The relationship between gene flow and population difference during secondary contact between butterfly sister species

Erik D. Nelson*, Qian Cong and Nick V. Grishin

Department of Biophysics, University of Texas Southwestern Medical Center,

6001 Forest Park Blvd., Room ND10.124, Dallas, Texas 75235–9050

∗ E-mail: nelsonerikd@gmail.com

Abstract

Closely related species of butterfly sampled from southern suture zones in North America exhibit a continuous pattern of gene flow and population difference measures (index values) for autosomes, but not for the Z chromosome; When populations are compared through their Z chromosomes, index values obtained from samples of the same species are separated from those of closely related species by a gap of "missing" values, suggesting a discrete "on–off" criterion for species delimitation. Here, we explore the possibility that some, or all of the index data for suture zones reflects secondary contact between species formed in glacial refugia. We simulate fusion of butterfly populations limited by negative fitness interactions between genes in hybrids, assuming that interactions between autosomes and the Z chromosome are stronger than those among autosomes, and that hybrid fitness effects conform to Haldane's rule. We find that weakly interbreeding populations trace out a path toward equilibrium consistent with the data for butterfly suture zones, in which index values for the Z chromosome lag behind those for autosomes, leading to a similar gap of missing values when species become indistinguishable through their autosomes, but no evidence of a sudden change in index values for the Z chromosome on longer timescales. As a result, we find that the gap can be explained by a process in which the pattern of index data for the Z chromosome is, ultimately, continuous.

Introduction

A species can be defined as a population of organisms that maintains its genomic integrity despite interbreeding, and exchange of genetic material with other populations [1]. In practice, species are delineated by comparing genomic sequences using simple measures of genetic distance; Genomes sampled from a population describe members of the same species if they are significantly more similar to each other than they are to genomes from closely related populations [2]. A large amount of effort has been devoted to the improvement of methods for comparing and characterizing natural populations [3–7]. However, since speciation and fusion of populations take place over extremely long time scales, the dynamics of these events can only be inferred from simulations, and samples of existing organisms.

It is worth noting the resemblance of this situation to the one encountered in a very different problem from the field of astrophysics – specifically, in describing the early evolution of stars toward their positions on the main sequence [8]. Long ago, it was found that by observing stars at various stages of their evolution in young galactic clusters, their evolutionary paths could be interpreted from the resulting pattern of points on a plot of stellar luminosity versus temperature, known as a Hertzsprung–Russell diagram [9,10]. In the present context, stars in a galactic cluster might be likened to pairs of interbreeding species belonging to a particular family of organisms, and the pattern of data exhibited by such pairs could provide similar insight into the conditions that regulate species integrity.

Recently, Cong et al. have collected exactly this sort of data in their study of butterfly sister species across suture zones in North America [2]. To determine the status of populations, they likewise employed two basic indicators: (i) The index of gene flow, I_{gf}, a measure of their own invention, which describes the fraction of genes in a population imported from its sister population, and (ii) the fixation index, F_{st}, a standard genetic

measure of population difference [5]. Multiple genomic samples were collected from each population, and separate indices were computed for autosomes and the Z chromosome. The main results of their work are shown in Fig. 1;

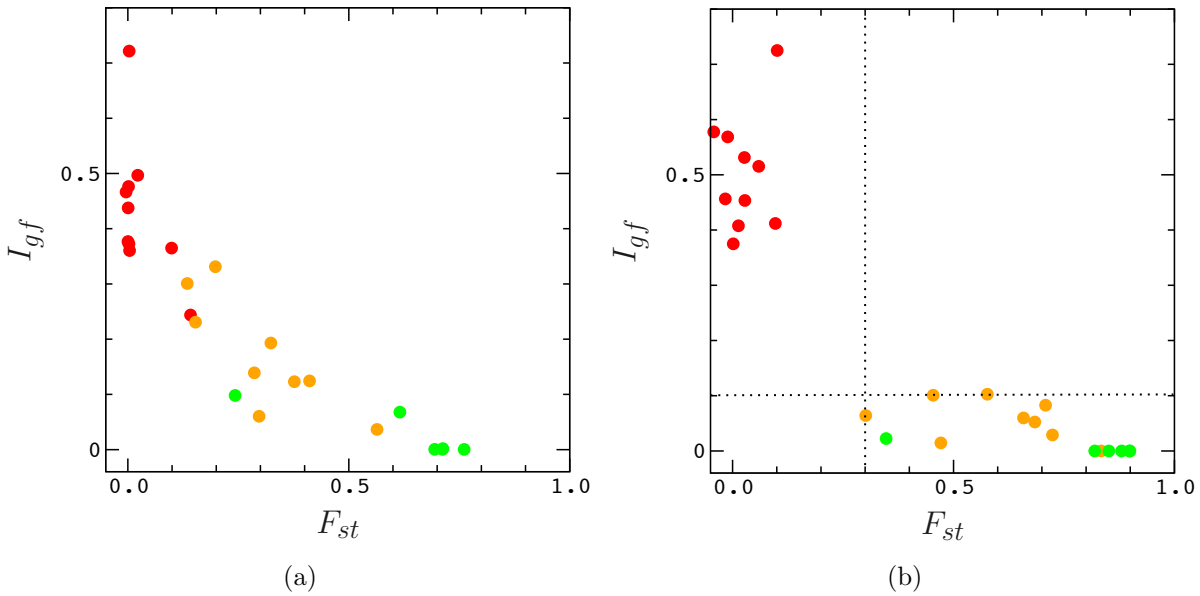


Fig. 1. Index of gene flow (I_{gf}) and fixation index (F_{st}) values for autosomes (A) and Z chromosomes (B) of sister species sampled by Cong et al. The data for I_{gf} is multiplied by a factor of 4 for agreement with the definition of I_{gf} used in this work. Data points describe pairs of organisms that have been classified as different species in the literature (green), closely related organisms for which classification is uncertain (yellow), and organisms of the same species (red). The dotted lines in panel B are included to guide the eye.

The panels in this figure describe index values for pairs of organisms that have been classified as different species in the literature (green), closely related organisms for which classification is uncertain (yellow), and organisms from the same species (red). When populations are compared through their autosomes (Fig. 1A), the data exhibit a continuous pattern across the entire range of index values; However, for the Z chromosome (Fig. 1B), the data obtained from samples of the same species (red) are separated from those of closely related species by a gap of "missing" values, suggesting a sudden transition be-

tween "phases" – for example, as might result from a sudden gain or loss of incompatible interactions with the Z chromosome in a population destined for speciation, or in sister populations destined for fusion, respectively. The suture zones sampled by Cong et al. appear to result from the migration of species from glacial refugia along the coastlines bordering the gulf of California and the gulf of Mexico [11], which suggests that the latter scenario (secondary, or higher order contact between sister species) may have more relevance to their study. For different species (i.e., green and yellow data points), fixation index values for the Z chromosome are always larger than those for the autosomes (Fig. 2). At the same time, the fraction of divergent positions in butterfly sister genomes is roughly the same for autosomes and the Z chromosome (see Fig. 5A of reference [2]).

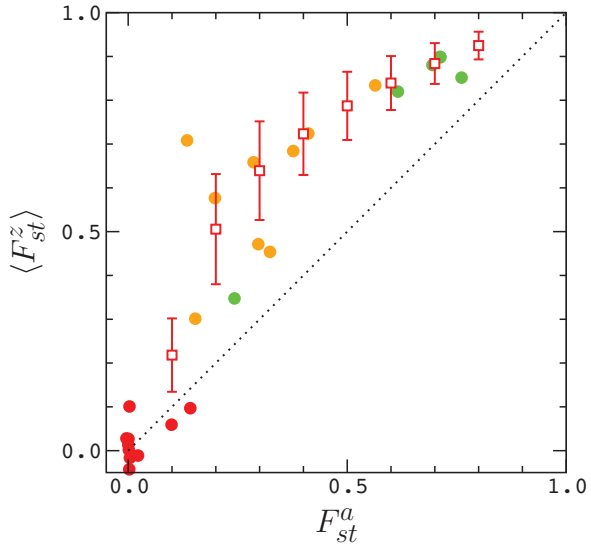


Fig. 2. Correlation between fixation indices for the Z chromosome, F_{st}^z, and the autosomes, F_{st}^a. The data and the color scheme is from Fig. 1. The dotted line, $F_{st}^z = F_{st}^a$, is included to guide the eye. Included for comparison is a plot of the average, $\langle F_{st}^z \rangle$, versus F_{st}^a (squares) for a model in which hybridization is limited by negative interactions between genes the autosomes and the Z chromosome (see below). Error bars indicate standard deviations in the model data; Errors in the means (not shown) fit within the symbols.

Together, these results indicate a slower rate of introgression among Z chromosomes, and suggest that mutations on the Z chromosome play a larger role in hybrid incompatibility (i.e., consistent with the large X effect [6, 12]).

In this work, we consider the possibility that some, or all of the data sampled by Cong et al. reflects secondary contact between species formed in glacial refugia. To explore the relationship between gene flow and population difference, we simulate fusion [13] of butterfly populations using a model developed by Orr and Turelli to describe the fitness costs of interactions between incompatible genes in hybrids [14–16]. In accord with the butterfly data, we assume that hybrid interactions between genes in autosomes and the Z chromosome are stronger than those among autosomes, and that hybrid fitness effects conform to Haldane's rule [12]. We find that weakly interbreeding populations trace out a path to equilibrium consistent with the data, in which indices for the Z chromosome approach their equilibrium values more slowly than those for autosomes, leading to a similar gap of missing index values when populations become indistinguishable through their autosomes, but without a sudden change in index values for the Z chromosome on longer timescales. As a result, we find that the missing values can be explained by a process in which the pattern of data for I^z_{gf} and F^z_{st} is, ultimately, continuous.

Model

The dynamics of interbreeding between butterfly populations are obviously very complicated, involving effects due to geography, climate, resource availability, mobility, and predation, to name a few [17]; Populations can undergo repeated episodes of contact and isolation, and significant changes in size that can affect the rate of change of indices for the Z chromosome relative to the autosome [6]. Here, we neglect many of these effects in favor of a more intelligible model that still incorporates some of the basic features of

butterfly genetics.

In our model, sister species are represented by populations (demes) of equal size that exchange randomly selected pairs of individuals in each generation. Populations evolve by plain Wright–Fisher dynamics, with random mating between male and female individuals: In each generation, a number of individuals (with mean m) are exchanged between demes. In each deme, male and female individuals are selected randomly (with replacement) in proportion to fitness to undergo mating. The genomes of selected individuals undergo explicit meiosis [18], with random recombination in males (meiosis is achiasmatic in female butterflies [6, 19]), producing a single tetrad of gametes for each individual. A single offspring is generated in each mating event by random union of male and female gametes, and the process continues until the initial populations are replenished.

To describe the effects of hybridization on fitness, we use a well known model of hybrid incompatibility developed by Orr [15]. The model is an extension of the Dobzhansky–Muller model to arbitrary numbers of interacting loci that Orr used to estimate the rate of increase of negative interactions between pairs of genes in first order hybrids formed from diverging populations. To obtain his estimate, Orr traced the acquisition of fixed mutations at random loci in two, initially identical, isolated populations, assuming that populations are monomorphic between fixation events (e.g. as in the weak mutation / strong selection limit [20]). To explain this process, let $\mathsf{g}(k)$ and $\mathsf{g}'(k)$ denote genomes representing the two populations after k substitutions. Initially, $\mathsf{g}(0) = \mathsf{g}'(0)$. It is assumed that mutations in g fix at different loci than mutations in g', and that g and g' acquire fixed mutations (substitutions) at the same rate; Multiple substitutions at the same locus are neglected. After the first substitution, each additional substitution in g (for example) occurs on a different background than it would in g', which leads to potential negative interactions with alleles at loci in g' that differ from those in g when g and g'

are combined in hybrids (i.e., because the substitution in g has not been "tested" in the background of alleles that differ in g'); Similar reasoning also holds for substitutions in g'. It was shown by Orr that the number of such interactions increases as the square of the total number of substitutions, in agreement with results for Drosophila [21]. The process is described graphically in Fig. 1 of reference [15].

In this work, we simulate the equilibration of weakly interbreeding populations starting from an advanced point in Orr's model. For simplicity, additional mutations are neglected. To construct the genomes $\mathsf{g}(k)$ and $\mathsf{g}'(k)$ used to initialize our simulations, we assign an equal number of fixed mutations to random loci in $\mathsf{g}(0)$ and $\mathsf{g}'(0)$ as prescribed above; Each mutation is assigned a random time index, and the temporal order of mutations is used to determine the set of potentially incompatible pairings between alleles in hybrids. Following Orr, we assume that each pairing has a small probability p of being incompatible. This smaller set of pairings (after random selection with probability p) is then assumed to exist in all hybrids formed during a simulation, consistent with the assumed independence of pair interactions in Orr's model [15].

To define the set of selected pairings, let G denote a general hybrid genome formed during a simulation, and let G_{kl} denote the allelic state of the gene at locus l in chromosome k, where $k \in [1, n]$, and n is the total (diploid) number of chromosomes in G; For simplicity, assume that homologous chromosomes are specified by sequential pairs of indices, $k \in \{1, 2\}, .. \{i-1, i\}, .. \{n-1, n\}$, with sex chromosomes specified by the last pair of indices, $k \in \{n-1, n\}$. To simplify the notation, we alternately use a single greek subscript to denote the pair of indices specifying the location of a particular gene – i.e., such that G_{μ} (for example) means the same thing as G_{kl}. Using this notation, we express the set of pairings between locations of genes in hybrids by the matrix $\mathsf{p}_{\mu\nu}$, where $\mathsf{p}_{\mu\nu} = 1$ when a pairing between alleles in g and g' connects gene location μ in g to gene location ν

in g', and $\mathsf{p}_{\mu\nu} = 0$ otherwise. We can then express the existence of a negative interaction between alleles G_μ and G_ν at locations μ and ν by the function,

$$\Delta_{\mu\nu}(\mathsf{G}) \;=\; \mathsf{p}_{\mu\nu}\,\delta_{\mathsf{G}_\mu,\,\mathsf{g}_\mu}\delta_{\mathsf{G}_\nu,\,\mathsf{g}'_\nu} \;+\; \mathsf{p}_{\nu\mu}\,\delta_{\mathsf{G}_\nu,\,\mathsf{g}_\nu}\delta_{\mathsf{G}_\mu,\,\mathsf{g}'_\mu} \tag{1}$$

where $\delta_{i,j}$ is the Kronecker function ($\delta_{i,j} = 1$ if $i = j$, and $\delta_{i,j} = 0$ otherwise). Note that if $\mathsf{p}_{\mu\nu} = 1$ then $\mathsf{p}_{\nu\mu} = 0$, since substitutions in g and g' occur at different loci. As a result, $\Delta_{\mu\nu} = 1$ when an interaction exists between alleles G_μ and G_ν, and $\Delta_{\mu\nu} = 0$ otherwise.

Below, we model butterfly genomes by $n = 40$ binary chromosomes of equal length L (similar to Heliconius genomes [19]) where $L = 100$ loci. Mutant and ancestral alleles in Orr's model are represented by binary alleles (1 and 0, respectively) in our model. To construct initial genomes for a simulation, we assign 10 fixed mutations to randomly selected loci in each pair of chromosomes in each genome, subject to the conditions described above. In selecting a set of pairings, we assume that hybrid interactions between the autosomes and the Z chromosome are dominant; All other pairings are neglected, and the individual fitness effects of mutations are considered neutral; In particular, the W chromosome acts only to determine the sex of an individual.

To define the fitness costs of interactions between selected pairs of alleles in hybrids, we explored two different models: In model (i), each pair of interacting alleles in a male genome G is assigned the same log fitness cost s, representing a typical, or average cost for incompatible genes in hybrids; The fitness of a male genome is then given by $w(\mathsf{G}) = \exp -s \sum_{\mu\nu} \Delta_{\mu\nu}(\mathsf{G})$. To compute the fitness of a female genome, alleles on the Z chromosome are treated as fixed alleles on the two Z chromosomes in a male genome; In this case, each term with $\Delta_{\mu\nu} = 1$ in a female genome contributes an amount $-2\,s$ rather than $-s$ to the exponent of $w(\mathsf{G})$. In model (ii), fitness is computed the same

way as in model (i) except that the effect of a pairing is recessive when both loci in $\Delta_{\mu\nu}$ are heterozygous; In that case, terms with $\Delta_{\mu\nu} = 1$ contribute an amount $-\epsilon\, s$ to the exponent of $w(\mathsf{G})$ where $\epsilon < 1$. This condition clearly affects male genomes only, since in female genomes, the locus on the Z chromosome is treated as homozygous. As a result, hybrid females are typically less fit than hybrid males in accordance with Haldane's rule. Since model (ii) is more realistic for butterfly populations, we focus on that model exclusively in this work. In all of the results below, the dominance factor is $\epsilon = 0.25$, the mean migration rate is $m = 1$ exchange per generation, and the mean recombination rate is $r = 1$ crossover per chromosome per meiosis, similar to the recombination rate in Heliconius [19].

Results

To compare model (ii) with the data in Fig. 1, we conduct simulations for various population sizes (N) and interaction frequencies (p), with fitness costs (s) selected so that F_{st} values for the model roughly approximate the data in Fig. 2. To generate data for a given N, p and s we conduct multiple simulations in parallel on 128 nodes of a high performance computer: On each node, we generate a different random interaction model (with frequency p), and for each model we conduct 10 trials; In each trial, populations begin in monomorphic states ($F_{st} = 1$) and are allowed to equilibrate via genetic drift and selection until they reach a state with (typically) $F_{st} \simeq 0.1$ for the autosomes, as in Fig. 1A (yellow and green data points).

The fixation index, and the index of gene flow are computed using methods similar to those described by Cong et al.; To define these objects, let $D_k(\mathsf{G}, \mathsf{G}')$ denote the number of pairwise differences (Hamming distance) between chromosome k in genome G and chromosome k in genome G', and let

$$D(\mathsf{G},\mathsf{G}') \;=\; D_1(\mathsf{G},\mathsf{G}') \;+\; D_3(\mathsf{G},\mathsf{G}') \;+\; ..\; D_{n-3}(\mathsf{G},\mathsf{G}') \tag{2}$$

denote the total distance between autosomes (for simplicity, we restrict genome comparisons to odd numbered chromosomes). The fixation index for autosomes is then

$$F_{st} \;=\; 1 \;-\; \frac{\langle D\rangle_{11} \;+\; \langle D\rangle_{22}}{2\,\langle D\rangle_{12}} \tag{3}$$

where braces, $\langle\ \rangle_{ij}$, denote averaging over genomes G and G' sampled from populations i and j, respectively, and $i,j \in \{1,2\}$. Due to the lower complexity of our model, we define the index of gene flow (I_{gf}) as the fraction of chromosomes (i.e., as opposed to transcript windows [2]) with $G_{min} \leq 0.25$, where

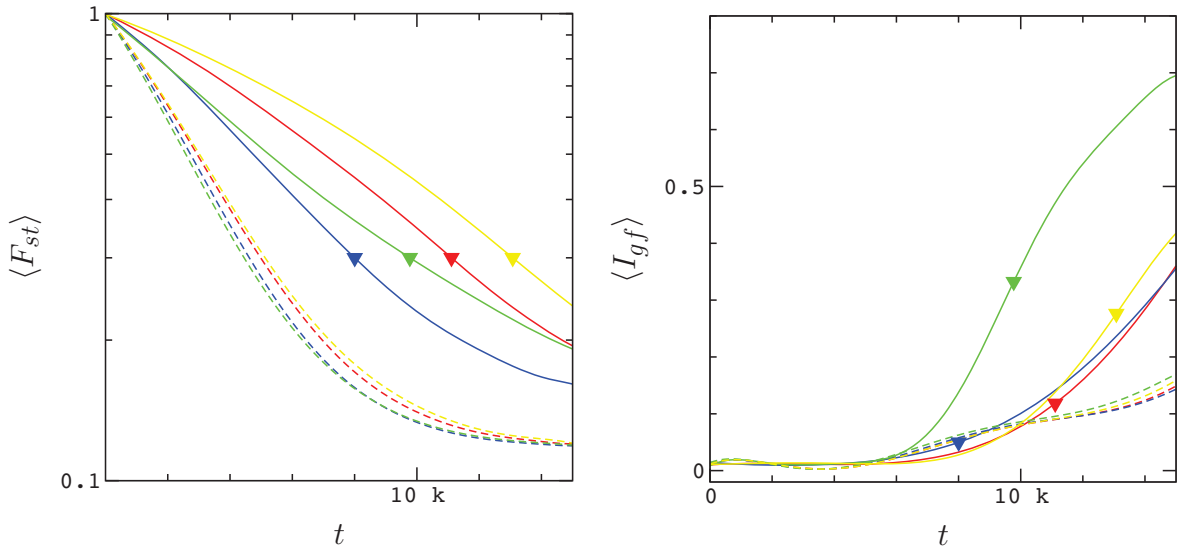
$$G_{min,k} \;=\; \frac{\mathsf{min}\, D_k}{\langle D_k\rangle_{12}} \tag{4}$$

and $\mathsf{min}\, D_k$ is the minimal value of $D_k(\mathsf{G},\mathsf{G}')$ obtained for the samples of genomes used to compute $\langle D_k\rangle_{12}$. Indices for the Z chromosome are obtained by substituting D_{n-1} for D and D_k in Eq.s (3) and (4) (by convention, the W chromosome is always the last chromosome in a female genome, so it plays no part in these expressions). Further details are provided in the Appendix.

We selected the model parameters s and p so that the data for F^z_{st} conforms, to varying degrees, with the pattern of data in Fig. 2 (see Fig. 5); Under these conditions, populations equilibrate more rapidly through their autosomes than their Z chromosomes leading to a pattern of data similar to that in Fig. 1;

To demonstrate this, we first averaged the index values over trials in a simulation at

regular points in time. Figure 3 provides a sample of these averages for simulations with $N = 8 \times 10^3$, where, for example, $\langle F^x_{st} \rangle(t)$ is the average value of F^x_{st} over all 1280 trials in a simulation after t generations, and $x \in \{z, a\}$;



(a) (b)

Fig. 3. Index of fixation $F_{st}(t)$ (A) and index of gene flow $I_{gf}(t)$ (B) averaged over trials sampled after t generations. Index values for the Z chromosome (autosomes) are described by solid (dashed) lines. The data describe simulations with $p = 0.00065$ (blue), $p = 0.0013$ (red), $p = 0.0026$ (yellow), and $p = 0.0065$ (green) corresponding to about 2, 5, 10, and 23 negative pairings per Z chromosome (on average) of strength $Ns = 92$, 60, 30, and 2.9, respectively. In both panels, colored triangles indicate the points in time when $\langle F^z_{st} \rangle(t) = 0.3$. For clarity, the results are represented by high precision polynomial fits to the averages (see supplementary Fig.s S1–S2 for a description of the fits and data statistics).

The data in Fig. 3 describe models with about 2 (blue), 5 (red), 10 (yellow), and 23 (green) negative pairings per Z chromosome (on average) of strength $Ns = 92$, 60, 30, and 2.9 respectively. It is evident by inspection of this data that, for certain sets of parameters, $\langle I^z_{gf} \rangle(t)$ is much smaller than its final value when $\langle F^z_{st} \rangle(t) \sim 0.3$ (triangles), the smallest value of F^z_{st} attained for different species in Fig. 1B. For example, for $p = 0.0013$ and

$Ns = 60$ (solid red lines), $\langle I^z_{gf}\rangle(t)$ is about 1/3 of its final value when $\langle F^z_{st}\rangle(t) \sim 0.3$ (red triangles). At the same time, $\langle F^a_{st}\rangle(t)$ is close to 0.1 and $\langle I^a_{gf}\rangle(t)$ is close to its final value, consistent with Fig. 1A. Note, however, that the final value of $\langle I^a_{gf}\rangle(t)$ reached in Fig. 3B is clearly smaller than expected from Fig. 1A. This discrepancy appears to result from the low complexity of our model – specifically, from the use of chromosomes rather than transcripts to compute G_{min} and I_{gf} in Eq. (4). We discuss this problem in more detail later below.

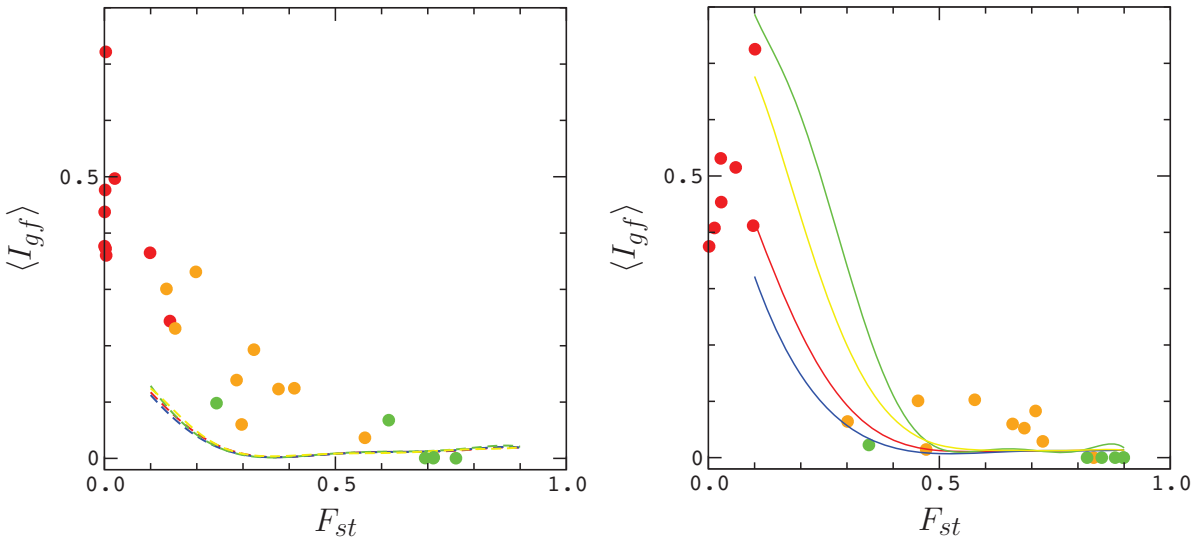


Fig. 4. Index of gene flow I_{gf} for the autosomes (A) and the Z chromosome (B) averaged over samples with a given value of F_{st}. The color scheme and simulation parameters correspond with those in Fig. 3. Butterfly data from Fig. 1 is included in respective panels for comparison. For clarity, the results are represented by high precision polynomial fits to the averages (see supplementary Fig. S2–S3 for a description of the fits and data statistics).

Similar results are obtained by averaging the index data over samples irrespective of time, as shown in Fig.s 4–5. In particular, for the simulation discussed above (red data), when $F^a_{st} \simeq 0.13$ in Fig. 5 we obtain $\langle F^z_{st}\rangle \simeq 0.3$, similar to Fig. 3A, and when $F^z_{st} \simeq 0.3$

in Fig. 4B, $\langle I^z_{gf}\rangle|_{F^z_{st}=0.3} \simeq 0.1$ is only about 1/4 of its maximum, or "same species" value, $\langle I^z_{gf}\rangle|_{F^z_{st}=0.1} \simeq 0.4$. Again, $\langle I^a_{gf}\rangle$ is smaller than expected from Fig. 1A when $F^a_{st} \lesssim 0.5$. Interestingly, there is a steep drop in $\langle F^z_{st}\rangle$ with F^a_{st} for all simulations when $F^a_{st} \lesssim 0.2$ in Fig. 5, while there is no sudden change in $\langle F^z_{st}\rangle(t)$ with time in Fig. 3A. Data statistics, errors in the means, and fits to the data in Fig.s 3–5 are described in the Supplementary Material.

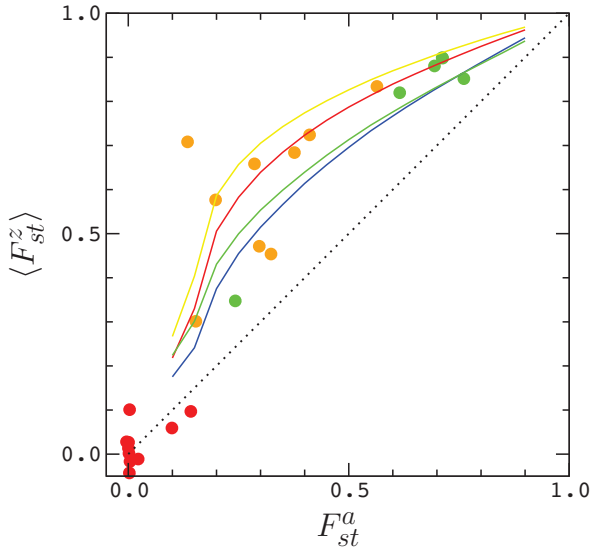


Fig. 5. Index of fixation for the Z chromosome, F^z_{st}, averaged over samples with a given value of F^a_{st}. The color scheme and the simulation parameters correspond with those in Fig. 3. Butterfly data from Fig. 2 is included for comparison. Errors in the means, and data widths are similar to those described for the model data in Fig. 2.

Because the correct population size needed to model suture zone data is unknown, and could be very large, it is also important to understand how the results of the model vary with N. For population sizes $N \sim 10^5$ (a small but reasonable value for Heliconius populations [6]), the amount of time needed to complete a simulation is about one month. Thus, for practical reasons, we decided to limit our study to a smaller range of population sizes, $N \lesssim 10^4$ – again, with s and p selected to approximate the data in Fig. 2 (recall that

the migration rate is fixed at $m = 1$ exchange per generation on average, regardless of N). Under these conditions, we find that the number of generations needed for a simulation to reach a state with $\langle F_{st}^a \rangle(t) \simeq 0.1$ is on the order of N. The value of s needed to maintain a gap between F_{st}^z and F_{st}^a increases with N, and increases more rapidly for smaller values of p (i.e., for fewer pairings between loci in the Z chromosome and the autosomes). For example, when $p = 0.0065$, each Z chromosome in a male F1 hybrid is paired with about 20 random loci in the autosomes, as noted above; In this case, the value of s needed for agreement with Fig. 2 increases only slightly with increasing population size, from about $Ns \sim 2$ when $N = 10^3$ to about $Ns \sim 3$ when $N = 8 \times 10^3$. By contrast, when $p = 0.00065$, the required value of s increases from about $Ns \sim 20$ when $N = 10^3$ to about $Ns \sim 100$ when $N = 8 \times 10^3$. As a result, the gap between F_{st}^z and F_{st}^a in the model is roughly related to the total fitness cost of hybridization. For the case $p = 0$ (no pairings), there is no gap between F_{st}^z and F_{st}^a (on average), as expected. Overall, the results seem to indicate that a pattern of data similar to Figs 3–5 will emerge for larger populations. Data for smaller population sizes, and C++ code used to generate the data in this work are available from the authors on request.

Discussion

We have shown that under certain conditions, the statistical relationship between I_{gf} and F_{st} for model butterfly species, interbreeding with negative interactions between genes in the autosomes and Z chromosomes in hybrids, agrees qualitatively with the data for butterfly sister species along suture zones; This suggests one possible explanation for the "missing" values in Fig. 1B – namely, that sister species have not been in contact long enough to traverse the index region $F_{st}^z < 0.3$. However, there are some significant differences between the model and real populations that are important to discuss.

First, as we have noted above, model averages for I_{gf} do not always conform well to the data for I_{gf} in Fig. 1. In particular, $\langle I^a_{gf}\rangle(t)$ and $\langle I^a_{gf}\rangle(F^a_{st})$ are both smaller than expected from Fig. 1A, while $\langle I^z_{gf}\rangle(t)$ and $\langle I^z_{gf}\rangle(F^z_{st})$ are somewhat larger than expected when $p \gtrsim 0.003$. These discrepancies appear to result from the low complexity of our model genomes: Clearly, smaller values of $G_{min} \leq 0.25$ (and therefore larger values of I_{gf}) will occur more frequently if the distribution of distance values is skewed toward smaller values of D. To compute I_{gf}, we have measured the distance between whole chromosomes. However, chromosomes are subject to a much higher rate of recombination than genomic transcripts. As a result, the probability of sampling a pair of individuals from different populations with the same, or similar sequences (i.e., in the tail of the distribution, $D \leq 0.25\,\langle D\rangle$) is probably larger for transcripts, since they maintain their integrity longer, and are usually exchanged between populations intact. Similarly, for $p \gtrsim 0.003$ in the model, the number of negative pairings per Z chromosome is comparable to the number of mutated loci. In this case, only those Z chromosomes with certain patterns of mutated loci (i.e., in which loci with strong negative effects have been removed by purifying selection) will have the ability to propagate in both populations, which should again increase the probability of sampling sequences from different populations with D in the tail of the distribution, $D \leq 0.25\,\langle D\rangle$. For a number of reasons (see below), we defer further investigation of G_{min} to future work.

Finally, it is important to note that real populations of butterflies are often much larger than the populations simulated in this work. For example, the effective population size for species of Heliconius butterflies, which we have used as a kind of model organism in this work, ranges from about 10^5 to as many as 10^7 individuals [6]. Based on the results in hand, there seems to be no reason to expect a different pattern of data to emerge from the model for larger populations. However, the interbreeding dynamic used in the model

completely neglects the spatial structure of butterfly populations. In real populations, interbreeding is limited by distance and mobility, whereas in the model, male and female individuals can interbreed as long as they occupy the same deme. For this reason, the parameters of the model can not be compared consistently with those of real populations, except perhaps in the case of populations that completely overlap. Accordingly, the model should probably be interpreted as a description of the suture zone itself, which typically contains a small fraction of the total population. For models that include the effects of geography and fluctuations in population size, it may still be possible that index values for the Z chromosome undergo some kind of sudden transition within butterfly suture zones. More detailed work is needed to understand these processes.

Acknowledgments. It is a pleasure to thank Jing Zhang for helpful comments during the completion of this work. This study is supported in part by a grant (to NVG) from the National Institutes of Health (GM127390).

Appendix

Equations (3) and (4) are computed by sampling an equal number genomes at random from each population. To compute F_{st} from Eq. (3) we sample 100 genomes per population. Alternatively, we compute F_{st} using the expression given by Bhatia et al. for individual sites [5],

$$F_{st}^{site} = \frac{(p_1 - p_2)^2 - p_1(1-p_1)/(n_1-1) - p_2(1-p_2)/(n_2-1)}{p_1(1-p_2) + p_2(1-p_1)} \tag{5}$$

where p_i is the frequency of a mutated allele in population i, and n_i is the number

of genomes sampled from population i (see Eq. (10) of reference). In our work, p_i is computed by sampling the entire population ($n_i = N$). To compute the genomic value of F_{st}, we average the numerator and denominator of Eq. (5) over mutated sites, as recommended by Bhatia et al. in reference [5]. We find that the two methods agree to a precision of about 10^{-3}. Thus, for practical purposes, the methods are interchangeable. Because the probability of obtaining a genome comparison with $G_{min} \leq 0.25$ will depend on the number genomes sampled, we compute I_{gf} by sampling only 10 genomes per population [3], about twice the number sampled from butterfly populations by Cong et al. [2].

References

[1] Sperling F (2003) *Butterfly molecular systematics: From species definitions to higher level phylogenies* eds. Boggs CL, Ehrlich PR. (University of Chicago Press), pp. 431–458.

[2] Cong Q, Zhang J, Grishin N (2019) Genomic determinants of speciation in butterflies (https://www.biorxiv.org/content/10.1101/837666v1).

[3] Geneva AJ, Muirhead CA, Kingan SB, Garrigan D (2015) A new method to scan genomes for introgression in a secondary contact model. *PLoS ONE* 10:e0118621.

[4] Rosenzweig BK, Pease JB, Besansky NJ, Hahn MW (2016) Powerful methods for detecting introgressed regions from population genomic data. *Mol. Ecol.* 25:2387–2397.

[5] Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting fst: The impact of rare variants. *Gen. Res.* 23:1514–1521.

[6] Van Belleghem SM et al. (2017) Patterns of z chromosome divergence among heliconius species highlight the importance of historical demography. *Mol. Ecol.* 27:3852–3872.

[7] Berner D (2019) Allele frequency difference afd – an intuitive alternative to fst for quantifying genetic population differentiation. *Genes* 10:308–322.

[8] Wikipedia (2020) Hayashi track (https://en.wikipedia.org/wiki/Hayashi track).

[9] Hayashi C (1961) Stellar evolution in early phases of gravitational contraction. *Publ. Astron. Soc. Jpn.* 13:450–422.

[10] Henyey LG, Lelevier R, Levée RD (1955) The early phases of stellar evolution. *Publ. Astron. Soc. Pac.* 67:154–160.

[11] Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in north america. *Am. Nat.* 166:581–591.

[12] Presgraves DC (2008) Sex chromosomes and speciation in drosophila. *Trend. Genet.* 24:336–343.

[13] Harris K, Nielsen R (2016) The genetic cost of neanderthal introgression. *Genetics* 203:881–891.

[14] Orr HA (1993) A mathematical model of haldane's rule. *Evolution* 47:1606–1611.

[15] Orr HA (1995) The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics* 139:1805–1813.

[16] Orr HA, Turelli M (2001) The evolution of postzygotic isolation: accumulating dobzhansky–muller incompatibilities. *Evolution* 55:1085–1094.

[17] Arias M et al. (2018) Crossing fitness valleys: empirical estimation of a fitness landscape associated with polymorphic mimicry. *Proc. R. Soc. B* 283:20160391.

[18] Veller C, Kleckner N, Nowak MA (2018) A rigorous measure of genome–wide genetic shuffling that takes into account crossover positions and mendel's second law. *Proc. Natl. Acad. Sci. USA* 116:1659–1668.

[19] Edelman NB et al. (2019) Genomic architecture and introgression shape a butterfly radiation. *Science* 366:594–599.

[20] Gillespie JH (2004) *Population genetics, a concise guide*. (Johns Hopkins University Press, Baltimore, MD).

[21] Matute DR, Butler IA, Turissini DA, Coyne JA (2010) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329:1518–1521.