UC Davis
UC Davis Previously Published Works

Title

The temporal and genomic scale of selection following hybridization

Permalink

https://escholarship.org/uc/item/3b50j1q9

Journal

Proceedings of the National Academy of Sciences of the United States of America, 121(12)

ISSN

0027-8424

Authors

Groh, Jeffrey S
Coop, Graham

Publication Date

2024-03-19

DOI

10.1073/pnas.2309168121

Copyright Information

Peer reviewed

The temporal and genomic scale of selection following hybridization

Jeffrey S. Groh[a,1] (ID) and Graham Coop[a,1]

Genomic evidence supports an important role for selection in shaping patterns of introgression along the genome, but frameworks for understanding the evolutionary dynamics within hybrid populations that underlie these patterns have been lacking. Due to the clock-like effect of recombination in hybrids breaking up parental haplotypes, drift and selection produce predictable patterns of ancestry variation at varying spatial genomic scales through time. Here, we develop methods based on the Discrete Wavelet Transform to study the genomic scale of local ancestry variation and its association with recombination rates and show that these methods capture temporal dynamics of drift and genome-wide selection after hybridization. We apply these methods to published datasets from hybrid populations of swordtail fish (*Xiphophorus*) and baboons (*Papio*) and to inferred Neanderthal introgression in modern humans. Across systems, upward of 20% of variation in local ancestry at the broadest genomic scales can be attributed to systematic selection against introgressed alleles, consistent with strong selection acting on early-generation hybrids. Signatures of selection at fine genomic scales suggest selection over longer time scales; however, we suggest that our ability to confidently infer selection at fine scales is likely limited by inherent biases in current methods for estimating local ancestry from contiguous segments of genomic similarity. Wavelet approaches will become widely applicable as genomic data from systems with introgression become increasingly available and can help shed light on generalities of the genomic consequences of interspecific hybridization.

hybridization | introgression | wavelet transform | Neanderthal

The greater recognition in recent decades that introgression is a common feature of eukaryotic genomes has led to the view that species boundaries are semipermeable (1). In this view, differential introgression along the genome is the result of selective filtering, with some neutral or widely favored alleles able to permeate into the genome of a hybridizing species, while others are restricted by genetic linkage to alleles with deleterious effects in hybrids (2–4). While selection and assortative mating have long been thought to play an important role in maintaining species integrity in the face of gene flow, advances in genomic sequencing and analysis have brought forth the possibility of reconstructing a more complete picture of genomic exchange between hybridizing species, forcing us to reckon with the vast complexity of how the genomic outcomes of hybridization are shaped by a dynamic interplay between recombination, genetic drift, and selection (5, 6).

Following a hybridization event, recombination over multiple generations progressively breaks up contiguous segments of DNA inherited from the original source populations (ancestry tracts) into finer segments. Numerous genomic methods are now available to identify these tracts through genomic similarity to proxies for source populations, and use the clock-like breakdown of tracts (or linkage disequilibrium between introgressed alleles) to make inferences about the timing of past mixture events (7–10). The changing spatial scale of coinherited genetic material along the genome through time is simultaneously shaped by drift and selection acting at the population level and will in turn influence how selection plays out in hybrid populations.

Genetic drift in a hybrid population shapes the ancestry proportion along the genome by increasing either ancestry state at random. Along the length of a chromosome, these deviations in the ancestry proportion will be autocorrelated due to the fact that an allele from one ancestry background that drifts to high frequency will tend to carry with it linked alleles from the same ancestry. For instance, if genetic drift is rapid during the early generations of hybridization, while ancestry tracts are long, broad contiguous portions of the genome might randomly fix for either ancestry. Conversely, in a large population where genetic drift is slow, by the time an allele of one ancestry reaches fixation, it will have been unlinked from all but the closest neighboring alleles from the same source population. The progressive shortening of ancestry tracts is slowed and ultimately stopped

Significance

Selection against hybrids is thought to play an important role in limiting gene flow between species that hybridize in nature. While genomic methods are now routinely used to detect signals of hybridization, methods to connect these patterns with underlying evolutionary dynamics of selection in hybrid populations have been lacking. To understand the spatial genomic scale of hybridization signals, we apply tools from signal processing (the wavelet transform) and show how this approach can track temporal dynamics of evolutionary forces within hybrid populations. Applying these methods to published datasets, we find support for selection playing a large role in shaping ancestry patterns in the early generations after hybridization.

by genetic drift; once a genomic segment fixes for either ancestry state, recombination events within the segment will cease to create ancestry breakpoints in the descendant chromosomes (11–14).

Selection acting on hybrids will further shape variation in levels of introgressed ancestry along the genome. Hybrids often experience strong selection, as their genomes can contain genetic incompatibilities or encode maladaptive phenotypes (15). Selection for or against an introgressed allele in an admixed population will lead to an excess or depletion of the corresponding ancestry in the surrounding region, the length of which depends on the strength and timing of selection relative to the timing of admixture (16). This concept has been leveraged to identify selected loci in recently admixed populations (17, 18) and date the onset of selection on introgressed alleles (19). Similarly, strong selection against introgressed alleles in specific genomic regions is thought to have contributed to the formation of the so-called 'introgression deserts' (20, 21). Analogously, variable patterns of divergence between species along the genome are thought to form at least in part through barriers to gene flow, i.e., selection preferentially removing introgressed haplotypes in regions harboring incompatibilities or loci contributing to divergent adaptation (22–25).

Increasing attention has focused on forming a more general understanding of how the distribution and frequency of introgressed ancestry along the genome has been shaped by natural selection (20, 26). Various studies have identified genome-wide correlations between minor parent ancestry proportion (the source population contributing <50% of total ancestry) and recombination rate, indicating that selection has acted at many loci throughout the genome to remove alleles from the minor parent (27–31). Such correlations emerge due to the slower decay of linkage disequilibrium (LD) between deleterious alleles carried on introgressed segments in low recombination regions, allowing for selection to more efficiently remove linked introgressed alleles in these regions (32, 33). Whereas these correlations can represent a snapshot of the cumulative effects of selection over tens to thousands of generations, theoretical work has shown that the strength of selection acting on hybrids likely varies dramatically through time. Due to extensive admixture LD in early-generation hybrids, the selective effects of many introgressed alleles combine, creating very strong selection on individuals carrying introgressed haplotypes (34). Thus, under a model of selection against introgressed alleles at many loci throughout the genome, the rate of removal of introgressed ancestry is greatest in the first several generations following hybridization (33, 35).

We now have clear genomic evidence that selection plays a role in maintaining species in the face of hybridization, but thus far have lacked a methodology to disentangle the temporal effects of selection, and to understand how selection shapes spatial ancestry patterns in the genome. Here, we develop genomic methods for analyzing temporal dynamics of drift and selection in hybrid populations based on the Discrete Wavelet Transform (DWT), a tool commonly used in time series analysis. After introducing important features of the DWT, we show how the ancestry variance present at different genomic scales can be captured by the wavelet variance decomposition and how this captures the time scale of evolutionary processes. We further show that a wavelet decomposition of the correlation between introgressed ancestry proportion and recombination rate, which decomposes the correlation into contributions of different scales, tracks temporal dynamics of genome-wide selection acting against alleles from one source population. Finally, we apply these methods to three empirical datasets: time-series data from a hybrid population of swordtail fish (*Xiphophorus*), a hybrid swarm between yellow and anubis baboons (*Papio*), and Neanderthal ancestry in modern humans. Across all datasets, we find patterns consistent with selection beginning early after hybridization and continuing throughout multiple generations.

Spatial Decomposition of Genomic Signals Using the Discrete Wavelet Transform

A common goal in the analysis of temporal or spatial signals is to understand the scale of variation present in the signal. Here, we make use of the Discrete Wavelet Transform (DWT), which like the Fourier transform is used to understand the scale of variation in a signal, but in addition captures information about local signal features and so is appropriate for nonstationary signals. Although widely used in physical sciences, the wavelet decomposition has seen more limited application in population genomics (but see refs. 36–39 for examples).

We start with a signal of interest, $x(\ell)$, such as ancestry state x measured at a set of evenly spaced locations $\ell = 1, \ldots, L$ along a contiguous chromosome of length L. If the data are not evenly spaced, we first interpolate to obtain evenly spaced measurements, along either a genetic or physical map of a chromosome. We will use the DWT to decompose the variation in this signal—that is, deviations in the signal around its chromosome-wide average—into components associated with a discrete set of spatial genomic scales. This is accomplished through multiplying our signal with a set of wavelets, functions written as $\psi_{\lambda,i}(\ell)$ that capture changes in the signal over varying spatial scales and locations. Informally, each wavelet resembles a finite wave, oscillating equally between positive and negative values over some characteristic scale λ centered on some location i. While many such functions exist, we use Haar wavelets (examples shown as black lines in Fig. 1*A*), which take a positive constant value for a stretch of sequence of length λ, switching to a negative constant value at location i for another stretch of sequence of length λ, and are zero everywhere else.
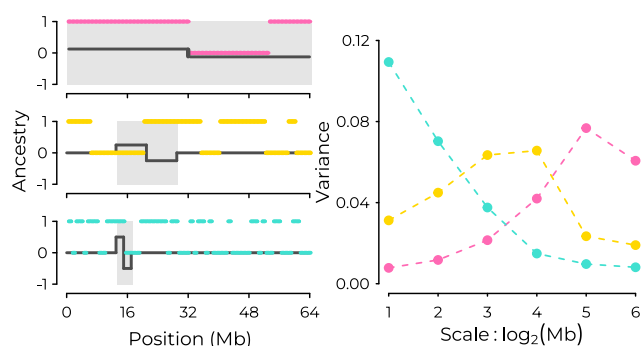

Fig. 1. (*Left*) Ancestry states $x(\ell) \in \{0, 1\}$ along three hypothetical chromosomes, with examples of Haar wavelets overlaid in dark gray. From *Top* to *Bottom*, ancestry tracts are shorter, representing different histories of recombination. Shaded intervals highlight the portion of the ancestry signal contributing to the resulting wavelet coefficient. (*Top Left*) Positive covariance between a $\psi_{\lambda=6}$ wavelet and x within the shaded interval yields a positive wavelet coefficient corresponding to a change in average ancestry state over the two halves of the chromosome. (*Middle Left*) Negative covariance between a $\psi_{\lambda=4}$ wavelet and x within the shaded interval yields a negative wavelet coefficient. (*Bottom Left*) Positive covariance between a $\psi_{\lambda=2}$ wavelet and x within the shaded interval gives a positive wavelet coefficient. (*Right*) The complete set of squared wavelet coefficients determines the power spectrum for the three ancestry signals, correspondence indicated in color.

The DWT transforms our signal into a set of wavelet coefficients, $w_{\lambda,i}$, each of which measures the strength of association between the signal and a corresponding wavelet:

$$w_{\lambda,i} = \sum_{\ell} \psi_{\lambda,i}(\ell)x(\ell) \propto \text{cov}(\psi_{\lambda,i}, x). \qquad [1]$$

Because the mean of a wavelet is zero, a wavelet coefficient is proportional to the covariance between the signal and the corresponding wavelet and thus measures the extent to which the wavelet captures variation in the signal over scale λ and location i. In the case of Haar wavelets, each wavelet coefficient measures a deviation between two adjacent windowed averages of the signal for a specific window size and location.

The set of wavelet coefficients produced by the DWT retains all of the information in the original signal while avoiding redundancy. To achieve this, the wavelet scales λ are chosen to form a doubling series (e.g., 1 kb, 2 kb, 4 kb, 8 kb, ...), and any two wavelets of different scales have zero covariance. Thus, variation in the signal at each scale is measured independently from variation measured at any other scale. This property distinguishes our approach from window-based approaches commonly used in genomics research, where statistics calculated from genomic windows of varying sizes are confounded, due to the fact that smaller windows are nested within larger windows.

The wavelets at a given scale are simply shifted versions of each other, with their nonzero portions covering different portions of the sequence. In the traditional DWT, these are placed such their nonzero portions cover the entire sequence without any overlap. Thus, any two wavelets of the same scale also have zero covariance, avoiding redundancy between neighboring wavelet coefficients. We instead use a modified and more flexible version of the traditional DWT known as the Maximum Overlap DWT, which, for a given scale, uses wavelets placed at all positions in the sequence (40). This yields less noisy estimates of the wavelet variance and covariance (described below) and retains the property of measuring variation independently across scales. Further details on the wavelet transform are given in *SI Appendix, Text 1*.

Wavelet Variance Decomposition. The average of the squared wavelet coefficients at a given scale, called a wavelet variance, $\hat{\sigma}_{\lambda}^2$, has the interpretation of the variance in our signal associated with changes in the value of the signal occurring at that scale:

$$\hat{\sigma}_{\lambda}^2(x) \propto \sum_i w_{\lambda,i}^2. \qquad [2]$$

Each squared wavelet coefficient $w_{\lambda,i}^2$ is a sum over pairs of loci of a pairwise ancestry product weighted by a pairwise wavelet product:

$$w_{\lambda,i}^2 = \sum_{\ell} \sum_{\ell'} \psi_{\lambda,i}(\ell)\psi_{\lambda,i}(\ell')x(\ell)x(\ell'). \qquad [3]$$

For Haar wavelets, this wavelet product is only nonzero over scale λ in the region centered on i; thus, $w_{\lambda,i}^2$ measures local variation in x in the corresponding region. The total variance of the original signal along the sequence, $\hat{\sigma}^2(x)$, can be decomposed as the sum of wavelet variances across scales:

$$\hat{\sigma}^2(x) = \sum_{\lambda} \hat{\sigma}_{\lambda}^2 \qquad [4]$$

(see also *SI Appendix,* Eqs. **S9** and **S12**). The set of wavelet variances above is known as the power spectrum. Since the wavelet scales λ are powers of two, if our sequence length is not itself a power of two, then there will be an additional component of leftover variance—referred to in the wavelet literature as scaling variance—due to the largest-scale wavelet not covering the entire sequence. As we average across chromosomes of different lengths, we fix the resolution of measurement (e.g., 50 kb or 2^{-10} Morgans) such that L is not a power of two, and so we are left with scaling variance. To simplify interpretation, we omit this scaling variance from the results shown in the main text as it represents only a minor component of the total variance in all our analyses.

When applied to ancestry state of a haploid copy of a chromosome, the power spectrum provides a summary of the length distribution of ancestry tracts (Fig. 1), with long ancestry tracts generating variance at broad scales (pink lines) and shorter ancestry tracts generating variance at finer scales (blue lines). This property is leveraged in the wavelet-based admixture dating methods of Pugach et al. (37) and Sanderson et al. (39) which rely on simulations. While recombination is the primary force determining the lengths of admixture tracts for single chromosomes, genetic drift and natural selection acting in a hybrid population will cause homologous pairs of the same chromosome within a population to covary in their ancestry state, which is our focus here. The spatial extent of this covariance can be captured by applying the power spectrum to mean ancestry averaged over multiple copies of the same chromosome, i.e., the ancestry proportion, which we will demonstrate using theory and simulations.

To obtain a complete spatial decomposition of the variance in ancestry proportion along the whole genome, we apply the wavelet transform separately to each chromosome (e.g., the ancestry proportion along chromosomes 1, 2, ..., N) and take a chromosome length-weighted average of wavelet variances across chromosomes at each scale. Due to heterogeneity in chromosome sizes, variation at the largest scales is present only on some chromosomes, and so we estimate wavelet variances using only the chromosomes for which a given scale is present.

Finally, since wavelet variances give only the within-chromosome portion of ancestry variance representing fluctuations around the mean for each chromosome, we account for the among-chromosome variance contribution by calculating a weighted variance of chromosome ancestry means. This among-chromosome variance will be labeled 'chrom' in the results. We can separately calculate the proportion of total genomic variance explained by each component; if all chromosomes have the same length, these quantities will be the same up to a constant. These components combined (wavelet variance, scaling variance, and among-chromosome variance) form a complete variance decomposition of our measured ancestry signal across the genome.

Wavelet Correlation Decomposition. Wavelet methods can also be applied to examine the scale of covariation between two signals (e.g., ancestry state and recombination rate), which we will use to examine temporal dynamics of selection. The overall correlation between two signals x and y (measured at the finest resolution available) can be decomposed into the sum of contributions from each scale:

$$Cor(x, y) = \sum_{\lambda} c_{\lambda} \rho_{\lambda}(x, y), \qquad [5]$$

where the wavelet correlation $\rho_\lambda(x, y)$ between our two signals at scale λ is weighted by an average proportion of variance explained by scale λ in the two signals, c_λ (see, e.g., Text S2 in 36). The correlation at scale λ is computed from the wavelet coefficients of x and y at scale λ and measures the strength of association between localized directional changes in x and y around their mean values at that scale. As with the variance decomposition, a complete decomposition of the overall genome-wide correlation will include additional terms due to leftover portions of the chromosome not covered by the largest wavelets, as well as an among-chromosome component.

Results

The Wavelet Variance Captures the Time Scale of Neutral Ancestry Change. We first illustrate the effects of genetic drift on the wavelet variance decomposition of ancestry state in a hybrid population in the absence of selection. Following a pulse of hybridization as recombination shortens ancestry tracts, genetic drift meanwhile generates deviations in the ancestry proportion along the genome away from the initial mixture proportion. The timescale of drift relative to recombination in hybrids determines the spatial scale of autocorrelation in these deviations—that is, the spatial scale of variance in mean ancestry along the genome. If a genetic bottleneck occurs shortly after the mixture event, the large deviations in ancestry proportion it causes will happen while ancestry tracts are long, so these deviations will be autocorrelated over broad scales (maroon, Fig. 2, *Top* panel). In contrast, drift in a large constant-sized population generates comparable ancestry deviations over much longer timescales (blue, Fig. 2, *Top* panel), by which point recombination has had more time to whittle down ancestry tracts, such that the variance in mean ancestry along the genome due to drift is on finer spatial scales.

The wavelet power spectrum of mean ancestry provides an elegant summary that captures these effects. A strong bottleneck concurrent with a hybridization pulse generates large wavelet variances at broad scales (maroon, Fig. 2, *Bottom* panel). Importantly, this broad-scale variance is maintained through time; even after 1,000 generations, these large wavelet variances at broad scales retain the signature of the early bottleneck. So long as ancestry remains polymorphic at many loci, genetic drift continues to generate variance in mean ancestry at progressively finer scales through time. Thus, following the bottleneck and population expansion, wavelet variances build at finer scales according to the rate of drift in the larger population. Contrast this to the case of a large, constant-sized population where variance along the genome only starts to become apparent at fine scales many generations after mixture (blue, Fig. 2, *Bottom* panel).

The wavelet variance of mean ancestry is determined by the pairwise product of ancestry at two loci (Eq. **3**) averaged across pairs of haplotypes i and j (*SI Appendix*, Eq. **S15**). Under a neutral model with random mating, the expectation of this product depends only on the recombination distance between two loci and population size and can be derived using coalescent theory. We describe here the simple case of a single pulse with admixture proportion α and no genetic drift. For alleles on the same haplotype ($i = j$), the product of ancestry states depends only on whether the haplotype recombined between the two loci between the present (with rate $r_{\ell,\ell'}$ per generation) and the time of admixture, t generations ago. For alleles on different chromosomes ($i \neq j$), the expected product of ancestry states at two loci depends only on whether they are independently inherited from the same source population, as in the absence of drift they cannot coalesce:

$$\mathbb{E}[x_i(\ell) x_j(\ell')] = \begin{cases} \alpha e^{-r_{\ell,\ell'}t} + \alpha^2(1 - e^{-r_{\ell,\ell'}t}) & \text{if } i = j \\ \alpha^2 & \text{if } i \neq j. \end{cases} \quad [6]$$

We derive the full result with genetic drift in a time-varying population size in *SI Appendix*, Text 2 (*SI Appendix*, Eq. **S17**). Our
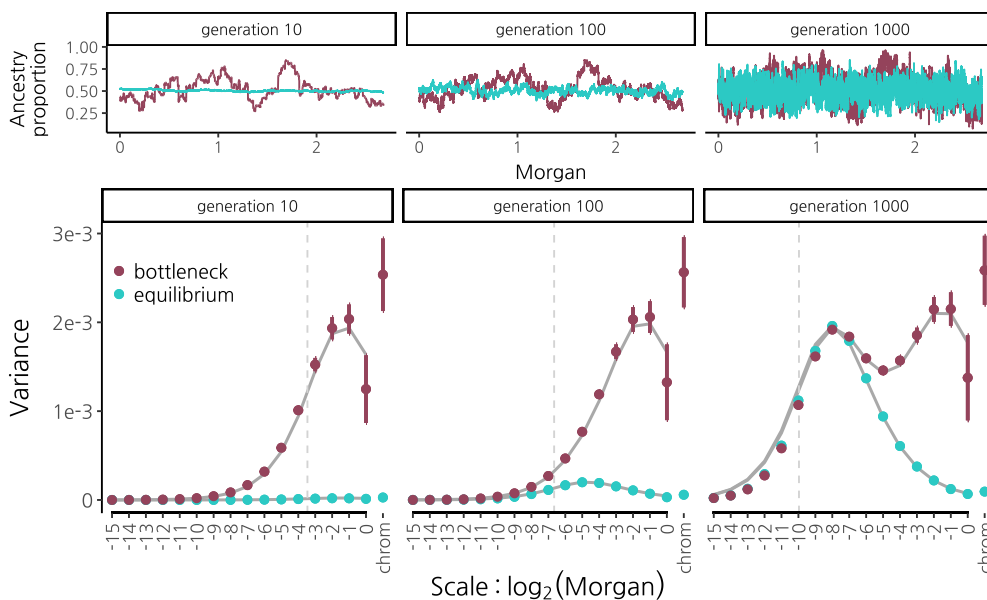

Fig. 2. Wavelet variance decomposition through time of ancestry proportion in a 50/50 population mixture undergoing genetic drift with recombination and no selection. We simulated using SLiM, (41) a population of constant size 2N = 20,000 (blue) and a population that undergoes a bottleneck to 2N = 200 for just the first 10 generations of recombination in hybrids, then expands to 2N = 20,000 (maroon). (*Top*) Ancestry proportion along human chromosome 1 from a single simulation run. From left to right, shown after 10, 100, and 1,000 generations of recombination in hybrids. (*Bottom*) Wavelet variance decomposition showing the spatial scale of variance in ancestry proportion. Points and error bars show means and 95% CIs across 20 replicate simulations. Solid gray lines show theoretical expectations. Vertical dotted gray lines indicate the expected distance between recombination breakpoints that have accrued along a single chromosome since the hybridization pulse. Note that since results are shown on the genetic map, recombination rate variation does not influence these patterns, other than through interpolation error.

theoretical expectations show good agreement with simulation results (Fig. 2, solid gray lines in the *Bottom* panel). We note however that subtle biases in our simulated wavelet variances (e.g., downward bias at fine scales in generation 1000) result from the interpolation of ancestry state between simulated loci that are evenly spaced on a physical map to locations that are evenly spaced on a genetic map (*SI Appendix*, Figs. S2 and S3).

Measuring the Timescale of Selection on Introgressed Ancestry. As with drift, selection generates deviations in mean ancestry along the genome away from the initial mixture proportion, with the extent of autocorrelation in these deviations determined by the timing and strength of selection relative to the timing of mixture, as well as variation in recombination rate along the sequence. To explore the role of selection in shaping ancestry variation, we performed forward simulations of a hybridization pulse followed by selection acting additively against alleles fixed in one ancestry background at many loci (10,000) genome-wide on a genetic map modeled on the human autosomes (for example representing selection due to polygenic adaptation of one source population to the local environment). We chose this highly polygenic model in part to provide fine-scale variation for

selection to act upon, but also discuss models with selection on fewer loci below.

We find that selection acting to remove introgressed alleles at multiple loci distorts the power spectrum toward proportionally greater ancestry variance at broad scales relative to the neutral expectation (Fig. 3*A*). This effect is seen across a range of numbers of loci (10 to 10,000) under selection in hybrids, with greater broad-scale variance generated when the same total additive selection strength is distributed across fewer loci (*SI Appendix*, Fig. S4). Separately, selection can decrease levels of ancestry variance across some or all scales (depending on the recombination map) relative to the neutral expectation by virtue of moving the introgressed ancestry proportion toward to zero (*SI Appendix*, Fig. S5). Note that widespread weak selection may lead to departures from the neutral power spectrum on the same order as those caused by interpolation. Thus, in comparing empirical results to neutral expectations it may in some cases be preferable to use simulations that incorporate the interpolation procedure.

Critically, variance in mean ancestry along the genome produced by systematic selection against one ancestry can be distinguished from that produced by genetic drift; we expect introgressed ancestry depletions against the genome-wide
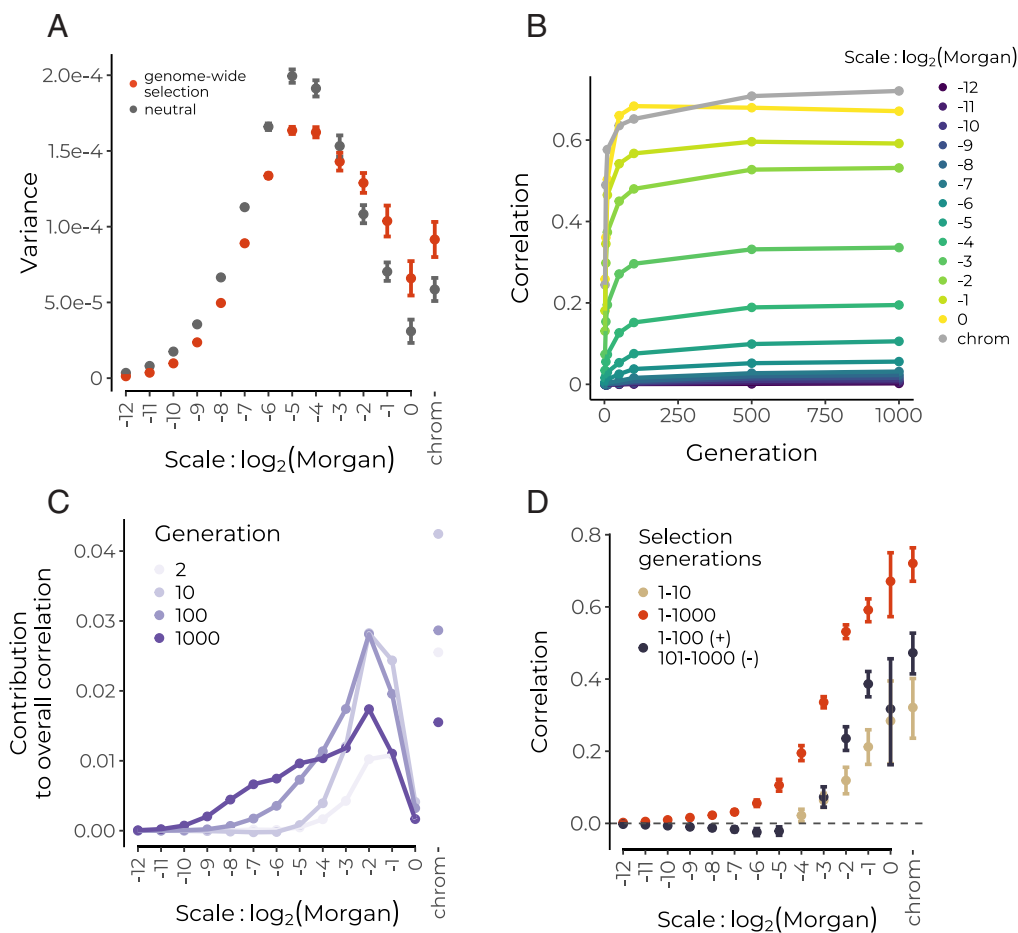

Fig. 3. Simulations of selection following a pulse of hybridization starting from a 50/50 population mixture. (*A*) Genome-wide selection in conjunction with broad-scale variation in recombination rates leads to differential removal of introgressed ancestry at broad scales, thereby biasing the power spectrum toward greater variance at these scales compared to the neutral expectation (viewed after 100 generations). (*B*) Selection against many alleles on one ancestry background rapidly establishes broad-scale correlations between recombination and minor parent ancestry in early generation hybrids. (*C*) Through time with continued selection, the overall correlation remains dominated by broad scales, but finer scales contribute increasingly more. (*D*) Selection acting only on the first 10 generations of recombinant hybrids generates significant positive wavelet correlations only at broad scales (brown) (viewed in generation 1000), whereas continuous selection over 1000 generations continues to generate correlations on finer scales (red). When selection acts continuously but reverses direction after 100 generations to favor the alternate ancestry, positive broad-scale correlations persist as negative correlations establish at finer scales (see also *SI Appendix*, Fig. S7*B*). Only significant correlations are shown; error bars represent 95% CIs across 20 replicate simulations.

background to be spatially associated with features that govern the strength of selection, namely the recombination rate and the density of selected sites. Indeed, positive correlations between minor parent ancestry and recombination rate or negative correlations with coding density are often the basis for inferring genome-wide selection against alleles from one ancestry, e.g., refs. 27 and 28.

As selection over multiple generations establishes these correlations against a shifting backdrop of the spatial scale of ancestry variance due to recombination, the wavelet spatial decomposition of the correlation between ancestry state and recombination (Eq. **5**) can be used to track the effects of selection through time. In the early generations after hybridization, when selection sees numerous selected alleles linked in long ancestry tracts, strong correlations rapidly establish at the broadest genomic scales (Fig. 3*B*). Correspondingly, as the bulk of variance in mean ancestry at this stage is also at broad scales, these scales contribute most to the total correlation (Fig. 3*C*). Wavelet correlations at finer scales increase more gradually through time as selection continues to operate on an increasingly fine-scale mosaic of ancestry tracts. As drift increases variance at finer scales and selection generates correlations on these scales (selection may be increasing or decreasing variances at these scales, *SI Appendix,* Fig. S5), fine scale correlations contribute more to the overall correlation through time (Fig. 3*C*). In simulations where the same strength of selection is spread over fewer loci under selection in hybrids, we find that correlations are weaker and do not continue to establish at finer scales through time (*SI Appendix,* Fig. S6). This makes intuitive sense, as we only expect selection to create fine-scale correlations with recombination if selected loci are distributed over those scales. In each case, broad scales continue to constitute a significant portion of the overall correlation through time, indicating that early selection has an outsized influence on overall patterns of ancestry in hybrids.

As we observed that the power spectrum of mean ancestry could preserve a memory of an early bottleneck after hybridization, we next asked whether the wavelet correlation decomposition could similarly be used to detect temporally localized effects of selection. We thus simulated a scenario where selection acted only on the first 10 generations of recombinant hybrids after admixture. In that case, correlations between recombination and ancestry proportion remained restricted to only the largest scales and remained present after 1,000 generations of recombination in hybrids (Fig. 3*D*, brown). This suggests that observing significant fine-scale wavelet correlations between recombination and ancestry proportion on the genetic map indicates that selection continues to have widespread effects on ancestry in later generations of mixture. We next simulated two additional scenarios, where 1) selection begins only after 500 generations of neutral mixture, and 2) where selection acts in every generation following the hybridization pulse but switches directions after 100 generations to favor the alternate ancestry allele at each locus (these model were chosen for illustrative purposes and not to necessarily reflect biologically realistic scenarios). We find that while selection acting only in later generations also readily generates significant broad-scale correlations with recombination (scenario 1), broad-scale correlations that are established by early selection are not reversed even by subsequent generations of selection acting in the opposing direction (scenario 2) (Fig. 3*D* and *SI Appendix,* Fig. S7). Thus, the wavelet decomposition of the correlation between mean ancestry and recombination is capable of revealing disparate effects of selection in different time periods.

In the analyses above, we modeled a single pulse of admixture to provide intuition for how the wavelet decompositions capture temporal dynamics of drift and selection on introgressed ancestry. In reality, hybrid populations may receive multiple influxes of parental-type individuals, or exist as a hybrid zone between populations exchanging migrants. Under such models, the lengths of introgressed segments will have a mixture distribution reflecting the cumulative effects of multiple hybridization events, and the resulting wavelet decompositions will capture the effects of selection and drift on this combined distribution of segments. While we demonstrated the wavelet decompositions using a specific model of demography and selection, the methods themselves are agnostic to any assumptions about the underlying model of hybridization. Indeed, these methods could be applied in alternative scenarios such as stable hybrid zones. Here, researchers may wish to compute wavelet summaries for subsets of individuals across space, or according to average ancestry proportions.

We next apply the wavelet methods illustrated above to previously published ancestry calls from several empirical datasets. In our theoretical and simulation work, "ancestry" can be defined precisely, as we directly track the descent of haplotypes from either of two well-defined populations that form a mixture at a specified time in the past. In reality, this information is not known and may be poorly defined, and ancestry for hybrid (or admixed) populations is defined with respect to genetic similarity to sets of reference samples (A and B) that are thought to be representative of the variation present in the original mixing groups. Thus, in describing specific analyses, we use terminology such as A-like haplotypes to refer to regions of the genome that have been computationally identified as more similar to reference sample A than sample B (42, 43). We use the term ancestry when we discuss the inferences we draw from these analyses, e.g., that selection acts against alleles from the species A ancestry, reflecting the fact that our inferences are placed in a conceptual model of two divergent populations mixing upon secondary contact.

Application to Hybrid Swordtail Fishes. To examine the roles of recombination, drift and selection in shaping ancestry patterns in a recently formed hybrid population, we analyzed time-series whole genome data from a hybrid population of swordtail fish from Acuapa river in Hidalgo, Mexico. This population is thought to have formed from hybridization between *Xiphophorus birchmanni* and *X. malinche* within the last 100 y (44). In several independently formed hybrid zones between the same species, Schumer et al. (27) observed positive correlations between recombination rate and minor parent ancestry proportion, implicating selection in shaping ancestry patterns genome-wide. Furthermore, numerous incompatibilities are known to segregate in these hybrid populations (45, 46).

We made use of previously inferred local ancestry patterns in a set of temporally staggered samples ranging from 2006 to 2018 (44). Genotype posterior probabilities of matching allopatric *X. birchmanni*/*X. malinche* alleles were called using a Hidden Markov Model (HMM) at a set of loci along the genome that are highly differentiated between the species, and we interpolated the average marginal posterior probability of matching the *X. malinche* allele to yield the sample proportion of malinche-like haplotypes at evenly spaced intervals along the genome.

To illustrate the effects of recombination across the sampling interval, we visualized the power spectrum of the proportion of malinche-like haplotypes on the genetic map for each collection year. As expected by the shortening of ancestry tracts through time, the power spectrum shifts toward greater variance at finer scales across the temporal sampling interval (Fig. 4*A*). Increasing
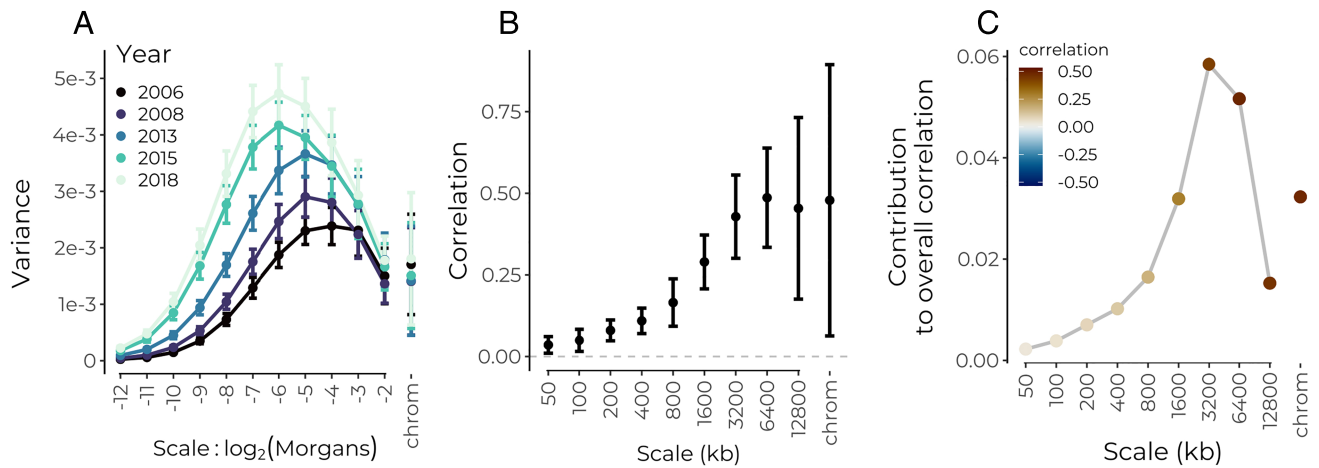
Fig. 4. Wavelet analysis of hybrid genomes in a population of *X. birchmanni* × *X. malinche* swordtail fish from Acuapa River in Hidalgo, Mexico. (*A*) Power spectrum of the malinche-like ancestry proportion for five time points between 2006 and 2018. Points are weighted averages across chromosomes, and error bars are 95% jackknife CIs. (*B*) Correlations at each spatial genomic scale (on the physical map) between wavelet coefficients for the proportion of malinche-like haplotypes and wavelet coefficients for recombination rate. Squared values give an estimate of the proportion of variance in ancestry state explained by systematic selection against malinche-like alleles. Data shown only for 2006, patterns similar across years. (*C*) Contribution of each genomic scale to the overall correlation.

levels of overall variance in the signal through time are not due to differences in sample size and are consistent with expected effects of genetic drift following hybridization (*SI Appendix*, Fig. S8; and may also reflect selection).

Turning to the role of selection in shaping ancestry patterns, we next applied our wavelet methods to examine the spatial structure of the correlation between the proportion of malinche-like haplotypes and the recombination rate along the genome. Recall that genetic drift will not produce any systematic associations between ancestry state and recombination rate, but that selection acting systematically against alleles from one ancestry will generate positive correlations between the recombination rate and the proportion of the ancestry being selected against. To match previously reported correlations, we analyze these correlations on the physical map of the genome. We find an overall positive correlation between the malinche-like proportion and recombination that is largely driven by broad-scale patterns (Fig. 4 *B* and *C*). While finer scales contribute progressively more to the overall correlation across the sampling interval (*SI Appendix*, Fig. S9), this pattern appears to be driven by the increasing variance at finer scales rather than increases in the correlations at these scales (Fig. 4*A* and *SI Appendix*, Fig. S10). Thus, while we observe strong broad-scale correlations consistent with selection acting against malinche alleles early on in the formation of the hybrid zone, we do not find evidence that much of the change in genome-wide ancestry patterns across this sampling interval (2006 to 2018) reflects selection against malinche alleles. Nonetheless, selection may still be shaping patterns of local ancestry; Powell et al. (44) reported evidence of contemporary selection against the malinche allele at a QTL for tail length in this population.

When viewed on the genetic map, we find that strong positive correlations are largely restricted to broad scales, consistent with a recent origin of the hybrid population (*SI Appendix*, Fig. S10*B*). We note however that we also detect weak but significant positive fine-scale correlations on the genetic map, e.g., at 2^{-10} Morgans, which would suggest selection acting on ancestry variation at finer scales than expected given the estimated age of the hybrid population. These finer scale correlations could possibly reflect older admixture present in the source populations that formed

this hybrid population, or errors in either HMM similarity calls or recombination rate inference at fine scales.

Genetic drift alone does not generate any systematic association between local recombination rates and local ancestry proportion, but this pattern is readily generated by genome-wide selection against introgressed alleles. Thus, the percent of variance explained from a regression of ancestry wavelet coefficients on recombination rate wavelet coefficients at a given scale can be interpreted as the percent of ancestry variance at that scale that can be attributed to selection under such a model. Applying this logic, we find that ~20% of variance at the broadest scales on both the physical and genetic map (i.e., >3.2 Mb, >0.125 Morgans) can be attributed to selection against minor parent ancestry. We consider this a lower bound estimate, given that it treats genomic similarity to *X. malinche* as a proxy for the locations of selected loci and that it assumes a model where selection always acts against malinche alleles. This approach can easily be extended by including wavelet coefficients for other genomic features such as the density of coding base pairs as predictors in the regression. For simplicity and consistency across datasets, we only present results using recombination rate as a predictor, although we note that in this dataset, including wavelet coefficients for coding sequence density as a predictor did not generally improve model fit.

Application to Hybrid Baboons in Amboseli. Genome-wide selection against hybrids has also been inferred in the case of hybrids between yellow baboons (*Papio cynocephalus*) and anubis baboons (*P. anubis*). Vilgalys et al. (2022) (31) analyzed whole genome sequence data from 442 baboons sampled near the center of a hybrid zone in the Amboseli basin of Kenya. This population is comprised of a complex mixture of early and late generation hybrids (potentially reflecting recurrent admixture over hundreds of generations), with most individuals having majority yellow-like ancestry. Although overall levels of anubis-like ancestry have been gradually increasing in this region over time due to immigration of anubis-like individuals into the hybrid zone (31, 47, 48), Vilgalys et al. (31) reported positive genome-wide correlations between recombination rate and anubis-like ancestry, consistent with widespread selection against alleles carried on this ancestry background within the hybrid zone.

We took advantage of the complex history of hybridization in this system to further demonstrate how the power spectrum can reveal demographic trends in hybrid zones. We apply the wavelet variance decomposition to an interpolated estimate of the proportion of anubis-like haplotypes within diploids at a set of loci differentiated between allopatric reference panels of yellow and anubis baboons. Overall, the majority of ancestry variance is present at finer scales on the genetic map compared to the swordtail hybrid population, suggesting hybridization occurring over much deeper timescales (Fig. 5*A*). We also find that the power spectrum varies according to an individual's genome-wide proportion of anubis-like ancestry, with the most anubis-like individuals showing greater ancestry variance at broad genomic scales. This pattern is consistent with the most anubis-like individuals in the sample being recent descendants of migrants with longer contiguous tracts of anubis-like ancestry, and implies multiple bouts of admixture rather than a single pulse. In sum, the power spectrum for ancestry state is consistent with this population being the product of both recent and historical hybridization events as previously inferred.

We next applied the wavelet methodology to examine the structure of a previously reported positive correlation between recombination rates and the proportion of anubis-like haplotypes along the genome (31). We again find positive correlations across multiple scales on the physical map, with broad scales contributing the most to the overall correlation (Fig. 5*B*). Comparable to the swordtail population, we find that ~20% of broad-scale variance in the proportion of anubis-like ancestry can be explained by systematic selection against anubis alleles, using just recombination rate as a predictor (Fig. 5*B*, squared values). When viewed on the genetic map, we find positive correlations across multiple scales consistent with ongoing selection against anubis-like ancestry, again with the strongest correlations at broadest scales (*SI Appendix*, Fig. S11).

Application to Neanderthal Ancestry in Modern Humans. As modern humans expanded out of Africa ~60,000 y ago, they interbred with Neanderthals present in Eurasia (49). Despite some introgressed Neanderthal variants having been adaptive (50), a number of studies have inferred that Neanderthal ancestry was on average deleterious in the modern human lineage, as Neanderthal-like haplotypes in modern humans are relatively depleted near conserved elements and in regions of low recombination (20, 26, 27, 51).

We applied wavelet methods in order to understand the relative contributions of drift and selection in shaping variation in Neanderthal ancestry along the genome. Here, we use three different call sets of Neanderthal-like haplotypes, inferred for the CEU 1,000 genomes samples (20, 52) and a large sample of modern Icelanders (53). Analyzing the power spectrum of the sample proportion of Neanderthal-like haplotypes on the autosomes for each set, we find reasonable agreement between our estimates and theoretical expectations under a model of a single pulse of neutral admixture 2,000 generations ago (Fig. 6*A*). Thus, while previously identified deserts of Neanderthal ancestry likely reflect early purifying selection against Neanderthal haplotypes, overall most of the variance in the frequency of Neanderthal-like haplotypes is on finer scales, largely consistent with the long-term effects of genetic drift. While this is perhaps surprising given previously inferred negative fitness costs of introgression, we found in simulations of dispersed weak selection assuming 10,000 loci contribute to a 20% fitness reduction in human–Neanderthal F1s, e.g., refs. 35 and 26) generates only subtle deviations from the neutral power spectrum (*SI Appendix*, Fig. S12). The sample average calls of different methods are in good agreement at broad scales (upward of several Mb), but are only weakly correlated at finer scales of measurement (e.g., tens to hundreds of kb, Fig. 6*B*). This is expected given the low inferred proportion of Neanderthal ancestry and the old age of the admixture event. All of the methods have substantially reduced power to detect short fragments, particularly in the presence of recombination rate heterogeneity (54).

Turning to correlations between the proportion of Neanderthal-like haplotypes and recombination rate, we find a consensus across datasets of positive correlations at broad scales on the physical map (Fig. 6*C* and *SI Appendix*, Fig. S13). However, at the finest scales, we find significant negative correlations for some of the call sets. These are not due to a confounding correlation between recombination and gene density, which are negatively correlated at these scales (*SI Appendix*, Fig. S14). The
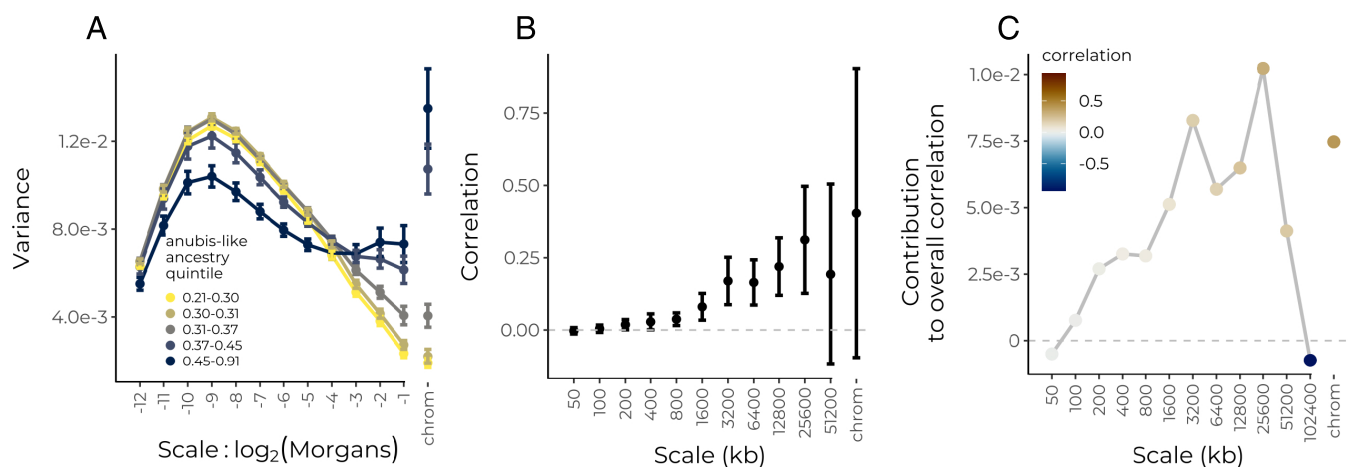

Fig. 5. Wavelet analysis of hybrid genomes between yellow and anubis baboons in Amboseli, Kenya. (*A*) Power spectrum of the proportion of anubis-like haplotypes within diploids on the genetic map, stratified by quintile of genome-wide average anubis-like ancestry. Error bars are 95% CIs using the SE of the wavelet variance across individuals within each quintile. (*B*) Correlations at each spatial genomic scale (on the physical map) between wavelet coefficients for sample proportion of anubis-like haplotypes and wavelet coefficients for recombination rate. Squared values give an estimate of the proportion of variance in ancestry state explained by systematic selection against anubis-like alleles. Weighted-jackknife 95% CIs shown for all but the largest scale which is present only on a single chromosome. (*C*) Contribution of each scale to the overall correlation. Although there is a strong negative correlation at scale 102,400 kb, this scale is only present on chromosome 1 and does not contribute substantially to the overall positive correlation.
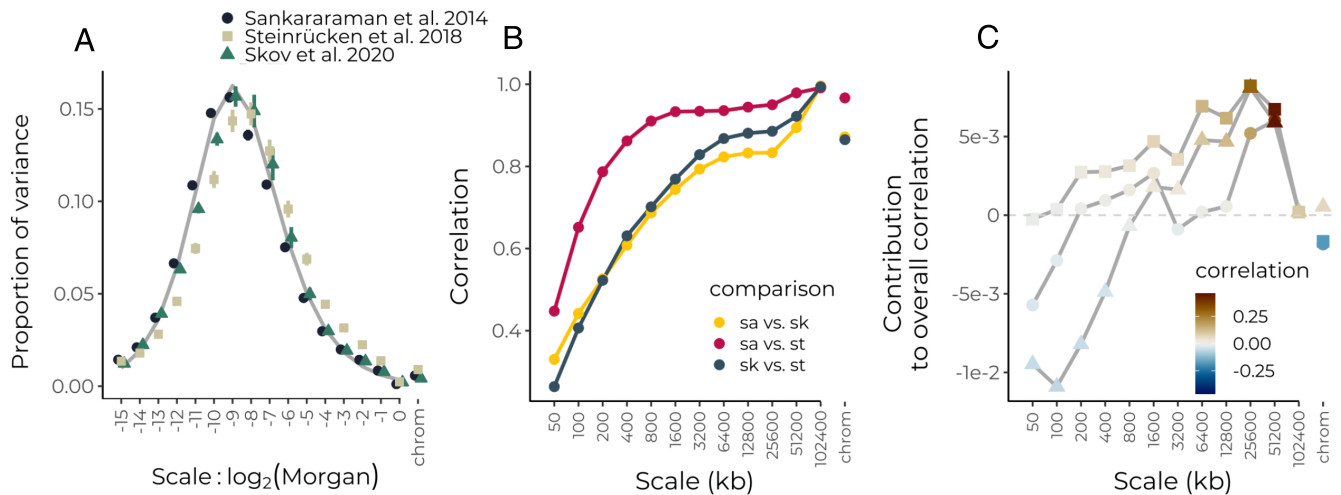
https://doi.org/10.1073/pnas.2309168121

Fig. 6. Wavelet analysis of inferred Neanderthal ancestry in modern humans. (*A*) Normalized power spectra of the proportion of Neanderthal-like tracts for three different studies (colored points) compared to a neutral expectation for 2,000 generations of admixture in a population size of 10,000 diploids (gray line). Error bars represent 95% CIs from a weighted jackknife across chromosomes. (*B*) Correlations across scales between different similarity calls from different datasets. (*C*) Contribution of each scale to the overall observed correlation between the proportion of Neanderthal-like haplotypes and log-transformed recombination rates. Shapes correspond to different studies as indicated in panel (*A*).

negative correlations could conceivably be generated by selection in late generations systematically favoring Neanderthal ancestry (as seen in our simulations of the case where introgressed variants are first deleterious but then later favored, Fig. 3*D*). However, we suggest they more likely reflect an inherent bias toward detecting Neanderthal-like haplotypes in regions of lower recombination. The estimated correlations at fine scales are variable across calling methods and appear at genomic scales where different sets of calls are only weakly correlated, suggesting these negative correlations are an artifact of the methods rather than underlying biological signal. Furthermore, we note that correlations at fine scales were highly sensitive to the posterior threshold applied in at least one dataset, further suggesting that inference of ongoing or recent selection based on fine-scale patterns of Neanderthal-like haplotypes is limited with current methods (*SI Appendix,* Fig. S15).

Discussion

Wavelet analyses are a promising tool for studying evolutionary forces acting in hybrid populations using genomic data. We have shown with theory and simulations how a decomposition of the variance in ancestry state along the genome contains signatures of demographic history under a simple population genetic model of admixture. Further, we used simulations to illustrate how selection against introgressed alleles impacts the scale of both the variance in ancestry and the correlation between introgressed ancestry and recombination rate. In total, these methods offer a compact summary of genome-wide admixture signals and can inform a more general understanding of the role of selection in shaping patterns of introgression across the genome.

In applying the method to several systems, we can observe generalities in the genomic consequences of hybridization. Most notably, the observed positive correlations between introgressed ancestry and recombination rate are largely dominated by patterns at broad genomic scales. Simulations indicate that these patterns establish rapidly in the earliest generations after hybridization when selection acts against multiple alleles from one ancestry across the genome. Thus, this evidence accords with our understanding that selection is strongest in the early

generations after a hybridization event (33–35), and suggests that this effect can be detected even hundreds or thousands of generations later. In all three cases, a reasonable proportion of the variation in ancestry state along the genome at the broadest scales is attributable to recombination rate variation (roughly 10 to 20%). As these estimates are a lower bound on the contribution of selection, we can say that selection plays a key role in shaping the genomic composition of early generation hybrid populations, with lasting effects. Weaker correlations at finer genomic scales also agree with our understanding that the strength of selection on hybrids dramatically decreases over generations (33), and suggests that genetic drift may be the dominant force shaping fine-scale genomic ancestry patterns. A related approach to estimating the contribution of selection to ancestry patterns across scales would be to apply these analyses to the correlation in ancestry state between independent replicates of hybrid populations derived from the same source populations. This approach may be more powerful in that locations and effects of selected alleles are internally matched when similar parental sources repeatedly hybridize under similar ecological conditions.

In demonstrating the effects of selection on wavelet-based statistics, we considered a simple additive model of selection against alleles at multiple loci carried in one ancestry background. However, the wavelet methods themselves are agnostic to the form of selection, and in real hybrid populations multiple forms of selection likely interact (6). Alternative models of selection that incorporate dominance and epistasis could generate different signatures in the wavelet statistics, such that these methods could potentially be applied toward distinguishing among models. For instance, Harris and Nielsen (35) found that if deleterious mutations in human and Neanderthal haplotypes are largely partially recessive, the direction of selection on introgressed Neanderthal variants in human populations could change through time due to the contrasting effects of purifying selection and selection for heterosis (see also ref. 55). We have shown here that such a reversal in the direction of selection on one ancestry can in principle be detected using the wavelet decompositions (Fig. 3*D*). One parameter explored here is the total number of loci under selection in hybrids. Using the same additive model with the total additive strength of selection held constant, we found that

the wavelet decompositions vary with the overall number—and correspondingly the density—of sites under selection in hybrids. Importantly, a high density of sites under selection is required to generate fine-scale correlations between introgression and recombination. Thus, the scale at which correlations between introgressed ancestry and recombination rate are generated is indicative of the scale at which selected loci are distributed in the genome.

An important caveat to these methods is that any systematic biases in the input data, including estimation of recombination rates and local ancestry inference, will be propagated into the wavelet transform. For example, biased detection of introgressed fragments toward low recombination regions may generate spurious signals of selection favoring introgressed ancestry genomewide. We have suggested this may be the case for Neanderthal introgression into humans, where we see negative correlations over fine genomic scales (Fig. 6C). These potential biases, together with inconsistent patterns across sets of calls and high sensitivity to posterior probability thresholds, also suggest more generally that inferences of selection on Neanderthal ancestry relying on fine-scale genomic patterns might need to be revisited (26, 56). Other issues such as the need for phased reference panel haplotypes and the potential for model mis-specification limit wider applicability of local ancestry inference with HMM-based methods in nonmodel systems.

One promising direction to overcoming these limitations is that the wavelet transform can be applied to single-locus admixture statistics (37, 39) to avoid directly inferring the boundaries of ancestry tracts. While these statistics contain additional noise unrelated to the admixture process, future work could develop a theoretical framework for applying these methods to single-locus admixture statistics, thereby avoiding biases arising from local ancestry inference with HMM methods. Additional theory is also needed to better understand the impact of selection on the wavelet variance and correlation decompositions, and to extend these ideas to scenarios with ongoing gene flow and hybrid zones.

We now appreciate that introgression is a common feature of eukaryotic genomes, and the proliferation of genomic sequence data presents an opportunity to study hybridization events across systems and time scales. Combined with the increasing availability of more complete genome assemblies and recombination maps, wavelet approaches should enable patterns in the strength and time scale of selection on hybrids to emerge across systems.

Materials and Methods

All code used to produce results shown in this manuscript can be found at https://github.com/jgroh/selection-against-introgression.

Simulation and Data Processing. All simulations were performed in SLiM 4.0 (41). Each locus in our simulations represented a genomic window of fixed physical length (e.g., 50 kb). Recombination rates between adjacent windows were modeled off a genetic map of the human autosomes (57). In simulations with selection, we fixed deleterious alleles in one source population at 10,000 loci placed uniformly at random on the physical map in each replicate run. We used a model of polygenic selection against introgressed ancestry where the fitness of individual i, (w_i), declines linearly with the fraction of introgressed alleles the individual carries (p, with the slope given by S: $w_i = 1 - pS$; similar models were studied in refs. 34, 2, and 33). For the analyses shown in the main text, we set $S = 1$, corresponding to F1 hybrids having a relative fitness of 0.5. Simulations used tree sequence recording (58), and ancestry along the genome was extracted from the tree sequences using *tskit* (https://tskit.dev/software/tskit.html).

For downstream wavelet analysis, we require signal values at evenly spaced positions, either on a physical or genetic map. For our simulations where simulated loci represented genomic windows of fixed physical length, we thus performed interpolation of ancestry and recombination signals to a grid of evenly spaced positions on the genetic map.

Wavelet Analysis. For all wavelet analyses, we used the Maximal Overlap DWT with Haar wavelets (40) implemented in the R package *waveslim* (59). Further background on the wavelet methods is provided in *SI Appendix*, Text 1.

As the wavelet transform only operates on contiguous signals, we perform the variance and correlation decompositions to each chromosome separately and then combine results across chromosomes in one of two ways. Due to heterogeneity in chromosome length, not all scales will be present on all chromosomes. Thus, for estimating wavelet variance magnitudes, we average over only those chromosomes for which a given scale is present, taking a weighted average where the weight is chromosome length. Separately, we calculate the proportion of total genomic variance contributed each scale by assigning a variance value of zero to those chromosomes for which a given scale is not present, and then performing the same weighted average across chromosomes. Values were then adjusted to account for the proportion of total genomic variance due to variance among chromosomes in their mean values of a signal. The among-chromosome portion is also a weighted average weighted by chromosome length.

Generic functions to perform the main analyses shown in this paper are contained in the R package *gnomwav* available at https://github.com/jgroh/gnomwav. The function *gnom_var_decomp* returns the genome-wide variance decomposition for any signal, including both forms of averaging over chromosomes. Likewise, *gnom_cor_decomp* returns genome-wide wavelet correlations for each scale. The contribution of each scale to the overall correlation between signals can be obtained from the output of these two functions. These functions are not specific to the investigation of admixture and should be broadly applicable to many genomic signals of interest.

Swordtail Analysis. We used genomic similarity calls from Powell et al. (44), consisting of posterior probabilities for diploid genotypes matching reference panels of *X. malinche* or *X. birchmanni* at a set of SNPs that are highly differentiated between the species. The frequency of the minor parent allele (*X. malinche*) for each individual was taken as a weighted average of the posterior probabilities of being homozygous and heterozygous for matching *X. malinche*, i.e., $\hat{p}_A = \mathbb{P}(AA) + \frac{1}{2}\mathbb{P}(Aa)$. We interpolated these estimates to distance of 50 kb and separately 2^{-12} Morgans and averaged across individuals for the admixture proportion. For analyses on the genetic map, we used an LD-based recombination map (Schumer lab, pers. comm.). Since these estimates were in units of $2N_e r$, we converted distances to Morgans using an estimate of $2N_e$ from the slope of a regression between the genetic lengths of chromosomes estimated from a cross-over map and those estimated from the LD map. As we observed extreme outliers in values of $2N_e r$, we first truncated the distribution of $2N_e r$ at 0.005, corresponding to 1.6% of the total genome (a threshold was also applied in ref. 27). We applied the lowest threshold possible beyond which we saw relatively stable estimates of $2N_e$, and we also observed a significant improvement in the fit of the above regression using the chosen value. We also found that results applying this threshold show better agreement with a previously inferred age of the hybrid zone (Schumer lab, pers. comm.).

Amboseli Baboon Analysis. We used genomic similarity calls from Vilgalys et al. (31), i.e., posterior probabilities for diploid genotypes matching reference panels of *P. anubis* or *P. cynocephalus* at a set of SNPs that are differentiated between the species. Estimates were interpolated as described above using an LD-based recombination map. As we again observed extreme outliers in values of $2N_e r$, we applied a threshold at 0.01 (corresponding to 1.4% of the total genome). To visualize the wavelet variance spectrum on a genetic map, we converted genetic lengths to units of Morgans as described above, using genetic lengths of chromosomes (60).

Neanderthal Introgression Analysis. We compared results using three separate estimates of Neanderthal-like haplotype frequency (20, 52, 53). We

interpolated the estimates as described above. For the frequency estimates from ref. 53, we used the sum across identified archaic fragments of a weighted average of the frequency of each fragment in the sample of Icelanders, with the weight being the portion of the window covered by the fragment. For the data from refs. 20 and 52, we used Neanderthal allele frequency estimates in the CEU sample of the 1000 genomes project (2N = 170). For these two studies, we tried two separate measures of Neanderthal-like haplotype frequency. First, we directly used marginal posterior probabilities of a site matching Neanderthal, averaged across individuals in the sample, e.g., column 11 in the output files provided by ref. 20. Next, following the analyses of the original authors, we applied a threshold to the marginal posterior probabilities, calling sites with marginal posterior probability above 0.90 from ref. 20 and above 0.42 from ref. 52 as Neanderthal-like, then taking the average of calls across haplotypes, e.g., column 15 in the files provided by Sankararaman et al. (20). Recombination rate estimates are from ref. 57. We ran the analyses separately using both hg19

and hg38 assembly coordinates. Results were similar in both cases, results are shown for hg38 coordinates. We note that recombination rates in humans are likely estimated with greater resolution relative to the datasets above; consistent with this, we found considerably greater variance in the recombination rate across scales and thus used log-transformed recombination rates for the analyses shown in this paper.

Data, Materials, and Software Availability. Previously published data were used for this work (20, 31, 44, 52, 53).

1. R. G. Harrison, E. L. Larson, Hybridization, introgression, and the nature of species boundaries. *J. Heredity* **105**, 795–809 (2014).
2. N. Barton, B. O. Bengtsson, The barrier to genetic exchange between hybridising populations. *Heredity* **57**, 357–376 (1986).
3. L. H. Rieseberg, K. Whitton, K. Gardner, Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**, 713–727 (1999).
4. C. I. Wu, The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001).
5. B. Ravinet et al., Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *J. Evol. Biol.* **30**, 1450–1477 (2017).
6. B. M. Moran et al., The genomic consequences of hybridization. *eLife* **10**, e69016 (2021).
7. S. Sankararaman, S. Sridhar, G. Kimmel, E. Halperin, Estimating local ancestry in admixed populations. *Am. J. Hum. Genet.* **82**, 290–303 (2008).
8. A. L. Price et al., Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLOS Genet.* **5**, e1000519 (2009).
9. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).
10. P. R. Loh et al., Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
11. R. A. Fisher, *The Theory of Inbreeding* (Oliver and Boyd, 1949).
12. P. Stam, The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genet. Res.* **35**, 131–155 (1980).
13. S. Gravel, Population genetics models of local ancestry. *Genetics* **191**, 607–619 (2012).
14. M. Liang, R. Nielsen, The lengths of admixture tracts. *Genetics* **197**, 953–967 (2014).
15. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Associates, 2004).
16. J. Maynard Smith, J. Haigh, The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
17. H. Tang et al., Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am. J. Hum. Genet.* **81**, 626–633 (2007).
18. I. Hamid, K. L. Korunes, S. Beleza, A. Goldberg, Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. *eLife* **10**, e63177 (2021).
19. S. Yair, K. M. Lee, G. Coop, The timing of human adaptation from Neanderthal introgression. *Genetics* **218**, iyab052 (2021).
20. S. Sankararaman et al., The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
21. M. Hajdinjak et al., Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. *Nature* **592**, 253–257 (2021).
22. P. Nosil, D. J. Funk, D. Ortiz-Barrientos, Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**, 375–402 (2009).
23. D. Berner, M. Roesti, Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. *Mol. Ecol.* **26**, 6351–6369 (2017).
24. K. Samuk et al., Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Mol. Ecol.* **26**, 4378–4390 (2017).
25. S. Aeschbacher, J. P. Selby, J. H. Willis, G. Coop, Population-genomic inference of the strength and timing of selection against gene flow. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7061–7066 (2017).
26. I. Juric, S. Aeschbacher, G. Coop, The strength of selection against Neanderthal introgression. *PLOS Genet.* **12**, e1006340 (2016).
27. M. Schumer et al., Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656–660 (2018).
28. S. H. Martin, J. W. Davey, C. Salazar, C. D. Jiggins, Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biol.* **17**, e2006288 (2019).
29. N. B. Edelman et al., Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594–599 (2019).
30. E. Calfee et al., Selective sorting of ancestral introgression in maize and teosinte along an elevational cline. *PLOS Genet.* **17**, e1009810 (2021).
31. T. P. Vilgalys et al., Selection against admixture and gene regulatory divergence in a long-term primate field study. *Science* **377**, 635–641 (2022).
32. M. W. Nachman, B. A. Payseur, Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. B: Biol. Sci.* **367**, 409–421 (2012).
33. C. Veller, N. B. Edelman, P. Muralidhar, M. A. Nowak, Recombination and selection against introgressed DNA. *Evolution* **77**, 1131–1144 (2023).
34. N. H. Barton, Multilocus clines. *Evolution* **37**, 454–471 (1983).
35. K. Harris, R. Nielsen, The genetic cost of Neanderthal introgression. *Genetics* **203**, 881–891 (2016).
36. C. C. A. Spencer et al., The influence of recombination on human genetic diversity. *PLOS Genet.* **2**, e148 (2006).
37. I. Pugach, F. Matveyev, A. Wollstein, M. Kayser, M. Stoneking, Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biol.* **12**, R19 (2011).
38. A. H. Chan, P. A. Jenkins, Y. S. Song, Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1003090 (2012).
39. J. Sanderson, H. Sudoyo, T. M. Karafet, M. F. Hammer, M. P. Cox, Reconstructing past admixture processes from local genomic ancestry using wavelet transformation. *Genetics* **200**, 469–481 (2015).
40. D. B. Percival, A. T. Walden, *Wavelet Methods for Time Series Analysis* (Cambridge University Press, 2000), vol. 4.
41. B. C. Haller, P. W. Messer, Slim 3: Forward genetic simulations beyond the Wright-Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
42. G. Coop, Genetic similarity versus genetic ancestry groups as sample descriptors in human genetics. arXiv [Preprint] (2023). https://arxiv.org/abs/2207.11595v2 (Accessed 8 January 2023).
43. National Academies of Sciences, Engineering, and Medicine, Using population descriptors in genetics and genomics research: A new framework for an evolving field (2023).
44. D. L. Powell et al., The genetic architecture of variation in the sexually selected sword ornament and its evolution in hybrid populations. *Curr. Biol.* **31**, 923–935 (2021).
45. D. L. Powell et al., Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science* **368**, 731–736 (2020).
46. B. M. Moran et al., A lethal mitonuclear incompatibility in complex I of natural hybrids. *Nature* **626**, 119–127 (2024).
47. A. Samuels, J. Altmann, Immigration of a *Papio anubis* male into a group of *Papio cynocephalus* baboons and evidence for an anubis-cynocephalus hybrid zone in Amboseli, Kenya. *Int. J. Primatol.* **7**, 131–138 (1986).
48. J. Tung, M. J. Charpentier, D. A. Garfield, J. Altmann, S. C. Alberts, Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Mol. Ecol.* **17**, 1998–2011 (2008).
49. R. E. Green et al., A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
50. F. Racimo, S. Sankararaman, R. Nielsen, E. Huerta-Sánchez, Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359–371 (2015).
51. B. Vernot, J. M. Akey, Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017–1021 (2014).
52. M. Steinrücken, J. P. Spence, J. A. Kamm, E. Wieczorek, Y. S. Song, Model-based detection and analysis of introgressed Neanderthal ancestry in modern humans. *Mol. Ecol.* **27**, 3873–3888 (2018).
53. L. Skov et al., The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* **582**, 78–83 (2020).
54. L. Skov et al., Detecting archaic introgression using an unadmixed outgroup. *PLOS Genet.* **14**, e1007641 (2018).
55. B. Y. Kim, C. D. Huber, K. E. Lohmueller, Deleterious variation shapes the genomic landscape of introgression. *PLOS Genet.* **14**, e1007741 (2018).
56. N. Telis, R. Aguilar, K. Harris, Selection against archaic hominin genetic variation in regulatory regions. *Nat. Ecol. Evol.* **4**, 1558–1566 (2020).
57. B. V. Halldorsson et al., Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363**, eaau1043 (2019).
58. B. C. Haller, J. Galloway, J. Kelleher, P. W. Messer, P. L. Ralph, Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Mol. Ecol. Res.* **19**, 552–566 (2019).
59. B. Whitcher, waveslim: Basic Wavelet Routines for One-, Two-, and Three-Dimensional Signal Processing (2022). R package version 1.8.4.
60. J. Rogers et al., A genetic linkage map of the baboon (*Papio hamadryas*) genome based on human microsatellite polymorphisms. *Genomics* **67**, 237–247 (2000).