aptara | EVO | evo_943 | Dispatch: 1-13-2010 | CE: XHV
Journal | MSP No. | No. of pages: 19 | PE: Sonia

ORIGINAL ARTICLE

doi:10.1111/j.1558-5646.2010.00943.x

# THE EFFICACY OF DIVERGENCE HITCHHIKING IN GENERATING GENOMIC ISLANDS DURING ECOLOGICAL SPECIATION

**Jeffrey L. Feder[1,2,3] and Patrik Nosil[2,4,5]**

[1]*Dept. of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556*

[2]*Institute for Advanced Study, Wissenschaftskolleg, Berlin, 14193, Germany*

[3]*E-mail: feder.2@nd.edu*

[4]*Department of Ecology and Evolutionary Biology, University of Boulder, Colorado, 80309*

[5]*E-mail: patrik.nosil@colorado.edu*

Received August 7, 2009
Accepted November 9, 2009

**Genes under divergent selection flow less readily between populations than other loci. This observation has led to verbal "divergence hitchhiking" models of speciation in which decreased interpopulation gene flow surrounding loci under divergent selection can generate large regions of differentiation within the genome (genomic islands). The efficacy of this model in promoting speciation depends on the size of the region affected by divergence hitchhiking. Empirical evidence is mixed, with examples of both large and small genomic islands. To address these empirical discrepancies and to formalize the theory, we present mathematical models of divergence hitchhiking, which examine neutral differentiation around selected sites. For a single locus under selection, regions of differentiation do not extend far along a chromosome away from a selected site unless both effective population sizes and migration rates are low. When multiple loci are considered, regions of differentiation can be larger. However, with many loci under selection, genome-wide divergence occurs and genomic islands are erased. The results show that divergence hitchhiking can generate large regions of differentiation, but that the conditions under which this occurs are limited. Thus, speciation may often require multifarious selection acting on many, isolated and physically unlinked genes. How hitchhiking promotes further adaptive divergence warrants consideration.**

**KEY WORDS: Divergent selection, gene flow, recombination, simulations, speciation islands.**

Genomic divergence may be heterogeneous during population divergence and speciation, during which genetic differentiation accumulates in some regions (genomic islands of divergence; Turner et al. 2005), whereas the homogenizing effects of gene flow or inadequate time for random differentiation by genetic drift precludes divergence in other regions (Feder 1998; Via 2001, 2009; Wu 2001; Berlocher and Feder 2002; Gavrilets and Vose 2005; Noor and Feder 2006; Via and West 2008; Nosil et al. 2009a). Divergent selection contributes to such variable genomic differentiation by causing specific loci and those physically linked to them to flow between populations less readily than others, thereby resulting in accentuated genetic divergence of regions affected by selection (Lewontin and Krakauer 1973; Via 2001, 2009; Wu 2001; Noor and Feder 2006). These ideas have a long history in studies of hybrid zones (Barton 1979, 1983, 2000; Templeton 1981; Barton and Bengtsson 1986; Harrison and Rand 1989; Gavrilets and Cruzan 1998; Avise 2000; Wu and Ting 2004; Noor and Feder 2006) and sympatric speciation (Feder 1998; Via 2001; Berlocher and Feder 2002). However, it is only with recent technical and analytical advances that allow genetic divergence at many loci to be screened in most any organism that numerous studies attesting to the porous nature of the genome have emerged

(Beaumont 2005; Noor and Feder 2006). For example, the last decade has seen the emergence of population genomic studies reporting "outlier loci" whose genetic divergence greatly exceeds that observed for the rest of the genome, putatively because such loci are differentiated beyond neutral expectations and affected by divergent selection (see Butlin 2008 and Nosil et al. 2009a for reviews). These observations have led to discussions of the implications of the heterogeneous nature of genomic differentiation for speciation.

A recent verbal theory of speciation via "divergence hitchhiking" ties the above ideas together to generate a mechanism by which speciation in the face of gene flow may be easier than previously thought (Smadja et al. 2008; Via and West 2008; Via 2009). The premise is that divergent selection reduces interbreeding between populations in different habitats, for example, by causing ecologically based selection against immigrants and hybrids (Nosil et al. 2005; Smadja et al. 2008; Via and West 2008; Via 2009). This reduces interpopulation recombination, and even if recombination occurs, selection reduces the frequency of immigrant alleles in advanced generation hybrids (Gavrilets 2004). This reduction in effective gene flow might allow large regions of genetic differentiation to build up in the genome around the few loci subject to divergent selection at the initiation of speciation. The idea rests on the assumption that a site under divergent selection will create a relatively large window of reduced gene flow around it, enhancing the potential to accumulate differentiation (both neutral and selected) at linked sites. In turn, these few genomic islands of divergence might provide a seed that can be expanded upon to cover even larger areas of chromosomes. As an alternative to divergence hitchhiking in a few genomic regions, speciation may be initiated and driven by "multifarious" selection acting on many different traits. By inference, these traits are likely affected by many, rather than one or a few, independent genetic changes throughout the genome, some of which fortuitously cause reproductive isolation (Rice and Hostert 1993; Nosil et al. 2009b). These two views represent different ends of a continuum and are not entirely mutually exclusive, as fortuitous physical linkage of different loci under multifarious selection could enhance speciation.

The extent to which divergence hitchhiking promotes speciation will depend, in part, on the size of the genomic region affected: the larger the region, the larger the proportion of the genome resistant to gene flow, and the greater the possibility that genes within the region contribute to further reducing gene flow. Empirical evidence concerning the size of differentiated regions in the genome is mixed. There are several examples in which regions were inferred to be large (Hawthorne and Via 2001; Emelianov et al. 2004; Harr 2006; Rogers and Bernatchez 2007; Via and West 2008). For example, in host races of *Acyrthosiphon* pea aphids and lake ecotypes of *Coregonus* whitefish, outlier loci from genome scans reside near QTL for phenotypic traits more often than expected by chance, yet the average distance between an outlier and the apparently nearest QTL is relatively large (10.6 and 16.2 centiMorgans, respectively) (Rogers and Bernatchez 2007; Via and West 2008). Indirect, although not conclusive, evidence that regions of differentiation are large also stems from the observation that numerous population genomic studies have easily detected outliers, despite having genomic coverage too sparse to be likely to detect the actual direct targets of selection (Nosil et al. 2009a), suggesting extensive linkage disequilibrium along chromosomes (Payseur et al. 2004). Q1

However, a number of other observations suggest that regions of differentiation can be small, including: (1) the tendency for accentuated divergence to be observed only in regions of extensively reduced recombination such as near centromeres (Turner et al. 2005; Geraldes et al. 2006) or near breakpoints of chromosomal inversions (Machado et al. 2007; Noor et al. 2007; White et al. 2007; Yatabe et al. 2007; Strasburg et al. 2009), (2) a lack of strong genetic divergence at neutral markers physically proximate to sites of divergent selection (Mäkinen et al. 2008a), and (3) the rapid decay of genetic divergence (i.e., $F_{ST}$) away from genomic islands (Dopman et al. 2005; Machado et al. 2007; Noor et al. 2007; Turner and Hahn 2007; Mäkinen et al. 2008b; Wood et al. 2008). Of course, the size of a differentiated region will depend on how such regions are delineated (Via 2009), but it seems unlikely that this could explain all the variation. What then are the evolutionary factors that predict the size of differentiated regions within the genome? Formal theory is required to address this question.

Some insight into the anticipated size of genomic islands accompanying local adaptation stems from an analytical and simulation study in which the verbal model of divergence hitchhiking is rooted. Specifically, Charlesworth et al. (1997) investigated population differentiation of neutral sites at increasing recombination distance from a single divergently selected locus. Their models considered a small local population ($n = 1000$) exchanging a low level of migrants ($m = 0.001$ per generation) compared to the magnitude of selection ($s = 0.1$ and 0.5). Consequently, the long-term persistence of hybrid genotypes in populations was uncommon, providing the opportunity for substantial neutral differentiation to accumulate around selected sites by drift. In contrast, when thinking about the demography of speciation-with-gene flow, we generally envision a pair of taxa with larger effective population sizes and higher levels of gene flow at the time of initial population divergence. Additionally, we expect more than a single locus to be under selection during speciation, even if the number of loci is modest (Rice and Hostert 1993; Coyne and Orr 2004).

Here, we use a combination of analytical and simulation approaches to develop single- and multilocus models of divergence hitchhiking for a wide range of parameter values, including those likely to be commonly observed for speciation-with-gene flow.

The theory of divergence hitchhiking actually concerns two main factors: (1) the ability for reduced gene flow surrounding selected sites to generate accentuated regions of neutral differentiation ($F_{ST}$) around the sites and (2) the opportunity for these reductions in gene flow to facilitate further divergence and thus act as nuclei for speciation. The current article focuses on the first issue. This is a reasonable starting point not only from a modeling perspective, but is also justified from an empirical standpoint because population genomic studies testing for islands of divergence tend to rely heavily or exclusively on divergence ($F_{ST}$) at putatively neutral markers to identify genomic islands and estimate their size (Via and West 2008; Nosil et al. 2009a; Via 2009). Future work on the second issue is in progress and is focusing on the implications of further divergence stemming from new mutations, as well as prestanding genetic variation (Barrett and Schluter 2008).

We report that although in certain circumstances high $F_{ST}$ may be observed at relatively large recombination distances from a selected site, this is not always (usually) expected during the formative stages of speciation-with-gene flow. Moreover, we find that if many loci are under selection, genomic-wide divergence occurs easily, erasing accentuated divergence near or at selected sites. Thus, the de novo build up of large islands of neutral differentiation in just a handful of genomic regions may often be the exception during initial stages of speciation. Our results therefore raise questions concerning the role of regions of differentiation caused by divergence hitchhiking as being "foci" for speciation, because multifarious selection on numerous unlinked and independently fixed loci can also drive the formative stages of population divergence (Rice and Hostert 1993).

# Methods

## OVERVIEW

We used analytical approaches and computer simulations to estimate effective migration rates ($m_e$) for neutral mutations in various linkage relationships to a locus or loci under divergent selection (see Table 1). These effective migration rate estimates were then substituted in place of migration rate $m$ into formula for calculating fixation indexes ($F_{ST}$) between populations. Comparisons of $F_{ST}$ values generated across a range of migration rates ($m$), selection intensities ($s$), numbers of loci under selection ($g$), recombination rates between neutral and selected loci ($r$), and effective

**Table 1. Variables in two-deme island model used to estimate effective migration rates ($m_e$) and $F_{ST}$ values between populations for a neutral locus 1 linked at varying recombination distances ($r$) to a second locus 2 under divergent selection with selection coefficient ($s$) and partial dominance ($h$=0.5). Bidirectional, symmetrical migration is assumed to occur between the two demes at rate $m$.**

| Deme 1 | | | | Deme 2 | | | |
|---|---|---|---|---|---|---|---|
| Neutral Locus 1 | | | | Neutral Locus 1 | | | |
| Alleles | | Genotypes | Fitness | Alleles | | Genotypes | Fitness |
| B | | BB | 1 | B | | BB | 1 |
| b | | Bb | 1 | b | | Bb | 1 |
| | | bb | 1 | | | bb | 1 |
| symmetrical migration ($m$) | | | | | | | |
| recombination at rate ($r$) between Locus 1 and 2 | | $m_{12}$ | | recombination at rate ($r$) between Locus 1 and 2 | | | |
| | | $m_{21}$ | | | | | |
| | | $m_{12} = m_{21}$ | | | | | |
| Divergent Selection Locus 2 | | | | Divergent Selection Locus 2 | | | |
| Alleles | Frequency | Genotypes | Fitness | Alleles | Frequency | Genotypes | Fitness |
| A | $p_1$ | AA | 1 | A | $p_2$ | AA | 1-s |
| a | $q_1$ | Aa | 1-$hs$ | a | $q_2$ | Aa | 1-$hs$ |
| | | aa | 1-$s$ | | | aa | 1 |

population sizes ($n_e$) provided metrics for assessing the level of neutral differentiation expected to accumulate around a selected locus during speciation-with-gene flow when a balance is reached between genetic drift, selection, and migration. They also serve as a useful summary statistic for comparing effective migration rates ($m_e$) among sites. We note that our analysis does not examine the extent and duration of linked neutral differentiation that will transiently be elevated between taxa when a new adaptive mutation arises and sweeps through one population (Hermisson and Pennings 2005; Nielsen 2005). This question will be the focus of future analysis, but is considered in the discussion. Our current work helps extend related models of barriers to gene flow (Petry 1983; Bengtsson 1985; Barton and Bengtsson 1986; Charlesworth et al. 1997; Gavrilets 2004) by considering the size of differentiated regions when many loci are under selection. The latter point is important because speciation usually requires genetic change at several loci (Rice and Hostert 1993; Coyne and Orr 2004; Gavrilets 2004; Wu and Ting 2004). Consequently, estimates of neutral differentiation predicted for a site linked to a single locus under disruptive selection, although informative, may reflect only the very earliest stages of speciation. To gain a clearer understanding of differentiation accompanying speciation therefore requires examining cases when multiple genetic differences have accumulated and are contributing to gene flow barriers between populations.

## EFFECTIVE MIGRATION RATE

Petry (1983), Bengtsson (1985), and Barton and Bengtsson (1986) originally derived formulas for estimating effective migration rates and developed the concept of barrier strength to describe the degree to which gene flow of a neutral marker was reduced between hybridizing populations due to divergent selection (reproductive isolation). Specifically, barrier strength ($b$) was defined as $m/m_e$ the product of the gross migration rate and the probability that a neutral allele survives selection following migration into the alternate population. Here, we are interested in $m_e$ as a measure for the net level of genetic interchange between populations. Estimates of $m_e$ can be used to derive $F_{ST}$ values for neutral sites linked to a locus under disruptive selection. The lower the value of $m_e$, the greater the barrier strength and the restriction to gene flow, and thus the higher the estimated $F_{ST}$ value.

## ESTIMATING $m_e$ And $F_{ST}$ WHEN A SINGLE LOCUS IS UNDER SELECTION

We focused on algebraic and simulation estimates of effective migration rates and $F_{ST}$ for two clear reasons. First, because they form much of the theoretical underpinning for the divergence hitchhiking hypothesis. Second, because empirical genomic scan studies almost exclusively quantify genetic differentiation between populations based on $F_{ST}$ (Beaumont 2005). The effective migration rate from deme 1 into deme 2 for a neutral allele $B$ at a neutral locus 1 in deme 1 linked to a single locus 2 under disruptive selection has been previously estimated as

$$m_e = m(q_1(1-hs) + p_1(1-s)r/(1-(1-hs)(1-r))) \quad \text{(modified from Charlesworth et al. 1997)} \quad \frac{n!}{r!\,(n-r)!}, \quad (1)$$

where,

$m$ = gross migration rate between deme 1 and 2.

$p_1$ = frequency of the big $A$ allele at the selected locus 2 in deme 1 at selection/migration equilibrium. The $A$ allele is favored in deme 1 and disfavored in deme 2.

$q_1$ = frequency of the little $a$ allele at the selected locus 2 in deme 1. The allele $a$ is disfavored in deme 1 and favored in deme 2.

$r$ = recombination rate between the neutral locus 1 and the selected locus 2.

$s$ = selection coefficient against disfavored $AA$ homozygote at locus 2 in deme 2. Symmetric selection coefficients are assumed for genotypes between demes.

$h$ = dominance coefficient for the $Aa$ heterozygote at locus 2.

The degree of genetic differentiation expected to accumulate at equilibrium between demes for a neutral locus can be determined by using the estimate for effective migration rate in the standard, two subpopulation equations for migration and drift (Hudson 1990; Slatkin 1991; see Charlesworth et al. 1997):

$$\pi_{T-S} = 1/8N_e m_e (T = \text{total population and } S = \text{subpopulations}), \quad (2)$$

where,

$N_e$ = total effective population size of the two demes together = $2n_e$ for two subpopulations.

The fixation index $F_{ST}$ can be calculated as the ratio of the between subpopulation to total population genetic differentiation

$$F_{ST} = \pi_{T-S}/\pi_T, \quad (3)$$

where,

$$\pi_T = (1-q_1) + 1/2Nr + \pi_{T-S}. \quad (4)$$

Equation (1) assumes that once a neutral allele $B$ at locus 1 emigrating from deme 1 has recombined away from the disfavored $A$ allele-containing chromosome for locus 2 in deme 2, the $B$ neutral allele will persist in deme 2. This is a reasonable assumption when the migration rate is low and selection strong. Under these circumstances, $A$ allele-containing chromosomes in deme 2 will be rare, and thus the neutral introgressing $B$ allele will be unlikely to recombine back to such a chromosome. However, when migration rates are high relative to selection, this will not be the case. In these instances, equation (1) will tend to overestimate

the effective migration rate, as back recombination will remove the neutral *B* marker at locus 1 from the favored *a*-containing chromosome in deme 2 and back migration will return the neutral *B* marker to deme 1.

To estimate $F_{ST}$ under conditions when migration is high relative to selection (as expected during early stages of speciation), we developed a computer simulation approach to estimate $m_e$ (see Table 1). We recently used a similar approach to examine the role of chromosomal inversions in the maintenance of genetic differentiation between hybridizing populations at loci that are under divergent selection or cause reproductive isolation (Feder and Nosil 2009). In the current study, the approach incorporated bidirectional migration and the potential for back recombination. The computer simulations were based on a two-island population genetic model with demes 1 and 2 having equal population sizes and exchanging migrants at a symmetrical rate *m*, conditions that can favor the maintenance of variation between populations. The simulations were started with the frequency of the favored *A* allele at locus 2 in deme 1 being fixed and absent from deme 2. The two populations were then allowed to attain selection/migration equilibrium at the selected locus 2. We note that this is not an allopatric model. We only started the demes with fixed allele frequency differences for convenience. Starting conditions do not affect our results in any way because initiating the simulations with any starting frequency for allele *A* at locus 2 would result in the same equilibrium frequency for the *A* allele being quickly reached between demes.

Following the attainment of equilibrium (designated time $t_0$), a single episode of migration at rate *m* was modeled in which all migrants from deme 1 into deme 2 were homozygous for a new *b* mutation at the neutral locus 1, whereas all other individuals in both demes were homozygous for the *B* allele (i.e., we simulated a pulse-chase population genetic experiment in which all migrants into deme 2 at time $t_0$ were uniquely genetically labeled with the mutation *b* at locus 1 and the ultimate fate of the *b* allele was determined to estimate the effective introgression rate of the neutral marker). The dynamics of the *b* mutation were followed until it equilibrated at the same frequency ($q_{eq}$) in the two demes, allowing for continued disruptive selection, bidirectional migration at rate *m*, and recombination between the neutral and selected loci at rate *r*. At equilibrium, we estimated the effective migration rate ($m_e$) as two times the frequency of the neutral *b* mutation in deme 2 ($2q_{eq}$). The equilibrium frequency of the *b* allele was multiplied by a factor of two because only half of the initial immigrant neutral alleles actually have the potential to introgress from deme 1 into deme 2 given symmetrical migration rates between demes (e.g., if the frequencies of two neutral alleles *B* and *b* are initially differentially fixed between subpopulations, then they will eventually equilibrate at 0.5:0.5, setting an upper bound of 0.5 for introgression in the absence of selection). The simulation values for $m_e$ were then substituted into equations 2–4 to estimate $F_{ST}$. This allowed us to directly compare results from our simulations to analytical approximations. We also used the equilibrium frequencies *p* and *q* from the simulations for the *A* and *a* alleles at locus 2 in our analytic estimates of $m_e$ derived from equation (1).

The computer simulations were written in MATLAB(tm) and based on a life cycle with selection following migration and preceding mating (newborns > dispersal > selection > meiosis/recombination > mating > zygotes) following Feder and Nosil (2009). Population sizes were assumed to be equal and independently regulated in the two demes. Selection was modeled to be soft, with both demes large (infinite) and contributing equally to the migrant pool. Individuals were assumed to migrate or remain in their natal population independent of genotype. Mating was assumed to occur at random within the two subpopulations following migration and selection. Three different intensities of disruptive selection symmetric between demes 1 and 2 were considered in the simulations, $s = 0.01$ (weak), $s = 0.1$ (moderate), and $s = 0.5$ (strong). Three levels of migration were considered ($m = 0.001$ [low], $m = 0.01$ [moderate], and $m = 0.1$ [high]). Seven recombination rates were considered between the neutral locus 1 and the selected locus 2, ranging from extremely tight linkage ($r = 0.001$) to unlinked ($r = 0.5$). Selection was modeled to affect viability between juvenile and adult life stages with segregating alleles interacting in a partially dominant manner such that the relative fitness of the two alternate *AA* and *aa* homozygotes and the *Aa* heterozygote for locus 2 were 1, $1 - s$, and $1 - hs$, respectively, where $h = 0.5$. We note that under these conditions, the equilibrium frequencies for the *A* and *a* alleles at locus 2 will be the mirror images of one another in demes 1 and 2. Consequently, the frequency $p_2$ for the *A* allele at locus 2 in deme 2 will be equal to $1 - p_1$ and the frequency $q_2$ of the *a* allele at locus 2 at equilibrium in deme 2 will be equal to $1 - q_1$, resulting in $p_1 = q_2$ and $q_1 = p_2$. These same considerations also hold at equilibrium when deriving multilocus estimates for $m_e$ and $F_{ST}$ below.

### ESTIMATING $m_e$ AND $F_{ST}$ WHEN MULTIPLE LOCI ARE UNDER SELECTION

We first considered the effects of multiple loci by expanding our computer simulations to include additional, unlinked genes under disruptive selection. For these simulations, fitness interactions were assumed to be multiplicative between loci, with per locus selection coefficients uniform across genes (either $s = 0.01$, 0.1, or 0.5 for each contributing locus). The resulting estimates of $m_e$ were then used to derive $F_{ST}$ values between populations, as described for the single locus analysis.

Practical computational considerations limited our simulation approach to a maximum of five loci under disruptive

selection. To examine the consequences of greater numbers of loci therefore required a second approach: analytical approximations to estimate $m_e$. A number of models have been developed to estimate $m_e$ with multiple loci under selection (Gavrilets 1997; Pialek and Barton 1997; Navarro and Barton 2003). For example, Bengtsson (1985) showed that if there are a number of unlinked loci with equal effects contributing to viability in a multiplicative way, then the effective migration rate for a neutral locus unlinked to the loci under selection is approximately equal to

$$m_e = m(W_{hyb})^2, \tag{5}$$

where,

$W_{hyb}$ = mean relative fitness of hybrids formed between immigrant individuals from deme 1 and resident individuals in deme 2.

With a life cycle in which selection follows migration and precedes mating, equation (5) can be modified to take into account the initial lower fitness of immigrants to become

$$m_e = m(W_m/W_r)(W_{hyb})^2, \tag{6}$$

where,

$W_m/W_r$ = ratio of the relative fitness of migrant versus resident individuals in deme 2 following migration.

To account for linkage of the neutral locus to one of the multiple genes under selection, a composite formula combining equations (1) and (6) and can be derived to estimate effective migration rate in which

$$m_e = m(q_1(1-hs) + p_1(1-s)r/(1-(1-hs)(1-r))) \times (W_m/W_r)(W_{hyb})^2, \tag{7}$$

or more accurately to account for fitness differences between homozygous and heterozygous genotypes at the selected gene of interest linked to the neutral locus

$$m_e = m\big(q_1^2 + p_1q_1(1-hs) + \big(p_1q_1(1-hs)r + p_1^2(1-s)r\big)\big/(1-(1-hs)(1-r))\big)(W_m/W_r)(W_{hyb})^2. \tag{8}$$

To solve equation (8) it requires estimating equilibrium allele and genotype frequencies when multiple loci in the genome are under divergent selection. We used an iterative approach to estimate $p_1$, $q_1$, $p_2$, and $q_2$ allele frequencies at each of the selected loci in demes 1 and 2 based on the premise that at equilibrium, gene frequencies within demes 1 and 2 prior to migration will be constant across generations. In addition, given random mating within populations, genotype frequencies for each selected locus before migration will roughly conform to Hardy–Weinberg expectations of $p_1^2$, $2\,p_1q_1$, and $q_1^2$ in deme 1 with associated fitness for that locus of 1, $1-hs$, and $1-s$ in deme 1, whereas the mirror image will hold for deme 2. Expanding from a single selected locus to an additional number of $g$ unlinked loci, the mean relative fitness of $AA$, $Aa$, and $aa$ genotypes in deme 1 are approximately $v$, $(1-hs)v$, and $(1-s)v$, respectively, where

$$v = \left[p_1^2 + 2p_1q_1(1-hs) + q_1^2(1-s)\right]^g. \tag{9}$$

As noted above, when selection coefficients and migration rates are symmetrical between subpopulations, allele and genotype frequencies in deme 2, as well as the relative fitness of these genotypes, will be the mirror image of those in deme 1. Consequently, genotype frequencies in demes 1 and 2 can be mixed in the appropriate ratio of $m$ to $1-m$ to determine the genotypic compositions of the populations following migration. These genotype values can subsequently be used in conjunction with the relative fitness coefficients for multilocus genotypes in the demes to derive allele frequency estimates for a locus after selection. Setting these postselection allele frequencies equal to the values prior to migration in the preceding generation results in an approximation of the equilibrium values for the allele $A$ ($p_1$) and the allele $a$ ($q_1$) at a selected locus in deme 1 (or for any of the $g$ other selected genes in the genome) under multilocus disruptive selection. We used successive iterations on the computer to solve these equations and estimate $p$ and $q$ allele frequencies in demes 1 and 2. These equilibrium values for $p$ and $q$ were then used to calculate $W_m/W_r$. Moreover, the allele frequencies $p$ and $q$ after selection for migrant versus resident individuals were also used, assuming random mating within demes, to calculate the relative fitness of multilocus hybrid offspring compared to residents ($W_{hyb}$). These values for $p$, $q$, $W_m/W_r$, and $W_{hyb}$ were substituted into equation (8) to estimate $m_e$ and $F_{ST}$ for multiple loci under selection.

We note that our approach for considering multiple loci under selection was equivalent to analyzing the effects of increasing amounts of overall (= total), genome-wide divergent selection on patterns of neutral genetic differentiation at loci linked to one of the selected loci. Thus, when such total selection is strong, there will be less opportunity for recombination as more migrant and hybrid individuals will die before reproduction and so the number of loci and total strength of selection are closely related. However, the effects of spreading the same total amount of selection across loci can be roughly approximated, for example, by comparing $F_{ST}$ values when $(1–0.5)^{ns}$ for strong selection equals $(1–0.1)^{nm}$ for moderate selection, where $n$s and $n$m are the number of loci under strong ($s = 0.5$) and moderate selection ($s = 0.1$), respectively. This translates into approximately 6.58 loci under moderate selection for each locus under strong selection.

## Results

The overall patterns observed in both single- and multilocus models (Figs. 1–6), under different parameter values, are summarized

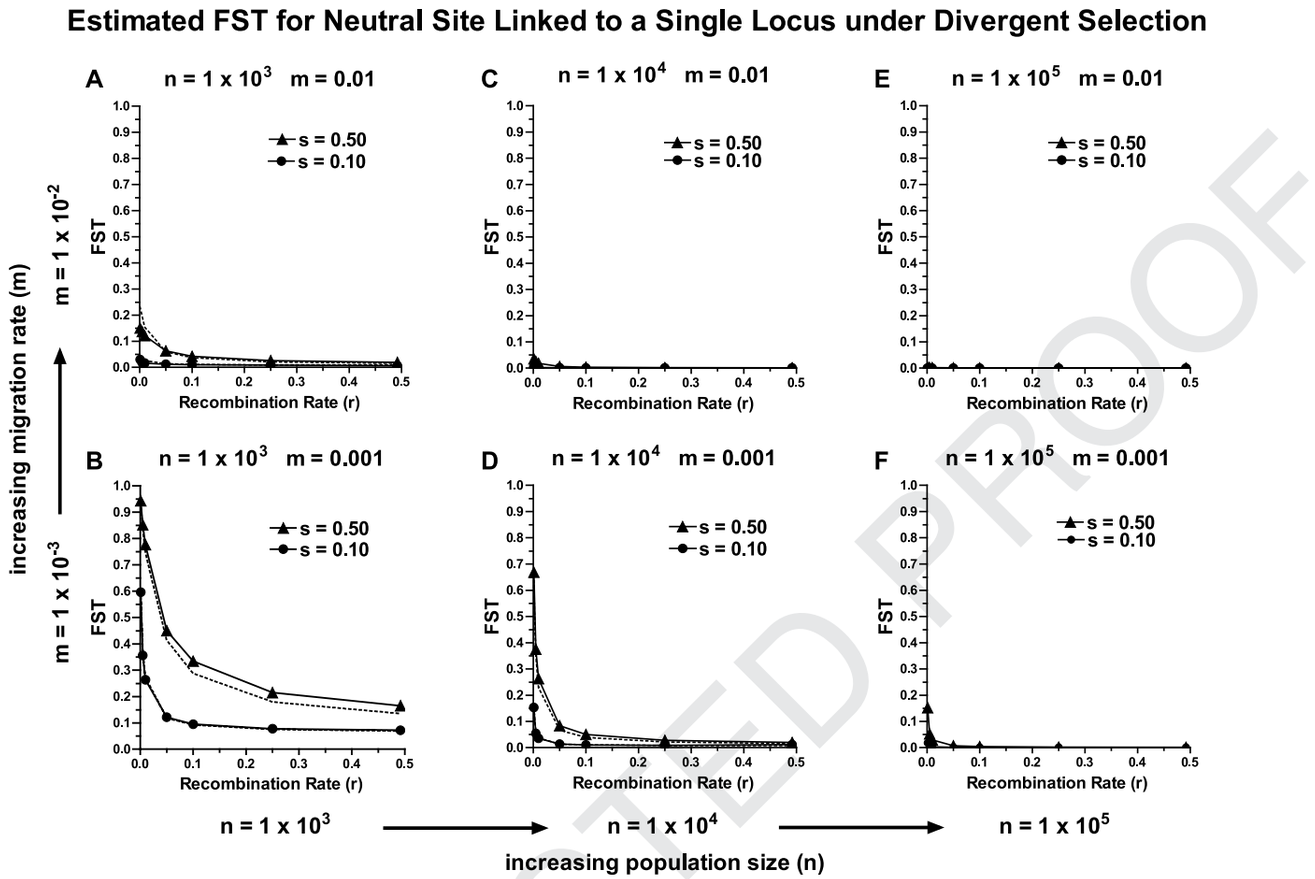


**Figure 1.** Estimated $F_{ST}$ for a neutral site linked at various recombination rates ($r$) to a single locus under divergent selection between two populations. Solid lines represent $F_{ST}$ values calculated for moderate ($s = 0.1$, circle symbols) and strong ($s = 0.5$, triangles) selection derived from population genetic computer simulations, as discussed in Methods section. Stippled lines in panels A, B, and D are $F_{ST}$ values based on analytical equation (1) in Methods section. For higher migration rates and larger population sizes, $F_{ST}$ values were essentially zero, similar to panel E.

in Table 2 and depicted in Figure 7. In general, divergence hitchhiking did not generate large regions of neutral differentiation, except for certain scenarios such as when migration rates and population sizes were small, or when multiple, but not numerous, loci were under selection (Figs. 1–7, S1, and S2). Specific results were as follows.

## SINGLE LOCUS UNDER DISRUPTIVE SELECTION: PATTERNS OF NEUTRAL DIFFERENTIATION

Estimated $F_{ST}$ values were very similar between the analytical and simulation single-locus approaches (compare stippled [analytical] and solid [simulation] lines in Fig. 1A,B,D). As expected, simulation $F_{ST}$ values were generally slightly higher than those from the analytical formula, because the simulations account for back recombination and migration, resulting in lower $m_e$ estimates.

The general conclusion from simulation and analytical analyses was that estimated $F_{ST}$ values for a neutral locus linked to a single gene under divergent selection are not expected to be large (Fig. 1). For example, based on a Lewontin–Krakauer distribution (1973) and a Beaumont-type outlier analysis (2005) "outlier status" would require $F_{ST}$ values greater than approximately five times baseline levels (when $r = 0.5$). Such levels of genetic divergence were not widespread. The exception was when migration rates and population sizes were both low ($m = 0.001$, $n_e = 1{,}000$), and fitness trade-offs strong ($s = 0.5$). In this circumstance, elevated neutral differentiation would accumulate between populations even relatively far away from the selected site (Fig. 1B), as previously described in figure 8A of Charlesworth et al. (1997). However, for migration rates $\geq 0.01$ and population sizes of 10,000–100,000, $F_{ST}$ values were low (near zero) and indistinguishable from baseline expectations except when the neutral locus was closely physically linked (e.g., $r = 0.005$–$0.001$) to the selected gene (Figs. 1C,E and 6A; compare Fig. 7A vs. 7D). Based on a rough equivalent of 1 megabase

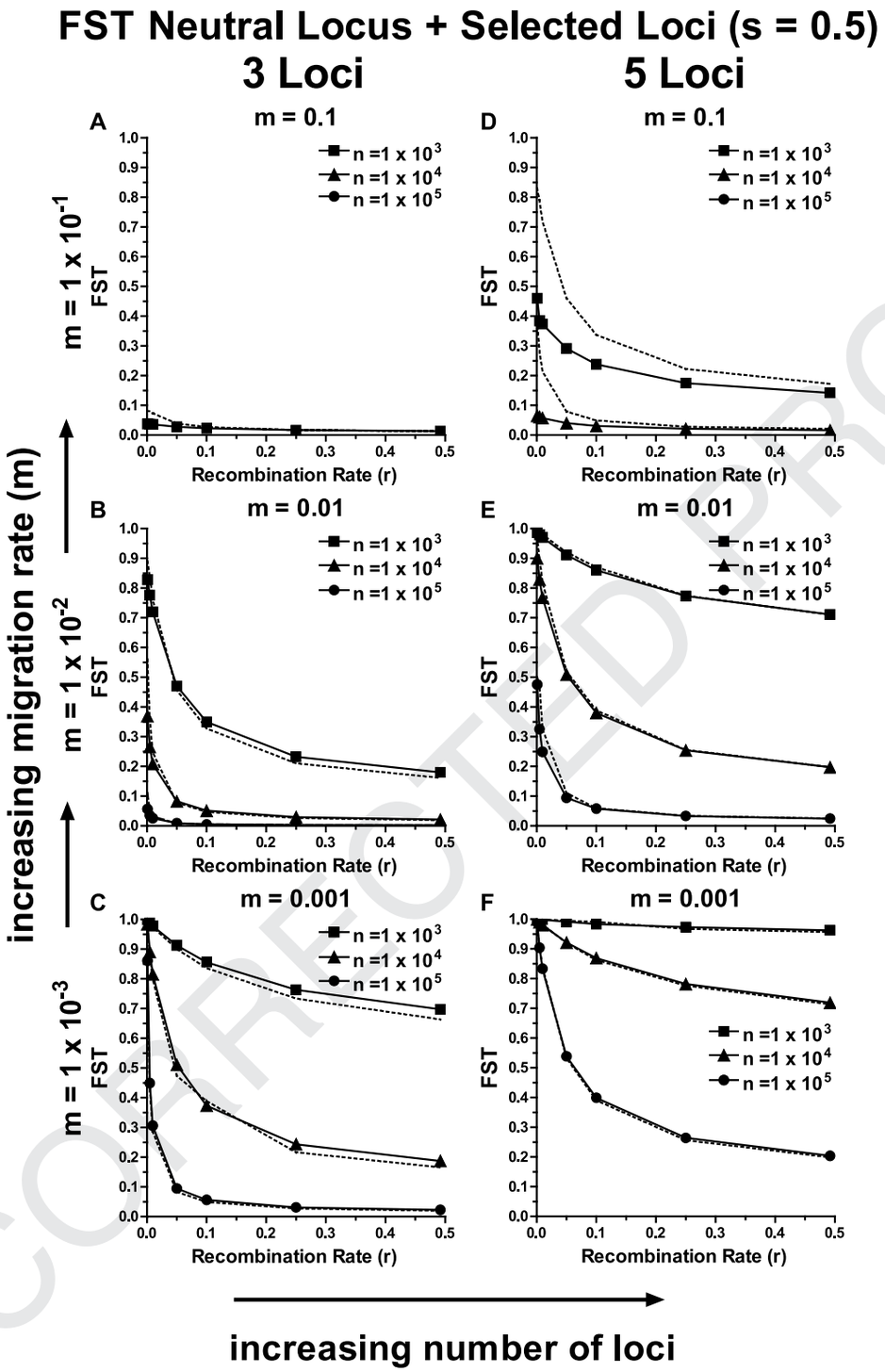


**Figure 2.** Estimated $F_{ST}$ for a neutral site linked at various recombination rates ($r$ = 0.001, 0.005, 0.01, 0.05, 0.10, 0.25, and 0.50) to one of a total of three or five loci under strong disruptive selection ($s$ = 0.5). Solid lines represent $F_{ST}$ values derived from population genetic computer simulations, whereas stippled lines depict values calculated from the composite analytical approach for multiple loci. Results are given for migration rates ($m$) of 0.001, 0.01, and 0.1 per generation coupled with effective population sizes ($n_e$) of 1 × $10^3$, 1 × $10^4$, and 1 × $10^5$.

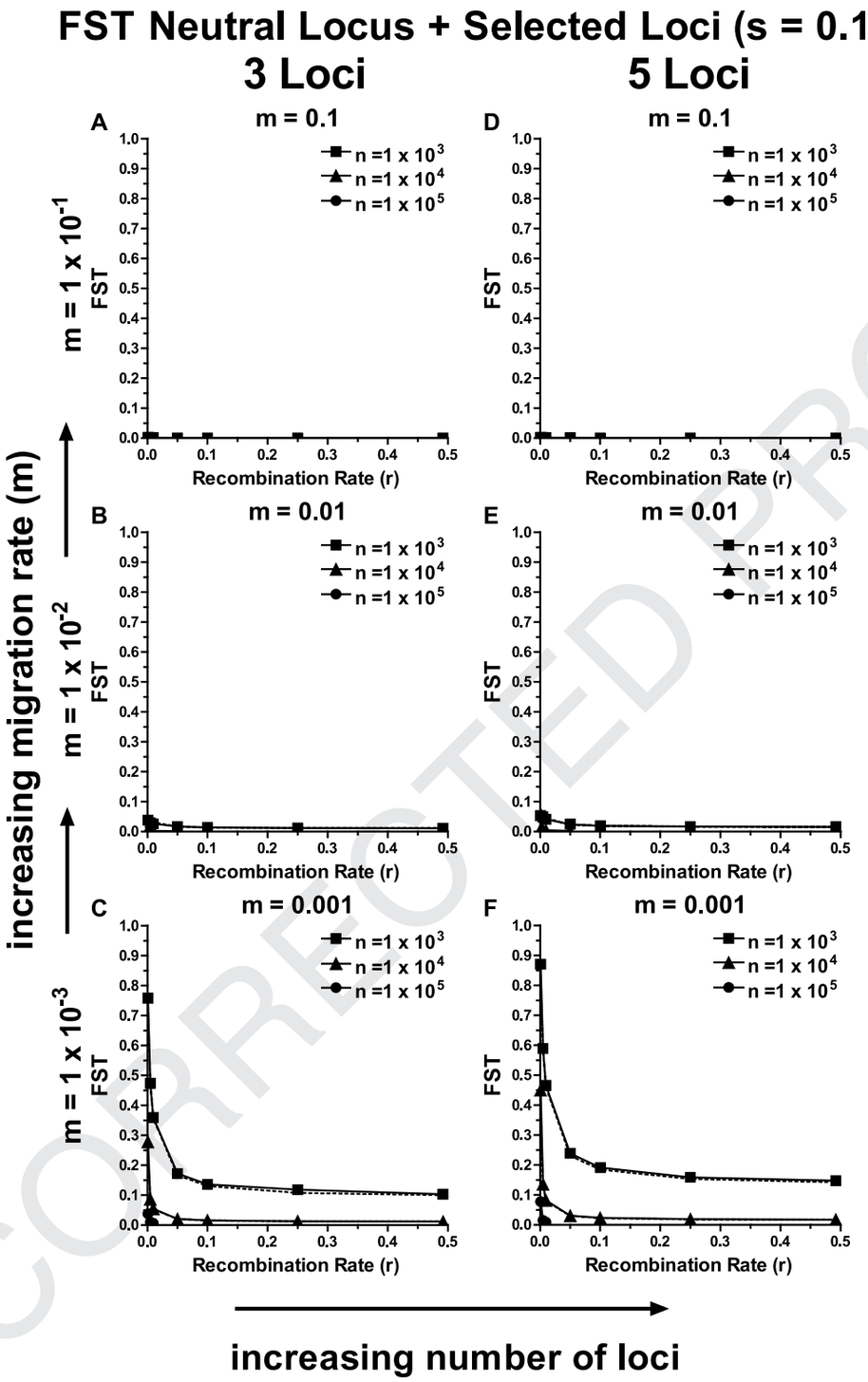


**Figure 3.** Estimated $F_{ST}$ for a neutral site linked at various recombination rates ($r$ = 0.001, 0.005, 0.01, 0.05, 0.10, 0.25, and 0.50) to one of a total of three or five loci under moderate disruptive selection ($s$ = 0.1). Solid lines represent $F_{ST}$ values derived from population genetic computer simulations, whereas stippled lines depict values calculated from the composite analytical approach for multiple loci. Results are given for migration rates ($m$) of 0.001, 0.01, and 0.1 per generation coupled with population sizes ($n_e$) of 1 × $10^3$, 1 × $10^4$, and 1 × $10^5$.

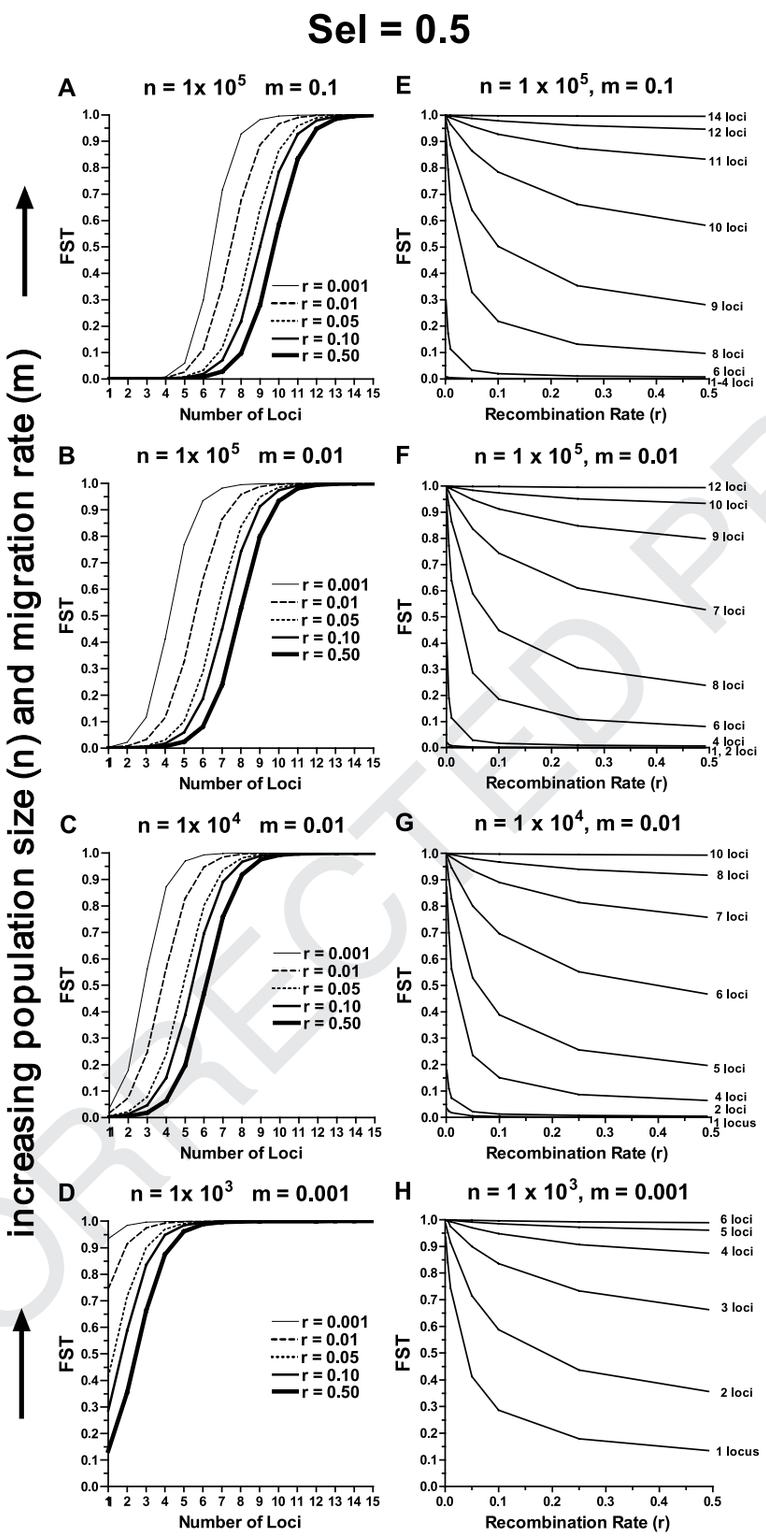


**Figure 4.** Estimated $F_{ST}$ for a neutral site linked at various recombination rates ($r$) to one of a given number of loci under strong divergent selection ($s = 0.5$), as determined by the composite analytical approach for multiple loci. Column panels display results in different orientations with number of loci or recombination rate representing the x-axis.

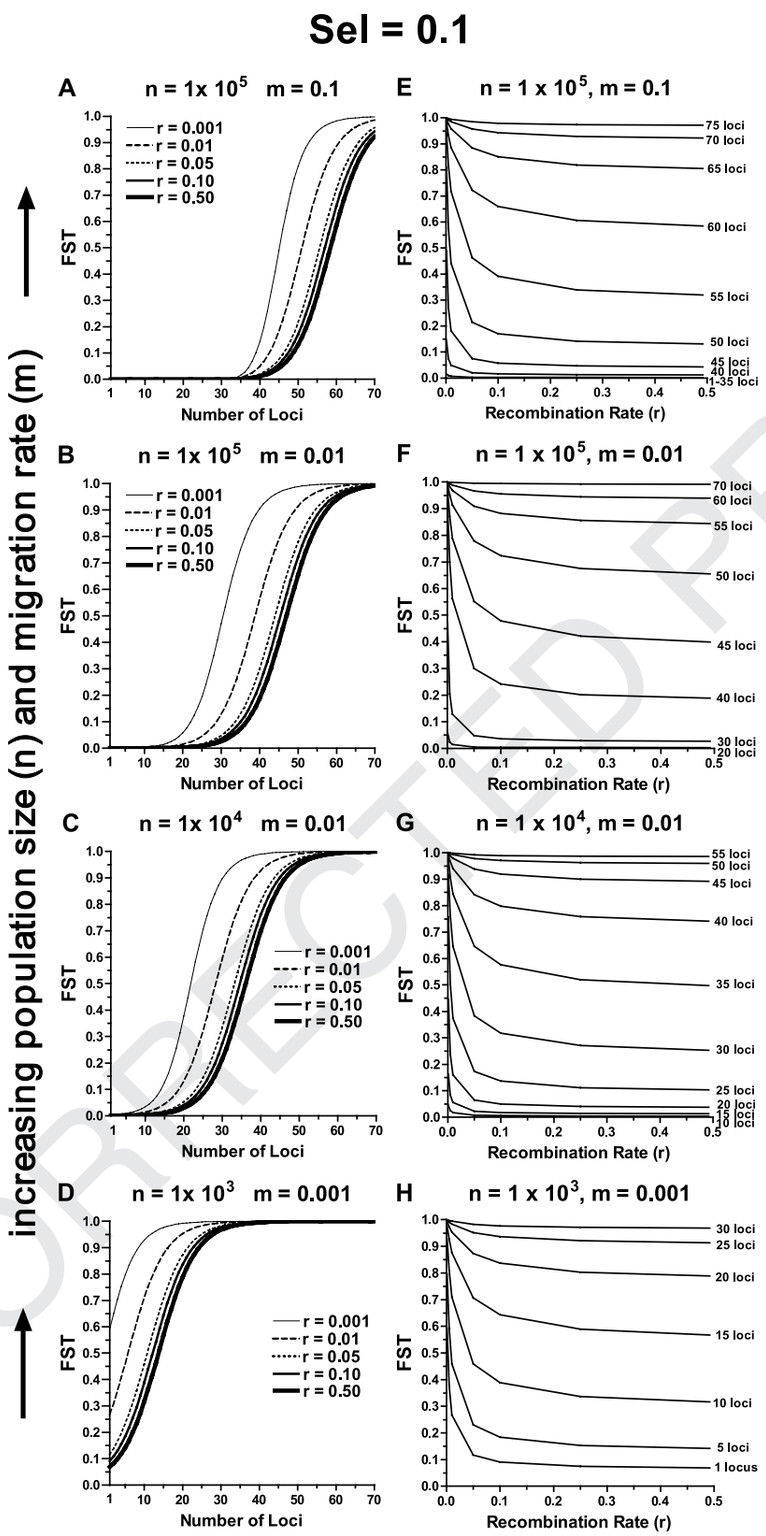


**Figure 5.** Estimated $F_{ST}$ for a neutral site linked at various recombination rates ($r$) to one of a given number of loci under moderate divergent selection ($s = 0.1$), as determined by the composite analytical approach for multiple loci.

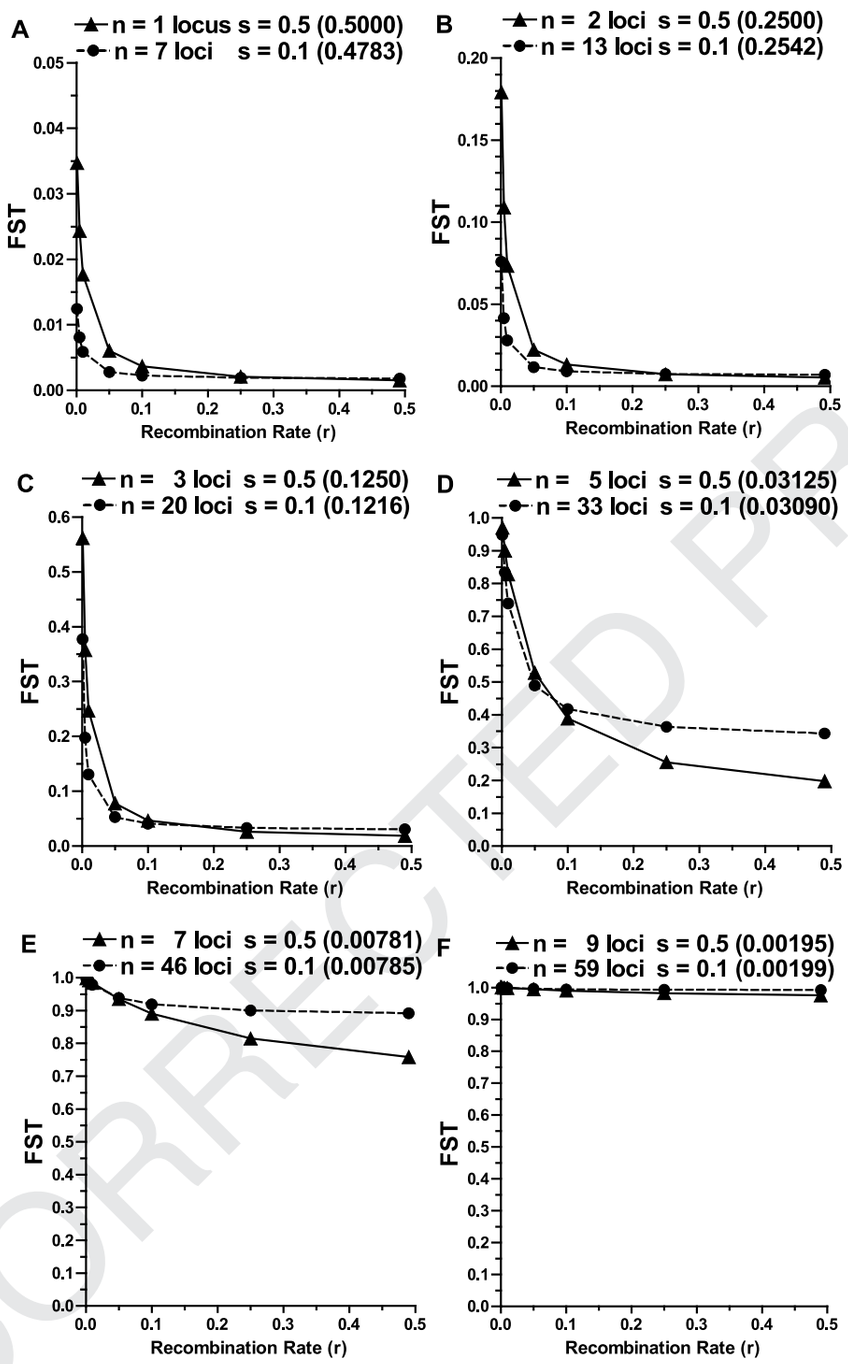


**Figure 6.** Comparisons of $F_{ST}$ values estimated by the composite analytical approach for a neutral site linked at various recombination rates ($r$) to a selected site when the total strength of selection is similar for the indicated numbers of loci under strong ($s = 0.5$) and moderate ($s = 0.1$) divergent selection. Shown are results for a migration rate ($m$) of 0.01 and an effective population sizes ($n_e$) of $1 \times 10^4$.

pairs (Mbp) of DNA per centiMorgan, this would suggest a region of from 1,000 to 5,000 bp. These findings imply that divergence hitchhiking will usually not be substantial during the early stages of speciation if only a single locus is under disruptive selection.

We note that our conclusions concerning the scope of neutral genetic differentiation surrounding a single selected site mirror those reached by Petri (1983) based on diffusion approximations of a two-locus, island–continent migration model.

**Table 2. Summary of the results from single- and multilocus models. See also Figure 7.**

| Scenario | Results | Conclusion | Figure |
|---|---|---|---|
| **Single locus:** small population sizes, low migration rates ($n_e$=1,000, $m$=0.001) | Neutral differentiation can extend away from a selected site far along a chromosome, resulting in a relatively large region of genetic differentiation (particularly if selection is strong) | Effects of divergence hitchhiking can be appreciable | Figs. 1B, 7A |
| **Single locus:** larger population sizes and migration rates | Little or no neutral differentiation unless the neutral locus is very closely physically linked (e.g., $r$=0.001) to a strongly selected gene | Effects of divergence hitchhiking generally weak | Figs. 1, 7D |
| **Multilocus:** multiple loci | When multiple (but not "numerous") loci are under strong selection, neutral differentiation can be accentuated near selected sites and some differentiation can extend far along a chromosome | Effects of divergence hitchhiking can be appreciable (so long as numerous loci are not under selection) | Figs. 2, 4, 7B, S1 |
| **Multilocus:** numerous loci | When numerous loci are subject to selection, genome-wide divergence occurs irrespective of linkage. The term numerous is relative and how many loci are required for genome-wide differentiation depends on parameter values such as selection strength | Effects of divergence hitchhiking weak or absent, genome-wide divergence occurs | Figs. 2, 5, 7C, F, S1, S2 |

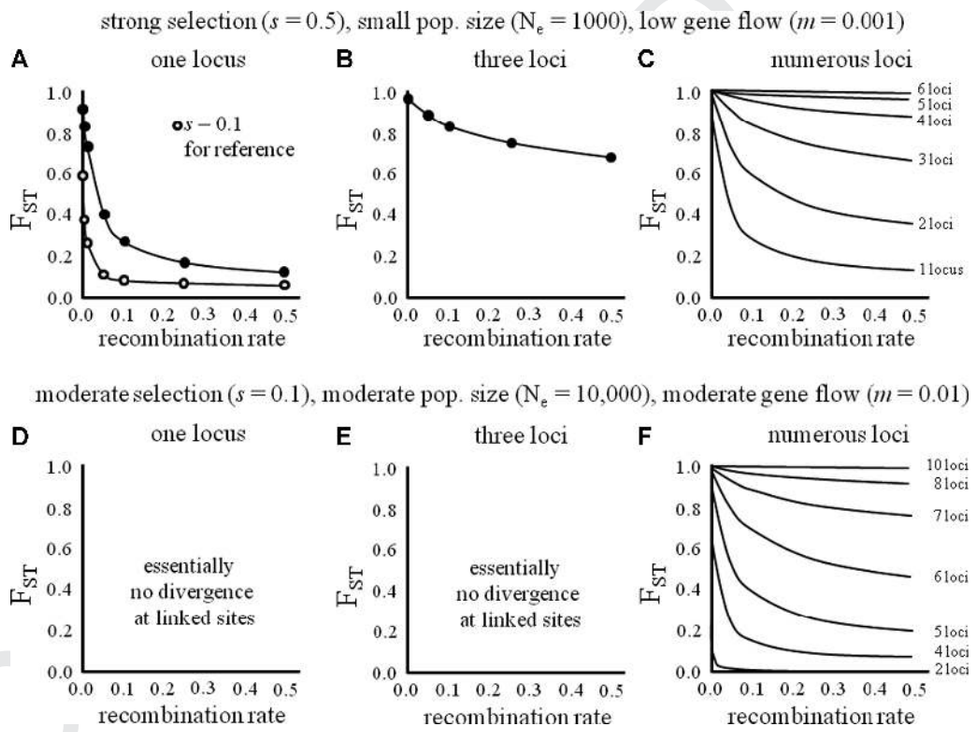


**Figure 7. A visual summary of our general findings. Divergence hitchhiking can generate regions of neutral differentiation extending away from a selected site (genomic islands), but only under certain conditions. Specifically, when a single locus is under divergent selection, large regions of differentiation extending away from a selected site are only observed when selection is strong, effective population sizes small, and gene flow low (contrast panels A and D). When multiple loci are under selection, regions of differentiation can be larger and more readily observed (panel B). However, this pattern is tempered via two strong caveats. First, even with selection on multiple loci, regions of neutral differentiation still occur only under certain regions of parameter space, for example, generally requiring strong selection (contrast panels B and E, and also Figs. 2 and 3) or very tight linkage to a selected site. Second, when numerous loci are under selection, genome-wide divergence can occur such that genomic islands are erased. This effect can be seen in panel C by comparing scenarios with 1–3 loci to those with 4–6 loci under selection, and in panel F by comparing scenarios with 1–6 loci to those with 7–10 loci under selection. pop. = population.**

### MULTIPLE LOCI UNDER DISRUPTIVE SELECTION: PATTERNS OF NEUTRAL DIFFERENTIATION

Estimated $F_{ST}$ values were generally very similar between the analytical and simulation multilocus approaches (compare stippled [analytical] solid [simulation] lines in Figs. 2 and 3). The exception was when the migration rate was high ($m = 0.1$) for strong selection ($s = 0.5$), where the composite analytical formula predicted higher levels of neutral differentiation than the simulations (Fig. 2D).

The multilocus $F_{ST}$ estimates implied that within certain ranges of loci under disruptive selection, conditions exist where divergence hitchhiking can generate reasonably large regions of accentuated neutral differentiation (Figs. 2–6, S1, and S2). However, this finding is tempered by three caveats: (1) the conditions are narrow when selection is strong because populations can potentially rapidly advance to a phase in which genome-wide differentiation occurs regardless of linkage, in essence erasing genomic islands (see Fig. 7C,F), (2) neutral differentiation is often accentuated only for sites very closely linked to the target genes under selection (e.g., $\leq$1 centiMorgan), and (3) the initial number of selected loci needed for observing elevated neutral differentiation around a site under moderate to weak selection can be large, suggesting a reduced role for divergence hitchhiking on only a few genomic regions during early stages of speciation.

The consequences of strong, multilocus selection ($s = 0.5$) are depicted in Figures 2, 4, 6, 7, and S1. If migration rates and population sizes in the early stages of divergence-with-gene flow speciation are $\geq$0.01 and $1 \times 10^4$, respectively, the simulation and analytical models predicted that from two to eight selected loci are needed to first detect effects of divergence hitchhiking on elevating neutral differentiation. At this stage, the neutral locus generally has to reside within around 1 centiMorgan ($r = 0.01$) or so of the selected gene for the effect to be marked (e.g., to be five times above the baseline level for $r = 0$). Increasing the number of selected loci could enlarge this region slightly. However, the role of divergence hitchhiking in accentuating neutral divergence quickly closes when from seven to 13 loci were under strong divergent selection. At this point, fixed (or nearly fixed) allele differences begin to accumulate between populations for all loci, irrespective of their linkage to selected sites. In essence, genomic islands become "submerged" beneath a high genome wide level of genetic divergence (isolation-by-adaptation across the genome; Nosil et al. 2008). Although, $s = 0.5$ is very strong selection, comparable selection coefficients have been described in nature, for example, in cases of selection on cryptic (Nosil 2004) and mimetic color-pattern loci (Mallet 2006). In addition, we assumed multiplicative fitness effects in our analyses. If fitness were to decrease more slowly than multiplicative, then the number of loci under selection required for reaching genome-wide divergence would be larger. In contrast, if fitness were to increase faster than multiplicative (positive epistatic fitness interactions), then fewer loci may be required.

Moderate selection ($s = 0.1$) combined with reasonable migration rates and population sizes increased the initial number of selected loci required for divergence hitchhiking up to 10–50 (Figs. 3, 5, 6, and S2). Under moderate selection, very tight linkage of the neutral locus to a selected locus (around 1 centiMorgan or less) was needed to detect a pronounced hitchhiking effect. Populations showed uniformly high $F_{ST}$ across the genome when 45–80 loci experienced moderately strong selection.

Under weak selection ($s = 0.01$), the number of loci defining when divergence hitchhiking may act and the requirement for tight linkage were even greater (not graphed). For example, with $m = 0.01$ and $n_e = 10{,}000$ around 350 loci under selection were needed to generate a region of increased neutral differentiation (and only for neutral sites within 0.1 centiMorgan of the selected locus). Roughly, 500–600 loci generated uniformly high $F_{ST}$ across the genome.

### COMPARISONS FOR SIMILAR LEVELS OF TOTAL SELECTION

The general conclusions reached above concerning divergence hitchhiking were not greatly affected by whether similar amounts of total selection were concentrated on just a few genes under strong selection versus spread across several loci experiencing moderate selection. However, there were some quantitative effects on $F_{ST}$ that for the case of $m = 0.01$ and $n_e = 10{,}000$ are graphically depicted in Figure 6 that could potentially influence outlier locus detection in a genome scan. Similar effects were seen across the range of $m$ and $n$ parameter values analyzed in the study. Most importantly, except for when the fitness of migrants was very low (>0.002) and genetic differentiation nearly complete across the genome (Fig. 6F), $F_{ST}$ increased more sharply with decreasing recombination rate for a neutral site linked to a locus under strong than moderate selection (Fig. 6A–E). As a consequence, it would generally be easier to statistically detect divergence for a particular neutral marker linked to a locus under strong than moderate selection when the total amount of selection is similar across the genome. The degree to which this difference affects the total number of outlier loci identified in a genome scan is a question requiring further analysis.

## *Discussion*

Our results imply that although divergence hitchhiking may sometimes create large regions of differentiation around a selected site, this generally requires relatively specific conditions. In essence, what is required is that effective recombination is significantly reduced locally in the genome without being substantially reduced globally. Specifically, when a single locus is

under strong selection, large regions of differentiation are only expected when migration rates are low relative to the strength of selection and population sizes are small, resulting in low effective recombination rates around the selected site. When multiple loci are under selection, regions of differentiation can be larger. However, when enough loci are under selection, gene flow becomes extensively reduced across the whole genome, allowing genome-wide genetic divergence (when exactly this occurs depends on parameter values). A pattern that might result under this multilocus scenario is a positive association among population pairs between their level of neutral genetic and adaptive phenotypic divergence. Such a pattern is analogous to isolation-by-distance, but where gene flow becomes increasingly reduced by increasing adaptive divergence, rather than greater geographic distance (isolation-by-adaptation) (Nosil et al. 2008). The overall scenario of genome-wide divergence due to selection on many loci is consistent with the "multifarious selection" hypothesis of Rice and Hostert (1993), in which speciation is promoted by a multitude of different selection pressures acting on many genes/traits (see Nosil et al. 2009b for review). These overall findings lead us to recognize that the often-continuous process of speciation-with-gene flow might often have three "stages" during which different evolutionary processes predominate.

## STAGES OF ECOLOGICAL SPECIATION

The first stage involves the establishment of initial genetic differentiation at one or a few loci (c.f. Via 2009). This most likely occurs via moderate to strong disruptive selection on these loci, because weaker selection may be unable to counter the then still high rates of gene flow (Nosil et al. 2009b; Via 2009). Divergence hitchhiking is not expected to play a major role in this initial stage, because high migration rates (and/or large effective population sizes) during this period preclude widespread neutral differentiation. Nonetheless, under certain conditions, divergence hitchhiking may play a role, for example, if there is fortuitous tight linkage (among selected genes or among selected and neutral loci), the genes under selection reside in regions of extensively reduced recombination (e.g., within-chromosomal rearrangements), or demographic factors (e.g., inbreeding, assortative mating, cyclical parthenogenesis) accentuate the effectiveness of selection. For example, when hybridization follows migration, selection on the subsequent asexual, clonal hybrid generations of a cyclic parthenogen would reduce gene flow by a factor of $(W_{hyb})^{n-1}$ compared to an obligate sexually reproducing species, where $n$ represents the number of asexual generations in the parthenogen's life cycle.

The second stage represents a period in which enough genetic changes have accumulated to reduce effective migration at sites physically proximate to those under selection. Thus, effective migration around such sites may be low enough to allow neutral differentiation. At this time, it is also possible that fortuitous linkage of new mutations to loci already subject to divergent selection can facilitate further divergence between populations. This second stage is thus the period during which divergence hitchhiking could most strongly promote genetic differentiation and speciation, and during which regions of differentiation might build in clusters within the genome. However, this second phase can be transitory, because once numerous loci diverge, phase three is initiated, and even if genomic islands built up during the second phase, they may rapidly become erased (i.e., submerged via a high genome-wide level of genetic differentiation).

This final, third phase is in which gene flow becomes extensively reduced across the whole genome because many loci are now under selection and diverging between populations. Thus, the genome generally closes to introgression and genome-wide divergence occurs. During this period, linkage relationships and divergence hitchhiking are not expected to be important for predicting genetic differentiation, as selected and neutral differences accumulate across genome.

## EXTENSIONS AND CAVEATS TO OUR RESULTS

### *Epistasis, habitat preference, mating preference, and life cycle*

Our models considered divergent selection generating fitness trade-offs between populations in different environments, a common scenario in nature (Schluter 2000). Future work could consider the effects of epistasis between loci, because such epistasis might favor physical linkage between interacting loci (Kimura 1956; Charlesworth and Charlesworth 1975; Kouyos et al. 2006), thereby potentially affecting the opportunity for divergence hitchhiking. Other factors not considered, such as habitat specific and assortative mating, might also enhance the potential for divergence hitchhiking, although if they hasten genome-wide divergence, they could reduce the window during which divergence hitchhiking facilitates speciation. Finally, we considered a life history with selection immediately following migration. This scenario enhances the effectiveness of disruptive selection in reducing effective migration rates. Cases in which selection occurs on juvenile offspring prior to migration, disruptive selection will be less effective as a gene flow barrier (Fry 2003). Thus, windows of opportunity for divergence hitchhiking can be reduced during the early stages of speciation for life cycles in which selection follows mating. Q2

### *Islands as seeds for further divergence and selective sweeps*

The current results represent the first component of a larger development of theory of genomic architecture accompanying speciation-with-gene flow. Our $F_{ST}$ values reflect the degree to

which physical linkage to a selected site might influence patterns of linked neutral differentiation. Our results do not encapsulate absolute probabilities for gain or loss of new variation, or how such probabilities vary with linkage to loci under disruptive selection. Additionally, our current results are important for assessing patterns of $F_{ST}$ at neutral markers, but our analyses did not consider how reduced gene flow around divergently selected loci encourages further divergence. This second issue requires further attention, and might particularly consider where further divergence is at linked selected versus neutral loci.

We also note that our analysis of $F_{ST}$ focused on levels of differentiation expected when a balance is reached between migration, selection, and recombination. We did not consider the possibility that strong selection associated with adaptation to a new habitat could rapidly sweep linked neutral markers to high frequency. In this case, $F_{ST}$ could initially be at a high maximum over an extended area and then decay through gene exchange over time (Nielsen 2005). Moreover, even when ecological races are old, new mutations may occasionally sweep through one or the other of the races, generating new hitchhiking events. However, initial adaptation to a new habitat may often be predicated on preexisting rather than new mutational variation (Barrett and Schluter 2008), which can significantly reduce the magnitude of neutral differentiation surrounding selected sites resulting in "soft sweeps" (Orr and Batencourt 2001; Hermisson and Pennings 2005; Prezworski et al. 2005). Extending the current model to scenarios which consider further divergence at sites linked to those under selection is a logical next step, one which might address more broadly the types of genomic architecture likely to underlie speciation-with-gene flow.

### *Allopatric and parapatric divergence*

Our study pertains to speciation-with-gene flow. The implications under an allopatric context are likely to differ, at least slightly (Via 2001, 2009; Butlin 2008; Via and West 2008). Under allopatric divergence, gene flow does not oppose genetic differentiation. Even upon secondary contact, baseline levels of gross migration may be reduced relative to taxa that did not undergo a period of allopatric divergence (i.e., some barriers to gene flow may have evolved in allopatry and act as such upon secondary contact). Thus, a greater role for divergence hitchhiking immediately following secondary contact between already divergent populations may exist compared to that at the earliest stages of speciation-with-gene flow. The importance of this issue for explaining speciation in nature depends strongly on what stage of divergence populations commonly come into secondary contact. Extensive genetic differentiation and a high degree of reproductive isolation built up in allopatry can take a long period of time to decay upon secondary contact, especially if migration rates are low, thereby obscuring any role that divergence hitchhiking may have played in generating differences following contact. These aspects concerning secondary contact require further development.

In addition, our analysis of divergence hitchhiking largely considered habitats to be discrete and sympatric. However, ecological speciation-with-gene flow may also commonly be initiated where habitats are partially separated and have transitional biomes between them. When habitats are adjacent, hybrid zone theory becomes relevant, and future work could explore the consequences of different spatial structures for divergence hitchhiking.

## IMPLICATIONS FOR FINDING GENES UNDER SELECTION

Our findings also have implications for evolutionary genetics, particularly for the search for "speciation genes" causing reproductive isolation (Wu 2001; Coyne and Orr 2004; Wu and Ting 2004; Noor and Feder 2006). Because regions of differentiation are often predicted to be small, it could be hard to find such regions in nature, unless genomic coverage is dense or many such regions are distributed across the genome. However, once such a region is found, it might be near the target of selection. In some conditions, however, such as when selection acts on many loci and genome wide divergence occurs, it may be difficult to statistically differentiate a neutral locus linked to a selected locus from the baseline level of genetic differentiation observed throughout the genome. In this respect, the key issue is where a neutral locus falls on the scale between the maximal amount of differentiation expected due to complete linkage to a selected site ($r = 0$) and the background level when unlinked ($r = 0.5$; Note that the background level in a genome scan does not reflect the gross migration rate [$m$] but the genome-wide effective rate [$m_e$]). When total selection is strong across the genome, maximal and background points may both be high and not very different. Thus, for 11 loci under strong selection ($s = 0.5$) when $m = 0.01$ and $n_e = 100{,}000$, the expected $F_{ST}$ values at equilibrium for a neutral site completely linked ($r = 0$) versus unlinked ($r = 0.5$) to a target selected gene are 0.9999 versus 0.9806, respectively. In other circumstances when total selection is weaker, the expected value for a neutral site may often be low and close to the baseline unless the site resides near a target gene under selection. Thus, for a single locus under strong selection when $m = 0.01$ and $n_e = 100{,}000$, the expected $F_{ST}$ values for a neutral site 1 centiMorgan versus unlinked to a target selected gene are only 0.0018 and 0.00015, respectively, differences which could be difficult to statistically distinguish due to sampling variance.

## GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

We report patterns of genetic differentiation expected for neutral loci under the divergence hitchhiking hypothesis. We find that divergence hitchhiking can generate regions of differentiation under some, but not all, conditions. Thus, rather than discount the

effects of divergence hitchhiking, our findings generate some predictions about when it is likely to be of greatest importance: when migration rates are low, populations are small, multiple but not numerous loci are subject to strong selection, other factors such as chromosomal inversions reduce recombination, and perhaps during secondary contact. Because these conditions will not always occur, or the duration in which they occur can be brief, regions of differentiation created by divergence hitchhiking will not necessarily act as seeds for divergence in the genome. Instead, whenever selection on loci is greater than migration, we might often expect new divergence to crop up more evenly throughout the genome, as observed in several population genomic studies (Scotti-Saintagne et al. 2004; Achere et al. 2005; Grahame et al. 2006; Butlin 2008; Egan et al. 2008; Nosil et al. 2008, 2009a; Turner et al. 2008; Wood et al. 2008). Indeed, there may be only transient stages of speciation when it is possible to observe substantial neutral differentiation localized around islands of divergence (Via 2009), and where such divergence is important for speciation. Further work on divergence hitchhiking should focus on the extent to which regions of divergence that are generated can "grow" during the speciation process, including when multiple loci under selection are themselves linked, and the significance of such growth for causing the reduced gene flow that characterizes the formation of new species.

## ACKNOWLEDGMENTS

We thank N. Barton, J. Mallet, A. Meyer, D. Funk, M. Noor, D. Ortiz-Barrientos, H. Collin, J. Galindo, M. Doebeli, and an anonymous reviewer for comments on previous versions of this manuscript and for discussions pertaining to genomic differentiation and speciation. During the preparation of the manuscript, the authors were hosted by the Institute for Advanced Study, Wissenschaftskolleg, Berlin. This work was supported by grants to JLF from the National Science Foundation and the United States Department of Agriculture.

## LITERATURE CITED

Achere, V., Favre, J. M., G. Besnard, and S. Jeandroz. 2005. Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). Mol. Ecol. 14:3191–3201.

Avise, J. 2000. Phylogeography: the history and formation of new species. Harvard Univ. Press, Cambridge, MA.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends Ecol. Evol. 23:38–44.

Barton, N. H. 1979. Gene flow past a cline. Heredity 43:333–340.

———. 1983. Multilocus clines. Evolution 37:454–471.

———. 2000. Genetic hitchhiking. Philos. Trans. R. Soc. Lond. B 355:1553–1562.

Barton, N. H., and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridizing populations. Heredity 57:357–376.

Beaumont, M. A. 2005. Adaptation and speciation: what can F-st tell us? Trends Ecol. Evol. 20:435–440.

Bengtsson, B. O. 1985. The flow of genes through a genetic barrier. Pp. 31–42 *in* J. J. Greenwood, P. H. Harvey, and M. Slatkin, eds. Evolution essays in honor of John Maynard Smith. Cambridge Univ. Press, Cambridge.

Berlocher, S. H., and J. L. Feder. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? Annual Review of Entomology 47:773–815.

Butlin, R. K. 2008. Population genomics and speciation. Genetica online Sept. 6. DOI 10.1007/s10709-008-9321-3.

Charlesworth, D., and B. Charlesworth. 1975. Theoretical genetics of Batesian mimicry II. Evolution of supergenes. J. Theor. Biol. 55:305–324.

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70:155–174.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc. Sunderland, MA.

Dopman, E. B., L. Perez, S. M. Bogdanowicz, and R. G. Harrison. 2005. Consequences of reproductive barriers for genealogical discordance in the European corn borer. Proc. Natl. Acad. Sci. USA 102:14706–14711.

Egan, S. P., P. Nosil, and D. J. Funk. 2008. Selection and genomic differentiation during ecological speciation: isolating the contributions of host-association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. Evolution 62:1162–1181.

Emelianov, I., F. Marec, and J. Mallet. 2004. Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proc. R. Soc. Lond. B 271:97–105.

Feder, J. L. 1998. The apple maggot fly, *Rhagoletis pomonella*: flies in the face of conventional wisdom about speciation? Pp.130–144 *in* D. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, London.

Feder, J. L., and P. Nosil. 2009. Chromosomal inversions and species differences: where are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? Evolution in press.

Gavrilets, S. 1997. Single locus clines. Evolution 51:979–983.

———. 2004. Fitness landscapes and the origin of species. Princeton Univ. Press, Princeton, NJ.

Gavrilets, S., and M. B. Cruzan. 1998. Neutral gene flow across single locus clines. Evolution 52:1277–1284.

Gavrilets, S., and A. Vose. 2005. Dynamic patterns of adaptive radiation. Proc. Natl. Acad. Sci. USA 102:18040–18045.

Geraldes, A., N. Ferrand, and N. W. Nachman. 2006. Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). Genetics 173:919–933.

Grahame, J. W., C. S. Wilding, and R. K. Butlin. 2006. Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. Evolution 60:268–278.

Harr, B. 2006. Genomic islands of differentiation between house mouse subspecies. Genome Res. 16:730–737.

Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp.111–133 *in* D. Otte and J. A. Endler, eds. Speciation and its consequences. Sinauer, Sunderland, MA.

Hawthorne, D. J., and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412:904–907.

Hermisson, J., and P. S. Pennings. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. 7:1–45.

Kimura, M. 1956. A model of genetic system which leads to closer linkage by natural selection. Evolution 10:278–287.

Kouyos, R. D., S. P. Otto, and S. Bonhoeffer. 2006. Effect of varying epistasis on the evolution of recombination. Genetics 173:589–597.

Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. Genetics 74:175–195.

Machado, C. A., T. S. Haselkorn, and M. A. F. Noor. 2007. Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. Genetics 175:1289–1306.

Mäkinen, H. S., J. M. Cano, and J. Merilä. 2008a. Identifying footprints of directional and balancing selection in marine and freshwater threespine stickleback (*Gasterosteus aculeatus*) populations. Mol. Ecol. 17:3565–3582.

Mäkinen, H. S., T. Shikano, J. M. Cano, and J. Merilä. 2008b. Hitchhiking mapping reveals a candidate genomic region for natural selection in three-spined stickleback chromosome VIII. Genetics 178:435–465.

Mallet, J. 2006. What has *Drosophila* genetics revealed about speciation? Trends Ecol. Evol. 21:186–193.

Navarro, A., and N. H. Barton. 2003. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. Evolution 57:447–459.

Nielsen, R. 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39:197–218.

Noor, M. A. F., and J. L. Feder. 2006. Speciation genetics: evolving approaches. Nat. Rev. Genet. 7:851–861.

Noor, M. A. F., D. A. Garfield, S. W. Schaeffer, and C. A. Machado. 2007. Divergent between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. Genetics 177:1417–1428.

Nosil, P. 2004. Reproductive isolation caused by visual predation on migrants between divergent environments. Proc. R. Soc. Lond. B 271:1521–1528.

Nosil, P., T. H. Vines, and D. J. Funk. 2005. Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation-by-adaptation' and multiple roles for divergent selection. Evolution 62:316–336.

Nosil, P., D. J. Funk, D. Ortiz-Barrientos. 2009a. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.

Nosil, P., L. Harmon, and O. Seehausen. 2009b. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.

Orr, H. A., and Batencourt. 2001. Haldane's sieve and adaptation from standing genetic variation. Genetics 157:875–844.

Payseur, B. A., J. G. Krenz, and M. W. Nachman. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution 58:2064–2078.

Prezworski, M., C. Graham, and J. D. Wall. 2005. The signature of positive selection on standing genetic variation. Evolution 59:2312–2323.

Petry, D. 1983. The effect on netral gene flow of selection at a linked locus. Theor. Popul. Biol. 23:300–313.

Pialek, J., and N. H. Barton. 1997. The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. Genetics 145:493–504.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments in speciation: what have we learned in 40 years? Evolution 47:1637–1653.

Rogers, S. M., and L. Bernatchez. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae). Mol. Biol. Evol. 24:1423–1438.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford Univ. Press, Oxford, U.K.

Scotti-Saintagne, C., S. Mariette, I. Porth, P. G. Goicoechea, T. Barreneche, K. Bodenes, K. Burg, and A. Kremer. 2004. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. Genetics 168:1615–1626.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. Genet. Res. 58:167–175.

Smadja, C., J. Galindo, and R. K. Butlin. 2008. Hitching a lift on the road to speciation. Mol. Ecol. 17:4177–4180.

Strasburg, J. L., C. Scotti-Saintagne, I. Scotti, Z. Lai, and L. H. Rieseberg. 2009. Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. Mol. Biol. Evol. 26:1341–1355.

Templeton, A. 1981. Mechanisms of speciation – A population genetic approach. Annu. Rev. Ecol. Syst. 12:23–48.

Turner, T. L., and M. W. Hahn. 2007. Locus- and population specific selection and differentiation between incipient species of *Anopholes gambiae*. Mol. Biol. Evol. 24:2132–2138.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLOS Biol. 3:1572–1578.

Turner, T. L., M. T. Levine, M. L. Eckert, and D. J. Begun. 2008. Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. Genetics 179:455–475.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol. Evol. 16:381–390.

———. 2009. Natural selection in action during speciation. Proc. Natl. Acad. Sci. USA 106:9939–9946.

Via, S., and J. West. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. 17:4334–4345.

White, B. J., M. W. Hahn, M. Pombi, B. J. Cassone, N. F. Lobo, F. Simard, and N. J. Besansky. 2007. Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheleles gambiae*. PLoS Genet. 3:e217. Q7

Wood, H. M., J. W. Grahame, S. Humphray, J. Rogers, and R. K. Butlin. 2008. Sequence differentiation in regions identified by a genome scan for local adaptation. Mol. Ecol. 17:3123–3135.

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

Wu, C.-I., and C.-T. Ting. 2004. Genes and speciation. Nat. Rev. Genet. 5:114–122.

Yatabe, Y., N. C. Kane, C. Scotti-Saintagne, and L. H. Rieseberg. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. Genetics 175:1883–1893.

Associate Editor: M. Doebeli

## *Supporting Information*

The following supporting information is available for this article:

**Figure S1.** Estimated $F_{ST}$ for a neutral site linked at various recombination rates ($r = 0.001$, 0.005, 0.01, 0.05, 0.10, 0.25, and 0.50) to one of an increasing total number of loci (1–15) each under strong disruptive selection ($s = 0.5$), as calculated from composite analytical approach for multiple loci.

**Figure S2.** Estimated $F_{ST}$ for a neutral site linked at various recombination rates ($r = 0.001$, 0.005, 0.01, 0.05, 0.10, 0.25, and 0.50) to one of an increasing total number of loci (1–70) each under moderate disruptive selection ($s = 0.1$), as calculated from the composite analytical approach for multiple loci.

Supporting Information may be found in the online version of this article.
(This link will take you to the article abstract).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## *Queries*

**Q1** Author: Please provide the expansion of the term QTL in the sentence "For example, in host races of *Acyrthosiphon*. . .."

**Q2** Author: Fry (2003) has not been included in the Reference List. Please supply complete publication details.

**Q3** Author: Orr and Betancourt (2001) has been changed to Orr and Batencourt (2001) so that this citation matches the Reference List. Please check.

**Q4** Author: Prezeworski et al. (2005) has been changed to Prezworski et al. (2005) so that this citation matches the Reference List. Please check.

**Q5** Author: Acheré et al. (2005) has been changed to Achere et al. (2005) so that this citation matches the Reference List. Please check.

**Q6** Author: Please update the publication details of Feder and Nosil (2009).

**Q7** Author: Please provide the page range in White et al. (2007).