST} under complex demographic contexts. Our findings yield several key insights.

Divergence Time Modulates Signal Clarity. The ability to distinguish introgression from ILS improves markedly with increased divergence time. Intermediate divergence scenarios (e.g., T_{ABC} = 400,000, T_{AB} = 800,000 generations) produced sharper, more distinct distributions of both D-statistic and F_{ST} across migration rates than shallow divergence scenarios. This is consistent with theoretical predictions that deeper divergences reduce the stochastic noise introduced by ILS, thereby increasing the signal-to-noise ratio for introgression detection (Durand et al., 2011; Hibbins et al., 2020).

Threshold Effects in Migration Detection. Both D-statistic and F_{ST} exhibited threshold behaviors with respect to migration rate. Low to moderate gene flow (e.g., $m \leq 10^{-4}$) often failed to produce signals distinguishable from pure ILS, particularly under shallow divergence. This implies a lower bound on introgression detectability, beyond which the signal is masked by genealogical stochasticity. Recognizing this threshold is essential when interpreting ambiguous results from empirical studies.

Effective Population Size Shapes Signal Magnitude. Larger effective population sizes (N_e = 500,000) resulted in weaker D-statistic signals and smaller decreases in F_{ST} under the same migration intensities compared to smaller populations (N_e = 100,000). This reflects the persistence of ancestral polymorphisms in large populations, which buffers the impact of gene flow and amplifies ILS. Consequently, demographic scaling must be considered when applying these statistics to real populations.

Genomic Heterogeneity Is Inherent, Even Without Selection. Sliding window analyses of both D-statistic and F_{ST} showed substantial variation across the genome despite simulations being conducted under uniform neutral conditions. D-values ranged from -1 to $+1$, and F_{ST} spanned values from 0.15 to 0.50, underscoring the stochastic nature of genealogical variation. This cautions against over-interpreting genomic "islands" of differentiation without further evidence such as selective sweeps, functional annotations, or reduced diversity metrics.

Complementary Metrics Improve Inference Robustness. Joint analyses revealed that D-statistic

and F_{ST} exhibit opposing trends with increasing migration rates, providing complementary perspectives on gene flow. The D-statistic saturates at high migration intensities while F_{ST} continues to decline, making them useful in different signal regimes. The crossover points between these metrics—where D starts to dominate over F_{ST}—can help define quantitative thresholds for introgression detection under varying demographic conditions.

Table 11: Summary of Key Diagnostic Patterns in Mixed Introgression Scenarios

Factor	Observed Effect on Signal Interpretation
Divergence Time	Sharper signal separation and narrower distributions at intermediate divergence
Migration Rate	Detectable signal threshold at $m \gtrsim 10^{-4}$; saturation of D-statistic at high m
Effective Population Size	Larger N_e reduces signal magnitude due to stronger ILS and slower drift
Genomic Heterogeneity	Large window-to-window variation in D and F_{ST} even under uniform simulation
D vs. F_{ST}	Opposing signal trajectories; together they increase power and robustness

Conclusion. Taken together, these findings show that distinguishing introgression from ILS requires careful consideration of demographic context, signal thresholds, and genome-wide variability. Our results reinforce the value of combining summary statistics—especially D-statistic and F_{ST}—to increase inference power and reduce false positives due to stochastic processes.

5.4 Power and Detection Limits

To evaluate the ability of the D-statistic and F_{ST} to detect introgression under complex evolutionary conditions, we systematically analyzed their behavior across a range of migration rates and demographic parameters. This section focuses on the detection power, joint behavior, and scaling limits of both metrics, offering insight into when and how each signal can be reliably interpreted under mixed ILS and introgression.

5.4.1 Joint Distribution of D and FST

We begin by visualizing how D-statistic and F_{ST} jointly behave under varying migration rates using a two-dimensional kernel density estimation (KDE). This analysis was restricted to intermediate divergence times, where both incomplete lineage sorting (ILS) and introgression are likely to overlap.

The resulting KDE plot (Figure 24) reveals a clear inverse relationship between the two metrics: as D-statistic values increase, F_{ST} decreases. For example, the red kernel representing high migration (labelled "Super High") clusters at $D \approx 0.66$ and $F_{ST} \approx 0.30$, whereas the blue kernel for "No Migration" centers around $D \approx 0$ and $F_{ST} \approx 0.45$. This pattern is consistent with theoretical expectations—introgression increases allele sharing and thus elevates D, while simultaneously homogenizing populations and reducing F_{ST} (Green, 2010).

Importantly, the KDE also highlights the nonlinear nature of these metrics: intermediate migration levels (e.g., "Moderate" and "High") occupy transitional density regions, making it difficult to assign causal inference based on a single metric alone. This supports the case for joint interpretation of D and F_{ST} in mixed scenarios where both gene flow and ILS are present.

5.4.2 Detection Power Under Divergence Scenarios

We next quantified detection power for both metrics across migration rates, defining power as the proportion of replicates exceeding a detection threshold. For D, a fixed threshold of $|D| > 0.2$ was used, while the F_{ST} threshold was dynamically derived as the 95th percentile under the null (no migration) condition.

Intermediate Divergence

Under intermediate divergence times, the D-statistic shows strong sensitivity to increasing migration. As shown in Figure 25, power increases from 0.42 (low migration, 1×10^{-6}) to 0.81 (very high migration, 0.1). F_{ST}, in contrast, remains consistently low across all migration regimes, with power rarely exceeding 0.05. This result aligns with theoretical expectations: D is specifically designed to detect excess allele sharing caused by recent introgression, whereas F_{ST} is a more general measure of population differentiation and is less responsive to recent gene flow (Patterson, 2012).

These results highlight a nonlinear detectability pattern—where D is nearly uninformative at low migration but becomes highly powerful above a certain threshold. Meanwhile, F_{ST}'s lack of responsiveness emphasizes its limitation for intro-

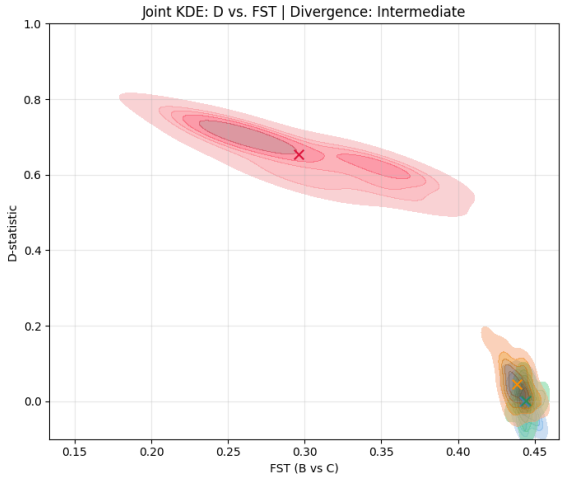

Figure 24: Joint kernel density estimation (KDE) of D-statistic and F_{ST} (B vs C) under intermediate divergence time. Each colored density contour represents a different migration rate regime, from no migration (blue) to super high migration (red). 'X' markers indicate the mean values of each distribution. The plot reveals a pronounced inverse relationship between the metrics: as migration increases, D-statistic values rise (reflecting greater allele sharing due to introgression), while F_{ST} declines (indicating increasing genetic homogenization). Transitional regimes such as "Moderate" and "High" exhibit overlapping and diffuse contours, underscoring the ambiguity in interpreting either statistic in isolation. This highlights the diagnostic value of combining D and F_{ST} to disentangle complex signal mixtures in scenarios where both ILS and introgression are active.

gression detection in moderately diverged populations.

Ancient Divergence with Large N_e

To test metric robustness under deep divergence and large population size, we analyzed power in scenarios with ancient migration (200,000 generations ago) and N_e = 500,000. Here, the D-statistic fails to detect gene flow across all migration regimes (Figure 26), with power values ≈ 0.0. In contrast, F_{ST} exhibits moderate power (0.5) at low to moderate migration rates, before dropping off at higher rates.

This pattern reflects the temporal erosion of the D-statistic signal: the deeper the gene flow occurs in the past, the more likely its signal is overwritten by lineage sorting and drift, especially in large populations. F_{ST}, while blunt, retains some discriminatory power due to its integration over time,

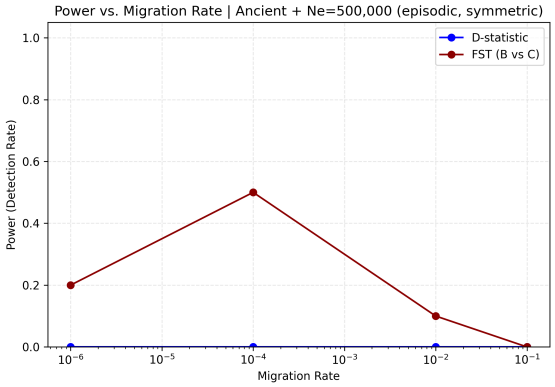

Figure 25: Detection power vs. migration rate under intermediate divergence. The D-statistic (blue) shows increasing power, peaking at 0.81 under very high migration. F_{ST} (red) remains nearly flat, with low power across all regimes. This suggests D is more sensitive in detecting recent introgression in intermediate divergence scenarios.

and thus emerges as the more reliable metric in ancient scenarios.

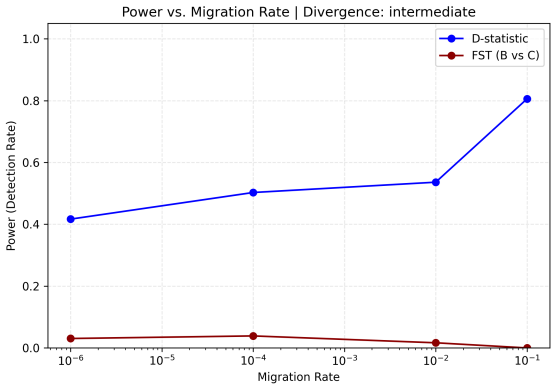

Figure 26: Detection power vs. migration rate under ancient migration with N_e = 500,000. F_{ST} (red) peaks at moderate migration but declines thereafter. The D-statistic (blue) shows negligible power, demonstrating its limitations in detecting introgression from distant time points.

5.4.3 Saturation and Scaling Behavior

Finally, we evaluated how the mean values of D and F_{ST} change with increasing gene flow under intermediate divergence and N_e = 100,000.

As shown in Figure 27, the D-statistic increases sharply up to a migration rate of 0.01, after which it plateaus. This plateau represents a saturation point, beyond which further gene flow does not proportionally increase D. Such behavior is well-

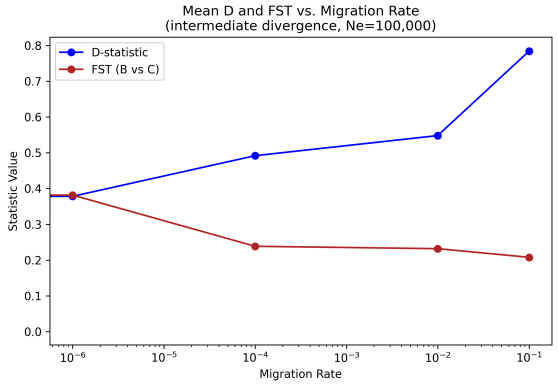

Figure 27: Mean D-statistic and F_{ST} vs. migration rate under intermediate divergence and N_e = 100,000. D (blue) increases rapidly and saturates near 0.78. F_{ST} (red) decreases consistently, indicating a smoother scaling with gene flow. This reveals D's ceiling effect and F_{ST}'s continuous sensitivity to migration.

documented and reflects D's upper bound, which is constrained by the mathematical structure of the ABBA-BABA test (Durand, 2011).

F_{ST}, on the other hand, shows a steady linear decline across the entire migration range. Unlike D, F_{ST} does not saturate and continues to decrease as allele frequencies become more homogenized due to gene flow. This divergence in scaling behavior reinforces the idea that D is optimal for detecting recent or moderate levels of introgression, while F_{ST} can still be informative in continuous migration scenarios where fine-scale divergence persists.

6 Discussion

The challenge of disentangling introgression from incomplete lineage sorting (ILS) is a central problem in evolutionary genomics. Both processes can generate similar gene tree discordance, yet their biological interpretations are fundamentally distinct. Our study systematically evaluated two commonly used statistics—D-statistic and FST—across a comprehensive landscape of demographic conditions, using simulations that independently and jointly modulate ILS and introgression. Here, we integrate the findings from Sections 5.1 to 5.4 to assess the power, sensitivity, and interpretability of these metrics, and consider their implications and limitations for empirical evolutionary analysis.

6.1 Summary of Metric Behavior Across Scenarios

Table 12 summarizes the key findings across four simulation scenarios. Broadly, our results confirm that D-statistic excels in detecting introgression, particularly under intermediate divergence and moderate-to-high migration. However, it saturates at high gene flow levels, becoming less informative. In contrast, FST exhibits a steady and interpretable decline with increasing migration and is more robust to ILS noise, especially under ancient divergence.

Table 12: Summary of metric behavior across scenarios (Sections 5.1–5.4).

Scenario	D-statistic Behavior	F_{ST} Behavior
Pure ILS	Centers near 0; no power	High variance; insensitive to introgression
Pure Introgression	Strong increase with gene flow; saturates at high migration	Steady decline with gene flow; sensitive across full range
Mixed Scenarios	Skewed distributions under moderate migration; high detection power under intermediate divergence	Fails to differentiate weak introgression from ILS; limited power under intermediate divergence
Joint Behavior	Inverse correlation with F_{ST}; clusters reveal clear signal separation under migration	Improves interpretability when jointly analyzed with D-statistic

The inverse relationship between these metrics under introgression creates a characteristic signature that can be leveraged for more accurate inference. This pattern emerges from their complementary sensitivities: D-statistics directly measure allele sharing asymmetries, making them powerful for detecting directional gene flow, while FST quantifies overall differentiation, providing a continuous measure across varying intensities of genetic exchange. This relationship aligns with coalescent theory predictions, where introgression simultaneously increases the frequency of ABBA patterns (elevating D) while homogenizing allele frequencies (reducing FST).

6.2 Implications for Empirical Studies

These results offer important practical insights. First, under scenarios with high ILS—such as shallow divergence or large effective population sizes—the D-statistic loses specificity, occasionally yielding non-zero values purely from allele sorting. Conversely, F_{ST} remains relatively stable in such regimes, though its lack of directionality limits its diagnostic utility when used alone.

Second, in high migration regimes, the D-statistic saturates, while F_{ST} continues to reflect

declining population differentiation. This plateauing of D limits its utility for estimating the magnitude of gene flow, even when it successfully detects its presence. These observations align with prior theoretical expectations that D is most effective in intermediate introgression regimes and loses resolution at extremes (Martin et al., 2015; Hibbins and Hahn, 2022).

Third, our joint KDE analysis (Figure 24) shows that D and F_{ST} exhibit a strong inverse correlation in mixed regimes. This relationship provides a useful diagnostic axis: increasing D coupled with decreasing F_{ST} supports introgression, while stable D with high F_{ST} suggests ILS. Importantly, our simulations show that neither metric alone can confidently resolve all cases; instead, combining the two offers a more nuanced and reliable framework for evolutionary inference.

Despite these strengths, real-world applications often include confounding factors not modeled here. For instance, ancestral population structure, linked selection, and recombination rate variation can each distort expectations for both D and FST. Such effects may generate D-statistic skew or FST depressions in the absence of introgression. Thus, care must be taken when extrapolating simulation-derived thresholds to empirical datasets.

Future empirical applications would benefit from using this joint metric framework alongside additional tools like f-branch statistics, local ancestry methods, or machine learning classifiers trained on simulations. Moreover, applying our framework to a real genomic dataset would be a natural next step to validate our simulation-based diagnostic boundaries in the presence of biological complexity.

6.3 Comparison with Alternative Methods

While our study focused on D-statistics and FST, several alternative approaches have been developed to address the ILS-introgression challenge. Branch-length-based methods like QuIBL (Edelman et al., 2019) leverage the expectation that coalescence times differ between ILS and introgression scenarios. Similarly, likelihood-based approaches such as PhyloNet (Wen et al., 2018) can explicitly model reticulate evolution by fitting network topologies to multilocus data.

Our two-metric framework offers several advantages over these alternatives. First, D-statistic and FST calculations are computationally efficient and scalable to whole-genome data, whereas network methods often become prohibitively expensive for large datasets. Second, our approach requires fewer assumptions about underlying evolutionary processes, making it robust to model misspecification. Third, the joint D-FST space provides an intuitive visual framework for interpretation that can be readily adopted by empirical researchers without specialized computational expertise.

However, our approach also has limitations compared to more sophisticated methods. Network-based approaches can simultaneously infer topology, divergence times, and migration parameters, providing a more complete evolutionary picture. Local ancestry methods offer finer-scale resolution of introgressed segments, which may be particularly valuable in identifying adaptive introgression. Future work should focus on integrating our two-metric framework with these complementary approaches to leverage their respective strengths.

6.4 On the Limits of Detection

Section 5.4 highlights important detection thresholds. We show that D-statistic power rises sharply once gene flow exceeds a critical level, particularly under intermediate divergence. In contrast, F_{ST} exhibits higher power only under low migration in ancient divergence settings, where long-term differentiation provides more dynamic range for introgression signals to manifest.

These results underscore the importance of accounting for demographic context when selecting metrics. Introgression detection is fundamentally constrained by the interplay of divergence time, effective population size (N_e), and the timing and strength of gene flow. Metrics like D are prone to false positives under strong ILS, while F_{ST} may miss weak or recent introgression events altogether. Our results suggest that meaningful detection requires situating observed values within a demographic landscape, and integrating multiple complementary metrics to parse ambiguous signals.

7 Practical Guide for Empirical Studies

Based on our comprehensive simulation analyses, we provide the following recommendations for researchers seeking to distinguish introgression from incomplete lineage sorting in empirical datasets.

7.1 Pre-Analysis Considerations

- **Demographic Context Matters:** Before interpreting D-statistics or F_{ST} values, estimate key demographic parameters such as divergence times and effective population sizes (N_e) for your study system, as these strongly influence expected distributions under the null hypothesis of pure ILS.
- **Signal Thresholds Are Context-Dependent:** Rather than applying universal significance thresholds (e.g., $|D| > 0.2$), establish system-specific expectations by simulating pure ILS scenarios matching your estimated demographic parameters. This provides a more accurate baseline for detecting introgression.
- **Genomic Heterogeneity Is Expected:** Substantial window-to-window variation in both D-statistics and F_{ST} will occur even under uniform introgression or pure ILS. This heterogeneity arises from neutral processes alone and should not be over-interpreted as evidence for differential selection.

7.2 Analytical Approach

- **Always Combine Multiple Metrics:** No single statistic reliably distinguishes introgression from ILS across all demographic scenarios. We recommend:
  - Calculate genome-wide D-statistics to detect and quantify directional gene flow.
  - Compute pairwise F_{ST} between all population pairs, not just sister taxa.
  - Analyze the inverse relationship between D and F_{ST} for putative introgression pairs.
- **Sliding Window Analysis:** Implement window-based analyses (40–50 kb windows recommended) to examine spatial patterns, but interpret "genomic islands" cautiously:
  - Windows with strongly positive D-values and reduced F_{ST} provide the strongest evidence for introgression.
  - Verify consistent signal across adjacent windows to reduce false positives from stochastic variation.
  - Compare window distribution patterns to expected neutral heterogeneity from simulations.
- **Consider Temporal Signals:** Recent introgression produces stronger signals than ancient events:
  - D-statistics decay over time due to recombination and drift.
  - F_{ST} shows stronger sensitivity to ongoing gene flow than ancient events.
  - Investigate temporal patterns by examining linkage patterns and haplotype block sizes.

7.3 Interpretation Guidelines

D-Statistic Interpretation:

- $D \approx 0$ with narrow distribution: Consistent with pure ILS or no gene flow.
- $D > 0$ with low F_{ST}: Strong evidence for introgression.
- D with high variance but centered at 0: Caution warranted; may reflect high ILS under small N_e.
- D saturates at high migration rates, limiting quantitative inference about migration magnitude.

F_{ST} Interpretation:

- Reduced F_{ST} between non-sister taxa relative to pure-ILS expectations indicates introgression.
- F_{ST} decreases linearly with increasing gene flow, even when D-statistics plateau.
- F_{ST} is less powerful for detecting ancient introgression unless divergence is deep.
- Check for divergence erosion effects on sister-taxa F_{ST} as a secondary signal.

Joint Distribution Analysis:

- Plot D vs. F_{ST} for all population pairs to identify clusters indicating different evolutionary histories.
- Intermediate divergence scenarios provide the clearest separation between ILS and introgression signals.
- Signals near the decision boundary require additional validation with orthogonal methods.

All simulation code, analysis scripts, and visualization tools used in this study are available at:

`https://github.com/chewyuenrachael/Evolutionary-Genetics-Research/tree/main`

Researchers can adapt these resources to generate null distributions specific to their study systems and implement the joint D-statistic and F_{ST} framework for empirical data analysis.

8 Conclusion

This study presents a systematic evaluation of D-statistic and F_{ST} under diverse evolutionary scenarios, revealing the conditions under which each metric succeeds or fails to distinguish introgression from incomplete lineage sorting (ILS). By simulating pure, mixed, and time-dependent regimes of gene flow and divergence, we show that the interplay of demographic parameters—not the metrics themselves—sets the limits of detectability.

Our key contribution lies in demonstrating that D-statistic and F_{ST} offer complementary, not redundant, insights into complex histories of gene exchange. The D-statistic excels in detecting recent or moderate introgression, while F_{ST} serves as a continuous indicator of long-term differentiation. When used in combination, these metrics provide a robust diagnostic framework for identifying introgression even in the presence of substantial ILS.

Crucially, our findings challenge the overreliance on single-metric heuristics and emphasize the need for joint statistical interpretation within a well-specified demographic model. Rather than seeking universal thresholds, we advocate for context-aware inference, where simulation-informed baselines guide hypothesis testing. This shift is essential for resolving ambiguity in empirical systems where introgression and ILS are not mutually exclusive, but co-occurring and entangled.

As genome-wide data becomes increasingly available across the tree of life, our framework offers a scalable foundation for parsing discordant gene trees in both model and non-model organisms. Future work should expand this simulation-based approach to incorporate selective sweeps, recombination heterogeneity, and more complex network topologies. Additionally, benchmarking these metrics on empirical datasets with known histories will be key to validating their diagnostic power beyond simulation.

In sum, distinguishing introgression from ILS is not a matter of choosing the right statistic, but of triangulating signal with an awareness of demographic nuance. Our study contributes the tools, thresholds, and interpretive strategies to make that distinction clearer and more scientifically rigorous.

Acknowledgments

I am deeply grateful to my advisors, Professor Jon Wilkins and Professor Trisha Stan for their invaluable guidance, encouragement, and thoughtful feedback throughout every stage of this capstone research project. Their mentorship not only shaped the direction of this work but also greatly enriched my understanding of evolutionary genetics.

I would also like to thank Professor Andrew J. Blumberg for his support and insightful discussions during the formative stages of this project, which began during my summer research internship at Columbia University in 2023. His guidance helped lay the conceptual foundations that this study builds upon.

References

R. Abbott, D. Albach, S. Ansell, J. W. Arntzen, S. J. Baird, N. Bierne, others, and D. Zinner. 2013. Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2):229–246. Doi:10.1111/j.1420-9101.2012.02599.x.

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

E. Anderson. 1949. *Introgressive hybridization*. John Wiley Sons.

M. L. Arnold. 1997. *Natural hybridization and evolution*. Oxford University Press.

M. L. Arnold and N. H. Martin. 2009. Adaptation by introgression. *Journal of Biology*, 8(9):82. Doi:10.1186/jbiol176.

P. D. Blischak, J. Chifman, A. D. Wolfe, and L. S. Kubatko. 2018. Hyde: a python package for genome-scale hybridization detection. *Systematic Biology*, 67(5):821–829.

L. Breiman. 2001. Random forests. *Machine Learning*, 45(1):5–32. Doi:10.1023/A:1010933404324.

B. Charlesworth. 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.a025961.

T. E. Cruickshank and M. W. Hahn. 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.

J. H. Degnan and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology Evolution*, 24(6):332–340. Doi:10.1016/j.tree.2009.01.009.

C. A. Deutsch, J. J. Tewksbury, R. B. Huey, K. S. Sheldon, C. K. Ghalambor, D. C. Haak, and P. R. Martin. 2018. Increase in crop losses to insect pests in a warming climate. *Science*, 361(6405):916–919. Doi:10.1126/science.aat3466.

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

D. A. R. Eaton and R. H. Ree. 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/syt038.

N. B. Edelman, P. B. Frandsen, M. Miyagi, B. Clavijo, J. W. Davey, R. B. Dikow, others, and J. Mallet. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science*, 366(6465):594–599. Doi:10.1126/science.aaw2090.

S. V. Edwards, Z. Xi, A. Janke, B. C. Faircloth, J. E. McCormack, T. C. Glenn, others, and A. J. Baker. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, 94:447–462. Doi:10.1016/j.ympev.2015.10.027.

A. Eriksson and A. Manica. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proceedings of the National Academy of Sciences*, 109(35):13956–13960. Doi:10.1073/pnas.1200567109.

T. Flouri, X. Jiao, B. Rannala, and Z. Yang. 2020. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution*, 37(4):1211–1223. Doi:10.1093/molbev/msz296.

M. C. Fontaine, J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, others, and N. J. Besansky. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347(6217):1258524. Doi:10.1126/science.1258524.

E. S. Forsythe, D. B. Sloan, and M. A. Beilstein. 2020. Divergence-based introgression polarization. *Genome Biology and Evolution*, 12(5):463–478. Doi:10.1093/gbe/evaa048.

J. H. Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232. Doi:10.1214/aos/1013203451.

P. R. Grant and B. R. Grant. 2016. Introgression of a transcription factor gene and the evolution of beak morphology in darwin's finches. *Molecular Ecology*, 25(21):5584–5596. Doi:10.1111/mec.13838.

R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, others, and S. Pääbo. 2010. A draft sequence of the neandertal genome. *Science*, 328(5979):710–722. Doi:10.1126/science.1188021.

C. Guo, Z.-H. Guo, and D.-Z. Li. 2019. Phylogenomic analyses reveal intractable evolutionary history of a temperate bamboo genus (poaceae: Bambusoideae). *Plant Diversity*, 41(4):213–219. Doi:10.1016/j.pld.2019.07.005.

B. C. Haller and P. W. Messer. 2019. Slim 3: Forward genetic simulations beyond the wright–fisher model. *Molecular Biology and Evolution*, 36(3):632–637. Doi:10.1093/molbev/msy228.

J. A. Hartigan and P. M. Hartigan. 1985. The dip test of unimodality. *Annals of Statistics*, 13(1):70–84. Doi:10.1214/aos/1176346577.

P. W. Hedrick. 2013. Adaptive introgression in animals: Examples and comparison to new mutation. *Heredity*, 107(3):205–213. Doi:10.1038/hdy.2011.81.

J. Heled and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, 27(3):570–580. Doi:10.1093/molbev/msp274.

J. Hey and R. Nielsen. 2007. Integration within the felsenstein equation for improved Markov chain monte carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, 104(8):2785–2790. Doi:10.1073/pnas.0611164104.

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

K. E. Holsinger and B. S. Weir. 2009. Genetics in geographically structured populations: Defining, estimating and interpreting f_st. *Nature Reviews Genetics*, 10(9) : 639 – −650. $Doi : 10.1038/nrg2611$.

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

E. D. Jarvis, S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, others, and G. Zhang. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346(6215):1320–1331. Doi:10.1126/science.1253451.

S. Joly, P. A. McLenachan, and P. J. Lockhart. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist*, 174(2):E54–E70. Doi:10.1086/600082.

J. Kelleher, A. M. Etheridge, and G. McVean. 2016. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Computational Biology*, 12(5):e1004842. Doi:10.1371/journal.pcbi.1004842.

J. F. C. Kingman. 1982. The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248. Doi:10.1016/0304-4149(82)90011-4.

J. Knox, T. Hess, A. Daccache, and T. Wheeler. 2012. Climate change impacts on crop productivity in africa and south asia. *Environmental Research Letters*, 7(3):034032. Doi:10.1088/1748-9326/7/3/034032.

S. Kong and L. S. Kubatko. 2021. Comparative performance of popular methods for hybrid detection using genomic data. *Systematic Biology*, 70(5):891–907. Doi:10.1093/sysbio/syab030.

J. T. Lovell, A. H. MacQueen, S. Mamidi, J. Bonnette, J. Jenkins, J. D. Napier, others, and J. Schmutz. 2021. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature*, 590(7846):438–444. Doi:10.1038/s41586-020-03127-1.

S. M. Lundberg and S. I. Lee. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30 (NIPS 2017):4765–4774.

W. P. Maddison. 1997. Gene trees in species trees. *Systematic Biology*, 46(3):523–536. Doi:10.1093/sysbio/46.3.523.

M. Malinsky, M. Matschiner, and H. Svardal. 2021. Dsuite - fast d-statistics and related admixture evidence from vcf files. *Molecular Ecology Resources*, 21(2):584–595. Doi:10.1111/1755-0998.13265.

M. Malinsky, H. Svardal, A. M. Tyers, E. A. Miska, M. J. Genner, G. F. Turner, and R. Durbin. 2018. Whole-genome sequences of malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology Evolution*, 2(12):1940–1955. Doi:10.1038/s41559-018-0717-x.

J. Mallet. 2005. Hybridization as an invasion of the genome. *Trends in Ecology Evolution*, 20(5):229–237. Doi:10.1016/j.tree.2005.02.010.

J. Mallet, N. Besansky, and M. W. Hahn. 2016. How reticulated are species? *BioEssays*, 38(2):140–149. Doi:10.1002/bies.201500149.

Y. Mao, C. R. Catacchio, L. W. Hillier, D. Porubsky, R. Li, and E. E. Eichler. 2021. A high-quality bonobo genome refines the analysis of hominid evolution. *Nature*, 594(7861):77–81. Doi:10.1038/s41586-021-03519-x.

S. H. Martin, J. W. Davey, and C. D. Jiggins. 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.

S. H. Martin, J. W. Davey, C. Salazar, and C. D. Jiggins. 2019a. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, 17(2):e2006288. Doi:10.1371/journal.pbio.2006288.

S. H. Martin, M. Möst, W. J. Palmer, C. Salazar, W. O. McMillan, F. M. Jiggins, and C. D. Jiggins. 2019b. Natural selection and genetic diversity in the butterfly heliconius melpomene. *Genetics*, 213(4):1237–1251. Doi:10.1534/genetics.119.302751.

M. W. Nachman and B. A. Payseur. 2012. Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B*, 367(1587):409–421. Doi:10.1098/rstb.2011.0249.

R. Nielsen and M. A. Beaumont. 2009. Statistical inferences in phylogeography. *Molecular Ecology*, 18(6):1034–1047. Doi:10.1111/j.1365-294X.2008.04059.x.

M. A. Noor and S. M. Bennett. 2009. Isolating mechanisms, sequence divergence, and the role of geography in speciation.