Manuscript type: Article

Recommended MBE section: Methods

Evaluating the use of ABBA-BABA statistics to locate introgressed loci

Simon H. Martin[‡1], John W. Davey[1], Chris D. Jiggins[1]

[1]Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

[‡] Corresponding Author: shm45@cam.ac.uk

ABSTRACT

Several methods have been proposed to test for introgression across genomes. One method tests for a genome-wide excess of shared derived alleles between taxa using Patterson's D statistic, but does not establish which loci show such an excess or whether the excess is due to introgression or ancestral population structure. Here, we use simulations and whole genome data from *Heliconius* butterflies to investigate the behavior of D when applied to small genomic regions, as has been attempted in several recent studies. We find that D is unreliable as it gives inflated values when effective population size is low, causing D outliers to cluster in genomic regions of reduced diversity. As an alternative, we propose a related statistic f_d, a modified version of a statistic originally developed to estimate the genome-wide fraction of admixture. f_d is not subject to the same biases as D, and is better at identifying introgressed loci. Finally, we show that both D and f outliers tend to cluster in regions of low genetic divergence, which can confound analyses aimed at differentiating introgression from shared ancestral variation at individual loci.

INTRODUCTION

Hybridization and gene flow between taxa play a major role in evolution, acting as a force against divergence, and as a potential source of adaptive novelty (Abbott et al. 2013). Although identifying gene flow between species has been a long-standing problem in population genetics, the issue has received considerable recent attention with the analysis of shared ancestry between humans and Neanderthals (for example, Yang et al. 2012; Wall et al. 2013). With genomic data sets becoming available in a wide variety of other taxonomic groups, there is a need for reliable, computationally tractable methods that identify, quantify and date gene flow between species in large data sets.

A sensitive and widely used approach to test for gene flow is to fit coalescent models using maximum-likelihood or Bayesian methods (Pinho and Hey 2010). However, simulation and model fitting are computationally intensive tasks, and are not easily applied on a genomic scale. A simpler and more computationally efficient approach that is gaining in popularity is to test for an excess of shared derived variants using a four-taxon test (Kulathinal et al. 2009; Green et al. 2010; Durand et al. 2011). The test considers ancestral ('A') and derived ('B') alleles, and is based on the prediction that two particular SNP patterns, termed 'ABBA' and 'BABA' (see Methods), should be equally frequent under a scenario of incomplete lineage sorting without gene flow. An excess of ABBA patterns is indicative of gene flow between two of the taxa, and can be detected using Patterson's *D* statistic (Green et al. 2010; Durand et al. 2011; see Methods for details). However, an excess of shared derived variants can arise from factors other than recent introgression, in particular non-random mating in the ancestral population due to population structure (Eriksson and Manica 2012). It is therefore important to make use of additional means to distinguish between these alternative hypotheses, for example, by examining the size of introgressed tracts (Wall et al. 2013), or the level of absolute divergence in introgressed regions (Smith and Kronforst 2013).

The *D* statistic was originally designed to be applied on a genome-wide or chromosome-wide scale, with block-jackknifing used to overcome the problem of non-independence between loci (Green et al. 2010). However, many researchers are interested in identifying particular genomic regions subject to gene flow, rather than simply estimating a genome-wide parameter. Theory predicts that the rate of gene flow should vary across the genome, both in the case of secondary contact after isolation (Barton and Gale 1993) as well as continuous gene flow during speciation (Wu 2001). Indeed, a maximum likelihood test for speciation with gene flow devised by Yang (2010) is based on detecting this underlying heterogeneity. Moreover, adaptive introgression might lead to highly localized signals of introgression, limited to the particular loci under selection.

Recent genomic studies have used F_{ST} to characterize heterogeneity in divergence across the genome, often interpreting the variation in F_{ST} as indicative of variation in rates of gene flow (for example, Hohenlohe et al. 2010). However, it is well established that, as a relative measure of divergence, F_{ST} is dependent on the within-population genetic diversity (Charlesworth 1998), and is therefore an unreliable indicator of how migration rates vary across the genome. In particular, heterogeneity in purifying selection and recombination rate could confound F_{ST}-based studies (Noor et al. 2009, Hahn et al. 2012, Roesti et al. 2012, Cruickshank and Hahn, 2014). Several studies have begun to explore new approaches to characterize heterogeneity in patterns of introgression among small genomic regions. Green et al. (2010) identified candidate Neanderthal introgression regions in human genomes by locating 50 kb windows that showed deep coalescences in human populations. On average, these windows showed an excess of Neanderthal alleles, consistent with introgression. Garrigan et al. (2012) developed a likelihood ratio test to identify genomic windows that have been shared between *Drosophila* species. Models with and without gene flow were evaluated over 1 and 5 kb genomic windows. Windows showing a significantly better fit to the gene flow model were widely distributed across the autosomes, but scarce on the Z chromosome. Roux et al. (2013) used an ABC framework to fit models of single and variable introgression rates among protein-coding loci in *Ciona* species. Their results indicated strong heterogeneity among loci, and unidirectional introgression consistent with multiple incompatibilities between species.

There have also been recent attempts to characterize heterogeneity in patterns of introgression across the genome using the *D* statistic, calculated either in small windows (Smith and Kronforst 2013, Kronforst et al. 2013) or for individual SNPs (Rheindt et al. 2014). However, it is not clear how reliably *D* can be used to find the location of individual introgressed loci, and what factors might influence it. Any inherent biases of the *D* statistic when applied to specific loci have implications for methods that assume its robustness. For example Smith and Kronforst (2013) made use of the *D* statistic in a proposed test to distinguish between the hypotheses of introgression and shared ancestral variation at wing-patterning loci of *Heliconius* butterflies. Two wing patterning loci are known to show an excess of shared derived alleles between co-mimetic populations of *Heliconius melpomene* and *Heliconius timareta* (*Heliconius* Genome Consortium, 2012). At one of these loci, phylogenetic evidence and patterns of linkage disequilibrium are consistent with recent gene flow (Pardo-Diaz et al. 2012). Nevertheless, Smith and Kronforst (2013) argue that this shared variation might represent an ancestral polymorphism that was maintained through the speciation event by balancing selection. Conceptually, this is not unlike the population structure argument of Eriksson and Manica (2012), except that here structure is limited to one or a few individual loci.

Smith and Kronforst proposed that these alternatives could be distinguished by calculating absolute divergence. Introgression should lead to more recent coalescence and reduced divergence at the affected loci, while the locus-specific structure hypothesis should lead to an excess of shared derived alleles, but no reduction in absolute divergence compared to other loci in the genome. Loci showing evidence of shared ancestry were located by calculating the *D* statistic for each 5 kb window, and identifying outliers using an arbitrary cutoff (the 10% of windows with the highest *D* values). The mean absolute genetic divergence (d_{XY}) was then compared between the outliers and non-outliers, and found to be significantly lower in outlier windows, consistent with recent introgression (Smith and Kronforst 2013). This method makes two assumptions: firstly, that the *D* statistic can accurately identify regions that carry a significant excess of shared variation, and secondly, that *D* outliers do not have inherent biases leading to their co-occurrence with regions of low absolute divergence.

The robustness of the *D* statistic for detecting a genome-wide excess of shared derived alleles has been thoroughly explored (Green et al. 2010; Durand et al. 2011; Yang et al. 2012; Eaton and Ree 2013; Martin et al. 2013). However, it has not been established whether *D* provides a robust and bias-free means to identify individual loci that have introgressed. In the present study, we first assess the reliability of the *D* statistic as a means to quantify introgression. Using simulations of small sequence windows, we compare *D* to a related statistic that was developed by Green et al. (2010) specifically for estimating *f,* the proportion of the genome that has been shared, and we propose improvements to this statistic. We then use whole-genome data from several *Heliconius* species to investigate how these statistics perform on empirical data, and specifically how they are influenced by underlying heterogeneity in diversity across the genome. Lastly, we use a large range of simulated data sets to test the proposal that recent gene flow can be distinguished from shared ancestral variation based on absolute divergence in *D* outlier regions.

RESULTS

The *D* statistic is not an unbiased estimator of gene flow

Patterson's *D* statistic was developed to detect, but not to quantify introgression. To test how sensitive the value of *D* is to various factors apart from the proportion of introgression, we used the derivation of Durand et al. (2011, Equation 5) to examine how various factors affect the expected value of D. Here, the proportion of introgression (*f*), corresponds to the proportion of haplotypes in the recipient population (P_2) that trace their ancestry through the donor population (P_3) at the time of gene flow (Figure 1A). The expected *D* value increases with the proportion of introgression (*f*), but not linearly (Figure 1B). Importantly, expected *D* increases as population size decreases (Figure 1B,C). The split times between populations also have a small effect, with a more recent split between P_1 and P_2 leading to higher expected values of *D* (Figure 1C). This indicates that the value of the *D* statistic is dependent on several parameters other than the amount of gene flow.

Direct estimators of *f* outperform *D* on simulated data

Analysis of simulated data confirmed that the *D* statistic is not an appropriate measure for quantifying introgression over small genomic windows, but that direct estimation of the proportion of introgression (*f*), provides a more robust alternative. The *D* statistic (Equation 1, Methods) was compared to three related estimators of *f* (Equations 4, 5 and 6, Methods). To compare the utility of these statistics for quantifying introgression in small genomic windows, we simulated sequences from four populations: P_1, P_2 and P_3 and outgroup O, with a single instantaneous gene flow event, either from P_3 to P_2 or from P_2 to P_3. Simulations were performed over a range of different values of *f* (the probability that any particular haplotype is shared during the introgression event), and with various window sizes, recombination rates and times of gene flow.

A subset of the results are shown in Figure 2, and full results are provided in Figure S1 A-I. In general, the *D* statistic proved sensitive to the occurrence of introgression, with strongly positive values for any non-zero value of *f*, but a poor estimator of the absolute value of *f* (Figure 2). Moreover, *D* values showed dramatic variance, particularly at low simulated values of *f*. Even in the absence of any gene flow, a considerable proportion of windows had intermediate *D* values. This variance decreased only marginally with increasing window size and recombination rate (Figures S1A-I).

We compared *D* to the *f* estimator of Green at al. (2010) (Equation 4), which is referred to as f_G below, along with two proposed modified versions of this statistic (Equations 5 and 6). The first, f_{hom} (Equation 5), is similar to f_G in that it explicitly assumes unidirectional gene flow from P_3 to P_2, but makes a further assumption that maximal introgression would lead to complete homogenization of allele frequencies in P_2 and P_3. This is a conservative assumption, as an extremely high rate of migration would be necessary to attain a maximal value of f_{hom}. The second, f_d (Equation 6), is dynamic in that it allows for bidirectional introgression on a site-by-site basis, setting the donor population at each site as that which has the higher frequency of the derived allele. In simulations of gene flow from P_3 to P_2, all three *f* estimators gave fairly accurate estimates of the simulated *f* value, provided gene flow was recent (Figures 2, S1A). When gene flow occurred further back in time, *f* estimators tended to give underestimates, but were nevertheless well correlated with the simulated *f* value (Figure S1A). In simulations of gene flow in the opposite direction, from P_2 to P_3, both f_G and f_{hom} showed considerable stochasticity, particularly when recombination rates were low and gene flow was recent (Figure S1A-I). The absolute size of the window had little effect on this behavior (S1A-I), implying that it was not an effect of the number of sites analyzed, but rather the level of independence among sites. Unlike these two statistics, f_d behaved predictably at all recombination rates and times of gene flow, giving estimates that were fairly well correlated with the simulated *f*, but underestimating its absolute value (Figure 2, Figure S1A-I). Generally, the variance in f_d was lower than in the other two *f* estimators (Figure S1A-I). Importantly, unlike the *D* statistic, f_d displayed minimal variance at low simulated values of *f* (Figure 2).

Although none of the examined measures was able to accurately quantify both forms of introgression in all cases, f_d showed some appealing characteristics as a measure to identify introgressed loci in a genome scan approach. It had low variance and was not prone to false-positives when gene flow was absent and recombination rare. In all cases, it provided estimates that were proportional to the simulated level of introgression. Although it tended toward underestimates, genome scans for introgressed loci would primarily be interested in relative rates of introgression across the genome, rather than absolute rates.

f estimation is robust to variation in nucleotide diversity across the genome

Analysis of published *Heliconius* whole-genome data confirmed that Patterson's *D* statistic was prone to extreme values in regions of low diversity, whereas *f* estimators were not. We re-analyzed published whole-genome sequence data from two closely-related *Heliconius* butterfly species, *Heliconius melpomene* and *Heliconius timareta*, and four outgroup species from the related silvaniform clade. The races *H. melpomene amaryllis* and *H. timareta thelxinoe* are sympatric in Peru, and show genome-wide evidence of gene flow (Martin et al. 2013), with particularly strong signals at two wing-patterning loci: *HmB*, which controls red pattern elements, and *HmYb* which controls yellow and white pattern elements (*Heliconius* Genome Consortium 2012, Pardo-Diaz et al. 2012). To determine whether heterogeneity in diversity across the genome may influence the *D* and *f* statistics, these were calculated, along with nucleotide diversity, in non-overlapping 5 kb windows across the genome. Variance in the *D* statistic was highest among windows with low nucleotide diversity (Figure 3A), and decreased rapidly with increasing diversity (Figure S3). Windows from the wing patterning loci were among those with the highest *D* values, but there were large numbers of additional windows with D=1 or thereabouts. By contrast, f_d, calculated for all windows with positive *D*, was far less sensitive to the level of diversity, with most outlying windows showing intermediate levels of diversity (Figures 3B). Notable exceptions were windows located within the wing-patterning regions, which tended to have high f_d values and below average diversity. This is consistent with the strong selection known to act upon the patterning loci. Most importantly, the vast majority of windows that had the highest *D* values were not among those with the highest *f* estimates, except for those at the wing patterning loci. Based on the simulation results above, we suggest therefore that most of the *D* outliers were spurious, and that f_d provides a better measure of whether a locus has been subject to introgression. Finally, we also tested the other two *f* estimators described here: f_G and f_{hom} (Equations 4 and 5). Both performed similarly to f_d except that both had higher variance (Figures S2,S3), and both gave a considerable number of values greater than one, confirming that f_d was the most conservative and stable statistic.

Taken together, these findings demonstrate that, when small genomic windows are analyzed, a high *D* value alone is not sufficient evidence for introgression. Many of the *D* outlier loci probably represent statistical noise, concentrated in regions of low diversity, whereas f_d outliers tend to be less biased.

This effect could also be observed on the scale of whole chromosomes. The variance in *D* among 5 kb windows for each chromosome was strongly negatively correlated with the average diversity per chromosome (r(19)=-0.936, p<0.001) (Figure 3C). This relationship was most clearly illustrated by the Z chromosome: it had the lowest diversity by some margin, as expected given its reduced effective population size, and the highest variance among *D* values for 5 kb windows, despite the fact that previous analyses suggest very limited gene flow affecting this chromosome (Martin et al. 2013). By contrast, the variance in f_d, estimated for all windows with positive *D*, had a weak positive correlation with the mean diversity per chromosome (r(19)=0.440, p<0.05). This was driven by the fact that the Z chromosome had the lowest diversity and also the lowest variance in f_d, as expected given the reduced gene flow affecting this chromosome. When the Z chromosome was excluded, there was no significant relationship between the variance in f_d values and average diversity (r(18)=0.092, p>0.05). In summary, these data show that extreme *D* values, both positive and negative, occur disproportionately in genomic regions with lower diversity, whereas f_d values are less biased by underlying heterogeneity in genetic variation.

Inherent biases in the *D* and *f* statistics confound a test to distinguish between introgression and shared ancestral variation

The biases associated with the *D* statistic described above may have important consequences for methods that use *D* to identify candidate introgressed regions. For example, Smith and Kronforst (2013) proposed a method to discriminate between gene flow and shared ancestral variation that relies upon *D* values calculated for short genomic windows. Briefly, the Smith and Kronforst test calculated *D* for all 5 kb windows. Absolute divergence (d_{XY}) was then compared between windows that were outliers (top 10%) for the *D* statistic and the remaining non-outlier windows. It makes intuitive sense that introgression between species at a specific genomic region should reduce the between-species divergence here as compared to the rest of the genome, while shared ancestry due to ancestral population structure would not lead to lower divergence. We first confirmed this prediction using simulations, and then assessed whether biases in the *D* statistic might affect the power of the method.

To test the prediction that introgression and ancestral population structure leave distinct footprints in terms of absolute divergence, 10 000 sequence windows for three populations and an outgroup were simulated. 9000 windows were defined as 'Background', having the topology (((P_1,P_2),P_3),O), without any gene flow or population structure. The remaining 1000 windows were defined as 'Alternate' and were subject to either gene flow or structure (see Methods for details). Ten percent of windows were defined as Alternate to match Smith and Kronforst's design, wherein the top 10% of *D* values are taken as outliers. Three different Alternate scenarios were considered: gene flow from P_2 to P_3, gene flow from P_3 to P_2 and ancestral structure leading to shared ancestry between P_2 and P_3. The ancestral structure scenario is intended to model a region of the genome undergoing balancing selection or some other process that maintains polymorphism at particular loci prior to the speciation event. This was achieved by setting the topology of the simulated Alternate windows to ((P_1,(P_2,P_3)),O) and altering the split times (Figure 4A-D). As a result, gene flow or structure in the Alternate windows can be considered to be complete (f=1). For example, under gene flow from P_2 to P_3, all P_3 alleles trace their ancestry through P_2 at the time of gene flow. This simplified design, where gene flow is absent in 90% of the sequences and complete in 10%, although biologically unlikely, allowed for the most straight-forward and predictable test of Smith and Kronforst's method; if the logic of the method does not follow in this design, it is unlikely to do so in more complex situations.

For each of the three evolutionary scenarios, 120 different permutations of split times and times of gene flow or structure were simulated (Table S1). To simplify our comparisons between models, we focused specifically on d_{XY} between P_2 and P_3, the most relevant parameter when testing for introgression between P_2 and P_3. As predicted, in simulations using a recombination rate parameter (4Nr) of 0.01, in all models simulating gene flow, average d_{XY} between P_2 and P_3 was significantly lower in Alternate windows compared to Background windows (p<4e-05 in all cases, 99% with Bonferroni correction over 240 tests; see Figure 4E,F for examples). In contrast, in all models simulating ancestral population structure, there was no significant difference in P_2-P_3 d_{XY} between the background and alternate windows, again in agreement with predictions (see Figure 4E,F for examples). These findings therefore demonstrate that the intuitive premise of Smith and Kronforst's (2013) method is justified.

We then tested whether introgression could be distinguished from shared ancestral variation where loci with shared ancestry are not known (as would be the situation with empirical data), but are instead inferred by selecting the top 10% of *D* values (outliers), following the Smith and Kronforst method. We also tested this method using the top 10% of *f* estimates among windows with positive *D* (using Equations 4, 5 and 6). Using the *D* statistic to identify outliers, mean d_{XY} between P_2 and P_3 was significantly reduced in outlier windows as compared to non-outlier windows in 239 of 240 gene flow models (p<4e-05, 99% with Bonferroni correction over 240 tests; Table S1, Figures 4E,4F,5). The single non-significant case had gene flow from P_2 to P_3, the most ancient possible t_{23} and the most recent possible t_{12}, with *D* Outliers identifying only 11.9% of the Alternate windows (Table S1). Using any of the three *f* estimators, mean d_{XY} between P_2 and P_3 was significantly reduced in outlier windows in all 240 gene flow models (for each estimator, p<4e-05, 99% with Bonferroni correction over 240 tests; Table S1, Figures 4E,4F,5).

However, mean P_2-P_3 d_{XY} was also significantly reduced in *D* and *f* outliers in many of the 120 models simulating ancestral population structure (*D*: 104 models, f_G: 67 models, f_{hom}: 65 models, f_d: 70 models, for each estimator, significant tests had p<4e-05, 99% significance level with Bonferroni correction over 240 tests; Table S1, Figures 4G,5). This reduction tended to be much weaker than those under most gene flow models (Figures 4G,5), but was nevertheless often significant. This demonstrates that a simple test for reduced divergence in P_2-P_3 d_{XY} among *D* or *f* outlier windows would, under a range of ancestral structure scenarios, produced results consistent with introgression. The fact that this bias was similar whether *D* or *f* estimators were used to identify outliers indicates that there is an inherent tendency in all of these statistics toward regions with below-average divergence between P_2 and P_3. To confirm this finding, we analyzed a set of simulations using a null model, with no gene flow or structure in any of the 10 000 windows, over 45 permutations of split times. Outlier windows showed significantly reduced d_{XY} between P_2 and P_3 in most or all of the null models (*D*: 39 models, f_G: 45 models, f_{hom}: 44 models, f_d: 44 models, for each estimator, significant tests had p<2e-04, 99% significance level with Bonferroni correction over 45 tests; Table S1, Figures 4H,5). Finally, we repeated all of these simulations with a lower within-window recombination rate parameter (4Nr) of 0.01. This exaggerated the problems, with at most 3 of the ancestral structure models showing non-significant drops in P_2-P_3 d_{XY} for outliers, whether they were defined by any of the statistics (for each statistic significant tests had p<4e-05, 99% significance level with Bonferroni correction over 240 tests; Table S1, Figures 4K,5). Similarly, 32 of the 45 null models showed significant drops in P_2-P_3 d_{XY} for *D* outliers and all 45 null models, showed significant drops for outliers defined by the three *f* estimators (for each statistic p<2e-04, 99% significance level with Bonferroni correction over 45 tests; Table S1, Figure 4L).

In summary, although shared ancestral variation and introgression can theoretically be distinguished based on the fact that only the latter should reduce d_{XY} between P_2 and P_3, an inherent bias in both the *D* and *f* statistics makes a simple test for a statistical difference in d_{XY} between outliers and non-outliers problematic. Both *D* and *f* outliers tended toward windows with lower P_2-P_3 d_{XY}, regardless of the underlying evolutionary history, and particularly when recombination rates were low. There were clear differences in the magnitude of the decrease in P_2-P_3 d_{XY} between outlier and non-outlier windows that could potentially be used to distinguish introgression from shared ancestral variation, although a more sophisticated model-fitting approach would be necessary.

DISCUSSION

With the advent of population genomics, studies of species divergence have moved from simply documenting inter-specific gene flow, towards the identification of specific genomic regions that show strong signals of either introgression or divergence (Heliconius Genome Consortium, 2012; Garrigan et al. 2012; Staubach et al. 2012; Roux et al. 2013). This is a useful goal for many reasons. It can permit the identification of large-scale trends, such as chromosomal differences, and the fine-scale localization of putative targets of adaptive introgression for further characterization. Therefore, simple and easily computable statistics that can be used to identify loci with a history of introgression have considerable appeal. We have shown here that Patterson's *D,* which has recently been used for this purpose, is not reliable when applied on a small scale. Instead, estimation of *f*, the proportion of introgression, particularly using our proposed statistic f_d, provides a better means to identify putatively introgressed regions. Nevertheless, both *D* and f_d tend to identify regions of reduced inter-species divergence, even in the absence of gene flow, which may confound tests to distinguish between recent introgression and shared ancestral variation based on absolute divergence in outlier regions.

Previous studies have explored the behavior of Patterson's *D* statistic, a test for gene flow based on detecting an inequality in the numbers of ABBA and BABA patterns, using whole genome analyses across large numbers of informative sites (Green et al. 2010; Yang et al. 2012; Eaton and Ree 2013; Martin et al. 2013; Wall et al. 2013). These studies have shown that *D* is a robust method to test for an excess of shared variation on a genome-wide scale. In particular, the non-independence among linked sites can be accounted for by block-jackknifing. However, the behavior of *D* in short genomic regions has not been previously investigated. Here, we document two main problems with the *D* statistic as a means to identify introgressed loci. Firstly, it is not an unbiased estimator of the amount of introgression that has occurred. In particular it is influenced by effective population size (N_e), leading to more extreme values when N_e is low. Secondly, when calculated over small windows, it is highly stochastic, particularly in genomic regions of low diversity and low recombination rate, such that *D* outliers will tend to be clustered within these regions. Local reductions in genetic diversity along a chromosome can come about through neutral processes, such as population bottlenecks, but also through directional selection. Therefore, these problems may be exacerbated in studies specifically interested in loci that experience strong selective pressures, as this would increase the likelihood of detecting chance outliers at such loci.

Direct estimation of *f,* the proportion of introgression, holds more promise as a robust method for detecting introgressed loci. Green et al. (2010) proposed that *f* could be estimated by comparing the observed difference in the number of ABBA and BABA patterns to that which would be expected in the event of complete introgression. As this expected value is calculated from the observed data, this method controls for differences in the level of standing variation, making it more suitable for application to small regions. In Green et al.'s approach, complete introgression from P_3 to P_2 was taken to mean that P_2 would come to resemble a subpopulation of lineage P_3. Here we make the conservative assumption that complete introgression would lead to homogenization of allele frequencies, such that the frequency of the derived allele in P_2 would be identical to that in P_3. Green et al.'s approach assumed unidirectional introgression from P_3 to P_2, but can lead to spurious values when introgression occurs in the opposite direction. We have therefore proposed a new, 'dynamic', estimator of *f,* in which the donor population can differ between sites, and is always the population with the higher frequency of the derived allele. Although this conservative estimator leads to slight underestimation of the amount of introgression that has occurred, it provides an estimate that is roughly proportional to the level of introgression, regardless of the direction. It is therefore a more suitable measure for identifying introgressed loci. This is supported by our analysis of whole-genome data from *Heliconius* butterflies, where many 5 kb windows had maximal D values (D=1) , but only a few had high f_d values, the vast majority of which were located around the wing patterning loci previously identified as being shared between these species through adaptive introgression (*Heliconius* Genome Consortium 2012, Pardo-Diaz et al. 2012).

The sensitivity of the D statistic to heterogeneous genomic diversity is likely to affect studies that have drawn conclusions from D statistics calculated for particular genome regions. For example, Wall et al. (2013) showed that long (8-100kb) haplotypes segregating in human populations showed evidence of a Neanderthal origin, as indicated by elevated D statistics. However, it may be that such haplotypes would be over-represented in low-recombination regions, which also tend to have reduced diversity in humans and many other species (Cutter and Payseur 2013). In another recent *Heliconius* study, F_{ST} was calculated for 5kb windows across the genome. Windows showing increased differentiation between *H. melpomene* and *H. pachinus* (according to F_{ST}) also showed significantly elevated *D* statistics in a test for introgression between the same species pair (Kronforst et al. 2013). This illustrates how the sensitivity of both of the statistics to within-species diversity can produce conflicting results. This may especially be the case for studies using small genomic regions; at the extreme, Rheindt et al. (2014) calculated *D* for single SNPs and predicted that genes linked to SNPs with outlying *D* values are more likely to have been introgressed.

In the present study, we asked whether biases in the *D* statistic could influence a recently proposed method to distinguish between introgression and shared ancestral variation (Smith and Kronforst 2013). The premise of this test is that introgression should result in an excess of shared derive alleles and a reduction in absolute divergence, whereas shared ancestral variation will exhibit the former but not the latter signature. Smith and Kronforst identified 5 kb windows that were outliers for *D*, and compared the absolute divergence (d_{XY}) among populations in these outlier windows with d_{XY} at non-outlier windows. In particular, reduced divergence between P_2 and P_3 among the outlier windows would be consistent with gene flow. Our simulations confirmed that the intuitive predictions of this method are valid, but also showed that this test can be misled by the use of *D* to identify outliers. Windows that were outliers for *D* exhibited below average d_{XY} between P_2 and P_3, even in many simulations where gene flow or ancestral population structure were absent. Somewhat surprisingly, all three *f* estimators showed a similar bias, implying that it does not simply reflect the more stochastic nature of the *D* statistic. We hypothesize that *D* would have additional problems in real genomes, where selective constraint leads to a correlation between within-species diversity and between-species divergence, causing *D* outliers to be even more strongly associated with reduced d_{XY}. It is notable that the reduction in divergence among *D* and *f* outliers was almost always greater in simulations with introgression, across a large range of split times and dates of gene flow. There may, therefore, be considerable information about the evolutionary history of DNA sequences present in the joint distribution of d_{XY} and f_d. On the other hand, in real data, levels of divergence can vary dramatically due to heterogeneity in selective constraint, mutation rate and recombination rate, which would exaggerate the problems described here.

Conclusions

In an era of increasing availability of genomic data, there is a demand for simple summary statistics that can reliably identify genomic regions that have been subject to selection, introgression and other evolutionary processes. It seems unlikely, however, that any of these processes can be reliably distinguished from the genomic variability caused by demography and drift by any single summary statistic. This is most convincingly illustrated by the literature on selective sweeps, where ever more complex inferences from sequence data are being developed to infer a history of selection (Li et al. 2012). Here we have shown that, while Patterson's *D* statistic provides a robust signal of shared ancestry across the genome, it should not be used for naïve scans to ascribe shared ancestry to small genomic regions, due to its tendency toward extreme values in regions of reduced variation. Estimation of *f,* particularly using f_d, provides a better tool for the identification of introgressed loci. Analysis of d_{XY} among *D* and *f* outliers to distinguish introgression from shared ancestral variation can be problematic. However, the joint distribution of d_{XY} and *f* statistics may be a useful summary statistic for model-fitting approaches to distinguish between these evolutionary hypotheses.

MATERIAL AND METHODS

Statistics used to detect shared ancestry

In this study, we focused on an approach to identify an excess of shared derived polymorphisms, indicated by the relative abundance of two SNP patterns termed "ABBAs" and "BABAs" (Green et al. 2010). Given three populations and an outgroup with the relationship (((P_1, P_2), P_3), O) (Figure 1A), ABBAs are sites at which the derived allele "B" is shared between the non-sister taxa P_2 and P_3, while P_1 carries the ancestral allele, as defined by the outgroup. Similarly, BABAs are sites at which the derived allele is shared between P_1 and P_3, while P_2 carries the ancestral allele. Under a neutral coalescent model, both patterns can only result from incomplete lineage sorting or recurrent mutation, and should be equally abundant in the genome (Durand et al. 2011). A significant excess of ABBAs over BABAs is indicative either of gene flow between P_2 and P_3, or some form of non-random mating or structure in the population ancestral to P_1, P_2 and P_3. This excess can be tested for, using Patterson's *D* statistic,

$$D(P_1, P_2, P_3, O) = \frac{\sum C_{ABBA}(i) - C_{BABA}(i)}{\sum C_{ABBA}(i) + C_{BABA}(i)} \qquad (1)$$

where $C_{ABBA}(i)$ and $C_{BABA}(i)$ are counts of either 1 or 0, depending on whether or not the specified pattern (ABBA or BABA) is observed at site *i* in the genome. Under the null hypothesis of no gene flow and random mating in the ancestral population, *D* will approach zero, regardless of differences in effective population sizes (Durand et al. 2011). Hence, a *D* significantly greater than zero is indicative of a significant excess of shared derived alleles between P_2 and P_3.

If population samples are used, then rather than binary counts of fixed ABBA and BABA sites, the frequency of the derived allele at each site in each population can be used (Green et al. 2010, Durand et al. 2011), effectively weighting each segregating site according to its fit to the ABBA or BABA pattern, with

$$C_{ABBA}(i) = (1 - \hat{p}_{i1})\,\hat{p}_{i2}\,\hat{p}_{i3}\,(1 - \hat{p}_{i4}) \qquad (2)$$

$$C_{BABA}(i) = \hat{p}_{i1}\,(1 - \hat{p}_{i2})\,\hat{p}_{i3}\,(1 - \hat{p}_{i4}) \qquad (3)$$

where p_{ij} is the frequency of the derived allele at site *i* in population *j*. These values are then used in equation 1 to calculate *D* (Durand et al. 2011).

Green et al. (2010) also proposed a related method to estimate *f*, the fraction of the genome shared through introgression (Green et al. 2010, Durand et al. 2011). This method makes use of the numerator of equation 1, the difference between sums of ABBAs and BABAs, which is called *S*. In the example described above, with $((P_1,P_2),P_3),O)$, the proportion of the genome that has been shared between P_2 and P_3 subsequent to the split between P_1 and P_2 can be estimated by comparing the observed value of *S* to a value estimated under a scenario of complete introgression from P_3 to P_2. P_2 would then resemble a lineage of the P_3 taxon, and so the denominator of equation 1 can be estimated by replacing P_2 in equations 2 and 3 with a second lineage sampled from P_3, or by splitting the P_3 sample into two,

$$\hat{f}_G = \frac{S(P_1, P_2, P_3, O)}{S(P_1, P_{3a}, P_{3b}, O)} \qquad (4)$$

where P_{3a} and P_{3b} are the two lineages sampled from P_3. Splitting P_3 arbitrarily in this way may lead to stochastic errors at individual sites, particularly with small sample sizes. These should be negligible when whole-genome data are analyzed but could easily lead to erroneous values of *f* (including *f*>1) when small genomic windows are analyzed, as in the present study. We therefore used a more conservative version, in which we assume that complete introgression from P_3 to P_2 would lead to complete homogenization of allele frequencies. Hence, in the denominator, P_{3a} and P_{3b} are both substituted by P_3:

$$\hat{f}_{hom} = \frac{S(P_1, P_2, P_3, O)}{S(P_1, P_3, P_3, O)} \qquad (5)$$

While this conservative assumption may lead to underestimation of the proportion of sites shared, it also reduces the rate of stochastic error. Moreover, in the present study, we are less concerned with the absolute value of *f*, and more with the relative values of *f* between genomic regions.

The *f* statistic assumes unidirectional gene flow from P_3 to P_2 (i.e. P_3 is the donor and P_2 is the recipient). Since the branch leading to P_3 is longer than that leading to P_2 (Figure 1A), gene flow in the opposite direction (P_2 to P_3) is likely to generate fewer ABBAs. Thus, in the presence of gene flow from P_2 to P_3, or in both directions, the *f* equation should lead to an underestimate. However, when small genomic windows are analyzed, the assumption of unidirectional gene flow could lead to overestimates, because any region in which derived alleles are present in both P_2 and P_3, but happen to be at higher frequency in P_2, will yield *f* estimates that are greater than 1. Thus, we propose a dynamic estimator in which the denominator is calculated by defining a donor population (P_D) for each site independently. For each site, P_D is the population (either P_2 or P_3) that has the higher frequency of the derived allele, thus maximizing the denominator and eliminating *f* estimates greater than 1:

$$\hat{f}_d = \frac{S(P_1, P_2, P_3, O)}{S(P_1, P_D, P_D, O)} \qquad (6)$$

Assessing the ability of *D* and *f* estimators to quantify introgression in small sequence windows

To assess how reliably Patterson's *D* statistic, and other estimators of *f* are able to quantify the actual rate of introgression, we simulated sequence datasets with differing rates of introgression using ms (Hudson 2002). For each dataset, we simulated 100 sequence windows for 8 haplotypes each from four populations with the relationship (((P_1,P_2),P_3),O). The split times t_{12} and t_{23} (as on Figure 1A) were set to 1 × 4N generations and 2 × 4N generations ago, respectively, and the root was set to × 4N generations ago. An instantaneous, unidirectional admixture event, either from P_3 to P_2 or from P_2 to P_3, was simulated at a time t_{GF} with a value *f*, which determines the probability that each haplotype is shared. We tested two different values for t_{GF}: 0.1 and 0.5 × 4N generations ago. For each direction of gene flow and each t_{GF}, 11 simulated datasets were produced, with *f* values ranging from 0 (no gene flow) to 1 (all haplotypes are shared). Finally, the entire set of simulations was repeated with three different window sizes: 1, 5 and 10 kb, and with three different recombination rates: 0.001, 0.01 and 0.1, in units of 4Nr, the population recombination rate. DNA sequences were generated from the simulated trees using Seq-Gen (Rambaut & Grass 1997), with the HKY substitution model and a branch scaling factor of 0.01. Simulations were run using the compare_f_estimators.r, which generates the ms and Seq-Gen commands automatically. An example set of commands to simulate a single 5kb sequence using the split times mentioned above, with gene flow from P_3 to P_2 at t_{GF} = 0.1 and f = 0.2, and with a recombination rate parameter of 0.01 would be:

```
ms 32 1 -I 4 8 8 8 8 -ej 1 2 1 -ej 2 3 1 -ej 3 4 1 -es 0.1 2 1 -ej 0.1 5 3 -r 50 5000 -T | tail -n +4 | grep -v // >
treefile
partitions=($(wc -l treefile))
seq-gen -mHKY -l 5000 -s 0.01 -p $partitions <treefile >seqfile
```

We then compared the mean and standard error for *D* (Equation 1) and the three *f* estimators (Equations 4, 5 and 6), calculated for all 100 windows in each dataset.

Analysis of *Heliconius* whole genome sequence data

To investigate how the *D* and *f* statistics are affected by underlying diversity in a given window, we re-analyzed whole genome data from Martin et al. (2013). For ABBA BABA analyses, populations were defined as follows: P_1 = *Heliconius melpomene aglaope* (4 diploid samples), P_2 = *Heliconius melpomene amaryllis* (4), P_3 = *Heliconius timareta thelxinoe* (4), O=*Heliconius hecale* (1), *Heliconius ethilla* (1), *Heliconius pardalinus sergestus* (1), and *Heliconius pardalinus* sp. Nov. (1). Patterson's *D* (Equation 1) and the three *f* estimators (Equations 4,5,6) were calculated, along with nucleotide diversity (π) and absolute divergence (d_{XY}), for non-overlapping 5 kb windows across the genome. Both π and d_{XY} were calculated as the mean number of differences between each pair of individuals, sampled either from the same population (π), or from separate populations (d_{XY}). Sites with missing data were excluded in a pairwise manner, and each pair of individuals contributed equally to the mean. Windows were restricted to single scaffolds and windows for which fewer than 3000 sites had genotype calls for at least half of the individuals were discarded. To calculate *D* and the *f* estimators only bi-allelic sites were considered. The ancestral state was inferred using the outgroup taxa, except when the four outgroup taxa were not fixed for the same allele, in which case the most common allele overall was taken as ancestral. The *HmB* locus was defined as postions 300 000 to 450 000 on scaffold HE670865 and the *HmYb* locus as positions 650 000 to 900 000 on scaffold HE667780. We also analyzed windows from each of the 21 chromosomes of the *H. m. melpomene* genome sequence separately. Scaffolds were assigned to chromosomes according to the *Heliconius* Genome Consortium (2012), and incorporating the improved assignment of Z-linked scaffolds by Martin et al. (2013) (details available in Dryad repositories http://dx.doi.org/10.5061/dryad.m27qq and http://dx.doi.org/10.5061/dryad.dk712). This analysis was performed using egglib_sliding_windows.py, and figures were generated using Figures_3_S2_S3.R.

Assessing a test to distinguish introgression from shared ancestral variation based on absolute divergence

Smith and Kronforst (2013) proposed a simple test to distinguish between the hypotheses of pre- and post-speciation shared ancestry based on absolute divergence. To assess this method on data of known history, we generated a large range of sequence datasets using ms (Hudson 2002) and Seq-Gen (Rambaut & Grass 1997). For the simplest ('null') model 10 000 5kb sequence windows were simulated for 8 haplotypes each from three populations and an outgroup, with the relationship (((P_1,P_2),P_3),O), without gene flow or population structure. To approximate a scenario in which a subset of the genome has a distinct phylogenetic history, either due to gene flow or genomically-localized ancestral population structure, we used a combined model approach. This entailed combining 9000 5kb windows from the null model (90% “Background” windows), with 1000 5kb windows simulated under with the topology ((P_1,(P_2,P_3)),O), consistent with shared ancestry between P_2 and P_3 (10% “Alternate” windows). By altering the split times, three distinct scenarios were emulated: Gene flow from P_2 to P_3, gene flow from P_3 to P_2, and ancestral structure (Figure 4A-D). Using entirely distinct topologies in this way is equivalent to making the probability of gene flow (or structure) equal to one in the 1000 Alternate windows. While this approach of partitioning each dataset into two somewhat arbitrarily-sized subsets with evolutionary histories at two extremes is biologically unrealistic, it provided a simple and powerful framework in which to evaluate Smith and Kronforst's approach, with clear expectations. Model combination datasets were generated using run_model_combinations.py and shared_ancestry_simulator.R, which generates the ms and Seq-Gen commands automatically, in a similar form to those given above. For example if t_{12} = 1, t_{23} = 2, 4Nr = 0.01 and gene flow from P_3 to P_2 at t_{GF} = 0.2, the ms calls for Background and Alternate models, respectively, would be:

```
ms 32 1 -I 4 8 8 8 8 -ej 1 2 1 -ej 2 3 1 -ej 3 4 1 -r 50 5000 -T
ms 32 1 -I 4 8 8 8 8 -ej 0.2 2 3 -ej 2 3 1 -ej 3 4 1 -r 50 5000 -T
```

We calculated Patterson's *D* (Equation 1) and the three *f* estimators (Equations 4,5,6) for all windows, and identified the top 1000 'outliers' (10%) with the most extreme values. For *D*, only positive values were included as outliers, as negative values indicate an excess of BABAs, consistent with introgression between P_1 and P_3. Similarly, for *f* estimators, only windows with D≥0 were considered, as these values only give meaningful quantification of introgression when there is an excess of ABBAs. To compare P_2-P_3 divergence between the Background and Alternate windows, or between outlier and non-outlier windows, we calculated d_{XY} for each window as described above, for each pair of populations. Average d_{XY} was compared between subsets of windows using a Wilcoxon rank-sum test, as values tended to be non-normally distributed (confirmed with Bonferroni-corrected Shapiro-Wilk tests).

These tests were repeated over a large range of split times. In all cases the root was set to 3.0 × 4N generations ago, and the other splits ranged from 0.2 to 2.0. Times of gene flow and structure also varied same scale. In total this gave 45 null models and 120 models each for the two gene flow scenarios and ancestral structure (405 overall). The analyzed models therefore covered a vast range of biologically relevant scales. In all cases, the Seq-Gen branch scaling factor was set to 0.01. Full parameters for all models are provided in Table S1. Finally, to examine the effects of recombination rate, the entire simulation study was repeated using population recombination rate (4Nr) values of 0.01 and 0.001. Summary statistics for all models were compiled using generate_summary_statistics.R.

Software

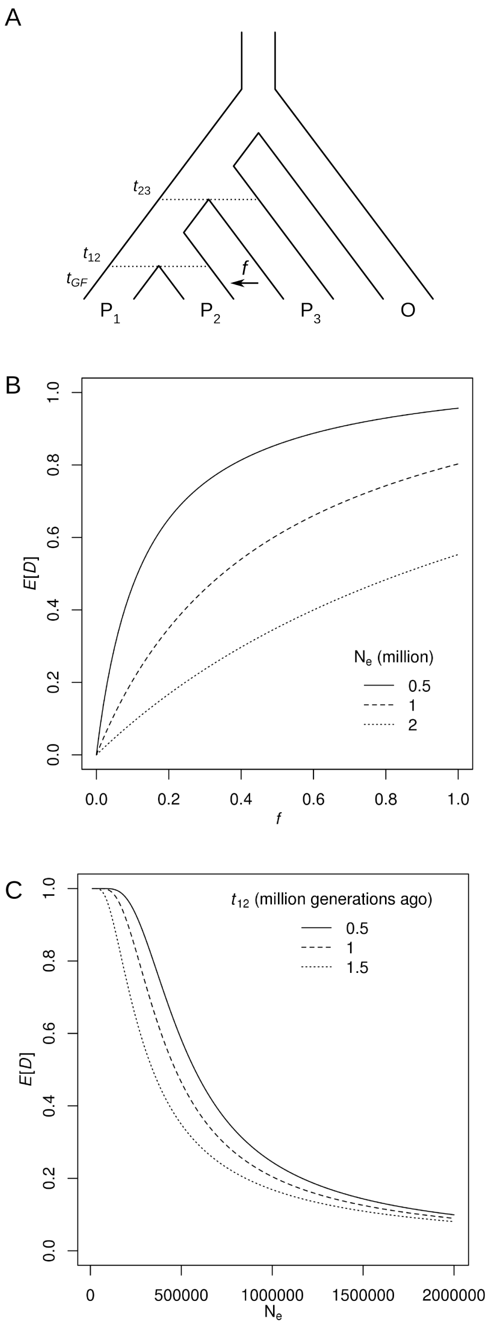
Code and data for this manuscript will be made available as a Data Dryad repository on acceptance. Most files, with instructions for running scripts to generate the results, are currently available on GitHub at https://github.com/johnomics/Martin_Davey_Jiggins_evaluating_introgression_statistics. Large data sets can be made available on request and will be made publicly available after review. This work was made possible by the free, open source software packages EggLib (De Mita and Siol 2012), phyclust (Chen 2011), R (R Core Team 2013), ggplot2 (Wickham 2009), plyr (Wickham 2011), reshape (Wickham 2007) and Inkscape (http://www.inkscape.org).

ACKNOWLEDGEMENTS

We are grateful to Milan Malinsky for helpful and detailed discussions about this work. We also thank Jim Mallet and Marcus Kronforst for comments on the manuscript. Richard Merrill and Richard Wallbank contributed to initial discussions that inspired the work. Finally, this work was supported by the Leverhulme Trust (F/09364/E to C.D.J.) and the Herchel Smith Fund (Postdoctoral Fellowship to J.W.D.).

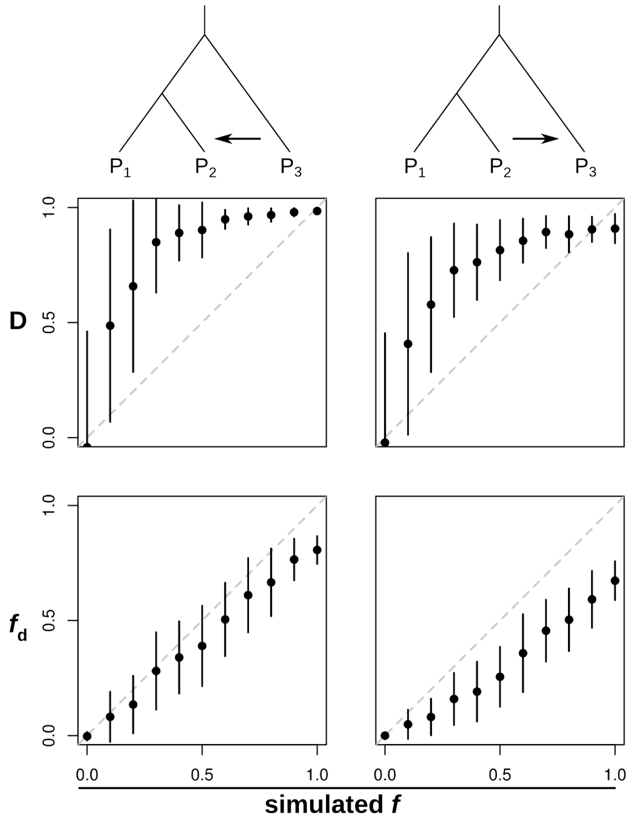
FIGURES

Figure 1. Expected value of the *D* statistic



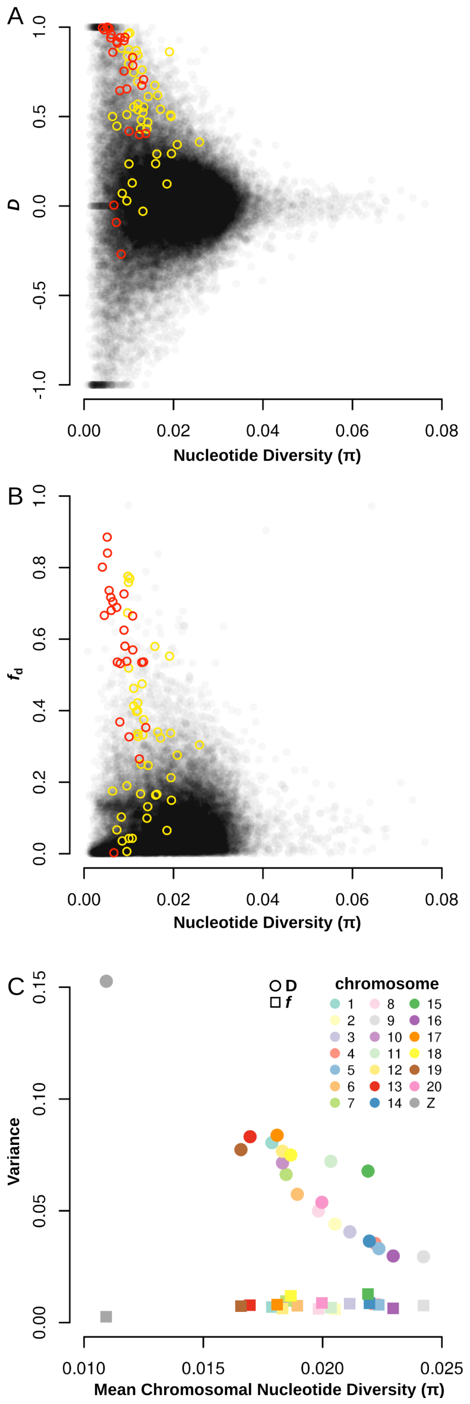
A. This derivation from Durand et al. (2011), of the expected value of Patterson'*s D* statistic *E*[*D*], depends on the two split times, t_{12} and t_{23}, separating populations P_1, P_2 and P_3. It assumes a single instantaneous admixture event from P_3 to P_2 at t_{GF}, after which a proportion *f* of P_2 individuals trace their ancestry through P_3. The effective population size, N_e, is constant through time and the same in all populations. **B.** The expected value of *D* as a function of *f*, the proportion of introgression, at three different effective population sizes: 0.5, 1 and 2 million. Split times are fixed at 1 million generations for t_{12} and 2 million generations for t_{23}. **C.** The expected value of *D* as a function of N_e, showing the effect of varying t_{12}. In all three cases, t_{23} is set at 2 million generations ago, and *f* is set to 0.1.

Figure 2. Comparing statistics to detect and quantify introgression



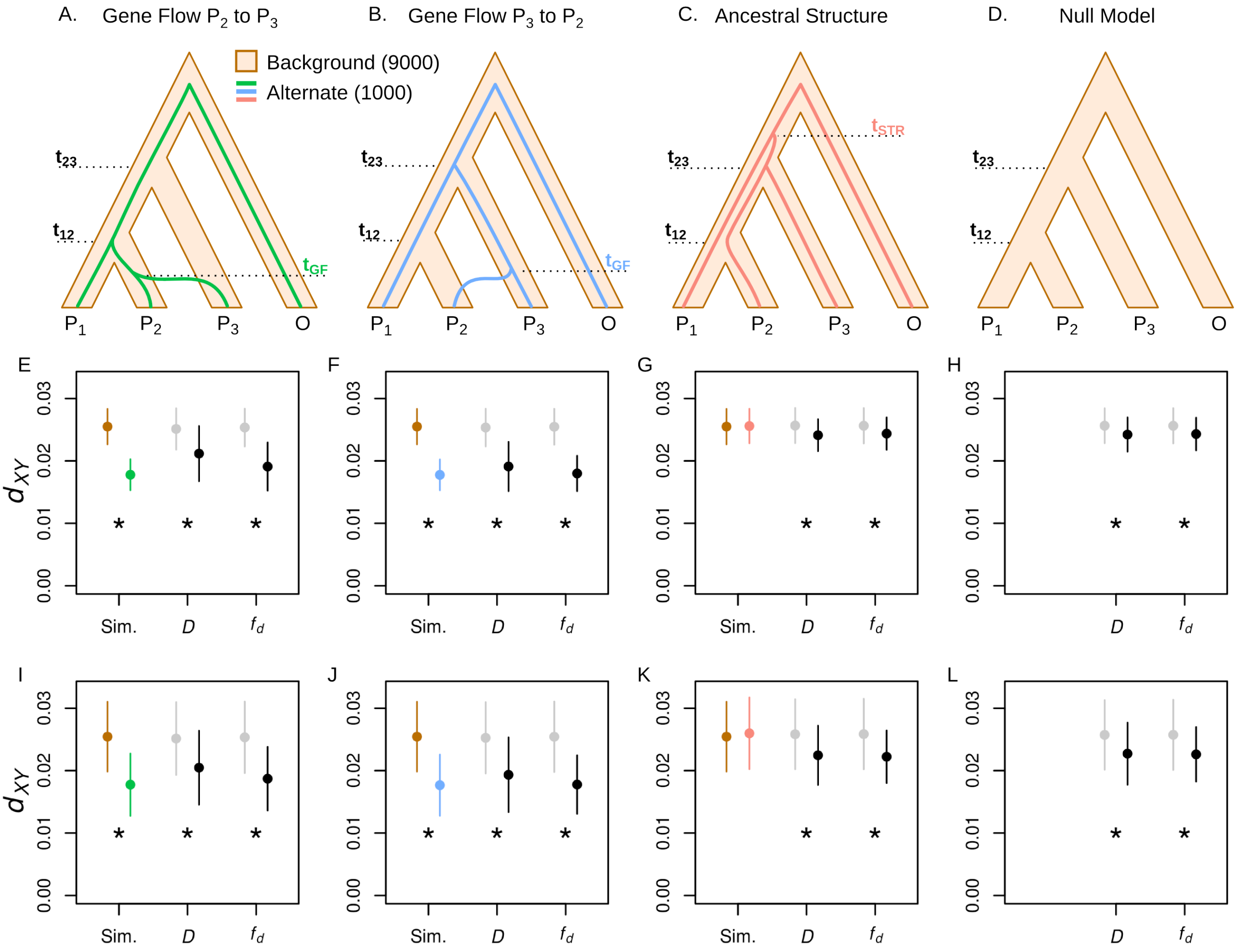
Results from a subset of the simulations: window size 5kb, time of gene flow (t_{GF}) 0.1 × 4N generations ago, and population recombination rate 0.01. See Figures S1A-I for full results. Plots show means and standard deviations for *D* and f_d, calculated over 100 simulated sequences (See Methods for details). Simulations covered 11 different values of *f,* the proportion of introgression. Gene flow was simulated either from P_3 to P_2 (conventional model) (left-hand column) or from P_2 to P_3 (right-hand column). Dashed diagonal lines show the expectation of a perfect estimator of *f*.

Figure 3. Effects of genetic diversity on the *D* and f_d in *Heliconius* whole genome data



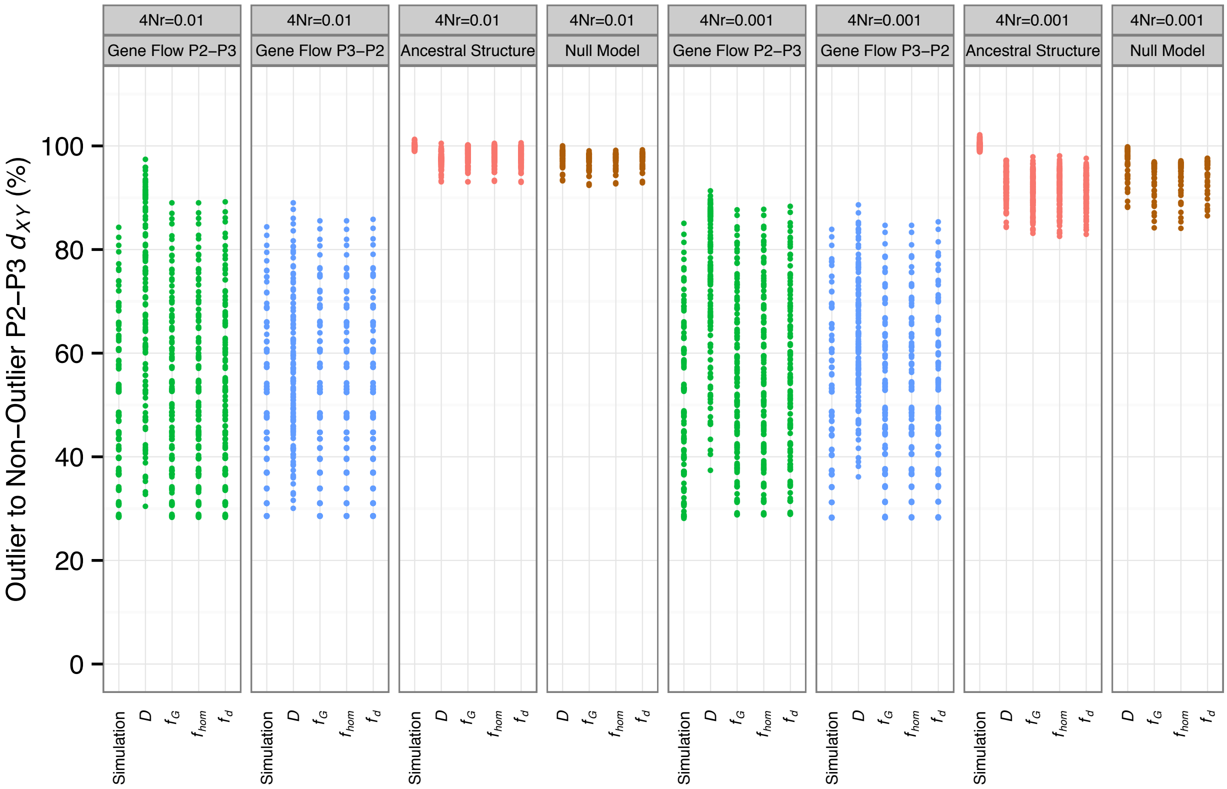
A,B. Values of *D* and f_d for non-overlapping 5 kb windows across the genome, plotted against nucleotide diversity. *f* values are only plotted for windows with D≥0. Data from Martin et al. 2013. Taxa used are as follows, P_1: *Heliconius melpomene aglaope*, P_2: *Heliconius melpomene amaryllis*, P_3: *Heliconius timareta thelxinoe*, O: four *Heliconius* species from the silvaniform clade. Colored points show windows located within the wing patterning loci *HmB* (red) and *HmYb* (yellow), see Methods. **C.** The variance among *D* and f_d values for each chromosome, plotted against the mean nucleotide diversity from all windows for each chromosome.

Figure 4. Simulations to evaluate a method to distinguish introgression from shared ancestral variation



A-C. Combined models were made up of 9000 sequence windows simulated under the "Background" topology (brown outline) and 1000 windows simulated under an "Alternate" topology (colored line). Three distinct evolutionary scenarios were simulated by varying the split times t_{12}, t_{23}, t_{GF} and t_{STR}; **A, E, I**: gene flow from P$_2$ to P$_3$, **B, F, J**: gene flow from P$_3$ to P$_2$, **C, G, K**: ancestral structure. **D, H, L.** Null models were made up of 10 000 sequences simulated under the Background topology only. **E-L.** Example data from a single simulated dataset for each of the four types of models. Split times (in units of 4N generations) were as follows: t_{12} = 0.6 in all four cases, t_{23} = 0.8 in all four cases, t_{GF} = 0.4 in both gene flow models and t_{STR} = 1.0. Points show mean and standard deviation for P$_2$-P$_3$ d_{XY} calculated over subsets of trees: simulated Background and Alternate trees (brown and colored points) or non-outliers and outliers (gray and black points) identified using the *D* and f_d statistics. A significant reduction in P$_2$-P$_3$ d_{XY} for the Alternate compared to Background windows, or for outliers compared to non-outliers, is indicated by astrices. **E-H** show results of simulations with a population recombination rate (4Nr) of 0.01. **I-L** show results for the same models, but with a population recombination rate (4Nr) of 0.001.

Figure 5. Mean d_{XY} between P_2 and P_3 in outlier windows as a percentage of P_2-P_3 d_{XY} in non-outlier windows.



Outlier windows defined by Alternate or Background topology (Simulation) or by outlying *D* and *f* values, as per Figure 4. Model types shown in color (gene flow from P_2 to P_3, green; gene flow from P_3 to P_2, blue; ancestral structure, red; null model, brown). Results for two different recombination rates are shown (4Nr=0.01, left; 4Nr=0.001, right).

REFERENCES

Abbott R, Albach D, Ansell S, et al. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Barton NH, Gale KS. 1993. Genetic analysis of hybrid zones. In: Price J, Harrison RG, editors. Hybrid Zones and the evolutionary process. USA: Oxford University Press.

Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. 15:538–543.

Chen W-C. 2011. Overlapping Codon model, Phylogenetic Clustering, and Alternative Partial Expectation Conditional Maximization Algorithm, Ph.D. Diss., Iowa Stat University.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. In Press

Cutter AD, Payseur B a. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14:262–274.

De Mita S, Siol M. 2012. EggLib: processing, analysis and simulation tools for population genetics and genomics. BMC Genet. 13:27.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239–2252.

Eaton D, Ree R. 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 682:689-706.

Eriksson A, Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. Proc. Natl. Acad. Sci. U. S. A. 109:13956–13960.

Garrigan D, Kingan SB, Geneva AJ, Andolfatto P, Clark AG, Thornton K, Presgraves DC. 2012. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. Genome Res. 22:1499–1511.

Green RE, Krause J, Briggs AW, et al. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.

Hahn MW, White BJ, Muir CD, Besansky NJ. 2012. No evidence for biased co-transmission of speciation islands in Anopheles gambiae. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367:374–384.

The *Heliconius* Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98.

Hohenlohe PA, Bassham S, Currey M, Cresko WA. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos. Trans. R. Soc. London B. Biol. Sci. 367:395–408.

Kronforst MR, Hansen MEB, Crawford NG, Gallant JR, Zhang W, Kulathinal RJ, Kapan DD, Mullen SP. 2013. Hybridization reveals the evolving genomic architecture of speciation. Cell Rep. 5:666–677.

Kulathinal RJ, Stevison LS, Noor MAF. 2009. The Genomics of Speciation in Drosophila: Diversity, Divergence, and Introgression Estimated Using Low- Coverage Genome Sequencing. PLoS Genet. 5(7):e1000550.

Li J, Li H, Jakobsson M, Li S, Sjödin P, Lascoux M. 2012. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Mol. Ecol. 21:28–44.

Mallet J, Barton N. 1989. Strong natural selection in a warning-color hybrid zone. Evolution 43:421–431.

Martin SH, Dasmahapatra KK, Nadeau NJ, et al. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23:1817–1828.

Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity. 103:439–444.

Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, McMillan WO, Jiggins CD. 2012. Adaptive introgression across species boundaries in Heliconius butterflies. PLoS Genet. 8:e1002752.

Pinho C, Hey J. 2010. Divergence with Gene Flow: Models and Data. Annu. Rev. Ecol. Evol. Syst. 41:215–230.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rambaut A, Grass NC. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Bioinformatics 13:235–238.

Rheindt FE, Fujita MK, Wilton PR, Edwards S V. 2014. Introgression and phenotypic assimilation in Zimmerius flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. Syst. Biol. 63:134–152.

Roesti M, Hendry AP, Salzburger W, Berner D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. Mol. Ecol. 21:2852–2862.

Roux C, Tsagkogeorga G, Bierne N, Galtier N. 2013. Crossing the Species Barrier: Genomic Hotspots of Introgression between Two Highly Divergent *Ciona intestinalis* Species. Mol. Biol. Evol. 30:1574–1587.

Smith J, Kronforst MR. 2013. Do *Heliconius* butterfly species exchange mimicry alleles? Biol. Lett. 9:20130503.

Staubach F, Lorenc A, Messer PW, Tang K, Petrov DA, Tautz D. 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (Mus musculus). PLoS Genet. 8:e1002891.

Wall JD, Yang MA, Jay F, et al. 2013. Higher levels of neanderthal ancestry in East Asians than in Europeans. Genetics 194:199–209.

Wickham H. 2007. Reshaping data with the reshape package. J. Stat. Softw. 21(12).

Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer (New York).

Wickham H. 2011. The Split-Apply-Combine Strategy for Data Analysis. J. Stat. Softw. 40(1):1-29.

Wu C. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Yang MA, Malaspinas A-S, Durand EY, Slatkin M. 2012. Ancient structure in Africa unlikely to explain Neanderthal and non-African genetic similarity. Mol. Biol. Evol. 29:2987–2995

Yang Z. 2010. A likelihood ratio test of speciation with gene flow using genomic sequence data. Genome Biol. Evol. 2:200–211.