Among-species rate variation produces false signals of introgression

Thore Koppetsch[1,*], Milan Malinsky[2,3] and Michael Matschiner[1,*]

[1] *Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, 0318 Oslo, Norway*
[2] *Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland*
[3] *Department of Fish Ecology and Evolution, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland*

**Correspondence to be sent to: Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, 0318 Oslo, Norway; Email: thore.koppetsch@nhm.uio.no, michael.matschiner@nhm.uio.no*

Abstract

The role of interspecific hybridization in the context of diversification dynamics has recently seen increasing attention. Genomic research has now made it abundantly clear that both hybridization and introgression – the exchange of genetic material through hybridization – are far more common than previously thought. Moreover, even highly divergent species were found to hybridize and backcross. These findings raise the question whether commonly used methods for the detection of introgression are applicable to such divergent hybridizing species, given that most of these methods were originally developed for analyses at the level of populations and recently diverged species. In particular, the assumption of constant evolutionary rates, which is implicit in many commonly used approaches, is more likely to be violated as evolutionary divergence increases. To test the limitations of introgression detection methods when being applied to divergent species, we simulated thousands of genomic datasets under a wide range of settings, with varying degrees of among-species rate variation and introgression. Using these simulated datasets, we were able to show that commonly applied statistical methods, including the D-statistic

and tests based on sets of phylogenetic trees, produce false-positive signals of introgression between highly divergent taxa when these have different rates of evolution. These misleading signals are caused by the presence of homoplasies that occur at different rates when rate variation is present. To distinguish between the patterns caused by rate variation and genuine introgression, we developed a new test that is based on the expected clustering of introgressed sites and implemented this test in the program Dsuite.

Key words: hybridization; introgression; rate variation; *D*-statistic; tree topology variation; branch lengths; phylogenetics; phylogenomics.

Recent research has made it abundantly clear that hybridization – the production of viable offspring between distinct species – is far more common than previously thought (Mallet, 2005; Taylor and Larson, 2019). Moreover, hybridization appears to be particularly frequent in rapidly diversifying clades (Meier et al., 2017; Patton et al., 2020; Mitchell and Whitney, 2021) and is increasingly linked to the emergence of hybrid species (Rieseberg et al., 1995; Lamichhaney et al., 2018; Runemark et al., 2018). Hybridization therefore appears to promote diversification in certain situations (Seehausen, 2004; Abbott et al., 2013), contrary to the traditional view in which hybridization is seen as inhibiting speciation (Mayr, 1942).

Besides the previously unexpected amount of hybridization in nature and its potential link to diversification, recent studies have also revealed that even highly divergent species are sometimes still able to hybridize and backcross. Apart from records of interspecific hybrids within a genus, such as crosses between fin whale *Balaenoptera physalus* and blue whale *B. musculus* (Pampoulie et al., 2021), also intergeneric hybrids are known, for example between yellow anaconda *Eunectes notaeus* and common boa *Boa constrictor* (Ernst et al., 2014). Furthermore, even cases of interfamilial hybrids have been reported. Several individuals have been claimed to be hybrids between chicken *Gallus gallus* and other galliform birds (McCarthy, 2006; Gray, 1958). An example of interfamilial

fowl hybrids are crosses between helmeted guineafowl *Numida meleagris* and chicken *G. gallus*, two species from the families Numididae and Phasanidae that diverged about 30 million years ago (Prum et al., 2015). Putative hybrids between those two taxa have been recorded already since the 1930s (Ghigi, 1936) and an F1-hybrid was recently identified genetically (Alfieri et al., 2023).

However, the perhaps most extreme example of hybridization between divergent lineages involves the Russian sturgeon *Acipenser gueldenstaedtii* and the American paddle fish *Polyodon spathula*, that have recently been crossed successfully in the lab (Káldy et al., 2020). The respective families Acipenseridae and Polyodontidae diverged as early as 140–180 million years ago (Betancur-R et al., 2017; Rabosky et al., 2018), a time frame comparable to the divergence of placental mammals and marsupials (Álvarez-Carretero et al., 2022). Even though some of these hybrids are known from captive specimens only, there are various examples of hybridization events between deeply divergent lineages occurring naturally in the wild, e.g., among coral reef fishes (Pomacanthidae) with over 10% mitochondrial divergence (Tea et al., 2020) or between colubrid snakes of different genera (LeClere et al., 2012).

Introgression — the transfer of genetic material between taxa, as the result of hybridization — has also been reported between divergent lineages. For example, evidence of ancient gene exchange was found between the Komodo dragon *Varanus komodoensis* and Australian monitor lizards (Varanidae), two lineages that separated around 13 million years ago (Pavón-Vázquez et al., 2021). Similarly, signatures of introgression were reported among different species of eels (genus *Anguillla*) that diverged up to 10 million years ago (Barth et al., 2020), for Palearctic water snakes (genus *Natrix*) (Schöneberg et al., 2023), and for sea turtles (Cheloniidae) (Vilaça et al., 2021), where the divergence between hybridizing taxa occurred in the latter two cases around 20 million years ago.

In plants, evidence of even more ancient hybridization has been identified in both ferns and angiosperms (Stull et al., 2023). In four different groups of ferns, phylogenetic

conflicts between nuclear and chloroplast genome sequences support relatively recent introgression between genera that diverged around 60 million years ago (Rothfels et al., 2015; Lehtonen, 2018). Finally, the angiosperm order Ericales has been reported to possess genes derived from either Cornales or Gentianidae, two lineages that diverged over 100 million years ago (Stull et al., 2020).

These findings raise the question whether methods to detect introgression from genomic data are still appropriate for such old groups (Hibbins and Hahn, 2022), given that key methods were originally developed for analyses at the level of populations and recently diverged species. One of the most commonly used approaches for introgression detection is the D-statistic, which was first applied to assess genetic exchange between Neanderthals and the ancestors of modern humans (Green et al., 2010). The D-statistic detects introgression through the so-called ABBA-BABA test (Green et al., 2010; Durand et al., 2011), based on an imbalance in the sharing of ancestral ('A') and derived ('B') alleles across the genomes of four populations or species. This test assumes that, in the absence of introgression but presence of incomplete lineage sorting (ILS), two sister species share an equal proportion of derived 'B' alleles with any third species. A statistically significant excess of allele sharing in either direction (an excess of 'ABBA' or 'BABA' sites) is then considered indicative of genetic exchange between non-sister taxa. Although misleading signals can under certain scenarios be created by population structure in ancestral species (Durand et al., 2011; Eriksson and Manica, 2012), the D-statistic is considered to be robust under a wide range of evolutionary scenarios when applied to genome-wide data (Zheng and Janke, 2018).

However, the violation of two assumptions implicit in the use of the D-statistic can cause misleading results when analyzing potential introgression among divergent groups: First, all variable sites are assumed to result from a single substitution, and thus homoplasies — caused by independent substitutions or reversals — are assumed to be absent. By themselves, homoplasies are not expected to produce a false signal of

introgression, because they should increase the numbers of 'ABBA' and 'BABA' sites equally. But when a second assumption — that of uniform substitution rates for all species — is violated in combination with the occurrence of homoplasies, this would lead to unequal numbers of 'ABBA' and 'BABA' sites and, potentially, a D-statistic falsely supporting introgression (Pease and Hahn, 2015; Amos, 2020).

Both violations, homoplasies and substitution-rate variation, are more likely to occur in older groups of species. Homoplasies require that sites are substituted on two different branches of the tree, which occurs more often when these branches are longer. Substitution-rate variation, on the other hand, is influenced by factors such as metabolic rate, generation time, longevity, or temperature, that are all expected to be similar among closely related species but may vary with increasing phylogenetic distance (Sayres et al., 2011; Bromham, 2020; Hua and Bromham, 2017; Ivan et al., 2022; Hua et al., 2015). A misleading effect of substitution-rate variation on the D-statistic, generating false-positive signals of introgression, has been suspected repeatedly (Pease and Hahn, 2015; Zheng and Janke, 2018; Hibbins and Hahn, 2022), and was recently supported by simulations under the birth-death-hybridization process (Justison et al., 2022; Frankel and Ané, 2023).

To avoid the effects of rate variation on introgression detection, tree-based tests have been suggested (Frankel and Ané, 2023). A tree-based equivalent of the D-statistic has been used in various studies (Barth et al., 2020; Ronco et al., 2021). In this approach, rooted phylogenetic trees are first built for a large number of loci (regions with hundreds to thousands of basepairs) across the genome, and the inferred set of trees is then analyzed for topological asymmetry in three-species subsets just like site patterns are in the ABBA-BABA test. Thus, the most frequent tree topology for a set of three species is assumed to represent their species tree, and the frequencies of the second- and third-most frequent topologies are compared to each other. A significant difference in these frequencies is then interpreted as evidence of introgression. The test statistic has been named D_{tree} in Ronco et al. (2021) (who were unaware that a non-normalized version of this statistic had

already been called Δ by Huson et al. 2005). Frequencies of tree topologies have also been used to infer introgression in other studies (Schumer et al., 2016; Gante et al., 2016; Figueiró et al., 2017; Suvorov et al., 2022). One might expect that this tree-based alternative of the D-statistic is more robust to homoplasies, given that the occurrence of one or few homoplasies per locus should not have an effect on the tree topology (Hibbins and Hahn, 2022; Frankel and Ané, 2023). On the other hand, homoplasies in combination with rate variation can lead to long-branch attraction (Felsenstein, 1978), which might bias tree-topology frequencies even if its effect on each individual tree is weak. The degree to which rate variation and homoplasies affect tree-topology based introgression tests has not been examined so far.

Here, we use simulations to test the robustness of introgression detection methods to the combined effects of homoplasies and rate variation, as expected to occur in divergent groups of species. We simulate genomic datasets under a wide range of settings, including varying population sizes, divergence times, recombination rates, mutation rates, introgression rates, and degrees of among-species rate variation. Besides the D-statistic and its tree-based equivalent D_{tree}, we apply three further tree-based methods to detect introgression in complementary ways: the phylogenetic network approach implemented in SNaQ (Solís-Lemus et al., 2017), the approach based on branch-length distributions implemented in QuIBL (Edelman et al., 2019), and a method based on divergence-time distributions in time-calibrated phylogenies. The latter method was presented by Meyer, Matschiner, and Salzburger (2017), and will henceforth be called "MMS17 method". We hypothesized that all of these methods could produce false signals of introgression when among-species rate variation is present, and that these signals would become stronger with increasing divergence times, mutation rates, and degree of rate variation. Our results confirm that most tested methods are affected by rate variation, even though they can perform reliably for certain combinations of parameters. To address this vulnerability of introgression detection methods and to distinguish between true signals of introgression

and the false signals resulting from rate variation, we developed a new test based on the distribution of 'ABBA' sites on the genome, and we implemented this test into the popular introgression analysis software Dsuite (Malinsky et al., 2021). We assess the performance of this test using our simulated datasets, and confirm its suitability across a broad range of parameters.

Materials and Methods

Simulations

To test the performance of commonly applied introgression-detection methods, genomic data were simulated under diverse scenarios. All simulations were conducted in the Python version 3.8.6 environment using the program msprime v.1.0 (Baumdicker et al., 2021). A four-taxon phylogeny was defined for the species P1, P2, P3, and P4, in which P1 and P2 were sister species and P4 was the outgroup to all others. The divergence of species P1 and P2, $t_{\mathrm{P1,P2}}$, was set to occur either 10, 20, or 30 million generations ago, with species P3 and P4 in each case set to branch off 10 and 20 million generations earlier, respectively. Thus, the most recent common ancestor of the four species dated to between 30 and 50 million generations ago (Fig. 1). All simulated species had identical and constant effective population sizes (N_{e}), set to either $N_{\mathrm{e}} = 10^4$ or $N_{\mathrm{e}} = 10^5$ in separate simulations. An effective population size of $N_{\mathrm{e}} = 10^6$ was used in exploratory simulations, but as these simulations were too computationally demanding and their results did not seem to differ from those based on smaller population sizes, final simulations were based on the two smaller population sizes. We conducted one set of simulations that did not include any genetic exchange between species while other simulations included introgression between species P2 and P3. In these cases, P2 and P3 exchanged migrants with a rate m of either $m = 10^{-9}$, $m = 10^{-8}$, $m = 10^{-7}$, or $m = 10^{-6}$ per individual per generation, which is equivalent to the exchange of a migrant on average every $10^2 - 10^5$ generations when

$N_e = 10^4$ or every $10 - 10^4$ generations when $N_e = 10^5$. In all simulations, migration between P2 and P3 occurred for the same period of time, beginning with the divergence of P1 and P2 and ending 2.5 million generations later (Fig. 1).

Based on this model of divergence and introgression, we simulated the evolution of the genomes of the four species, modeling these as a single chromosome with a length of 20 million basepairs (Mbp). The recombination rate r of this chromosome was set to $r = 10^{-8}$ and the mutation rate μ was set to either $\mu = 10^{-9}$ or $\mu = 2 \times 10^{-9}$ in separate simulations. Mutations were simulated under the Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al., 1985) with a transition-transversion rate ratio $\kappa = 2$. Finally, we implemented among-species variation in the mutation rate to model a decreased, unchanged, or increased rate in species P2, with the rate change taking place immediately after its divergence from P1. Because msprime does not allow mutation-rate variation, we used a work-around with the same outcome, extending or shortening the branch leading to P2 with a scale factor s. We repeated the simulations using each of the five scale factors $s = 0.25$ (very slow P2), $s = 0.5$ (slow P2), $s = 1$ (unchanged P2), $s = 2$ (fast P2), and $s = 4$ (very fast P2) to model varying degrees of among-species rate variation. For each of the four simulated species, we sampled ten haploid chromosomes to form diploid genomes for five individuals per species. To implement the above-mentioned work-around for $s = 0.25$ and $s = 0.5$, we sampled individuals from species P2 at a time point in the past so that the length of its branch was effectively divided by 2 or 4. For $s = 1$, all samples were taken at the present. With scale factors $s = 2$ and $s = 4$, P2 was sampled at the present, but all divergences were shifted into the past by the amount of generations by which P2 was extended, and P1, P3, and P4 were instead sampled in the past. In summary, we performed simulations with all possible combination of $t_{\mathrm{P1,P2}} \in \{1 \times 10^7, 2 \times 10^7, 3 \times 10^7\}$, $N_e \in \{10^4, 10^5\}$, $m \in \{0, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}\}$, $r = 10^{-8}$, $\mu \in \{1 \times 10^{-9}, 2 \times 10^{-9}\}$, and $s \in \{0.25, 0.5, 1, 2, 4\}$; a total of 300 parameter combinations. For each of these, we performed ten replicate simulations, recording the resulting genomic datasets in 3,000 files

in the variant call format (VCF).

The range of parameters used in our simulations was selected to be comparable with some of the study systems to which we referred in the Introduction as cases of highly divergent hybridizing lineages. In terms of numbers of generations, recombination rate, and mutation rate, our simulations are comparable to the interfamilial hybridization events between helmeted guineafowl *N. meleagris* and chicken *G. gallus* (Alfieri et al., 2023). The divergence of these two lineages can be placed around 30 million generations ago, assuming a generation time of 1 year (Mariadassou et al., 2021; Sanchez-Donoso et al., 2014) and a divergence about 30 million years ago (Prum et al., 2015). The mutation rate of galliforms is comparable to those chosen for our simulations ($\mu \in \{10^{-9}, 2 \times 10^{-9}\}$), as it was estimated around 1.91×10^{-9} per site and year (Nam et al., 2010).

Patterson's D-statistic

Patterson's D-statistic (Green et al., 2010) measures signals of introgression in a species trio P1, P2, and P3 by counting the numbers of sites at which these species share alleles. Denoting ancestral alleles as 'A' and derived alleles as 'B', 'ABBA' sites are those at which P2 and P3 share the derived allele, while P1 and P3 share the derived allele at 'BABA' sites. By definition, the allele carried by the outgroup P4 is considered the ancestral allele 'A'. The D-statistic is then calculated as the difference between the number of 'ABBA' sites C_{ABBA} and that of 'BABA' sites C_{BABA}, normalized by the sum of these two numbers:

$$D = \frac{C_{\text{ABBA}} - C_{\text{BABA}}}{C_{\text{ABBA}} + C_{\text{BABA}}} \tag{1}$$

In the absence of introgression, C_{ABBA} and C_{BABA} are expected to be equal, leading to an expectation of $D = 0$. However, this expectation is based on several assumptions, including that of equal rates for species P1 and P2, which we violated in part of our simulations. We therefore expected to find signals of introgression in these simulated

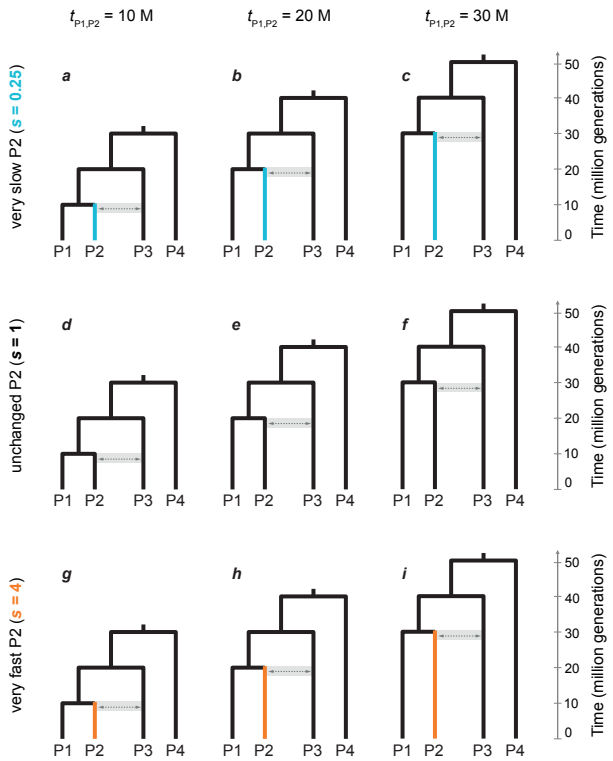

Fig. 1 Four-taxon phylogenies used in simulations. Divergence times of species P1 and P2 ($t_{\mathrm{P1,P2}}$) are 10 (**a**,**d**,**g**), 20 (**b**,**e**,**h**), and 30 million generations (**c**,**f**,**i**) in the past, with P3 and P4 branching off 10 and 20 million generations earlier, respectively. Species P2 evolved with a mutation rate that was either slower ($s = 0.25$; **a**–**c**), unchanged ($s = 1$; **d**–**f**), or faster ($s = 4$; **g**–**i**) than the mutation rate of all other species (the intermediate values $s = 0.5$ and $s = 2$ were tested but are not shown here). In simulations that included introgression, this occurred symmetrically between P2 and P3, beginning with the divergence of P1 and P2, and continuing for 2.5 million generations (gray arrows).

datasets even when it was absent.

We calculated the D-statistic for each of the 3,000 simulated genomic datasets with the program Dsuite v.r50 (Malinsky et al., 2021), using the program's "Dtrios" function. We provided the correct species tree topology to Dsuite with option "-t" so that a positive D would always support introgression between P2 and P3 while a negative D would support introgression between P1 and P3. By using the Dsuite implementation of the D-statistic, we were able to account for within-species variation in the calculation of C_{ABBA} and C_{BABA}. When multiple individuals are sampled per species, Dsuite calculates C_{ABBA} and C_{BABA} based on the frequencies of the ancestral and derived alleles within the species (see Equations 1 and 2 in Malinsky et al., 2021). With the frequency of the derived

allele 'B' at site i in the genome of species j denoted as $f_{\mathrm{B},j,i}$ and a total number of sites n,

$$C_{\mathrm{ABBA}} = \sum_{i=1}^{n} (1 - f_{\mathrm{B,P1},i}) \times f_{\mathrm{B,P2},i} \times f_{\mathrm{B,P3},i} \tag{2a}$$

$$C_{\mathrm{BABA}} = \sum_{i=1}^{n} f_{\mathrm{B,P1},i} \times (1 - f_{\mathrm{B,P2},i}) \times f_{\mathrm{B,P3},i} \tag{2b}$$

For both parts of Equation 2, Dsuite defines the derived allele as the one occurring at lower frequency in the outgroup P4 and multi-allelic sites are ignored. The significance of D was assessed with a block jackknife test, based on 20 equally sized subsets of the genomic data.

Tree-based Introgression Detection Methods

Besides the D-statistic, we applied four tree-based introgression-detection methods to all datasets simulated with a population size of $N_{\mathrm{e}} = 10^5$ and a mutation rate of $\mu = 2 \times 10^{-9}$, a total of 750 datasets. To infer local trees as input for these methods, variants were extracted from equally spaced windows across the simulated chromosome. We separately extracted 5,000 windows of 200 bp, 2,000 windows of 500 bp, and 1,000 windows of 1,000 bp from each of the 750 datasets. These window sizes were chosen as a compromise between too little phylogenetic information in shorter windows and the occurrence of within-window recombination in larger windows, which could bias any phylogenetic inference (Bryant and Hahn, 2020). With these selected window sizes and numbers, only 1 Mbp out of the 20 Mbp of the simulated chromosomes were used for phylogenetic analyses. Additionally, per species and variable site, only the first allele of the first individual was extracted. All invariable sites within windows were replaced with randomly selected nucleotides A, C, G, and T, thus forming a sequence alignment for each window. By using only one allele of one individual per species – instead of both alleles of the five simulated individuals – we again reduced the amount of data by a factor of ten. Consequently, for any given simulation, only 0.5% of the data used to calculate the

D-statistic were also used for tree-based analyses. This data reduction was required due to the computational demands of our phylogenetic analyses: Because we used 750 genomic datasets in total and extracted 8,000 (5,000 + 2,000 + 1,000) windows from each of these, 6 million alignments were produced. As each alignment was used for phylogenetic analyses with both maximum-likelihood and Bayesian inference (see below), a total of 12 million such analyses were required.

D_{tree} — Conceptually similar to Patterson's D-statistic, D_{tree} aims to detect introgression by comparing the counts of alternative rooted tree topologies for a given species trio, in a large set of local trees sampled across the genome. For any such trio P1, P2, and P3, three different rooted tree topologies can be found: One in which P1 and P2 are sister species, one in which P1 appears next to P3, and one in which P2 and P3 are sisters. Like for the D-statistic, the assumptions of no introgression and no among-species rate variation predict that, if the most frequent of these tree topologies represents the species-tree, the other two should occur in equal frequencies due to ILS. Any significant difference in the frequencies of the latter two topologies, assessed for example with a one-sided binomial test, can therefore be seen as support for introgression.

In its unconstrained version (Ronco et al., 2021), D_{tree} is calculated from the counts of the second- and third-most frequent rooted topologies for the species trio, C_{2nd} and C_{3rd}, as

$$D_{\text{tree}} = \frac{C_{\text{2nd}} - C_{\text{3rd}}}{C_{\text{2nd}} + C_{\text{3rd}}} \tag{3}$$

However, the use of this unconstrained version of D_{tree} may underestimate high levels of introgression when the most frequent tree topology of the three species does not reflect the species tree (due to very high levels of genetic exchange and/or very short internal branches). Therefore, we here applied a constrained version of D_{tree} to test explicitly for introgression between P2 and P3:

$$D_{\text{tree}} = \frac{C_{\text{P2,P3}} - C_{\text{P1,P3}}}{C_{\text{P2,P3}} + C_{\text{P1,P3}}}, \tag{4}$$

where $C_{\text{P2,P3}}$ is the count of trees in which P2 and P3 are sisters, and $C_{\text{P1,P3}}$ is the count of trees that place P1 and P3 next to each other.

To generate these counts, we inferred maximum-likelihood phylogenies from each window alignment of the 750 simulated genomic datasets using IQ-TREE v.2.1.2 (Minh et al., 2020), specifying the HKY substitution model (Hasegawa et al., 1985) and P4 as outgroup. The resulting tree sets were processed as in Ronco et al. (2021) by removing branches shorter than 0.001 substitutions per site. As these were considered uninformative, the two nodes connected by them were converted into polytomies using the function di2multi of the R package ape (v.4.0) (Paradis et al., 2004). Following this step, we obtained $C_{\text{P1,P3}}$ and $C_{\text{P2,P3}}$ by counting how often P1 and P3, or P2 and P3, respectively, were sister species in a set of trees, ignoring polytomic relationships. We did so separately for the sets of trees corresponding to each simulated genomic dataset and window size, by applying the Ruby script analyze_tree_asymmetry.rb (Ronco et al., 2021). From the counts $C_{\text{P1,P3}}$ and $C_{\text{P2,P3}}$, D_{tree} was calculated using Equation 4. Finally, a one-sided binomial test was used to identify whether $C_{\text{P2,P3}}$ was significantly larger than $C_{\text{P2,P3}}$ and thus supporting introgression between P2 and P3.

SNaQ — The SNaQ (Species Networks applying Quartets) method, implemented in PhyloNetworks (Solís-Lemus and Ané, 2016; Solís-Lemus et al., 2017), is representative of a class of methods based on the multi-species coalescent model with hybridization (Meng and Kubatko, 2009). This class also includes approaches implemented in PhyloNet (Yu et al., 2014; Than et al., 2008; Yu and Nakhleh, 2015) or HyDe (Blischak et al., 2018). From a set of local trees, SNaQ quantifies concordance factors for unrooted species quartets (either all possible quartets or a random sample) and calculates the likelihood for each of these quartets under the multi-species coalescent model with hybridization. By

multiplying these likelihoods over all quartets, SNaQ derives the pseudolikelihood for a given species network. A heuristic search then allows SNaQ to estimate the network that optimizes the pseudolikelihood for a given maximum number of hybridization events. Thus, by repeating the SNaQ analysis with a maximum number of 0 and 1 such events, support for hybridization can be evaluated from the difference of the resulting pseudolikelihoods.

The multi-species coalescent model with hybridization considers hybridization events on the species level that instantaneously copy part of the genome from one species to another. Thus, this model is violated by our simulations in a way in which it may also be violated by most empirical cases of hybridizing species: Our simulations model hybridization between individuals that over long time scales (2.5 million generations) of ongoing introgression and subsequent drift, recombination, and occasional fixation has a gradual effect on the genomes of the recipient species. We expected that – barring other model violations – fitting such a period of hybridization and introgression to the multi-species coalescent model with hybridization would lead to the inference of a single hybridization event between species. Nevertheless, we expected that the support for this inferred hybridization event would correlate with the truth – i.e., the presence and the rate of introgression used in our simulations. We quantified this support as the difference in the Akaike information criterion (dAIC) (Akaike, 1974) for models that did or did not include a hybridization event, and considered dAIC values above 10 as significant. It has been pointed out that criteria like the Akaike information criterion are not generally suitable for the pseudolikelihoods estimated by SNaQ (Hibbins and Hahn, 2022). However this criterion is applicable in our case, because with no more than four species (i.e., a single quartet), SNaQ estimates the actual likelihood and not the pseudolikelihood (Solís-Lemus and Ané, 2016). We calculated the dAIC supporting introgression separately for each simulated genomic dataset and each of the three window sizes (200, 500, and 1,000 bp), based on the maximum-likelihood phylogenies inferred for these windows with IQ-TREE, with short branches replaced by polytomies as described above. We used PhyloNetworks

v.0.14.2 for these analyses, providing the correct species tree as starting tree and specifying P4 as the outgroup when calling SNaQ.

QuIBL — QuIBL (Quantifying Introgression via Branch Lengths) is an approach to estimate proportions of introgressed loci based on the distribution of branch lengths in a species trio (Edelman et al., 2019). By using branch lengths as a source of information, QuIBL is complementary to SNaQ, as the latter is informed exclusively by the topologies of a set of local trees. All species trios in a given dataset are used by QuIBL and examined independently of each other. Per trio, QuIBL sorts the set of local trees into three subsets, one for each of the three possible rooted topologies of the triplet. For each of the three subsets, QuIBL then determines the distribution of the lengths of the internal branch (in numbers of substitutions per site), across all of the local trees within the subset. Applying an expectation-maximization algorithm in combination with the Bayesian information criterion (BIC) (Schwarz, 1978), it uses the shape of these distributions to determine whether they result from a single process (i.e., ILS) or additionally from a second process. This second process is interpreted either as lineage sorting within a common ancestor or introgression, depending on the relationships of the three species in a predetermined species tree. In the latter case, the number of local trees in the respective subset, multiplied by the proportion of them assigned to introgression rather than ILS, quantifies the overall introgression proportion. Like the multi-species coalescent with hybridization implemented in SNaQ, the assumptions behind QuIBL also include a single pulse of hybridization instead of continuous introgression over a period of time (Edelman et al., 2019).

We applied QuIBL to the sets of local trees generated using IQ-TREE for the 750 genomic datasets and each of the three window sizes, again with short branches replaced by polytomies. As the main version of QuIBL ignores entire input trees when they contain one or more polytomies, we generated and used a fork of the code (https://github.com/miriammiyagi/QuIBL/pull/16) that ignores those trees only for the species trios involved in the polytomy, but not for others. The QuIBL parameters included

a likelihood precision treshold ("likelihoodthresh") of 0.01, a limit of 50 steps for the expectation-maximization algorithm ("numsteps"), and a scale factor of 0.5 to reduce the step size when an ascent step fails ("gradascentscalar"), as recommended by the authors. We further specified P4 as the outgroup to the trio formed by P1, P2, and P3. The results of QuIBL analyses were processed with the quiblR library (https://github.com/nbedelman/quiblR). Following Edelman et al. (2019), we considered support for introgression significant when the difference in BIC values (dBIC) was greater than 10.

MMS17 method — A third class of tree-based introgression detection methods uses distributions of divergence times in a set of ultrametric, time-calibrated local trees (Marcussen et al., 2014; Fontaine et al., 2015; Meyer et al., 2017). Of this class, we here apply the method developed by Meyer, Matschiner, and Salzburger (2017) ("MMS17 method"). This method compares the mean divergence times for all three possible pairs of species within a species trio, repeating this comparison for all possible species trios of a given dataset. For the species trio P1, P2, and P3, the mean ages of the most recent common ancestor (MRCA) of the pairs P1–P2, P1–P3, and P2–P3 are calculated over all local trees. If P1 and P2 are sister species and no introgression occurred with P3, the P1–P3 and P2–P3 mean MRCA age estimates are expected to be similar in the absence of introgression. In contrast, any introgression occurring between non-sister species should reduce one of these two mean MRCA ages. The difference between these pairwise mean MRCA (dMRCA) ages is therefore informative about past introgression within the species trio – the larger dMRCA, the stronger the evidence for introgression (Meyer et al., 2017). On the other hand, the MMS17 method does not include a formal statistical test allowing one to reject the null hypothesis of no introgression.

We used the Bayesian program BEAST2 v.2.6.4 (Bouckaert et al., 2019) to infer sets of time-calibrated local trees using three alignment window sizes (200, 500, and 1,000 bp) for each simulated genomic dataset. Per alignment, an input file for BEAST2 was

produced with the babette library (Bilderbeek and Etienne, 2018), specifying the birth-death tree process as a tree prior (Gernhard, 2008) and the HKY substitution model (Hasegawa et al., 1985). Each tree was time-calibrated with a strict-clock model and an age constraint on the root. This constraint was defined as a log-normal prior distribution with a mean according to the true root age used in the simulation of the respective dataset (assuming a generation time of one year) and a narrow standard deviation of 0.001. Each BEAST2 analysis was performed with 5 million Markov-chain Monte Carlo iterations. Upon completion of each BEAST2 analysis, the resulting posterior tree distribution was summarized in the form of a maximum-clade-credibility tree with TreeAnnotator v.2.6.4 (Heled and Bouckaert, 2013). For each of the 750 genomic datasets and each window size, we used all produced summary trees jointly for the MMS17 method as described above.

'ABBA'-Site Clustering

As our application of the D-statistic and tree-based methods revealed that most of them are sensitive to among-species rate variation (see Results), we here propose a new test to discriminate between spurious and genuine signals of introgression. Because rate variation can produce false signals of introgression through unbalanced addition of homoplasies, this test aims to distinguish between homoplasious and introgressed sites, exploiting the fact that introgressed haplotypes typically do not get broken down into individual variable sites before fixation. Therefore, these unbroken introgressed haplotypes introduce clusters of multiple linked variable sites. On the other hand, homoplasies are expected to occur individually one by one. Our new test therefore examines whether the 'ABBA' sites that are used for the D-statistic cluster along the chromosome – which would support introgression – or whether they are distributed homogeneously – which would indicate they are homoplasies.

In our application of this test, we identify "strong ABBA sites" as those for which

$$(1 - f_{\mathrm{B,P1}})\ f_{\mathrm{B,P2}}\ f_{\mathrm{B,P3}}\ (1 - f_{\mathrm{B,P4}}) + f_{\mathrm{B,P1}}\ (1 - f_{\mathrm{B,P2}})\ (1 - f_{\mathrm{B,P3}})\ f_{\mathrm{B,P4}} > 0.5, \qquad (4)$$

where $f_{\mathrm{B,P1}}$, $f_{\mathrm{B,P2}}$, $f_{\mathrm{B,P3}}$, and $f_{\mathrm{B,P4}}$ are the frequencies of the derived allele 'B' in populations P1, P2, P3, and the outgroup (see Equation 4a in Malinsky et al. 2021). Let $\vec{g}$ be a vector of all polymorphic sites in a dataset. We then define another vector $\vec{i}$, where we record the indices of of "strong ABBA sites" sites within $\vec{g}$. We divide the values in $\vec{i}$ by the length of $\vec{g}$ (the number of polymorphic sites in the dataset), obtaining a normalized vector $\vec{i}_n$ on the interval [0,1). To test for clustering of "strong ABBA sites" we compare this normalized $\vec{i}_n$ to the standard uniform distribution using a one-sample Kolmogorov-Smirnov test (Kolmogorov, 1933; Simard and L'Ecuyer, 2011). A significant test statistic supports clustering of "strong ABBA sites" on scaffolds and thus the occurrence of introgression. We implemented this test in the software Dsuite, where it can be called with the function "--KS-test-for-homoplasy" of the Dtrio module (Malinsky et al., 2021). To test the power and reliability of this new test, we applied it to simulated genomic datasets.

Results

Simulations

The numbers of variable sites in simulated datasets ranged from 1.45–8.20 million, primarily depending on the mutation rate μ and the divergence time $t_{\mathrm{P1,P2}}$ (Table 1). Between 34,000 and 1.3 million of these sites were multi-allelic. The alignments of lengths 200, 500, and 1,000 bp had mean numbers of variables sites between 14.4 and 398.8 (Table 1).

The mean lengths of chromosomal regions unbroken by recombination – termed "c-genes" by Doyle (1995) – were between 18 and 20 bp; however, the lengths of chromosomal regions sharing the same species topology ("single-topology tracts") was far longer. Without introgression ($m = 0$), all but two of the simulated chromosomes had the same topology from beginning to end. The two exceptions were datasets simulated with

Table 1. Numbers of variable sites in genomic datasets and alignments.

N_e	μ	$t_{P1,P2}$	Variable sites ($\times 10^6$)	Multi-allelic sites ($\times 10^3$)	Variable sites in alignments 200 bp	500 bp	1,000 bp
10^4	1×10^{-9}	1×10^7	1.45 – 2.27	34 – 85	14.4 – 22.6	35.9 – 56.6	72.1 – 113.0
10^4	1×10^{-9}	2×10^7	2.04 – 3.47	68 – 205	20.3 – 34.4	50.8 – 86.1	101.5 – 172.1
10^4	1×10^{-9}	3×10^7	2.62 – 4.58	114 – 368	25.9 – 45.4	64.6 – 113.3	129.4 – 226.8
10^4	2×10^{-9}	1×10^7	2.80 – 4.28	132 – 318	27.6 – 42.4	68.9 – 106.1	137.7 – 211.8
10^4	2×10^{-9}	2×10^7	3.88 – 6.33	259 – 738	37.6 – 62.3	95.4 – 155.8	190.3 – 311.4
10^4	2×10^{-9}	3×10^7	4.90 – 8.12	425 – 1,282	48.2 – 79.4	120.6 – 198.3	241.3 – 396.3
10^5	1×10^{-9}	1×10^7	1.53 – 2.33	38 – 90	14.8 – 22.9	36.9 – 57.1	73.8 – 114.4
10^5	1×10^{-9}	2×10^7	2.12 – 3.52	73 – 211	20.5 – 34.7	51.6 – 86.8	102.4 – 173.2
10^5	1×10^{-9}	3×10^7	2.69 – 4.64	120 – 378	26.3 – 45.6	65.5 – 114.1	113.0 – 227.8
10^5	2×10^{-9}	1×10^7	2.94 – 4.39	146 – 336	28.2 – 42.9	70.3 – 107.0	141.4 – 214.4
10^5	2×10^{-9}	2×10^7	4.01 – 6.43	278 – 763	38.8 – 62.7	97.1 – 156.7	193.8 – 313.5
10^5	2×10^{-9}	3×10^7	5.02 – 8.20	447 – 1,312	48.8 – 79.8	121.7 – 199.2	243.5 – 398.3

the population size $N_e = 10^5$ that included three and two single-topology tracts, respectively. Therefore, we conclude that the simulated data based on the divergence model in Fig. 1 had very little or no ILS. On the other hand, with introgression rates increasing from $m = 10^{-9}$ to $m = 10^{-7}$, the mean lengths of single-topology tracts decreased from a minimum of 64,516 bp (and a maximum of the chromosome length) to a range between 7,132 and 10,010 bp with a population size of $N_e = 10^4$, and from 22,026–217,391 bp to 422–789 bp with $N_e = 10^5$. However, with the highest simulated rate of introgression $m = 10^{-6}$, the lengths of single-topology tracts mostly increased again, to 2,246–95,238 bp with $N_e = 10^4$ and to 86–1,277 bp with $N_e = 10^5$ (Supplementary Table 1). The reason for this was a dominance of regions with introgression in these chromosomes, causing them to form single-topology tracts.

While only 0–3.7% of the chromosome were affected by introgression with $m = 10^{-9}$, these proportions grew to 1.4–10.0%, 30.6–46.5%, and 95.7–99.5% with $m = 10^{-8}$, 10^{-7}, and 10^{-6}, respectively. Because of these extreme differences, we focus on the scenarios of weak introgression ($m = 10^{-8}$) and strong introgression ($m = 10^{-7}$), besides the scenario without introgression ($m = 0$), in the remainder of the Results section. We present all results, including those obtained with very weak ($m = 10^{-9}$) and very strong introgression ($m = 10^{-6}$) in the Supplementary Material.

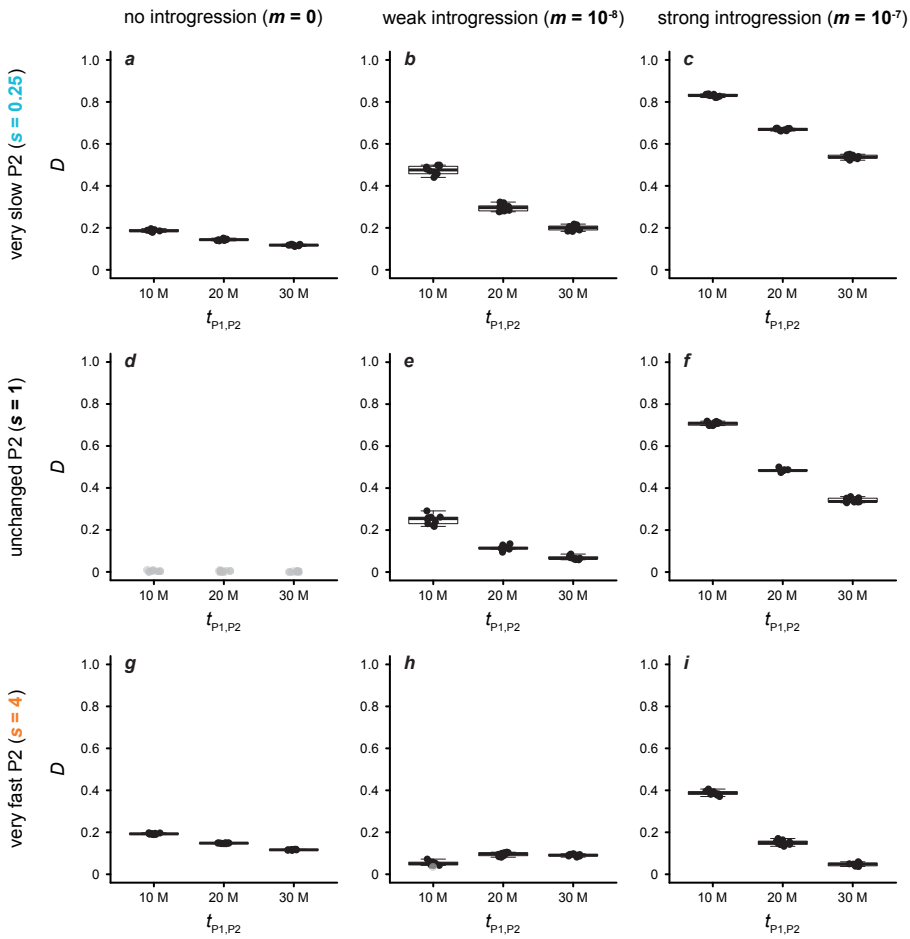

Fig. 2 Patterson's D-statistic for datasets simulated with a population size $N_e = 10^5$, a mutation rate $\mu = 2 \times 10^{-9}$, either no ($m = 0$; **a**,**d**,**g**), weak ($m = 10^{-8}$; **b**,**e**,**h**), or strong ($m = 10^{-7}$; **c**,**f**,**i**) introgression, and either a very slow ($s = 0.25$; **a**–**c**), unchanged ($s = 1$; **d**–**f**), or very fast ($s = 4$; **g**–**i**) rate of branch P2. All results obtained with other settings are shown in Supplementary Figures 1–4. Per divergence time $t_{\mathrm{P1,P2}} \in \{1 \times 10^7, 2 \times 10^7, 3 \times 10^7\}$, the D-statistic is shown for ten replicate simulations. Circles in black indicate significant results ($p < 0.05$), and only these are summarized with box plots.

Patterson's D-statistic

As expected, Patterson's D-statistic reliably indicated introgression when rate variation was absent ($s = 1$). With a population size $N_e = 10^5$ and mutation rate $\mu = 2 \times 10^{-9}$ (Fig. 2), the D-statistic was below 0.01 and insignificant ($p \geqslant 0.05$) for all replicate datasets when introgression was absent ($m = 0$), regardless of the divergence time $t_{\mathrm{P1,P2}}$ (Fig. 2d). With weak ($m = 10^{-8}$) or strong introgression ($m = 10^{-7}$), on the other hand, the D-statistic was in the ranges of 0.06–0.29 and 0.33–0.72, respectively, and in all

cases highly significant ($p < 10^{-11}$) (Fig. 2e,f). The D-statistic was lower (0–0.05) and in some cases not statistically significant in settings with very weak ($m = 10^{-9}$) introgression, and higher (0.59–0.87) and always significant ($p < 10^{-16}$) in settings with very strong ($m = 10^{-6}$) introgression (Supplementary Figure 2). In all cases the D-statistic decreased with increasing age of the phylogeny (i.e. with $t_{\mathrm{P1,P2}}$). This decrease was caused by homoplasies and reversals accumulating on the longer branches of the older phylogenies, augmenting both C_{ABBA} and C_{BABA} (Supplementary Note 1). Simulations with a lower population size ($N_{\mathrm{e}} = 10^4$) or a lower mutation rate $\mu = 1 \times 10^{-9}$ produced the same patterns (Supplementary Figure 1, Supplementary Figures 3,4).

In contrast to the results obtained without rate variation, the D-statistic was not a reliable indicator of introgression when rate variation was present. While the D-statistic was significant for nearly all datasets simulated with introgression (Supplementary Figures 1–4), it was also significant for all datasets simulated without introgression ($m = 0$) whenever rate variation was present. In these cases, the D-statistic ranged from 0.05 to 0.21 ($p < 10^{-9}$) (Fig. 2a,g, Supplementary Figures 1–4).

Like the decrease of the D-statistic with increasing age of the phylogeny, these false-positive signals were caused by homoplasies and reversals. This pattern can be illustrated focusing on the results obtained with a very fast rate of the P2 branch ($s = 4$) on the youngest phylogeny ($t_{\mathrm{P1,P2}} = 10$ million generations), shown in Fig. 2g. The high D-statistic of 0.19–0.20 for these simulated datasets resulted from a C_{ABBA} in the range of 42,679–43,257 and a C_{BABA} around 28,944–29,244. Perhaps contrary to expectations, this D-statistic does not support introgression between P2 and P3, but instead between P1 and P3 (Dsuite automatically rotates P1 and P2 so that $D \geqslant 0$). A detailed analysis of one replicate simulation output explains this result: As expected, the faster rate of evolution of P2 led to more homoplasies shared between P2 and P3 (11,583; considering only bi-allelic sites) than between P1 and P3 (3,314). However, as the outgroup P4 had a longer branch than P3, this difference was more than compensated for by a greater number of

homoplasies between P2 and P4 (16,856) compared to P1 and P4 (4,690). Additionally, far more reversals of substitutions shared by P1, P2, and P3 occurred on the branch leading to P2 (5,990) than on that leading to P1 (1,697), further increasing allele sharing between P1 and P3. The remaining difference between C_{ABBA} and C_{BABA} may be explained by multi-allelic sites, of which there were 21,123.

While the D-statistic is similarly high, in the range of 0.18–0.20, in datasets produced with a very slow rate of the P2 branch ($s = 0.25$), it supports introgression between P2 and P3 in these instances (Fig. 2a). As in the cases with an increased P2 rate, the imbalance between a C_{ABBA} of 11,565–12098 and a C_{BABA} of 7,877–8,211 is explained by homoplasies and reversals. While P1 and P3 shared more homoplasies (3,282) than P2 and P3 (1,046), P1 also shared even more homoplasies with P4 (4,679; compared to 1,493 homoplasies shared between P2 and P4). Additionally, more reversals of substitutions shared by P1, P2, and P3 occurred on the branch leading to P1 (1,637) compared to P2 (509), resulting in more allele sharing between P2 and P3 and thus the imbalance between C_{ABBA} and C_{BABA}.

The false signals of introgression were not exclusive to the datasets simulated with extreme rate variation ($s = 0.25$ and $s = 4$), but also affected the datasets with more modest rate variation ($s = 0.5$ and $s = 2$). While the D-statistic was lower in these cases (0.05–0.14), it remained highly significant for all these datasets ($p < 10^{-9}$) (Supplementary Figures 1–4).

Tree-based Introgression Detection Methods

D_{tree} — Sets of maximum-likelihood trees, generated for the simulated datasets, produced high D_{tree} values up to around 1, even when no introgression was present (Fig. 3d). This pattern did not seem to be affected by rate variation (Fig. 3a,g), and was found with the alignment lengths 200, 500, and 1,000 bp (Fig. 3d, Supplementary Figures 5–7). The applied binomial test did not support a significant difference ($p > 0.05$) between

$C_{\mathrm{P2,P3}}$ and $C_{\mathrm{P2,P3}}$ in almost all of these cases, and none of the values remained significant following Bonferroni correction (Bonferroni, 1935). These high but non-significant D_{tree} values result from small numbers of trees that are not concordant with the species phylogeny. With the youngest phylogeny ($t_{\mathrm{P1,P2}} = 10$ million generations) and an alignment length of 500 bp, no more than 10 out of 2,000 trees were discordant for any of the 10 replicates per P2 branch rate ($s \in \{0.25, 0.5, 1, 2, 4\}$; Fig. 3d, Supplementary Figure 6). With older phylogenies and the same alignment length, these numbers remained in the ranges of 4–35 and 14–125, for $t_{\mathrm{P1,P2}} = 20$ and $t_{\mathrm{P1,P2}} = 30$ million generations, respectively.

The lack of significance could indicate that unlike Patterson's D-statistic, D_{tree} might be robust to rate variation. Alternatively, however, it could also result from the reduced amount of data used in tree-based analyses (0.5% of the data used to calculate Patterson's D; see above). This is suggested by the comparison of the numbers of discordant trees in which either P2 or P1 appear as the sister to P3 (Supplementary Table 2). If rate variation would affect D_{tree} in the same way as Patterson's D, we would expect that most discordant trees group P3 and P2 when $s < 1$ and that P3 and P1 appear as sisters more often when $s > 1$. This was in fact the case for 11 out of 12 settings with $s < 1$ and for 9 out of 15 settings with $s > 1$, a pattern that is unlikely to arise by chance (one-sided binomial test $p < 0.01$; Supplementary Table 2). Thus, homoplasies and rate variation appear to bias tree topologies similar to how they influence site patterns, through weak but noticeable long-branch attraction.

When introgression was included in the simulations, this was reliably detected for migration rates $m \geqslant 10^{-8}$, regardless of divergence time $t_{\mathrm{P1,P2}}$ or alignment length (Fig. 3, Supplementary Figures 5–7). However, like Patterson's D-statistic, D_{tree} values were decreasing with increasing divergence times (e.g., Fig. 3e). This was apparently caused by added stochasticity in tree topologies resulting from homoplasious substitutions, as both types of discordant trees became more frequent with older divergence times (Supplementary Table 3).

SNaQ — The maximum-likelihood values reported by SNaQ were in all cases equally good or better for the model that included a hybridization event, compared to the

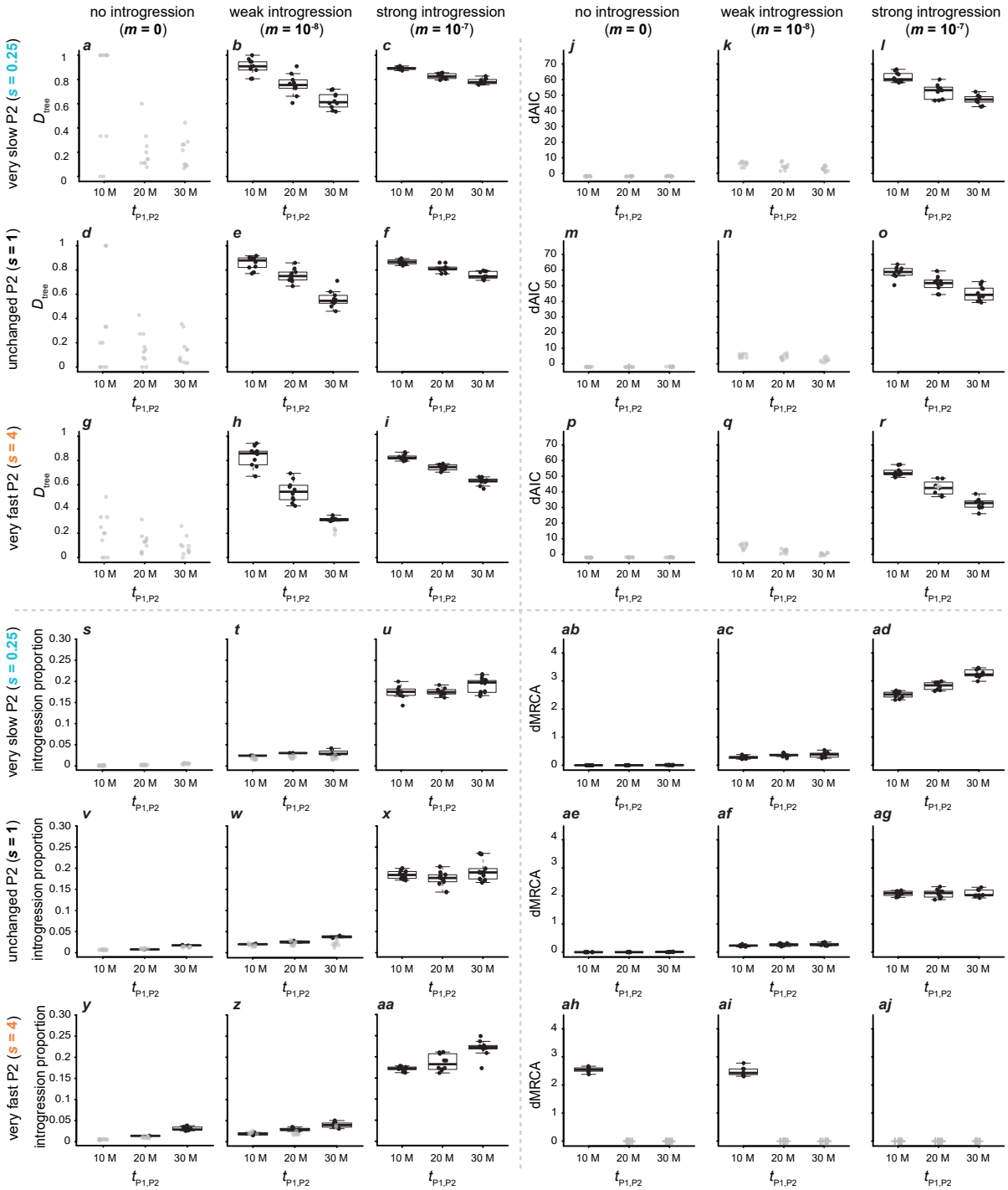

Fig. 3 (previous page) Signals of introgression detected with tree-based methods for datasets simulated with a population size $N_e = 10^5$, a mutation rate $\mu = 2 \times 10^{-9}$, either no ($m = 0$), weak $m = 10^{-8}$, or strong ($m = 10^{-7}$) introgression, either a very slow ($s = 0.25$), unchanged ($s = 1$), or very fast ($s = 4$) rate of branch P2, and an alignment length of 500 bp. All results obtained with other settings are shown in Supplementary Figures 5–16. Per divergence time $t_{P1,P2} \in \{1 \times 10^7, 2 \times 10^7, 3 \times 10^7\}$, results are shown for ten replicate simulations. **a**–**i**) D_{tree}; **j**–**r**) dAIC supporting introgression in networks produced with SNaQ; **s**–**aa**) introgression proportion estimated with QuIBL; **ab**–**aj**) dMRCA estimated with the MMS17 method, in units of million generations. In **a**–**aa**, circles in black indicate significant results ($p < 0.05$), and only these are summarized with box plots. In **ah**–**aj**, gray crosshairs indicate results based on incorrect topologies. As significance is not assessed with the MMS17 method, all other values are shown in black.

hybridization-free model (Supplementary Table 1). As these models differed by a single free parameter, this had the consequence that the dAIC comparing these models was always greater than or equal to -2. No effect of rate variation was recorded, and SNaQ correctly favored the model without hybridization when analyzing data simulated without introgression (Fig. 3j,m,p, Supplementary Figures 8–10): With the standard settings ($N_e = 10^5$, $\mu = 2 \times 10^{-9}$) and no introgression, dAIC values ranged from -2 to -1.31 for simulations with P2 branch rate $s = 0.25$ (Fig. 3j), from -2 to -1.41 for P2 branch rate $s = 1$ (Fig. 3m), and from -2 to -1.65 for P2 branch rate $s = 4$ (Fig. 3p). SNaQ had a low power to detect weak introgression ($m \leqslant 10^{-8}$) (Fig. 3k,n,q, Supplementary Figures 8–10). Only with a strong introgression rate in the simulations ($m \geqslant 10^{-7}$) did SNaQ detect significant signals of it (e.g., Fig. 3l,o,r). With the standard settings, dAIC values were non-significant and ranging from 0.66 to 6.82 when weak introgression ($m = 10^{-8}$) was present (Fig. 3n), but significant and ranging from 39.25 to 63.68 with strong introgression ($m = 10^{-7}$) (Fig. 3o). As with Patterson's D-statistic or D_{tree}, signals of introgression became weaker with increasing divergence times (Fig. 3l,o,r), probably because of the generally higher numbers of discordant trees inferred in those cases. The patterns described above were equally found with all tested alignment lengths (Supplementary Figures 8–10), and therefore seemed to be unaffected by alignment length.

QuIBL — QuIBL produced signals of introgression even when neither rate variation nor introgression were present. Analyzing sets of trees generated for alignments of 500 bp, one out of ten simulation replicates with $t_{P1,P2} = 20$ million generations and two

replicates with $t_{\mathrm{P1,P2}} = 30$ million generations produced significant results (Fig. 3v). This changed dramatically for different alignment lengths. With trees produced for alignments of 200 bp, QuIBL reported significant results for all ten replicates, regardless of phylogeny age (Supplementary Figure 11). On the other hand, with alignments of 1,000 bp, none of the results were significant. Like the level of significance, the introgression proportion estimated by QuIBL was higher with shorter alignments; ranging from 0.01–0.05 with alignments of 200 bp, from 0–0.01 with alignments of 500 bp, and remaining around 0 when alignments of 1,000 bp were used (Fig. 3v, Supplementary Figure 11).

Adding rate variation to the simulations, while still excluding introgression, influenced the results in opposite ways (Fig. 3s,y). With a very slow rate ($s = 0.25$) of the P2 branch, none of the replicate tree sets for alignments of 500 bp produced significant estimates (Fig. 3s). On the other hand, an increased rate of branch P2 ($s = 4$) led to even more false-positive, significant signals of introgression, particularly for older phylogenies ($t_{\mathrm{P1,P2}} = 30$ million generations; Fig. 3y). As before without rate variation, this pattern was affected by the length of the alignments used to produce the tree sets. With alignments of 200 bp, all analyses produced significant results, while alignments of 1,000 bp led to results that were in almost all cases non-significant (Supplementary Figures 11 and 13).

When introgression was simulated with $m \geqslant 10^{-8}$, QuIBL detected it largely reliably. The introgression proportion was then estimated between 0.01–0.09 with $m = 10^{-8}$ and between 0.11 and 0.28 with $m = 10^{-7}$, which was influenced only to a minor degree by rate variation (s), phylogeny age ($t_{\mathrm{P1,P2}}$), and alignment length (Fig. 3, Supplementary Figures 11–13).

MMS17 method — The MMS17 method performed as expected when neither rate variation nor introgression were present (Fig. 3ae, Supplementary Figures 14–16). Without introgression, the difference between the two oldest pairwise mean MRCA ages (dMRCA) was close to 0 (between 0 and 0.07 million generations). With increasing levels of introgression, dMRCA was continuously growing, to 0.14–0.42 million generations when

$m = 10^{-8}$ and 1.58–2.47 million generations when $m = 10^{-7}$. Phylogeny age ($t_{\mathrm{P1,P2}}$) had no noticeable influence on dMRCA in these cases, but dMRCA was slightly higher with shorter alignments compared to longer ones (Pearson's product-moment correlation, $p < 0.001$).

However, when rate variation was simulated, the MMS17 method became rather unreliable, particularly with faster rates ($s \geqslant 1$) of the P2 branch (Fig. 3ah, Supplementary Figures 14–16). With the very fast P2 rate ($s = 4$) and the youngest phylogeny ($t_{\mathrm{P1,P2}} = 10$ million generations), dMRCA increased to values between 1.48 and 4.04 myr, again depending on alignment length (Pearson's product-moment correlation, $p < 0.001$). These strong signals are the result of local trees in which P2 is incorrectly placed as the sister to a clade combining P1 and P3. As this placement extends the P2 branch length, the overall rate variation is lowered, improving the prior probability of the strict-clock model. With the older phylogenies ($t_{\mathrm{P1,P2}} \geqslant 20$ million generations), the two oldest mean pairwise MRCA ages were no longer those between P1 and P3 and between P2 and P3, leading to erroneous signals (Fig. 3ah).

In contrast, a slower rate ($s \leqslant 1$) of the P2 branch did not have a strong influence on dMRCA (Fig. 3ab). A false signal of introgression could nevertheless be observed when the tree set was based on short alignments of 200 bp (Supplementary Figure 14). In these cases, dMRCA ranged between 0.23 to 0.32 million generations.

'ABBA'-Site Clustering

Within the tested parameter space, our new test based on 'ABBA'-site clustering proved to be a robust indicator of introgression signals, unaffected by rate variation. With the standard settings ($N_{\mathrm{e}} = 10^5$, $\mu = 2 \times 10^{-9}$) and neither introgression nor rate variation, a single replicate with $t_{\mathrm{P1,P2}} = 30$ million years produced a marginally significant test result ($p = 0.044$) that became non-significant after Bonferroni correction (Bonferroni, 1935) (Fig. 4d). Adding rate variation did not change this pattern. Marginally significant

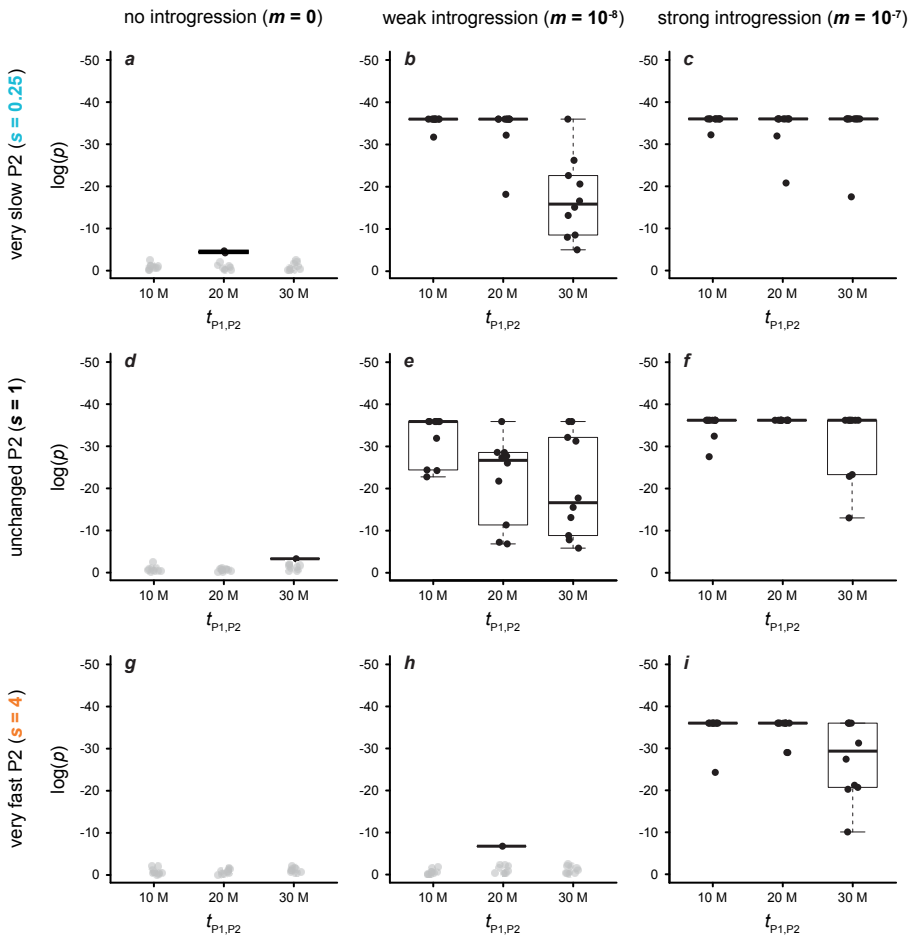

Fig. 4 Signals of introgression detected with the new 'ABBA'-site clustering approach for datasets simulated with a population size $N_e = 10^5$, a mutation rate $\mu = 2 \times 10^{-9}$, either no ($m = 0$; **a**,**d**,**g**), weak ($m = 10^{-8}$; **b**,**e**,**h**), or strong ($m = 10^{-7}$; **c**,**f**,**i**) introgression, and either a very slow ($s = 0.25$; **a**–**c**), unchanged ($s = 1$; **d**–**f**), or very fast ($s = 4$; **g**–**i**) rate of branch P2. All results obtained with other settings are shown in Supplementary Figures 17–21. Circles in black indicate significant results ($p < 0.05$), and only these are summarized with box plots.

results appeared when the P2 branch had a very slow rate ($s = 0.25$; 2 significant results with $p > 0.009$), or a slow rate ($s = 0.5$; 2 significant results with $p > 0.007$); however, all of these became insignificant after Bonferroni correction (Fig. 4a,g). Similar results were obtained with a smaller population size of $N_e = 10^4$, a lower mutation rate $\mu = 1 \times 10^{-9}$ (Supplementary Figures 17–20), or a variable recombination rate (Supplementary Note 2, Supplementary Figure 21).

Regardless of rate variation, the presence of introgression led to significant 'ABBA'-site clustering for all simulated datasets with the introgression rate of $m \geqslant 10^{-7}$,

and a large majority of datasets with $m = 10^{-8}$ (Fig. 4, Supplementary Figures 17–20). Particularly for simulations with the lower population size ($N_\mathrm{e} = 10^4$), our new approach regularly indicated introgression with the smallest possible p-value that Dsuite can report, $p = 2.3 \times 10^{-16}$ (Supplementary Figures 19 and 20). The cases that did not lead to significant 'ABBA'-site clustering despite the presence of introgression were limited to settings where the P2 rate was very fast ($s = 4$), the population size was large ($N_\mathrm{e} = 10^5$), and introgression was weak ($m \leqslant 10^{-8}$) (Fig. 4h).

While our main set of simulated datasets contained very little ILS, we considered the possibility that the 'ABBA'-site clustering test could be applied in settings where ILS is present or even prevalent. To test the performance of the test in these settings, and, specifically, to investigate the possibility of false-positives induced by fixation of unbroken ILS haplotypes, we performed a complementary set of simulations for which the length of the internal branch ancestral to P1 and P2 was reduced from 10 million generations to 1 million (low-to-moderate ILS) and to 100,000 generations (moderate-to-high ILS). The terminal branches leading to P3 and P4 were shortened accordingly, so that the phylogeny remained ultrametric in all cases. These complementary simulations were performed without introgression ($m = 0$) and rate variation ($s = 1$), testing the three divergence times ($t_{\mathrm{P1,P2}} \in \{1 \times 10^7, 2 \times 10^7, 3 \times 10^7\}$), both population sizes ($N_\mathrm{e} \in \{10^4, 10^5\}$), and the two mutation rates ($\mu \in \{1 \times 10^{-9}, 2 \times 10^{-9}\}$), with ten replicate datasets for each parameter combination. When the internal branch length was set to 1 million generations, between 42,174 and 47,652 bp ($0.21 - 0.24\%$) of the 20 Mbp-chromosomes simulated with a population size $N_\mathrm{e} = 10^5$ had a topology different from the species tree as the result of ILS. With the shorter internal branch length of 100,000 generations, the simulations with $N_\mathrm{e} = 10^4$ also led to ILS, affecting between 39,191 and 62,778 bp ($0.20 - 0.31\%$) of the 20-Mbp chromosomes. With $N_\mathrm{e} = 10^5$, as much as $3,972,624 - 4,058,703$ bp ($19.9 - 20.3\%$) carried ILS. The mean lengths of ILS-derived tracts ranged between 71.8 and 837.0 bp. These tract lengths are comparable for example to those introduced by weak

introgression in combination with the larger population size ($m = 10^{-8}$, $N_e = 10^5$), which ranged from 476.8 to 502.4 bp and led to highly significant results with the 'ABBA'-site clustering test (Fig. 2e). Crucially, these substantial levels of ILS did not lead to false-positive signals of introgression in the 'ABBA'-site clustering test. Only two out of the 120 simulated datasets with the internal branch length of 1 million generations produced significant p-values and none remained significant after Bonferroni correction. With the internal branch length of 100,000 generations, seven tests were just about statistically significant but, again, none remained significant after Bonferroni correction.

Discussion

As genome-wide data are becoming available for more and more species across the tree of life, these give us the opportunity to investigate the extent of between-species hybridization and introgression in unprecedented detail (Taylor and Larson, 2019). These data have already revealed an unexpected frequency of these processes among closely related species, and are beginning to uncover their occurrence also among more divergent lineages (Barth et al., 2020; Vilaça et al., 2021; Schöneberg et al., 2023). However, the results of studies on divergent taxa applying introgression-detection methods that were originally developed for closely related groups of species or populations must be critically evaluated (Pease and Hahn, 2015; Hibbins and Hahn, 2022; Zheng and Janke, 2018).

Our results confirm a recent report that demonstrated a sensitivity of Patterson's D-statistic, along with the related D_3 (Hahn and Hibbins, 2019) and HyDe (Blischak et al., 2018) tests, to among-species rate variation (Frankel and Ané, 2023). We extended these previous results to data simulated with a more diverse range of parameters, including different phylogeny ages, population sizes, mutation rates, and both homogeneous and variable recombination rates, corroborating that the D-statistic is generally sensitive to rate variation. Furthermore, we demonstrated that this sensitivity is not exclusive to summary statistics based on site patterns, but that various tree-based methods are

similarly affected. We identified a weak but noticeable effect of long-branch attraction that biases the topologies in sets of local trees and thus influences the D_{tree} statistic towards false signals of introgression. Additionally, methods using branch lengths are affected by rate variation, regardless of whether these branch lengths are measured in units of substitutions (QuIBL; Edelman et al. 2019) or time (the MMS17 method; Meyer et al. 2017).

To distinguish between false signals caused by rate variation and genuine indicators of introgression, we propose a new test for clustering of 'ABBA' sites along the genome. This pattern is expected to arise when several polymorphisms are located on the same introgressed haplotype, but not through homoplasies and rate variation. Indeed, our analyses demonstrated that this test is robust to rate variation, only indicating 'ABBA'-site clustering when introgression in fact occurred. Our implementation of this test in the program Dsuite is easy to use and comes with negligible added cost to Dsuite analyses. Given that Dsuite is among the fastest tools available for the calculation of D-statistics (Malinsky et al., 2021), the additional application of the 'ABBA'-site clustering test should be computationally feasible for all users.

In addition to introgression, clustering of 'ABBA' sites could in principle be expected to arise from ILS, as ILS-derived tracts can contain multiple genetic variants. Nevertheless, we showed here that the 'ABBA'-site clustering test is robust, at least for realistic use-cases with moderate levels of ILS. In the absence of introgression but presence of ILS, we did not observe any false-positives after multiple testing correction. This difference in sensitivity to introgression vs. ILS cannot be explained by the lengths of the tracts produced by these two processes, as these were comparable across the simulations. The explanation likely lies in the difference in numbers of 'ABBA' sites within introgressed vs. ILS tracts. Haplotypes introduced through introgression have had a long time, at least ten million generations in our simulations, to accumulate the mutations that produce 'ABBA' sites following introgression between P3 and P2. On the other hand, haplotypes

introduced through ILS had much less time for the accumulation of mutations that would ultimately produce 'ABBA' or 'BABA' sites – on average, one coalescent unit ($2N_e$ generations) – in the common ancestor of P1, P2, and P3. Thus, the tracts produced by ILS carry fewer 'ABBA' sites than those produced by introgression, which, as a consequence, renders the 'ABBA'-site clustering test robust to ILS when investigating potential introgression between highly diverged species.

We note that, although mutations are often considered random and approximated by a uniform distribution along the genome, like in our test and simulations, empirical studies may be affected by some degree of mutation rate variation along the genome (e.g., (Ségurel et al., 2014; Hodgkinson and Eyre-Walker, 2011). Because our test focuses only on polymorphic sites, it will be robust to straightforward regional rate variation – if a genomic region has a higher rate, it will increase both the number of homoplasies and the number of polymorphic sites in that region. It is conceivable that an elevated mutation rate at CpG sites within regions of ancestral high GC content (Ségurel et al., 2014) could locally increase the rate of homoplasies relative to polymorphic sites and lead to artificial clustering of 'ABBA' sites. However, such concern can easily be addressed by filtering out C/G $\rightarrow$ A/T mutations from the dataset before running Dsuite.

Our analyses revealed that tree-based methods can be useful for the detection of introgression when rate variation is present, and identified the conditions under which each approach performs reliably. While we observed an effect of long-branch attraction affecting the D_{tree} statistic, this effect was weak and only noticeable when all results were considered in aggregate. In fact, none of the datasets simulated without introgression produced a false-positive, significant D_{tree} statistic, even when D_{tree} itself reached the maximum value of 1.0 (Fig. 3a). On the other hand, D_{tree} was consistently large and significant even with weak introgression ($m = 10^{-8}$; Fig. 3b,e,h), suggesting that D_{tree} is a powerful detector of introgression. However, to account for the observed slight bias in tree topologies produced by rate variation, it may be wise to ignore weakly significant D_{tree}

results and only interpret the stronger ones, if present.

Besides D_{tree}, SNaQ appeared to be robust to rate variation across our simulated datasets (Fig. 3j,p). Given that SNaQ analyzes tree topologies, however, we caution that the same weak bias that affected D_{tree} might also be relevant for SNaQ. Like for D_{tree}, we thus advise that weaker signals reported by SNaQ might better be ignored. Furthermore, it has been pointed out that the use of AIC is inappropriate for the comparison of SNaQ results, due to the pseudolikelihood framework employed by SNaQ (Hibbins and Hahn, 2022). To avoid this issue, users of SNaQ may want to focus – like we did – on sets of four taxa with putative introgression, in which case SNaQ calculates and reports actual likelihood values (Solís-Lemus and Ané, 2016).

Finally, we found that while QuIBL produced false signals of introgression when the P2 branch had an elevated rate, it appeared to perform reliably when the rate was decreased instead of increased (Fig. 3s). Thus, when the level of rate variation in a clade is known and only single lineages have decreased rates, introgression signals reported by QuIBL may be considered.

In practice, the inference of ancient introgression between divergent species may often be hampered by the requirements of detection methods. Site-pattern-based methods (such as the D-statistic and the 'ABBA'-site clustering test) require datasets of single-nucleotide polymorphisms (SNP), that are typically obtained through read mapping towards a reference genome assembly. When investigating divergent taxa, however, it may no longer be possible to map all of them reliably to the same reference genome. As a result, SNP datasets produced for such taxa may be limited and prone to reference bias, which can generate misleading signals of introgression (Günther and Nettelblad, 2019). On the other hand, multi-marker sets of alignments, which are usually used for tree-based detection methods, have traditionally been produced through ortholog-identification approaches focusing on genes or ultra-conserved elements. While these approaches may be more suited for divergent taxa than read mapping, they are generally limited to certain

regions of the genome, corresponding to a set of input query sequences. Fortunately, two recent developments promise to overcome these limitations, rendering larger regions of the genome accessible for the detection of ancient introgression: First, methods for whole-genome alignment have finally matured to the degree that they can be applied to hundreds of genome assemblies of highly diverged taxa (Armstrong et al., 2020). By using assemblies instead of mapped reads, these whole-genome alignments are immune to reference bias, and allow the extraction of massive numbers of SNPs for site-pattern-based methods, or of alignment blocks for tree-based methods. Second, more and more genome assemblies are now highly contiguous, chromosome-resolved or nearly so (Rhie et al., 2021; Formenti et al., 2022). This is relevant for the completeness of whole-genome alignments, and reduces their fragmentation. Both will contribute to the utility of the new 'ABBA'-site clustering test, given that this test requires contiguous genomic blocks within which clustering can be observed.

In their combination, these new developments are now allowing us to push the limits of reliable introgression detection, enabling the inference of introgression even among species that have diverged many tens of millions of years ago. We are thus coming closer to being able to assess the true extent of hybridization and introgression across the tree of life.

Code Availability

Code for all our computational analyses is available on https://github.com/thorekop/ABBA-Site-Clustering

Acknowledgements

We are grateful to Fabrizia Ronco and Kjetil Lysne Voje for valuable help and support with R scripts. We thank Miriam Miyagi and Nathaniel Edelman for advice on how to interpret the QuIBL output when using trees with polytomies. All computations were performed on resources provided by Sigma2 – the National Infrastructure for High

Performance Computing and Data Storage in Norway.

References

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. Journal of Evolutionary Biology 26:229–246.

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

Alfieri, J. M., T. Johnson, A. Linderholm, H. Blackmon, and G. N. Athrey. 2023. Genomic investigation refutes record of most diverged avian hybrid. Ecology and Evolution 13.

Álvarez-Carretero, S., A. U. Tamuri, M. Battini, F. F. Nascimento, E. Carlisle, R. J. Asher, Z. Yang, P. C. J. Donoghue, and M. dos Reis. 2022. A species-level timeline of mammal evolution integrating phylogenomic data. Nature 602:263–267.

Amos, W. 2020. Signals interpreted as archaic introgression appear to be driven primarily by faster evolution in Africa. Royal Society Open Science 7:191900.

Armstrong, J., G. Hickey, M. Diekhans, I. T. Fiddes, A. M. Novak, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stiller, D. Genereux, J. Johnson, V. D. Marinescu, J. Alföldi, R. S. Harris, K. Lindblad-Toh, D. Haussler, E. Karlsson, E. D. Jarvis, G. Zhang, and B. Paten. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature 587:246–251.

Barth, J. M. I., C. Gubili, M. Matschiner, O. K. Tørresen, S. Watanabe, B. Egger, Y.-S. Han, E. Feunteun, R. Sommaruga, R. Jehle, and R. Schabetsberger. 2020. Stable species boundaries despite ten million years of hybridization in tropical eels. Nature Communications 11:1433.

Baumdicker, F., G. Bisschop, D. Goldstein, G. Gower, A. P. Ragsdale, G. Tsambos, S. Zhu, B. Eldon, E. C. Ellerman, J. G. Galloway, A. L. Gladstein, G. Gorjanc, B. Guo, B. Jeffery, W. W. Kretzschmar, K. Lohse, M. Matschiner, D. Nelson, N. S. Pope, C. D. Quinto-Cortés, M. F. Rodrigues, K. Saunack, T. Sellinger, K. Thornton, H. van Kemenade, A. W. Wohns, Y. Wong, S. Gravel, A. D. Kern, J. Koskela, P. L. Ralph, and J. Kelleher. 2021. Efficient ancestry and mutation simulation with msprime 1.0. Genetics 220:iyab229.

Betancur-R, R., E. O. Wiley, G. Arratia, A. Acero, N. Bailly, M. Miya, G. Lecointre, and G. Orti. 2017. Phylogenetic classification of bony fishes. BMC Evolutionary Biology 17:162.

Bilderbeek, R. J. C. and R. S. Etienne. 2018. babette: BEAUti 2, BEAST2 and Tracer for R. Methods in Ecology and Evolution 9:2034–2040.

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

Bonferroni, C. E. 1935. Il calcolo delle assicurazioni su gruppi di teste. Studi in Onore del Professore Salvatore Ortu Carboni. Pages 13–60.

Bouckaert, R. R., T. G. Vaughan, J. Barido-Sottani, S. Duchene, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, M. Matschiner, F. K. Mendes, N. F. Müller, H. A. Ogilvie, L. du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C.-H. Wu, D. Xie, C. Zhang, T. Stadler, and A. J. Drummond. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Computational Biology 15:e1006650.

Bromham, L. 2020. Causes of variation in the rate of molecular evolution. Pages 45–64 *in* The Molecular Evolutionary Clock (S. Ho, ed.). Springer, Cham.

Bryant, D. and M. W. Hahn. 2020. The concatenation question. Pages 3.4:1–3.4:23 *in* Phylogenetics in the Genomic Era (C. Scornavacca, F. Delsuc, and N. Galtier, eds.). No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE.

Doyle, J. J. 1995. The irrelevance of allele tree topologies for species delimitation, and a non-topological alternative. Systematic Botany 20:574–588.

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

Edelman, N. B., P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. Van Belleghem, N. Patterson, D. E. Neafsey, R. Challis, S. Kumar, G. R. P. Moreira, C. Salazar, M. Chouteau, B. A. Counterman, R. Papa, M. Blaxter, R. D. Reed, K. K. Dasmahapatra, M. Kronforst, M. Joron, C. D. Jiggins, W. O. McMillan, F. Di Palma, A. J. Blumberg, J. Wakeley, D. Jaffe, and J. Mallet. 2019. Genomic architecture and introgression shape a butterfly radiation. Science 366:594.

Eriksson, A. 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 USA 109:13956–13960.

Ernst, N., A. Schmitz, N. Chai, J. Rigoulet, A. Burgeois, M. Hohl, C. Hano, and I. Ineich. 2014. An unexpected occurrence—a case study of an intergeneric hybrid in giant snakes. Revue Suisse de Zoologie 121:293–317.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Biology 27:401–410.

Figueiró, H. V., G. Li, F. J. Trindade, J. Assis, F. Pais, G. Fernandes, S. H. D. Santos, G. M. Hughes, A. Komissarov, A. Antunes, C. S. Trinca, M. R. Rodrigues, T. Linderoth, K. Bi, L. Silveira, F. C. C. Azevedo, D. Kantek, E. Ramalho, R. A. Brassaloti, P. M. S. Villela, A. L. V. Nunes, R. H. F. Teixeira, R. G. Morato, D. Loska, P. Saragüeta, T. Gabaldón, E. C. Teeling, S. J. O'Brien, R. Nielsen, L. L. Coutinho, G. Oliveira, W. J. Murphy, and E. Eizirik. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Science Advances 3:e1700299.

Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, X. Jiang, A. B. Hall, F. Catteruccia, E. Kakani, S. N. Mitchell, Y.-C. Wu, H. A. Smith, R. R. Love, M. K. Lawniczak, M. A. Slotman, S. J. Emrich, M. W. Hahn, and N. J. Besansky. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347:1258524.

Formenti, G., K. Theissinger, C. Fernandes, I. Bista, A. Bombarely, C. Bleidorn, C. Ciofi, A. Crottini, J. A. Godoy, J. Höglund, J. Malukiewicz, A. Mouton, R. A. Oomen, S. Paez, P. J. Palsbøll, C. Pampoulie, M. J. Ruiz-López, H. Svardal, C. Theofanopoulou, J. de Vries, A.-M. Waldvogel, G. Zhang, C. J. Mazzoni, E. D. Jarvis, M. Bálint, G. Formenti, K. Theissinger, C. Fernandes, I. Bista, A. Bombarely, C. Bleidorn, F. Čiampor, C. Ciofi, A. Crottini, J. A. Godoy, J. Hoglund, J. Malukiewicz, A. Mouton, R. A. Oomen, S. Paez, P. Palsbøll, C. Pampoulie, M. J. Ruiz-López, H. Svardal, C. Theofanopoulou, J. de Vries, A.-M. Waldvogel, G. Zhang, C. J. Mazzoni, E. Jarvis, M. Bálint, S. A. Aghayan, T. S. Alioto, I. Almudi, N. Alvarez, P. C. Alves, I. R. Amorim, A. Antunes, P. Arribas, P. Baldrian, P. R. Berg, G. Bertorelle, A. Böhne, A. Bonisoli-Alquati, L. L. Boštjančić, B. Boussau, C. M. Breton, E. Buzan, P. F. Campos, C. Carreras, L. F. Castro, L. J. Chueca, E. Conti, R. Cook-Deegan, D. Croll, M. V. Cunha, F. Delsuc, A. B. Dennis, D. Dimitrov, R. Faria, A. Favre, O. D. Fedrigo, R. Fernández, G. F. Ficetola, J.-F. Flot, T. Gabaldón, D. R. Galea Agius, G. R. Gallo, A. M. Giani, M. T. P. Gilbert, T. Grebenc, K. Guschanski, R. Guyot, B. Hausdorf,

O. Hawlitschek, P. D. Heintzman, B. Heinze, M. Hiller, M. Husemann, A. Iannucci, I. Irisarri, K. S. Jakobsen, S. Jentoft, P. Klinga, A. Kloch, C. F. Kratochwil, H. Kusche, K. K. Layton, J. A. Leonard, E. Lerat, G. Liti, T. Manousaki, T. Marques-Bonet, P. Matos-Maraví, M. Matschiner, F. Maumus, A. M. Mc Cartney, S. Meiri, J. Melo-Ferreira, X. Mengual, M. T. Monaghan, M. Montagna, R. W. Mysłajek, M. T. Neiber, V. Nicolas, M. Novo, P. Ozretić, F. Palero, L. Pârvulescu, M. Pascual, O. S. Paulo, M. Pavlek, C. Pegueroles, L. Pellissier, G. Pesole, C. R. Primmer, A. Riesgo, L. Rüber, D. Rubolini, D. Salvi, O. Seehausen, M. Seidel, S. Secomandi, B. Studer, S. Theodoridis, M. Thines, L. Urban, A. Vasemägi, A. Vella, N. Vella, S. C. Vernes, C. Vernesi, D. R. Vieites, R. M. Waterhouse, C. W. Wheat, G. Wörheide, Y. Wurm, and G. Zammit. 2022. The era of reference genomes in conservation genomics. Trends in Ecology & Evolution 37:197–202.

Frankel, L. E. and C. Ané. 2023. Summary tests of introgression are highly sensitive to rate variation across lineages. bioRxiv, doi:10.1101/2023.01.26.525396.

Gante, H. F., M. Matschiner, M. Malmstrøm, K. S. Jakobsen, S. Jentoft, and W. Salzburger. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. Molecular Ecology 25:6143–6161.

Gernhard, T. 2008. The conditioned reconstructed process. Journal of Theoretical Biology 253:769–778.

Ghigi, A. 1936. Galline di Faraone e Tacchini. Hoepli, Milano, Italy.

Gray, A. P. 1958. Bird Hybrids. Commonwelth Agricultural Bureaux, Bucks, UK.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Y. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M. Siegemund,

A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, and S. Paabo. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.

Günther, T. and C. Nettelblad. 2019. The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLOS Genetics 15:e1008302.

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

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160–174.

Heled, J. and R. R. Bouckaert. 2013. Looking for trees in the forest: summary tree from posterior samples. BMC Evolutionary Biology 13:221.

Hibbins, M. S. and M. W. Hahn. 2022. Phylogenomic approaches to detecting and characterizing introgression. Genetics 220:iyab173.

Hodgkinson, A. and A. Eyre-Walker. 2011. Variation in the mutation rate across mammalian genomes. Nature Reviews Genetics 12:756–766.

Hua, X. and L. Bromham. 2017. Darwinism for the genomic age: connecting mutation to diversification. Frontiers in Genetics 8:12.

Hua, X., P. Cowman, D. Warren, and L. Bromham. 2015. Longevity is linked to mitochondrial mutation rates in rockfish: a test using poisson regression. Molecular Biology and Evolution 32:2633–2645.

Huson, D. H., T. Klöpper, P. J. Lockhart, and M. A. Steel. 2005. Reconstruction of reticulate networks from gene trees. Pages 233–249 *in* Research in Computational Molecular Biology. RECOMB 2005. Lecture Notes in Computer Science (S. Miyano, J. Mesirov, S. Kasif, S. Istrail, P. A. Pevzner, and M. Waterman, eds.) vol. 3500. Springer, Berlin, Heidelberg.

Ivan, J., C. Moritz, S. Potter, J. Bragg, R. Turakulov, and X. Hua. 2022. Temperature predicts the rate of molecular evolution in Australian Eugongylinae skinks. Evolution 76:252–261.

Justison, J. A., C. Solis-Lemus, and T. A. Heath. 2022. SiPhyNetwork: An R package for Simulating Phylogenetic Networks. bioRxiv, doi:10.1101/2022.10.26.513953.

Kolmogorov, A. N. 1933. Sulla determinazione empirica di una legge di distribuzione. Giornale dell'Instituto Italiano degli Attuari 4:83–91.

Káldy, J., A. Mozsár, G. Fazekas, M. Farkas, D. L. Fazekas, G. L. Fazekas, K. Goda, Z. Gyöngy, B. Kovács, K. Semmens, M. Bercsényi, M. Molnár, and E. Patakiné Várkonyi. 2020. Hybridization of Russian Sturgeon (*Acipenser gueldenstaedtii*, Brandt and Ratzeberg, 1833) and American Paddlefish (*Polyodon spathula*, Walbaum 1792) and evaluation of their progeny. Genes 11:753.

Lamichhaney, S., F. Han, M. T. Webster, L. Andersson, B. R. Grant, and P. R. Grant. 2018. Rapid hybrid speciation in Darwin's finches. Science 359:224–228.

LeClere, J. B., E. P. Hoaglund, J. Scharosch, C. E. Smith, and T. Gamble. 2012. Two naturally occurring intergeneric hybrid snakes (*Pituophis catenifer sayi* × *Pantherophis vulpinus*; Lampropeltini, Squamata) from the midwestern United States. Journal of Herpetology 46:257–262.

Lehtonen, S. 2018. ×*Lindsaeosoria flynnii* (Lindsaeaceae), another confirmed example of deep hybridization among the ferns. American Fern Journal 108:7–18.

Malinsky, M., M. Matschiner, and H. Svardal. 2021. Dsuite - Fast *D*-statistics and related admixture evidence from VCF files. Molecular Ecology Resources 21:584–595.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends in Ecology & Evolution 20:229–237.

Marcussen, T., S. R. Sandve, L. Heier, M. Spannagl, M. Pfeifer, International Wheat Genome Sequencing Consortium, K. S. Jakobsen, B. B. H. Wulff, B. Steuernagel, K. F. X. Mayer, and O.-A. Olsen. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092.

Mariadassou, M., M. Suez, S. Sathyakumar, A. Vignal, M. Arca, P. Nicolas, T. Faraut, D. Esquerré, M. Nishibori, A. Vieaud, C.-F. Chen, H. Manh Pham, Y. Roman, F. Hospital, T. Zerjal, X. Rognon, and M. Tixier-Boichard. 2021. Unraveling the history of the genus *Gallus* through whole genome sequencing. Molecular Phylogenetics and Evolution 158:107044.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York, NY, USA.

McCarthy, E. M. 2006. Handbook of Avian Hybrids of the World. Oxford University Press, New York, USA.

Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nature Communications 8:14363.

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

Meyer, B. S., M. Matschiner, and W. Salzburger. 2017. Disentangling incomplete lineage

sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. Systematic Biology 66:531–550.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37:1530–1534.

Mitchell, N. and K. D. Whitney. 2021. Limited evidence for a positive relationship between hybridization and diversification across seed plant families. Evolution 75:1966–1982.

Nam, K., C. Mugal, B. Nabholz, H. Schielzeth, J. B. Wolf, N. Backström, A. Künstner, C. N. Balakrishnan, A. Heger, C. P. Ponting, D. F. Clayton, and H. Ellegren. 2010. Molecular evolution of genes in avian genomes. Genome Biology 11:R68.

Pampoulie, C., D. Gíslason, G. Ólafsdóttir, V. Chosson, S. D. Halldórsson, S. Mariani, B. T. Elvarsson, M. H. Rasmussen, M. R. Iversen, A. K. Daníelsdóttir, and G. A. Víkingsson. 2021. Evidence of unidirectional hybridization and second-generation adult hybrid between the two largest animals on Earth, the fin and blue whales. Evolutionary Applications 14:314–321.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290.

Patton, A. H., M. J. Margres, B. Epstein, J. Eastman, L. J. Harmon, and A. Storfer. 2020. Hybridizing salamanders experience accelerated diversification. Scientific Reports 10:6566.

Pavón-Vázquez, C. J., I. G. Brennan, and J. S. Keogh. 2021. A comprehensive approach to detect hybridization sheds light on the evolution of Earth's largest lizards. Systematic Biology 70:877–890.

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

Prum, R. O., J. S. Berv, A. Dornburg, D. J. Field, J. P. Townsend, E. M. Lemmon, and A. R. Lemmon. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526:569–573.

Rabosky, D. L., J. Chang, P. O. Title, P. F. Cowman, L. Sallan, M. Friedman, K. Kaschner, C. Garilao, T. J. Near, M. Coll, and M. E. Alfaro. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature 559:392–395.

Rhie, A., S. A. McCarthy, O. Fedrigo, J. Damas, G. Formenti, S. Koren, M. Uliano-Silva, W. Chow, A. Fungtammasan, J. Kim, C. Lee, B. J. Ko, M. Chaisson, G. L. Gedman, L. J. Cantin, F. Thibaud-Nissen, L. Haggerty, I. Bista, M. Smith, B. Haase, J. Mountcastle, S. Winkler, S. Paez, J. Howard, S. C. Vernes, T. M. Lama, F. Grutzner, W. C. Warren, C. N. Balakrishnan, D. Burt, J. M. George, M. T. Biegler, D. Iorns, A. Digby, D. Eason, B. Robertson, T. Edwards, M. Wilkinson, G. Turner, A. Meyer, A. F. Kautt, P. Franchini, I. Detrich, H William, H. Svardal, M. Wagner, G. J. P. Naylor, M. Pippel, M. Malinsky, M. Mooney, M. Simbirsky, B. T. Hannigan, T. Pesout, M. Houck, A. Misuraca, S. B. Kingan, R. Hall, Z. Kronenberg, I. Sović, C. Dunn, Z. Ning, A. Hastie, J. Lee, S. Selvaraj, R. E. Green, N. H. Putnam, I. Gut, J. Ghurye, E. Garrison, Y. Sims, J. Collins, S. Pelan, J. Torrance, A. Tracey, J. Wood, R. E. Dagnew, D. Guan, S. E. London, D. F. Clayton, C. V. Mello, S. R. Friedrich, P. V. Lovell, E. Osipova, F. O. Al-Ajli, S. Secomandi, H. Kim, C. Theofanopoulou, M. Hiller, Y. Zhou, R. S. Harris, K. D. Makova, P. Medvedev, J. Hoffman, P. Masterson, K. Clark, F. Martin, K. Howe, P. Flicek, B. P. Walenz, W. Kwak, H. Clawson, M. Diekhans, L. Nassar, B. Paten, R. H. S. Kraus, A. J. Crawford, M. T. P. Gilbert, G. Zhang, B. Venkatesh, R. W. Murphy, K.-P. Koepfli, B. Shapiro, W. E. Johnson, F. Palma, T. Marques-Bonet, E. C. Teeling, T. Warnow, J. M. Graves, O. A. Ryder, D. Haussler,

S. J. O'Brien, J. Korlach, H. A. Lewin, K. Howe, E. W. Myers, R. Durbin, A. M. Phillippy, and E. D. Jarvis. 2021. Towards complete and error-free genome assemblies of all vertebrate species. Nature 592:737–746.

Rieseberg, L. H., C. Van Fossen, and A. M. Desrochers. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature 375:313–316.

Ronco, F., M. Matschiner, A. Böhne, A. Boila, H. H. Büscher, A. El Taher, A. Indermaur, M. Malinsky, V. Ricci, A. Kahmen, S. Jentoft, and W. Salzburger. 2021. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. Nature 589:76–81.

Rothfels, C. J., A. K. Johnson, P. H. Hovenkamp, D. L. Swofford, H. C. Roskam, C. R. Fraser-Jenkins, M. D. Windham, and K. M. Pryer. 2015. Natural hybridization between genera that diverged from each other approximately 60 million years ago. The American Naturalist 185:433–442.

Runemark, A., C. N. Trier, F. Eroukhmanoff, J. S. Hermansen, M. Matschiner, M. Ravinet, T. O. Elgvin, and G.-P. Sætre. 2018. Variation and constraints in hybrid genome formation. Nature Ecology & Evolution 2:549–556.

Sanchez-Donoso, I., J. Huisman, J. Echegaray, M. Puigcerver, J. D. Rodríguez-Teijeiro, F. Hailer, and C. Vilá. 2014. Detecting slow introgression of invasive alleles in an extensively restocked game bird. Frontiers in Ecology and Evolution 2:15.

Sayres, M. A. W., C. Venditti, M. Pagel, and K. D. Makova. 2011. Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. Evolution: International Journal of Organic Evolution 65:2800–2815.

Schöneberg, Y., S. Winter, O. Arribas, M. R. Di Nicola, M. Master, J. B. Owens, M. Rovatsos, W. Wüster, A. Janke, and U. Fritz. 2023. Genomics reveals broad

hybridization in deeply divergent palearctic grass and water snakes (*Natrix* spp.). Molecular Phylogenetics and Evolution 184:107787.

Schumer, M., R. Cui, D. L. Powell, G. G. Rosenthal, and P. Andolfatto. 2016. Ancient hybridization and genomic stabilization in a swordtail fish. Molecular Ecology 25:2661–2679.

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

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends in Ecology & Evolution 19:198–207.

Ségurel, L., M. J. Wyman, and M. Przeworski. 2014. Determinants of mutation rate variation in the human germline. Annual Review of Genomics and Human Genetics 15:47–70.

Simard, R. and P. L'Ecuyer. 2011. Computing the two-sided Kolmogorov-Smirnov distribution. Journal of Statistical Software 39.

Solís-Lemus, C. and C. Ané. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genetics 12:e1005896.

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

Stull, G. W., K. K. Pham, P. S. Soltis, and D. E. Soltis. 2023. Deep reticulation: the long legacy of hybridization in vascular plant evolution. The Plant Journal Epub ahead of print.

Stull, G. W., P. S. Soltis, D. E. Soltis, M. A. Gitzendanner, and S. A. Smith. 2020. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. American Journal of Botany 107:790–805.

Suvorov, A., B. Y. Kim, J. Wang, E. E. Armstrong, D. Peede, E. R. D'Agostino, D. K. Price, P. J. Waddell, M. Lang, V. Courtier-Orgogozo, J. R. David, D. Petrov, D. R. Matute, D. R. Schrider, and A. A. Comeault. 2022. Widespread introgression across a phylogeny of 155 Drosophila genomes. Current Biology 32:111–123.e5.

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

Tea, Y.-K., J.-P. A. Hobbs, F. Vitelli, J. D. DiBattista, S. Y. W. Ho, and N. Lo. 2020. Angels in disguise: sympatric hybridization in the marine angelfishes is widespread and occurs between deeply divergent lineages. Proceedings of the Royal Society B: Biological Sciences 287:20201459.

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

Vilaça, S. T., R. Piccinno, O. Rota-Stabelli, M. Gabrielli, A. Benazzo, M. Matschiner, L. S. Soares, A. B. Bolten, K. A. Bjorndal, and G. Bertorelle. 2021. Divergence and hybridization in sea turtles: Inferences from genome data show evidence of ancient gene flow between species. Molecular Ecology 30:6178–6192.

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

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

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