# Genomic rearrangements and the evolution of clusters of locally adaptive loci

Sam Yeaman[a,b,1]

[a]Department of Forest and Conservation Sciences and [b]Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Edited by Michael Lynch, Indiana University, Bloomington, IN, and approved April 1, 2013 (received for review November 8, 2012)

**Numerous studies of ecological genetics have found that alleles contributing to local adaptation sometimes cluster together, forming "genomic islands of divergence." Divergence hitchhiking theory posits that these clusters evolve by the preferential establishment of tightly linked locally adapted mutations, because such linkage reduces the rate that recombination breaks up locally favorable combinations of alleles. Here, I use calculations based on previously developed analytical models of divergence hitchhiking to show that very few clustered mutations should be expected in a single bout of adaptation, relative to the number of unlinked mutations, suggesting that divergence hitchhiking theory alone may often be insufficient to explain empirical observations. Using individual-based simulations that allow for the transposition of a single genetic locus from one position on a chromosome to another, I then show that tight clustering of the loci involved in local adaptation tends to evolve on biologically realistic time scales. These results suggest that genomic rearrangements may often be an important component of local adaptation and the evolution of genomic islands of divergence. More generally, these results suggest that genomic architecture and functional neighborhoods of genes may be actively shaped by natural selection in heterogeneous environments. Because small-scale changes in gene order are relatively common in some taxa, comparative genomic studies could be coupled with studies of adaptation to explore how commonly such rearrangements are involved in local adaptation.**

chromosomal rearrangement | genetic architecture | migration–selection balance | synteny

Understanding the genetic basis of adaptation is a central problem of evolutionary biology. In light of a rapidly growing body of empirical data, there has been considerable interest in how many genes contribute to local adaptation, the distribution of their effect sizes, and where they are located in the genome (1–3). Numerous studies have found evidence that the loci that are most strongly differentiated between populations are sometimes clustered together (i.e., in close physical linkage), in what have been termed "genomic islands of divergence" (4–11). However, other studies have found little or no evidence of divergent alleles clustering in large genomic islands (6, 10, 12). In light of these conflicting patterns, further research is necessary to clarify when and how genomic islands should be expected to evolve.

Genomic islands may evolve as a byproduct of natural selection or as direct response to natural selection (or some combination of the two). In the byproduct case, selection increases the frequency of a focal allele, causing alleles at linked neutral loci to also increase in frequency (13–17). This yields a signature of divergence at adjacent loci that decreases with increasing rates of recombination (i.e., distance from the selected site), and will gradually disappear following the initial selective sweep in populations that diverged in allopatry. If divergent selection occurs in the face of continued gene flow, genetic divergence at linked neutral loci will tend to persist, owing to a reduction in the effective rate of migration (18, 19), which is a limited "neutral" form of divergence hitchhiking (20).

Genomic islands can also evolve as a direct adaptive response to divergent selection with gene flow. If adaptation to a given environment is based on changes in allele frequency at multiple loci, then recombination between the chromosomes of a locally adapted parent and a maladapted immigrant will tend to break up combinations of alleles with high fitness that have been established by selection (21–24). As long as migration rates are low enough that the descendents of a locally adapted allele spend a large proportion of their time in the patch where they are fittest, then it is beneficial to be tightly linked to other alleles that are also locally adapted.

There are at least four different mechanisms for the evolution of genomic islands as a direct response to selection, based on the general advantage of reduced recombination between locally adapted alleles: (*i*) increased establishment probability of linked locally adapted alleles [divergence hitchhiking (5, 25, 26)], (*ii*) increased persistence time of linked locally adapted alleles following secondary contact (6, 27), (*iii*) competition among combinations of alleles with similar phenotypic effects but different linkage relationships [competition among genetic architectures (28)], and (*iv*) the fixation of genomic rearrangements that move locally adapted loci into close genetic linkage [competition among genomic architectures (28)]. For all four of these mechanisms, the fitness benefits due to linkage tend to decrease with increasing rates of recombination between locally adapted alleles (21, 24, 25, 28, 29). As a result, genomic islands are more likely to evolve in areas of the genome with low rates of recombination, such as coldspots or regions with segregating inversions (17, 30–33). It should be noted, however, that these regions are also more likely to have high levels of divergence due to the byproduct mechanism discussed above (17).

Unfortunately, it is often very difficult to evaluate which of these mechanisms may be involved in building genomic islands of divergence, especially because they are not necessarily mutually

## Significance

**Genome scans often find that the loci involved in local adaptation tend to cluster together on chromosomes. A leading explanation suggests that clusters evolve because the probability of a new mutation establishing is higher when occurring near another locally adapted mutation, because such architectures are seldom disrupted by recombination. I show that this theory is unlikely to explain empirically observed clusters. Instead, simulations show that clusters are more likely to form through genomic rearrangements that bring coadapted loci close together. This suggests that ecological selection may play an important role in shaping genome architecture, in contrast to many nonadaptive explanations.**

Author contributions: S.Y. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The source code for the simulations has been deposited in the Sourceforge Repository, https://sourceforge.net/projects/nemorearrange/.

[1]E-mail: yeaman@zoology.ubc.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1219381110/-/DCSupplemental.

exclusive. Very few studies have been able to identify whether loci with signatures of divergence are in fact contributing to local adaptation or just neutral alleles "along for the ride." Several empirical studies have identified multiple signatures of divergence that map to chromosomal inversions (34–37), but others have found that inversions were only weakly implicated in adaptive divergence (e.g., ref. 38). Beyond understanding the origins and maintenance of local adaptation, determining the importance of these various mechanisms also has implications for understanding the nature of genome evolution. If islands of divergence are actively built by the fixation of genomic rearrangements (mechanism *iv*), then adaptation-with-gene flow could be an important and underappreciated driver of evolution of genomic architecture.

This "competition among genomic architectures" mechanism could provide a partial explanation for the dramatic variation in the rates at which rearrangements accumulate in the genomes of different taxa. Rates of coarse-scale chromosomal rearrangement are more than three orders of magnitude higher in *Drosophila* than in mammals (39). Although relatively low rates of change in gene order have been found in Lepidopterans based on coarse-scale mapping data (40), fine-scale genomic analyses have shown rates of rearrangement as high as or higher than those in *Drosophila*, largely due to numerous small-scale transpositions and inversions (41, 42). However, genomic analyses of birds reveal rates of gene-order evolution even lower than those in mammals at both the micro- and macroscale (43, 44). Rates of rearrangement accumulation also vary across the genome; some highly conserved regions have very low rates of accumulation, relative to so-called fragile regions that experience higher rates of accumulation (45, 46). Explanations for this dramatic variation in rates of genome evolution have largely focused on nonadaptive arguments based on rearrangement rate, the activity of transposable elements, or demographic drift-based mechanisms that modify the efficacy of purifying selection to purge mildly deleterious rearrangements (39, 47–50). If rearrangements tend to be involved in local adaptation, then ecologically mediated positive selection may play a role in driving some of these differences in rates and patterns of genome evolution.

Here, I jointly consider the evolution of local adaptation and genome architecture in populations inhabiting heterogeneous environments. I develop a heuristic based on previously established analytical methods to explore the likelihood of genomic islands of divergence evolving under the standard divergence hitchhiking hypothesis (mechanism *i*). Using this heuristic, I show that increases in establishment probability due to divergence hitchhiking should rarely be large enough to expect strong statistical signatures of islands of divergence. I then use an analytical approximation and individual-based simulations to explore the potential importance of small genomic rearrangements as a way to create clusters of locally adaptive loci, which may be favored owing to reduced rates recombination between loci (mechanism *iv*). I find that clustering of locally adaptive loci evolves readily under a wide range of parameter space, suggesting that small-scale genomic rearrangements may play an important role in the evolution of islands of genomic divergence.

## Results

**Likelihood of Island Formation by Divergence Hitchhiking.** If genomic islands of divergence commonly evolve by divergence hitchhiking, then we would expect a statistical enrichment of the number of locally adapted alleles that occur in clusters, relative to the number of unlinked locally adapted alleles scattered throughout the genome. A simple expression for this ratio can be approximated by

$$\frac{n_{clustered}}{n_{unlinked}} = \frac{EP_{clustered}}{EP_{unlinked}} \frac{\mu_{clustered}}{\mu_{unlinked}}, \quad [1]$$

where $n_{clustered}$ and $n_{unlinked}$ represent the number of locally adapted alleles that establish in a cluster extending to 50 cM on either side of the focal locus (i.e., $r = 0.5$) vs. anywhere in the rest of the genome, $EP_{clustered}$ and $EP_{unlinked}$ represent the average establishment probabilities for clustered and unlinked mutations, and $\mu_{clustered}$ and $\mu_{unlinked}$ represent the genomic mutation rates averaged over all clustered and unlinked loci. The establishment probabilities, $EP_{clustered}$ and $EP_{unlinked}$, can be estimated by using the continent-island model of Bürger and Ackerman (24) to calculate the relative fitness of linked and unlinked mutations, and splicing this into Kimura's (51) equation for fixation probability, as shown by Yeaman and Otto (52). This approach assumes that a genomic island begins as a single differentiated allele (at a "focal locus") and expands as other linked locally adapted mutations establish (details are given in *Methods*).

Although the advantage due to divergence hitchhiking may be quite large for mutations that happen to be very tightly linked to the initially diverged focal locus, most genomes are very large, and the chance that a mutation would occur fortuitously in tight linkage is quite small, relative to the chance of a mutation occurring elsewhere in the genome. If we assume that the loci that could potentially contribute to local adaptation are randomly distributed throughout the genome, then relatively few of these loci would be found in tight linkage to the focal locus. The total advantage due to divergence hitchhiking can be quantified by averaging the ratio $EP_{clustered}/EP_{unlinked}$ over all possible linked sites (from $r = 0.5$ to $r \to 0$, around the focal locus). As shown in Fig. 1, for this model the average divergence hitchhiking advantage is rarely higher than threefold (and usually much lower) over a wide range of biologically realistic parameters. The only

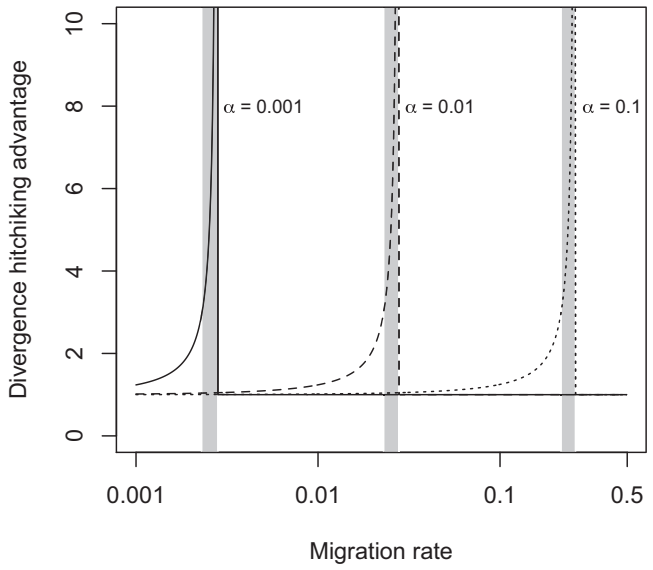


**Fig. 1.** On average, mutations linked to a locus with an established locally adapted allele have only a slight advantage over unlinked mutations in terms of increased establishment probability due to divergence hitchhiking, (divergence hitchhiking advantage = $EP_{clustered}/EP_{unlinked}$), except at very high migration rates, just below the critical threshold for the maintenance of local adaptation (gray shading). Curves show the divergence hitchhiking advantage for three strengths of selection on the new mutation (α), averaged over mutations occurring at a range of distances from the initial divergent locus, from 0.001 cM to 50 cM, using Eq. **3**. In all cases, selection on the established allele is β = 0.9; similar results are found for lower values of β, but peaks tend to be narrower.

exceptions to this pattern occur over very narrow ranges of migration rate falling just below the critical threshold above which no adaptive divergence occurs (Fig. 1, gray shading). In these cases, the divergence hitchhiking advantage is high because the establishment probability of small unlinked mutations is close to 0.

Under the assumption that the loci that can potentially contribute to local adaptation are uniformly distributed throughout the genome, the ratio $\mu_{linked}/\mu_{unlinked}$ will be a function of the fraction of the genome that occurs within 50 cM on either side of the focal locus (i.e., the region over which the divergence hitchhiking advantage is calculated in Fig. 1), relative to the fraction of the genome that is unlinked. To develop a simple heuristic under assumptions most favorable to the divergence hitchhiking hypothesis, this ratio can be approximated as $\mu_{clustered}/\mu_{unlinked} = 1/2/(k - 1/2) = 1/(2k - 1)$, where $k$ is the number of chromosomes, assuming chromosomes are all of equivalent size, and the region maintained as a cluster spans 50% of a single chromosome. If we assume the divergence hitchhiking advantage, $EP_{clustered}/EP_{unlinked} \leq 3$ (from Fig. 1) and $k = 21$ (as in the threespine stickleback), then using Eq. **1**, we would expect at least 13 unlinked alleles to establish for every clustered allele that establishes. Note that in most cases this ratio would be even larger, because the advantage of divergence hitchhiking is seldom as high as threefold (Fig. 1), and the region maintained as a cluster in most empirical contexts is typically much smaller than 50% of a chromosome.

Based on the above heuristic, very few of the alleles that contribute to local adaptation are expected to reside in a given island of divergence except under very limited regions of parameter space. It may be argued that a shortcoming of this heuristic is that it considers only the formation of a genomic island around a single focal locus, but there could be many such focal loci in different locations throughout the genome, each of which could form a genomic island (effectively increasing the mutational target and $\mu_{clustered}$). However, this issue is unlikely to be important because both the increase in establishment probability and the size of the genomic island that can be formed by divergence hitchhiking are proportional to the strength of selection on the focal divergent locus. If many such focal loci exist throughout the genome, then the maximum strength of selection at each focal locus will typically be lower, thereby reducing both the potential size of the cluster [due to $s$ vs. $r$ (19, 53)] and the divergence hitchhiking advantage for any linked mutations. Thus, divergence hitchhiking is unlikely to be a broadly important explanation for the existence of islands of divergence without some genomic mechanism that dramatically reduces the rate of recombination around the initially divergent locus, such as a large inversion (24, 32) or modifiers creating a recombination coldspot. Alternatively, if genomic rearrangements bring together the loci that contribute to local adaptation (thereby creating a genomic island), this would effectively increase the ratio of $\mu_{clustered}/\mu_{unlinked}$, and could facilitate divergence hitchhiking in subsequent bouts of local adaptation (if the alleles forming the genomic island happened to be lost owing to genetic homogenization among populations).

**Rearrangements Lead to Clustering of Locally Adaptive Loci.** Clusters of locally adaptive loci might be expected to evolve under migration–selection balance through the spread of genomic rearrangements, provided the rate of rearrangements and their probability of fixation are both sufficiently high. If the loci that contribute to local adaptation are initially randomly distributed throughout the genome, then the time required to evolve a configuration where these loci are organized in a single tightly linked cluster can be approximated by accounting for the rate of occurrence of suitable rearrangements and their probability of fixation [based on a model by Kirkpatrick and Barton (32)]. If $L$ is the number of base pairs over which $r \approx 0$, $G$ is the number of base pairs in the genome, $\tau$ is the rate of rearrangement per locus, per generation, $N_p$ is the total metapopulation census size, and $m$ is the migration rate, then the time (in generations) to evolve a clustered architecture can be approximated as

$$t \approx G/(N_p \tau m L) \quad [2]$$

(*Methods* gives the derivation). This coarse approximation will range over many orders of magnitude, depending on the parameters, and provides some intuition about their relative importance. It is insensitive to the number of loci involved in local adaptation ($n$), because when $n$ increases, the longer time required to rearrange a larger number of loci is offset by increases in the realized rate of rearrangement, because there are more loci that can potentially be rearranged into a cluster. However, the time required to evolve a partially clustered architecture (by Eq. **4**) may be several orders of magnitude less than the time to evolve a fully clustered architecture (by Eq. **2**; Fig. S1). These equations can be used to provide coarse predictions of whether clustering of genomic architecture might be expected in a given taxon. For example, if $N_p = 10^5$, $L = 10^5$, $G = 10^9$, and $m = 0.001$, then complete clustering might be seen at biologically realistic time scales if $\tau > 10^{-5}$ (as $t \approx 10^7$, by Eq. **2**), and partial clustering of $n_x = 5$ out of $n = 50$ loci might occur if $\tau > 10^{-7}$ (as $t \approx 2 \times 10^7$, by Eq. **4**). This model illustrates how slight variations in some parameters can yield large effects on the rate of genome evolution, but as a means of making quantitative predictions it should be seen as a coarse approximation, owing to the assumptions involved in its derivation (details are given in *Methods*).

To explore the evolution of genome architecture more explicitly, with fewer simplifying assumptions about the evolutionary dynamics, I used individual-based simulations of a two-patch model under migration–selection balance similar to that used in ref. 28. These simulations are initialized with 10 phenotype-affecting (hereafter, selected) loci equally spaced along a single chromosome, each separated by 49 neutral loci. Genomic rearrangements occur by moving a single selected locus to a new location according to either a cut-and-paste or a duplication model (*Methods*), at a per-locus rate of $\tau$. Because rearrangements can result in changes in the number of copies of each locus, small additional fitness costs were incurred for individuals that did not have exactly two copies of each of the 10 original selected loci (details are given in *Methods*).

When both populations inhabit the same type of environment, there is no benefit to having a tightly linked genomic architecture, and the distribution of selected loci on the simulated chromosome evolves very slowly, mainly under the influence of random genetic drift (Fig. 2*A*). By contrast, when populations inhabit different environments and experience a tension between migration and divergent selection, the architecture evolves rapidly through the successive fixation of single-locus rearrangements, eventually leading to clusters of tightly linked selected loci that are fixed across both populations (Fig. 2*B*). In this case, populations rapidly evolve genotypic divergence through the establishment of locally adapted alleles at each locus (usually within several thousand generations). Once genotypic divergence is established, rearrangements that move one selected locus with a locally adapted allele close to another tend to invade and replace less tightly clustered architectures, because of the advantage of reduced recombination between more tightly linked loci. In Fig. 2*B*, a nearly fully clustered architecture evolves in an amount of time similar to that predicted by Eq. **5** (for $m = 0.01$, $n = 10$, $L = 2$, $G = 490$, $\tau = 10^{-6}$, $N_p = 20{,}000$, then $t = 8.4 \times 10^5$).

In most cases, the rate of evolution toward clustered architectures increases with the strength of divergent selection (Fig. 3*A*) and migration rate (Fig. 3*B*). However, when migration rates are high and just below the critical threshold (24, 52), the rate of

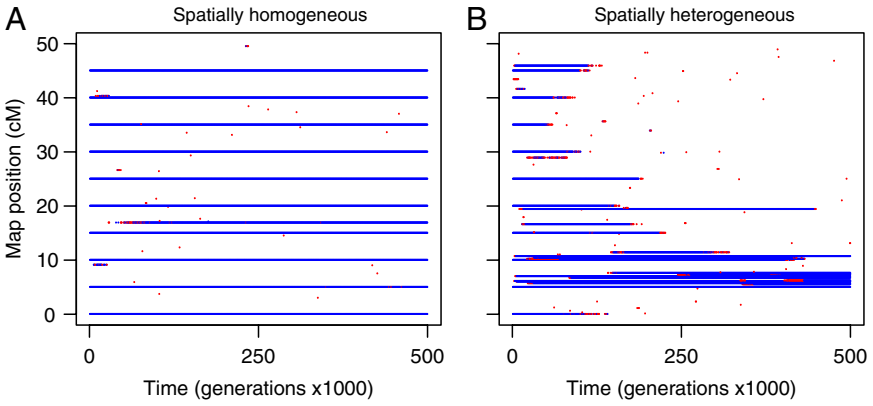


**Fig. 2.** The distribution of selected loci along a single chromosome over 500,000 generations of evolution when the environment is spatially homogeneous (*A*; optima are $\theta$ = +1 and +1) and spatially heterogeneous (*B*; optima are $\theta$ = +1 and −1). Populations are often polymorphic for arrangements, with selected and neutral loci both present in different individuals at the same chromosomal location. For a given chromosomal position at a given point in time, blue indicates that most individuals in both populations have a selected locus, red indicates that most individuals in one patch have a selected locus whereas most individuals in the other locus have a neutral locus, and white indicates that most individuals in both populations have a neutral locus; $m$ = 0.01, $\phi$ = 0.75, $N = 10^4$, $\tau = 10^{-6}$.

evolution to clustered architectures is somewhat reduced (e.g., $m$ = 0.1, Fig. 3*B*; $m_{crit}$ = 0.23). Although these general patterns are consistent with the predictions of Yeaman and Whitlock (28), the effect of the strength of selection on the rate of genome evolution is not accounted for by Eqs. **2** or **4**, likely in part because these approximations ignore rearrangements that yield clusters of loci with $r > 0$ (*Methods*). When $r > 0$, the amount of linkage disequilibrium between loci, and therefore the fitness benefits of linkage, should scale with the strength of selection (19, 53), such that under stronger selection there is effectively a larger target region ($L$) over which rearrangements can yield clusters of loci with fitness benefits owing to linkage.

The rate of evolution of clustering is also strongly determined by both the population size ($N$) and the rate of occurrence of rearrangements ($\tau$). Clustering evolves much more rapidly when $N = 10^4$ than when $N = 10^3$, but if $\tau$ is increased by an order of magnitude for $N = 10^3$, the rate of evolution of clustering is quite similar to when $N = 10^4$ (Fig. 3*C*). This shows that the rate of architecture evolution depends heavily on the population-level rate of occurrence of rearrangements ($N\tau$), as predicted by Eq. **2**.

Clustered genomic architectures also evolve when environments are both spatially and temporally heterogeneous (Fig. 4*A*) and when populations go through alternating periods of spatial heterogeneity and homogeneity (Fig. 4*C*). Fig. 4*A* shows the evolution of genomic architecture when populations inhabit a spatially heterogeneous environment (local optima are +1/−1) that also fluctuates over time (as shown in Fig. 4*B*). When the period of temporal fluctuations is short ($T$ =1,000, Fig. 4*A*), clustering evolves more slowly and stabilizes at an intermediate level. The amount of clustering at the end of the simulations varies nonmonotonically as a function of $T$, with the most clustering seen when $T$ = 5,000 (Fig. 4*A*). When the period is very long ($T = 10^5$), genomic architecture evolves toward highly clustered architectures but fluctuates due to transient increases in the number of unclustered selected loci (*SI Text* and Fig. S2).

When populations inhabit an environment that periodically alternates between spatially homogeneous and spatially heterogeneous (as shown in Fig. 4*D*), clustered genomic architectures still evolve, although the rate of evolution is slower for smaller $T$ (Fig. 4*C*). In these simulations, genetic divergence often collapses completely during the intervals of homogeneity and any islands of allelic divergence are lost, but clustered genomic architectures persist and can rapidly give rise to new islands of allelic divergence when heterogeneous environments are reestablished. Clustered architectures still evolve when the environment is homogeneous for more time than it is heterogeneous, but at a slower rate (*SI Text* and Fig. S3).

From the above results, it is clear that the rate of rearrangement is an important driver of architecture evolution, but is there also an effect of the type of rearrangement? There are numerous mechanisms that can give rise to the rearrangement of small fragments of the genome: (*a*) nonhomologous or ectopic recombination involving two closely spaced crossovers (54, 55), (*b*) repair following multiple double-stranded breaks (56–58), (*c*) intrachromosomal recombination forming circular DNA that reinserts elsewhere in the genome (59), (*d*) excision and reinsertion mediated by flanking transposons (60, 61), (*e*) transposon-mediated transduplication [e.g., by helitrons (62) or Pack-MULEs (63)], and (*f*) reverse transcription and retrotransposition of mRNA, which results in the loss of introns in the daughter copy (64). Of these, some result in the movement of a gene from one place to another (e.g., *a*–*d*), whereas others result in the creation of a duplicate copy in a new location (e.g., *e* and *f*). Unfortunately, there is still little information on the relative frequency of these different mechanisms in different species. Although all of the above simulations were run using a cut-and-paste model of rearrangement similar to mechanisms *a*–*d* (*Methods*), the rate of evolution toward clustered architectures was not substantially changed when a duplication model similar to mechanisms *e* and *f* was used instead (Fig. S4). However, unlike the cut-and-paste model, a low rate of gene deletion ($\eta$) was necessary for significant clustering to evolve in the duplication model, because otherwise there was no mechanism through which the less clustered parental copy of a locus could be lost (Fig. S4).

In these simulations, gene duplication, gene deletion, and recombination between rearranged and ancestral haplotypes can all result in changes in the number of copies of a given locus that are passed from parent to offspring. Empirical studies have found that changes in gene copy number often involve phenotypic changes or fitness costs, likely because changes in the dosage of some genes upset the overall balance of expression (65–67). All of the above simulations were run under the assumption of relatively weak selection against copy number variants (CNV; *Methods*). The rate of evolution toward clustered architectures tended to decrease with increasing strength of selection against CNV (Fig. S5), suggesting that clustered architectures are more likely to involve genes that have only minimal fitness effects due to changes in copy number (i.e., less potential to upset the balance of expression; refs. 65–67). When simulations were run excluding fitness costs for CNV, clustered architectures still evolved at rates comparable to those shown in Fig. 3, providing the rate of gene deletion was at least as high as the rate of rearrangement (Fig. S4).

## Discussion

**Implications for Genomic Islands of Divergence.** There has been considerable debate about the importance of divergence hitchhiking for the evolution of genomic islands of divergence (16, 20, 26), with recent work noting that the effect of divergence hitchhiking on establishment probability may often be quite limited relative to the effect of selection acting directly on beneficial

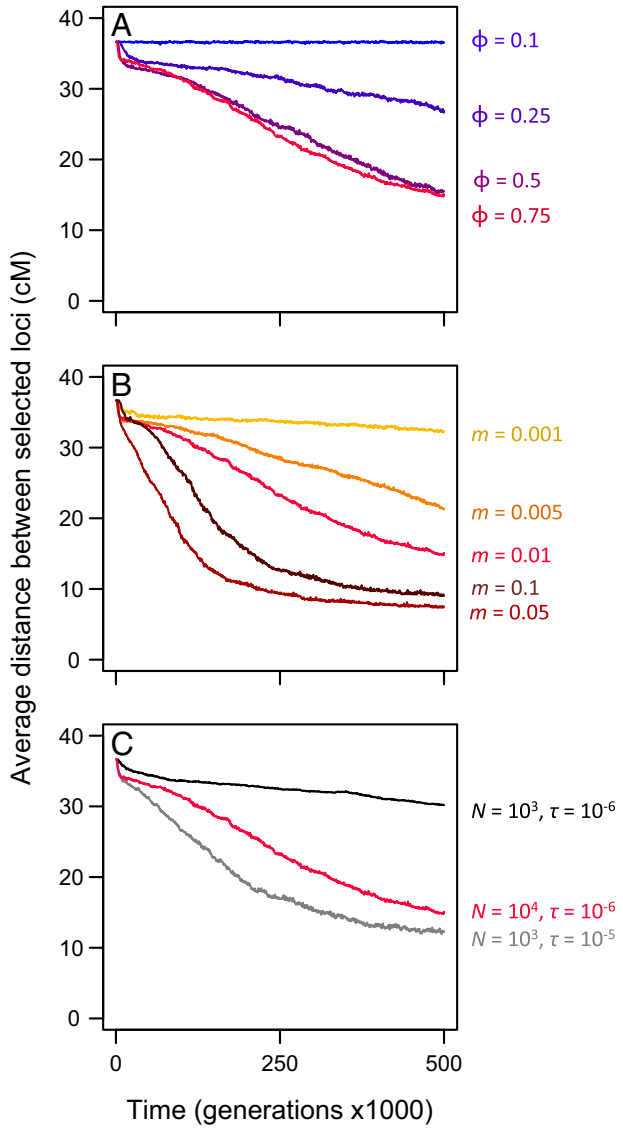


**Fig. 3.** The average pairwise distance between selected loci decreases over time at different rates, given different strengths of phenotypic selection (*A*), migration rates (*B*), and population sizes and rearrangement rates (*C*). Unless otherwise specified, $m = 0.01$, $\phi = 0.75$, $N = 10^4$, $\tau = 10^{-6}$, and all other parameters are as described in *Methods*.

mutations (25). Here, I present a heuristic model that explicitly quantifies the effect of divergence hitchhiking on establishment probability (Fig. 1) and show that it is unlikely to yield strong signatures of islands of divergence under the assumption that the loci that can contribute to local adaptation are randomly scattered throughout the genome. Instead, for each locally adapted allele that establishes in a genomic island due to divergence hitchhiking, there will likely be many alleles that establish at random locations throughout the genome. This occurs because the mutational target that resides within a potential genomic island is very limited relative to the mutational target across the whole genome, the fitness advantage for being tightly linked is small, and unlinked mutations still have relatively high probabilities of fixation.

Although the fitness advantage of being tightly linked may usually be too small to have much effect on the evolution of genomic islands on the short time scale of a single bout of local adaptation, the simulation results I present show that competition among genomic architectures can have pronounced effects over longer evolutionary time scales. Highly clustered genomic architectures evolve readily given sufficiently large populations, rearrangement rates, strengths of selection and rates of migration (Fig. 3). Within this context, the mutational target itself is rearranged, which would increase the proportion of clustered relative to unlinked mutations that could establish in any subsequent bouts of adaptation.

Divergence hitchhiking could potentially play an important role in forming genomic islands of divergence in conjunction with this rearrangement mechanism, if a species experiences temporal environmental or demographic fluctuations that result in repeated episodes of divergent selection and adaptive response followed by homogenization. The cycling of glacial and interglacial periods over millions of years provides an empirical example of one such source of environmental fluctuation, which has likely affected adaptation in numerous species (68). Here, I found that clustered genomic architectures evolve under a range of scenarios with environmental fluctuations (Fig. 4*A*), and these clustered architectures persist through periods of homogenization among populations (Fig. 4*C*), unlike the clustered allelic architectures described by Yeaman and Whitlock (28). The amount of clustering tended to vary with the period of temporal fluctuations (Fig. 4*A*), suggesting that glacial cycles might have quite different effects on the genomics of adaptation in long-lived species such as forest trees compared with short-lived species such as annual plants. Given the cyclical nature of climatic fluctuations and accompanying range shifts, it seems possible that for every speciation event in the history of some genera there have been multiple episodes of divergence and homogenization. Although the process of ecological speciation may be conceptualized as a continuum from early stages of divergence to "good" species with complete reproductive isolation (10), movement along this continuum is not unidirectional. The vagaries of history may result in repeated transitions back and forth along this continuum, increasing the window of opportunity for competition among architectures and divergence hitchhiking to play a role in adaption and, eventually, speciation.

Studies of the genetics of local adaptation and ecological speciation that find evidence of genomic islands should strive to test whether the loci involved have been brought together by rearrangements. This could be achieved by comparing how orthologs associated with local adaptation are distributed along chromosomes in a number of sister species spanning increasing taxonomic distances. Numerous studies have found evidence that segregating inversions are often involved in adaptive genetic divergence (36, 69, 70), but less focus has been placed on other types of rearrangements. It is worth noting that for the inversion-based mechanism of adaptation proposed by Kirkpatrick and Barton (32), inversions must be segregating to reduce recombination, whereas rearrangements discussed in this paper reduce recombination by changing the position of loci on chromosomes, and therefore tend to fix across populations. Another important prediction from this work is that islands of divergence with multiple selected alleles could be expected to form in allopatry, provided the species had a previous history of adaptation with gene flow (during which time clusters of loci could have evolved through rearrangements). Indeed, this could provide partial explanation for the similarity in size and number of genomic islands found in sympatric vs. allopatric pairs of populations of *Helianthus* (71).

**Implications for the Evolution of Genome Architecture.** With current advances in genome sequencing technologies, considerable research has focused on finding patterns in the organization of genes on chromosomes. Comparative genomic studies have found that the distribution of rearrangement breakpoints tends to be nonrandom throughout the genome, with higher numbers of rearrangements at "fragile sites" and lower numbers at "solid

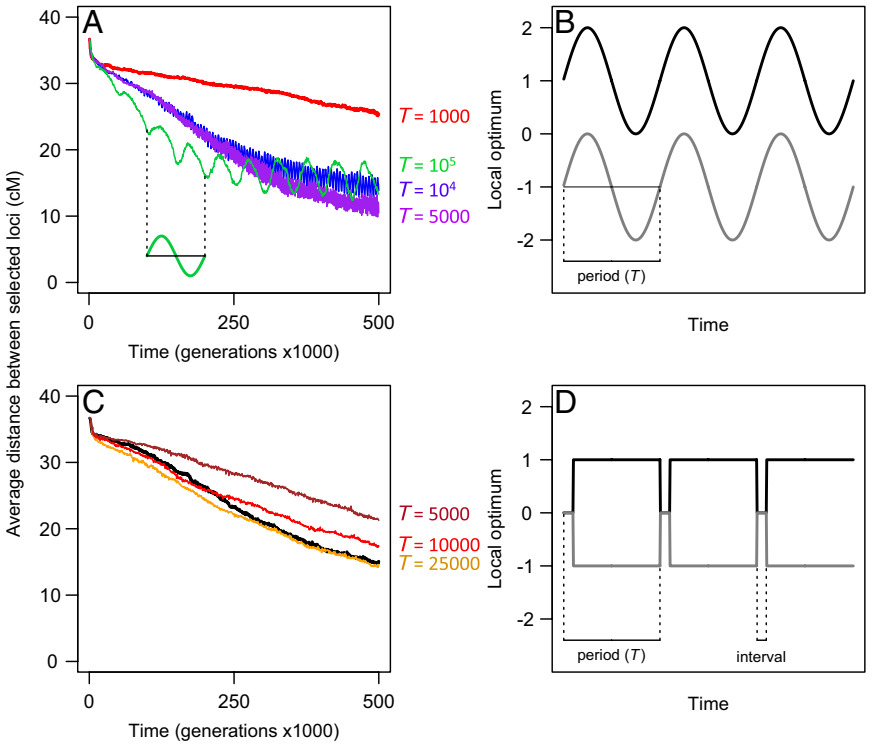


**Fig. 4.** Evolution of clustered architectures under spatial and temporal heterogeneity (*A*, as shown in *B* and spatial heterogeneity punctuated with periods of homogeneity (*C*, as shown in *D*). In C, the interval lasts 1,000 generations, during which $m = 0.5$; the black line shows the amount of clustering for a constant spatially heterogeneous environment with no periods of homogeneity. All other parameters are set as in Fig. 3.

sites" (45, 46). Studies of gene regulation have found evidence that neighboring genes sometimes have highly correlated expression profiles (72–75), or tend to be functionally related (76, 77). Evolutionary explanations for these patterns have been varied but are often related to the balance between rates of rearrangement, genetic drift, and purifying selection (50, 72). Where explanations invoke positive selection, the fitness benefits are typically based on changes that are endogenous to the organism and environment-independent, such as facilitated regulation of functionally related genes (78) or reduced expression noise (79).

The general perspective that arises from these studies is that the architecture of the genome evolves with little connection to the ecology of the species, beyond the impact of demography on genetic drift (50, 72). Lynch and coworkers (49, 80) have presented compelling evidence showing that species with smaller effective population sizes tend to have larger genomes, due the reduced efficiency of natural selection to eliminate insertions that increase the global mutation hazard. Beyond this link to ecology through demography, in most cases, the patterns that have been studied in comparative genomics are testable without going out into nature to study the ecology of the species. By contrast, the mechanism of genome evolution described in this study is explicitly tied to the ecology of the species. This mechanism can be seen as an extension of the inversion-based mechanisms of local adaptation and speciation that have received considerable study in recent years (30–32) and models for the evolution of supergenes (81), which both have much in common with previous studies on the evolution of recombination (82–86). This mechanism is also similar in many ways to models for the evolution of sex chromosomes; Charlesworth and Charlesworth (87) showed that rearrangements moving autosomal loci to sex chromosomes would be favored when different alleles were favored in each sex. By contrast, in the mechanism discussed here, rearrangements are favored because they reduce recombination between loci with alleles that are adapted to different environments. Although earlier models of chromosome evolution also noted that positive selection due to linkage could result in the spread of rearrangements, they typically focused on the effect of drift and population structure on the spread of large rearrangements with deleterious effects in the heterokaryotype (e.g., 88, 89).

Because the rearrangement-based mechanism discussed here is explicitly tied to ecology, studying it may require a different approach than is typically used in comparative genomics. In many cases, local adaptation involves genes from numerous pathways. For example, adaptation to freshwater in threespine sticklebacks includes responses to novel water chemistry, temperature regimes, visual environments, predator communities, and forage availability (89–93). Therefore, analyses of clustering based on gene ontology (e.g., 77) may fail to detect clusters of loci if they involve genes with such different proximate functions. Also, clusters that evolve in response to adaptation with gene flow might span physical distances much larger than those commonly analyzed in the coexpression cluster studies, which typically examine sliding windows of 2 to 15 genes (e.g., 73–75). For example, a genomic island spanning >10 cM has been observed in pea aphids adapting to different hosts (5), which presumably includes tens to hundreds of genes. One intriguing comparative genomic study of amniotes by Larkin et al. (93) found that rearrangement breakpoints tended to cluster in regions containing genes with gene ontology terms relating to inflammation and muscular contraction, whereas conserved regions with few breakpoints tended to include genes for development. They suggested that because inflammation and muscular contraction genes are more directly involved in interactions with the environment, evolution may have favored the spread of rearrangements in response to novel environmental challenges. Similarly, Freeling et al. (61) found that genes with products that interact more directly with rapidly changing biotic and abiotic environments (e.g., disease resistance or flowering) had higher

likelihoods of transposition in *Arabidopsis*. Although rather speculative, these interpretations of the data are consistent with the predictions for the spread of rearrangements due to local adaptation discussed here, so it would be interesting to test for similar patterns in other taxa.

The rate of rearrangement occurrence is one of the most important factors determining whether clustered architectures evolve (Eq. **2** and Fig. 3*C*), but very little is known about the magnitude of $\tau$ for different rearrangement mechanisms, despite mounting evidence that numerous small-scale rearrangements have accumulated within some lineages over time (42, 61). For example, in *Arabidopsis thaliana*, it is estimated that 21–27% of all genes have been rearranged since the common ancestor with Papaya, almost all of which occurred as single-gene rearrangements (61, 94). Whereas many of these were likely caused by retrotransposition (94), other mechanisms are implicated in at least 55% of the rearrangements that occurred since the *A. thaliana*–*Arabidopsis lyrata* split (59). Although highly speculative, it is possible that the 3D organization of chromosomes in the nucleus could facilitate the evolution of clustering. Spatial proximity of genes in the nucleus has been shown to influence both gene regulation (95) and rates of rearrangement (96), so if genes that tend to be involved in local adaptation are also brought into close proximity in the nucleus to facilitate coregulation, they might more readily be rearranged onto the same chromosome. Further research will help identify how commonly the various rearrangement mechanisms occur, and how this varies among species.

It is important to note that whereas spatial environmental heterogeneity will tend to favor tight clustering of the loci involved in local adaptation, some forms of temporal heterogeneity may favor less clustering. When a population evolves toward a new optimum by the fixation of novel beneficial mutations, tight linkage between selected loci will tend to slow the response to selection, due to the slower rate at which beneficial mutations on different chromosomes are brought together by recombination [i.e., the Hill–Robertson effect (86, 97, 98)]. Similarly, currently beneficial combinations of alleles might become detrimental following an environmental change, favoring increased recombination between the loci involved in local adaptation (23, 86, 99, 100). The results in Fig. 4*A* show that the relationship between period length ($T$) and amount of clustering is nonmonotonic, suggesting that further research is needed to determine how these different forms of selection will interact and whether some intermediate recombination rate may be optimal under certain types of environmental variation. More generally, any advantage of reduced recombination between locally adapted loci may be countered by a range of factors that favor increased rates of recombination (86). This brings us back to Turner (85), who asked, "Why does the genotype not congeal?" and suggested that clusters of loci might evolve due to locus-specific benefits of tighter linkage, despite global benefits of recombination. It seems likely some intermediate amount of clustering would yield rates of recombination low enough to maintain linkage disequilibrium between combinations of locally adapted alleles, but high enough to facilitate the purging of deleterious alleles and the spread of new beneficial mutations in times of environmental change. Given the importance of migration rates for determining the advantage of linkage (32), and mutation rates for determining the deleterious effects of linkage (86), it seems likely that the optimal amount of clustering will be determined to some extent by the ratio of mutation to migration. Regardless of which arrangement of adaptive loci is selectively optimal, drift and limited rearrangement rates might prevent the realization of such configurations. Comparative genomic studies of ecological adaptation will provide much-needed data to evaluate the relative importance of selective and stochastic processes in shaping the architecture of the genome.

## Methods

**Analytical Approximations of Establishment Probability.** The establishment probabilities of new linked ($EP_{clustered}$) and unlinked ($EP_{unlinked}$) mutations can be approximated using the splicing approach described by Yeaman and Otto (52) with the two-locus continent-island model of Bürger and Akerman (24). Consider a population experiencing natural selection and immigration, with two loci ($A$ and $B$) that recombine at rate $r$. If there is a stable polymorphism segregating at locus $B$ ($B_1$ and $B_2$ experience selection of $+\beta$ and $-\beta$, respectively), and a maladapted allele at locus $A$ ($A_2$ experiences selection of $-\alpha$), then a new locally adapted allele, $A_1$ (experiencing selection of $+\alpha$), will establish when the rates of recombination and immigration of the maladapted genotype ($m$; $A_2B_2$) are sufficiently low (24). Based on the derivations described by Bürger and Akerman (24), the rate at which this allele will invade can be approximated by evaluating the Jacobian matrix for their equation 2 at frequencies of $A_1B_1 = 0$, $A_2B_1 = 0$, $A_1B_2 = 1 - (m/\beta)$, and $A_2B_2 = m/\beta$ (equilibria from equation 3.9 in ref. 24). As shown by Yeaman and Otto (52), the leading eigenvalue of this matrix can be used to approximate the net strength of selection favoring the invasion of the allele $A_1$, and can be substituted for $s$ in population genetic approximations for fixation probability. Because the fixation probability of a new mutation is $\sim 2s$ in large populations, $A_1$ will establish with probability

$$EP = 2\alpha - \beta - r + \sqrt{\beta^2 + 2\beta r - 4mr + r^2}. \quad [3]$$

Therefore, the advantage due to divergence hitchhiking for a new mutation linked at some recombination rate $r = x$ can then be quantified as the ratio of $EP_{r\,=\,x}/EP_{r\,=\,0.5}$.

To parameterize this equation for Fig. 1, the average value of $EP_{clustered}$ was calculated across 50,000 loci, with recombination rates ranging linearly from 0.001 cM to 50 cM in increments of 0.001 cM (i.e., numerical integration). This approach assumes that gene density and recombination rates are uniform across the genome and is intended as a means of approximating $EP_{clustered}/EP_{unlinked}$ to parameterize Eq. **1**, rather than an attempt to make highly accurate predictions. Although the average of $EP_{clustered}/EP_{unlinked}$ would be higher if calculated over a smaller range of recombination rates, this would also imply a smaller mutational target (and therefore $\mu_{clustered}$), leading ultimately to a lower value for $n_{clustered}$.

**Evolution of Clustered Genomic Architectures.** If all $n$ loci that contribute to local adaptation begin in an unlinked configuration distributed randomly throughout the genome, the time required to rearrange $n_x$ of these loci into a single tightly linked cluster can be estimated as follows. If the per-locus rate of rearrangement is $\tau$ and the probability of insertion near a specific locus is $L/G$ (where $L$ is the size of the region around the locus and $G$ is the size of the genome), then rearrangements yielding a first pair of clustered loci will occur at probability of $2N_p\tau(L/G)n(n-1)$ per generation in a metapopulation of size $N_p$, because there are $n$ possible loci that could move into tight linkage with any of the $(n-1)$ other loci. Following this first step, the probability of rearrangements moving one of the remaining loci into tight linkage with this first cluster would be $2N_p\tau(L/G)(n-i)$, for the $i^{\text{th}}$ locus added to the cluster, where $i$ ranges from 2 to $(n_x - 1)$.

It is somewhat less straightforward to estimate the probability that any suitable rearrangements would spread through the population, but this can be approximated by considering only those rearrangements that result in extremely tight linkage, reducing the rate of recombination to $\sim 0$. For a model of local adaptation with immigration (at rate $m$), Kirkpatrick and Barton (32) showed that if a new chromosomal inversion captures two locally adapted loci, reducing the rate of recombination from $r = 0.5$ to $r = 0$, it should spread at a rate of $m$. Importantly, this rate is independent of the strength of selection on the loci within the inversion, so it should generalize to any rearrangements that reduce recombination to nearly 0 between two previously unlinked loci, regardless of the effect of sizes of the alleles involved (although at high migration rates, alleles with small $s$ would be lost). This approach neglects any direct fitness costs caused by the rearrangement itself owing to disrupted expression, problems with recombination, or imbalanced gene number, so $\tau$ should be interpreted as the rate of rearrangements that do not involve significant fitness costs. Following Yeaman and Otto (52), the probability of fixation of such rearrangements should be $\sim 2m$, and the expected time before the occurrence of a rearrangement that fixes to yield the first pair of clustered loci is the reciprocal of these combined probabilities: $t_1 = (2m \times 2N_p\tau(L/G)n(n-1))^{-1}$. Ignoring rearrangements that reduce clustering once it has evolved, because these will be disfavored, the expected time to evolve from an architecture where all $n$ loci

are unlinked to an architecture where $n_x$ of these loci are arranged in a single cluster can be approximated as

$$t = \left(4N_p\tau m(L/G)n(n-1)\right)^{-1} + \sum_{i=2}^{n_x-1}\left(4N_p\tau m(L/G)(n-i)\right)^{-1}. \quad [4]$$

If $n_x = n$, this reduces to

$$t = \left(4N_p\tau mL/G\right)^{-1}\left(\sum_{i=1}^{n-2} i^{-1} + 1/\left(n(n-1)\right)\right) = \frac{G(H_n + 1/(n(n-1)))}{4N_p\tau mL}, \quad [5]$$

where $H_n$ is the $(n-2)^{\text{th}}$ harmonic number (if $n = 2$, $H_n = 0$). Because $1 < H_n < 10$ for most biologically realistic values of $n$, this can be approximated to an order of magnitude as $t \approx G/(N_p\tau mL)$. Given the assumptions involved in applying the model from Kirkpatrick and Barton (32), parameterizing this model requires that $L$ represent the number of base pairs surrounding each locus over which $r \approx 0$. Note that $N_p$ represents the total size of the metapopulation, because rearrangements would be favored within whichever patch they occur and would likely eventually fix through the metapopulation (again, the dynamics involved here are oversimplified). Because this model ignores the contributions of rearrangements that yield less-tightly linked clusters (with $0 < r < 0.5$), it will tend to overestimate the time required to evolve a clustered architecture. A more complex model by Bürger and Akerman (24) suggests that the Kirkpatrick and Barton (32) model may underestimate the invasion rate, further indicating that Eqs. **4** and **5** should tend to overestimate the time required for the evolution of clustered architectures.

**Individual-Based Simulations.** Simulations were run using the version of Nemo (101) that was modified for ref. 28 (source code deposited in Sourceforge). Two populations exchanging migrants at rate $m$ experience divergent selection, where the fitness of an individual with a given phenotype ($Z$) is a function of the local optimum ($\theta = \pm 1$), strength of selection ($\phi$), and the curvature of the fitness function ($\gamma = 2$): $w = 1 - \phi(|\theta - Z|/|2\theta|)^{\gamma}$ (backward and forward migration rates are equal, given the constant densities). Unless otherwise indicated, $N = 10^4$, $\phi = 0.75$, $m = 0.01$, and the per locus mutation rate $\mu = 10^{-4}$. Unlike in ref. 28, mutations follow a house-of-cards model (102), with the new value of a mutation replacing any existing value. This mutation model was used so that no alleles of large effect could build up through multiple mutations at a single locus. Mutation effect sizes are drawn from a Gaussian distribution with mean = 0 and SD = 0.05 and are truncated so that all mutation sizes were $\leq |0.05|$. The phenotype is calculated by adding the mutation effects across all selected loci, so that a locally optimal phenotype in the patch with $\theta = 1$ could be built with 10 homozygous mutations of 0.05.

Genomic architecture is modeled by interspersing each of $n = 10$ loci that affect the phenotype (selected loci) with 49 neutral loci along a single chromosome (for a total of 500 loci) and setting the recombination rate between adjacent loci to 0.002, such that on average there would be one recombination event per individual per generation. Genomic rearrangements occur at rate $\tau$ per locus (set to $10^{-6}$ unless otherwise noted), according to one of two models of rearrangement. For the cut-and-paste model, rearrangements occurred by moving a single selected locus to a new location occupied by a neutral locus; the selected locus replaces the neutral locus in its new location and is replaced by a new neutral locus in its original location, to maintain chromosome size (this process was constrained so that one selected locus never replaced another selected locus). For the duplication model, the process was identical to the cut-and-paste model, except the "parent" locus retained its state as a selected locus. Selected loci could be lost (i.e., convert to neutral loci) at a per-locus rate of $\eta$, which represents the process of pseudogenization or gene deletion. Unless otherwise noted, all simulations were run using the cut-and-paste model with $\eta = 0$.

To represent the fitness costs arising from CNV, an additional fitness cost ($s_c$) was calculated for each individual based on a Gaussian function with $s_c = \sum_{i=1}^{n}(1 - \exp[-(2-c_i)^2/V_c])$, summed across contributions from $n$ loci, with $c_i$ representing the number of copies of the $i^{\text{th}}$ locus. When the width of this function, $V_c$, is small, selection against CNV is strong (e.g., for $V_c = 50$ and $c = 3$ at one locus, $s_c = 0.0198$); when $V_c$ is large, selection is weak (e.g., for $V_c = 5{,}000$ and $c = 3$ at one locus, $s_c = 0.0002$); for $V_c = \infty$, there is no selection against CNV. This approach is similar to a more complex model proposed by Gout et al. (103), which also assumes that fitness decreases with increasing deviations from optimal levels of gene expression, which in turn occur due to changes in gene copy number. The complete deletion of a gene will sometimes, but not always, result in lethality (104, 105); for simplicity, in all simulations any individual with $c = 0$ for any locus was given a fitness of 0 (irrespective of $V_c$). Unless otherwise stated, the simulations in this paper were run using $V_c = 5{,}000$.

Unless otherwise indicated, simulations were run for 500,000 generations, with statistics calculated every 500 generations, based on at least 20 replicates for each parameter combination. The environmental fluctuations shown in Fig. 4*A* were modeled using a sine wave with amplitude = 2 and period $T$, and these fluctuations were added onto the local optima in each population (i.e., +1/−1; as shown in Fig. 4*B*). The environmental fluctuations shown in Fig. 4*C* were modeled by changing the local optima from +1/−1 to 0/0 and the migration rate from 0.01 to 0.5 for 1,000 generations, every $T$ generations (as shown in Fig. 4*D*). For all other simulations shown in Fig. 4, migration rate was held constant at $m = 0.01$.

**ACKNOWLEDGMENTS.** I thank Mike Whitlock and Sally Otto for working with me on the projects that preceded this manuscript and providing helpful feedback and suggestions throughout. Discussions with L. Rieseberg, S. Aitken, J. Mee, D. Schluter, F. Guillaume, J. Mell, the Aitken, Rieseberg, and Whitlock laboratory group, and the Evolution of Genomic and Genetic Architecture journal club also made very important contributions to the development of this paper; helpful comments of two anonymous reviewers are also acknowledged. Special thanks to J. Mee for noting an error in equation 1 in ref. 28. Computations were run on the WestGrid Glacier High Performance Computing Consortium. This work was supported by Genome Canada and Genome BC funding to AdapTree (co-Project leaders S. Aitken and A. Hamann).

1. Allen Orr H (2001) The genetics of species differences. *Trends Ecol Evol* 16(7):343–350.
2. Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nat Rev Genet* 3(1):11–21.
3. Ehrenreich IM, Purugganan MD (2006) The molecular genetic basis of plant adaptation. *Am J Bot* 93(7):953–962.
4. Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 3(9):e285.
5. Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol* 17(19):4334–4345.
6. Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 18(3):375–402.
7. Nadeau NJ, et al. (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc Lond B Biol Sci* 367(1587):343–353.
8. Renaut S, et al. (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philos Trans R Soc Lond B Biol Sci* 367(1587):354–363.
9. Smadja CM, et al. (2012) Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution* 66(9):2723–2738.
10. Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends Genet* 28(7):342–350.
11. Rogers S, Bowles W, Mee JA (2012) The consequences of genomic architecture on ecological speciation in postglacial fishes. *Current Zoology*, in press.
12. Strasburg JL, et al. (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philos Trans R Soc Lond B Biol Sci* 367(1587):364–373.
13. Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* 23(1):23–35.
14. Hermisson J, Pennings PS (2005) Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* 169(4):2335–2352.
15. Przeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. *Evolution* 59(11):2312–2323.
16. Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity (Edinb)* 103(6):439–444.
17. Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc Lond B Biol Sci* 367(1587):409–421.
18. Petry D (1983) The effect on neutral gene flow of selection at a linked locus. *Theor Popul Biol* 23(3):300–313.
19. Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Heredity* 56:357–376.
20. Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64(6):1729–1747.
21. Charlesworth D, Charlesworth B (1979) Selection on recombination in clines. *Genetics* 91(3):581–589.
22. Pylkov KV, Zhivotovsky LA, Feldman MW (1998) Migration versus mutation in the evolution of recombination under multilocus selection. *Genet Res* 71(3):247–256.
23. Lenormand T, Otto SP (2000) The evolution of recombination in a heterogeneous environment. *Genetics* 156(1):423–438.
24. Bürger R, Akerman A (2011) The effects of linkage and gene flow on local adaptation: A two-locus continent-island model. *Theor Popul Biol* 80(4):272–288.
25. Feder JL, Gejji R, Yeaman S, Nosil P (2012) Establishment of new mutations under divergence and genome hitchhiking. *Philos Trans R Soc Lond B Biol Sci* 367(1587): 461–474.

26. Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos Trans R Soc Lond B Biol Sci* 367(1587):451–460.
27. Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature* 341(6242):497–503.
28. Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution* 65(7):1897–1911.
29. Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* 70(2):155–174.
30. Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol* 16(7):351–358.
31. Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* 98(21):12084–12088.
32. Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* 173(1):419–434.
33. Feder JL, Nosil P (2009) Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* 63(12):3061–3075.
34. Feder JL, Roethele JB, Filchak K, Niedbalski J, Romero-Severson J (2003) Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* 163(3):939–953.
35. Cheng C, et al. (2012) Ecological genomics of Anopheles gambiae along a latitudinal cline: A population-resequencing approach. *Genetics* 190(4):1417–1432.
36. Jones FC, et al.; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55–61.
37. Neafsey DE, et al. (2010) SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* 330(6003):514–517.
38. Strasburg JL, Scotti-Saintagne C, Scotti I, Lai Z, Rieseberg LH (2009) Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Mol Biol Evol* 26(6):1341–1355.
39. Ranz JM, Casals F, Ruiz A (2001) How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Res* 11(2):230–239.
40. Pringle EG, et al. (2007) Synteny and chromosome evolution in the lepidoptera: Evidence from mapping in *Heliconius melpomene*. *Genetics* 177(1):417–426.
41. d'Alençon E, et al. (2010) Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. *Proc Natl Acad Sci USA* 107(17):7680–7685.
42. Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487(7405):94–98.
43. Backström N, et al. (2008) A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics* 179(3):1479–1495.
44. Ellegren H (2010) Evolutionary stasis: The stable chromosomes of birds. *Trends Ecol Evol* 25(5):283–291.
45. Pevzner P, Tesler G (2003) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proc Natl Acad Sci USA* 100(13):7672–7677.
46. Murphy WJ, et al. (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* 309(5734):613–617.
47. Ono S (1973) Ancient linkage groups and frozen accidents. *Nature* 244(5414): 259–262.
48. Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L (2005) Chromosome evolution in eukaryotes: A multi-kingdom perspective. *Trends Genet* 21(12):673–682.
49. Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA* 104(Suppl 1):8597–8604.
50. Koonin EV (2009) Evolution of genome architecture. *Int J Biochem Cell Biol* 41(2): 298–306.
51. Kimura M (1962) On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.
52. Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* 65(7):2123–2129.
53. Barton NH (2000) Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci* 355(1403): 1553–1562.
54. Petes TD, Hill CW (1988) Recombination between repeated genes in microorganisms. *Annu Rev Genet* 22:147–168.
55. Bishop AJR, Schiestl RH (2000) Homologous recombination as a mechanism for genome rearrangements: Environmental and genetic effects. *Hum Mol Genet* 9(16): 2427–334.
56. Richardson C, Jasin M (2000) Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 405(6787):697–700.
57. Argueso JL, et al. (2008) Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc Natl Acad Sci USA* 105(33):11845–11850.
58. Mani R-S, Chinnaiyan AM (2010) Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nat Rev Genet* 11(12):819–829.
59. Woodhouse MR, Pedersen B, Freeling M (2010) Transposed genes in *Arabidopsis* are often associated with flanking repeats. *PLoS Genet* 6(5):e1000949.
60. Tonzetich J, Hayashi S, Grigliatti TA (1990) Conservatism of sites of tRNA loci among the linkage groups of several Drosophila species. *J Mol Evol* 30(2):182–188.
61. Freeling M, et al. (2008) Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. *Genome Res* 18(12):1924–1937.
62. Lai J, Li Y, Messing J, Dooner HK (2005) Gene movement by *Helitron* transposons contributes to the haplotype variability of maize. *Proc Natl Acad Sci USA* 102(25): 9068–9073.
63. Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431(7008):569–573.
64. Wang W, et al. (2006) High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* 18(8):1791–1802.
65. Springer M, Weissman JS, Kirschner MW (2010) A general lack of compensation for gene dosage in yeast. *Mol Syst Biol* 6:368.
66. Cooper GM, et al. (2011) A copy number variation morbidity map of developmental delay. *Nat Genet* 43(9):838–846.
67. Birchler JA, Veitia RA (2012) Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci USA* 109(37):14746–14753.
68. Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* 292(5517):673–679.
69. Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* 8(9):e1000500.
70. Joron M, et al. (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 477(7363):203–206.
71. Renaut S, et al. (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. In press.
72. Hurst LD, Pál C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5(4):299–310.
73. Williams EJB, Bowles DJ (2004) Coexpression of neighboring genes in the genome of Arabidopsis thaliana. *Genome Res* 14(6):1060–1067.
74. Schmid M, et al. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37(5):501–506.
75. Zhan S, Horrocks J, Lukens LN (2006) Islands of co-expressed neighbouring genes in Arabidopsis thaliana suggest higher-order chromosome domains. *Plant J* 45(3):347–357.
76. Lee JM, Sonnhammer ELL (2003) Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* 13(5):875–882.
77. Al-Shahrour F, et al. (2010) Selection upon genome architecture: conservation of functional neighborhoods with changing genes. *PLOS Comput Biol* 6(10):e1000953.
78. Batada NN, Urrutia AO, Hurst LD (2007) Chromatin remodelling is a major source of coexpression of linked genes in yeast. *Trends Genet* 23(10):480–484.
79. Batada NN, Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* 39(8):945–949.
80. Lynch M, Bobay LM, Catania F, Gout JF, Rho M (2011) The repatterning of eukaryotic genomes by random genetic drift. *Annu Rev Genomics Hum Genet* 12:347–366.
81. Charlesworth D, Charlesworth B (1975) Theoretical genetics of Batesian mimicry II. Evolution of supergenes. *J Theor Biol* 55(2):305–324.
82. Fisher R (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).
83. Kimura M (1956) A model of a genetic system which leads to closer linkage by natural selection. *Evolution* 10:278–287.
84. Nei M (1967) Modification of linkage intensity by natural selection. *Genetics* 57(3): 625–641.
85. Turner JRG (1967) Why does the genotype not congeal? *Evolution* 21:645–656.
86. Otto SP (2009) The evolutionary enigma of sex. *Am Nat* 174(Suppl 1):S1–S14.
87. Charlesworth D, Charlesworth B (1980) Sex-differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genetical Research* 35:205–214.
88. Wright S (1941) On the probability of fixation of reciprocal translocations. *Am Nat* 75:513–522.
89. Lande R (1985) The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity (Edinb)* 54(Pt 3):323–332.
90. Colosimo PF, et al. (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol* 2(5):E109.
91. Albert AYK, et al. (2008) The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution* 62(1):76–85.
92. Rogers SM, et al. (2012) Genetic signature of adaptive peak shift in threespine stickleback. *Evolution* 66(8):2439–2450.
93. Larkin DM, et al. (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res* 19(5):770–777.
94. Woodhouse MR, Tang H, Freeling M (2011) Different gene families in *Arabidopsis thaliana* transposed in different epochs and at different frequencies throughout the rosids. *Plant Cell* 23(12):4241–4253.
95. Lanctôt C, Cheutin T, Cremer M, Cavalli G, Cremer T (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet* 8(2):104–115.
96. Wijchers PJ, de Laat W (2011) Genome organization influences partner selection for chromosomal rearrangements. *Trends Genet* 27(2):63–71.
97. Martin G, Otto SP, Lenormand T (2006) Selection for recombination in structured populations. *Genetics* 172(1):593–609.
98. Hartfield M, Otto SP, Keightley PD (2010) The role of advantageous mutations in enhancing the evolution of a recombination modifier. *Genetics* 184(4):1153–1164.
99. Charlesworth B, Charlesworth D (1993) Directional selection and the evolution of sex and recombination. *Genet Res* 61(3):205–224.
100. Bürger R (1999) Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* 153(2):1055–1069.
101. Guillaume F, Rougemont J (2006) Nemo: An evolutionary and population genetics programming framework. *Bioinformatics* 22(20):2556–2557.
102. Turelli M (1984) Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor Popul Biol* 25(2):138–193.
103. Gout J-F, Kahn D, Duret L; Paramecium Post-Genomics Consortium (2010) The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet* 6(5):e1000944.
104. Tong AHY, et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 294(5550):2364–2368.
105. Carpenter AE, Sabatini DM (2004) Systematic genome-wide screens of gene function. *Nat Rev Genet* 5(1):11–22.