Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses

CK Griswold

Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

This paper uses computer simulations to determine how gene flow between populations affects (1) the genetic architecture of a local adaptation and (2) properties of alleles segregating in quantitative trait locus (QTL) mapping populations. Results suggest that the average magnitude of an allele that causes a phenotypic difference between populations declines as the migration rate increases, but with an increase in migration, alleles of larger magnitude cause proportionally more of the phenotypic difference between populations. Gene flow between populations that are used in a QTL study tends to cause the average magnitude and percent variance explained (PVE) of an allele in a mapping population to increase. Thus, although the average magnitude of an allele causing a difference declines with migration the average magnitude or PVE of an allele in a QTL mapping population may increase. The reason is that the probability an allele is sampled for a QTL mapping population is in direct proportion to its frequency and alleles of larger magnitude tend to segregate at relatively higher frequencies than alleles of smaller effect with an increased migration. As the rate of gene flow between populations increases, the proportion of the phenotypic difference explained by alleles that are segregating in a QTL mapping population (and therefore potentially detected) decreases. Lastly, results suggest QTL alleles of large effect ($>20\%$ PVE) should be commonly found, provided the divergence time between populations is not too long or optima of populations are not too far apart.
Heredity (2006) **96,** 445–453. doi:10.1038/sj.hdy.6800822; published online 12 April 2006

Introduction

When two populations experience different environments, theory predicts that gene flow between populations limits local adaptation. Although this is true, provided that selection is strong enough relative to gene flow, the two populations will adapt to their local environments to a certain extent (Felsenstein, 1976; Slatkin, 1978; Hendry *et al*, 2001). The local adaptation will manifest itself in some measurable phenotypic difference. Studies may then be carried out to find and measure the effect of alleles at genetic loci that cause a phenotypic difference between locally adapted populations. For instance, quantitative trait locus (QTL) mapping is now becoming commonly used to find loci with alleles that cause phenotypic differences between populations or species that experience gene flow (Hawthorne and Via, 2001; Peichel *et al*, 2001).

Classic theoretical studies by Wright (1940), Levene (1953) and Bulmer (1972) have shown that, depending on its selective effect, an allele may or may not be segregating in two locally adapted populations that exchange migrants. From a deterministic standpoint, an allele will either be fixed in both populations, extinct in

both populations, or segregating in both populations. The alleles that are segregating in both populations potentially cause phenotypic differences between populations. As the migration rate increases, the frequencies of segregating alleles are expected to converge between populations. This convergence in frequencies between populations leads to a reduction in the phenotypic difference caused by that allele. The rate of convergence in frequency is a function of the selective effects of the allele in the two populations. Furthermore, as migration rates vary, alleles with different selective effects are more or less likely to be fixed, segregating or extinct in both populations.

In summary, the works of Wright (1940), Levene (1953) and Bulmer (1972) are suggestive that gene flow not only determines the extent that populations phenotypically diverge, but also the genetic basis of the phenotypic difference (see also, Karlin 1982). A limitation of the work by Wright (1940), Levene (1953) and Bulmer (1972) is that their analysis focused on a single locus. In actuality, alleles at many loci are present when populations begin to diverge and new alleles arise at different loci by mutation during the course of divergence. Multilocus models of local adaptation along a cline (eg, Barton, 1983) have shown that linkage between alleles at multiple loci strengthens the barrier to dispersal between local environments relative to single locus models. Thus, to fully understand the genetic basis of a local adaptation, multilocus models are necessary that allow for physical linkage. Additionally, the genetic basis of a local adaptation is a function of the rates at which new alleles

arise by mutation *and* are retained by selection. Although alleles of small effect may arise often, they may be only weakly selected and lost or remain at low frequencies in both populations – thus causing little phenotypic difference. Rare new alleles of larger effect may be more strongly selected for in one population and against in another – thus contributing to a phenotypic difference – despite their rarity. In this paper, I seek to understand the genetic basis of a local adaptation when there is gene flow between two locally adapted populations and when the alleles that contribute to an adaptation are present at potentially many loci and have arisen by mutation during the course of adaptation.

QTL studies seek to find and estimate the properties of alleles that cause phenotypic differences between populations. Properties of alleles detected in QTL studies may change depending on whether populations are isolated or connected by gene flow. Whether an allele is detected in a QTL study is a function of the probability it is sampled for that study in the first place. The probability an allele is sampled in a QTL study is a function of its frequency in a population. The probability of sampling an allele of a given size may change depending on whether there is gene flow between populations or not. The change in probability arises because the expected frequency an allele segregates in a population may change with gene flow. The expected frequency an allele segregates is a function of its selective effect and the rate of gene flow. Thus, two alleles that both cause differences between populations, but have different effects on a phenotype may have different chances of being sampled in a QTL study depending on whether the populations used in the study are isolated or not. Accordingly, results from QTL studies may suggest that the genetic architecture of local adaptations is different when populations are isolated versus when they are connected, but this difference may, in part, be an artifact of the different sampling probabilities of alleles when there is gene flow versus when there is not. The extent to which this artifact is important has not been explored.

This paper uses computer simulations to quantify the genetic basis of local adaptations when there is gene flow versus when populations are isolated. It then takes samples from the two locally adapted populations to form mapping populations that may be used in a QTL experiment. Properties of the alleles present in the mapping populations are then compared to those present in the two locally adapted populations to see how well the properties correspond.

Computer simulation methods

Overview
The simulations are individual based. Each individual is diploid and consists of a genome of five paired chromosomes in which each chromosome consists of two million sites. There is initially a genetically homogeneous ancestral population that splits into two populations. A single character is modeled and undergoes selection in opposite directions in the two populations. Selection on individuals acts through fecundity. In some simulations individuals may migrate to the area occupied by the other population and potentially mate.

Mutations affecting the character(s) arise stochastically throughout the history of divergence.

Migration
Migration is modeled using forward migration rates such that an individual in population one migrates to population two with probability m_{12} per generation and an individual in population two migrates to population one with probability m_{21}. Traditionally, a ratio of migration rate (m) to selection coefficient (s) equal to 1.0 ($m/s = 1.0$) represents a threshold at which migration tends to overwhelm local adaptation (Crow and Kimura, 1970). In this article, I focus on average m/s ratios that are ≤ 1.0; although, I present results that suggest there is a potential for local adaptation even when $m/s > 1.0$.

Mutation and recombination
Mutations that give rise to new alleles occur at a rate of 10^{-10} per meiosis per site, and recombination occurs at each meiosis at a rate of 10^{-6} per meiosis per pair of adjacent sites. The mutation rate is within about an order of magnitude of rates estimated for mouse and maize morphological characters (Lynch and Walsh, 1998, 337). The average number of recombinations per chromosome per meiosis used here is slightly higher than an estimate of the average number of recombinations per chromosome in humans (2.0 versus 1.4) based on data provided by Kong *et al* (2002).

The distribution of effects of a mutation on the character is bidirectional (ie, mutation may either increase or decrease the character's value) and symmetric about zero (ie, mutations are neutral on average). The magnitude of a mutation is exponentially distributed. For this study, all mutations have codominant effects. Although there are no back mutations, the number of sites is so large relative to the number of changes over the course of the simulations that the mutational heritability remains nearly constant (at a phenotypic level). The mutation model assumes that mutations do not have epistatic phenotypic effects, accordingly results may not apply when there is a strong epistatic component to the effects of mutations (which can be deduced in a QTL study). The assumption of codominance may be important because empirical work shows that deleterious mutations are on average recessive (Peters *et al*, 2003). The recessivity may allow deleterious mutations to persist in populations longer (at low frequency) than if they were codominant. If these recessive deleterious mutations have effects that are small in magnitude, then they will not contribute much to the overall phenotypic difference between populations. Lastly, it is assumed that alleles act nonpleiotropically. With pleiotropy, the fate of an allele becomes less tied to the effect it has on a particular character. The consequences of pleiotropy are left for further study.

Mating
For each offspring, two parents are randomly chosen via fecundity selection in direct proportion to their fitness relative to all of the individuals in the population. Offspring are formed until a population size of 1000 is reached (exceptions will be noted). There is no survivorship selection.

Fitness

Fecundity is a symmetric linearly decreasing function about the optimum for the character. The initial phenotype of all individuals at the beginning of evolution is zero, and their fitnesses are set equal to one. The absolute fitness of individual i in population j, f_{ij}, is equal to $1 + y_j - \sigma_j |x_{ij} - o_j|$ when $|x_{ij} - o_j| < (1 + y_j)/\sigma_j$ and $f_{ij} = 0$ otherwise, where x_{ij} is the phenotype of individual i in population j, o_j is the optimum phenotype in population j, σ_j is the strength of selection in population j, and the term y_j scales f_{ij} such that an individual with a phenotypic value of zero for the character has a fitness of one in population j. Thus, the term y_j equals $\sigma_j |o_j|$. The phenotype of an individual is

$$x_{ij} = \sum_{k=1}^{a} \sum_{\ell=1}^{b_k} \sum_{n=1}^{2} \frac{\delta_{ijk\ell n}}{2}$$

where a is the number of chromosomes, b_k is the number of sites on the kth chromosome, n indexes over the pair of homologous chromosomes and $\delta_{ijk\ell n}$ is the homozygous effect of an allele at a particular site. The probability that individual i in population j is chosen to form an offspring is $f_{ij}/\sum_{i=1}^{N_j} f_{ij}$.

Analytic methods

Proportion of the difference: The average phenotypic difference that an allele at site ℓ on chromosome k with homozygous effect $\delta_{k\ell}$ causes between populations is $D_{k\ell} = \delta_{k\ell}(p_{k\ell,1} - p_{k\ell,2})$, where $p_{k\ell,1}$ is the frequency of the allele in population one and $p_{k\ell,2}$ is its frequency in population two. The formula for $D_{k\ell}$ assumes the alternate allele at the locus has a (scaled) homozygous effect of zero and that alleles act codominantly. The average difference between two individuals from the two populations is the sum of the average differences caused by each allele individually, $\sum_{k=1}^{a} \sum_{\ell=1}^{b_k} D_{k\ell}$. The proportion of the difference attributable to an allele on chromosome i at site j is thus, $D_{ij}/\sum_{k=1}^{a} \sum_{\ell=1}^{b_k} D_{k\ell}$.

Properties of QTL alleles in a mapping population

At the end of each replicate of the evolutionary divergence between the two populations, a mapping population was created according to either an outbred F_2 (assuming two F_1 individuals are crossed), outbred backcross (assuming a single F_1 individual is used in the backcross), or inbred design. In an outbred design the parents of F_1 individuals come from outbreeding populations. In an inbred design the parents of F_1 individuals come from inbred lines that were formed from outbreeding populations. If an allele was segregating in the mapping population, then it was assumed that the QTL experiment had enough power to detect that allele and the estimate of the effect of the allele was unbiased. Accordingly, this paper only analyzes how the segregating frequencies of alleles in the two populations – that are used to form a mapping population – affect what is potentially detected in a QTL experiment.

Percent variance explained (PVE)

Individual alleles as well as linked groups of alleles contribute to the genetic variance in the mapping population. In this analysis, I calculate the percent of the variance explained by individual alleles. An allele of size δ contributes $V(\delta) = p(1-p)\delta^2/2$ to the genetic

Table 1 Probability that an allele is segregating in a mapping population

Inbred F_2 design	$1-(p_1p_2+(1-p_1)(1-p_2))$
Outbred F_2 design	$1-(p_1^2p_2^2+(1-p_1)^2(1-p_2)^2)$
Inbred backcross design	$1-(p_1p_2+(1-p_1)(1-p_2))$
Outbred backcross design	$1-(p_1^2p_2+(1-p_1)^3(1-p_2))^a$

[a]Assuming the F_1 individual is backcrossed to an individual from population one.

variance in the mapping population, where p is its frequency in the mapping population. The PVE by an allele is $PVE = V(\delta)/V(total)$, where $V(total)$ is the total genetic variance, $\sum_{\Delta} V(\delta)$, such that Δ is the set of all alleles in the mapping population. Note that the PVE by an allele is relative to the overall genetic variance, not the phenotypic variance.

Performance of a QTL design

The measure of the performance of a QTL design used in this paper is the fraction of the difference between populations that is explained by the alleles segregating in a mapping population. Note that if an allele is not segregating in the mapping population, then it will not be detected. Depending on the frequencies of an allele in the two populations, it is more or less likely to be segregating in the mapping population of a QTL experiment. Table 1 gives the probability that an allele will be segregating in the mapping population and, as a consequence, potentially detected.

Confidence intervals

95% confidence intervals of the mean of a parameter estimate from simulations were based on 1000 bootstrap replicates.

Results

Magnitude of phenotypic differences between populations

In Figure 1, the magnitude of the phenotypic difference between populations relative to when there is no gene flow is plotted for divergence times of 10^3 and 10^4 generations. The figure suggests that phenotypic differences arise even when the average magnitude of the selective effect of an allele is equal to and sometimes less than the migration rate.

Fraction of the difference explained by locally fixed versus segregating alleles

The phenotypic difference between populations is explained more by segregating alleles than locally fixed alleles as the migration rate increases relative to selection (Table 2). Interestingly, locally fixed alleles sometimes contribute to the difference, even when there is migration (particularly if the divergence time is long). This is because populations have diverged to such an extent that migrants have very low fitness, on average, and the effective migration rate is nearly zero.

The populations in Table 2 were initially genetically homogeneous. Simulations were run for the case when populations diverged for 10^3 generations, but the ancestral population (consisting of 2000 individuals), prior to splitting into two populations (each consisting of

1000 individuals), was at approximately mutation-selection balance. When there was no migration between the descendant populations and the average selective magnitude of an allele was equal to that in Table 2, 48% (95%
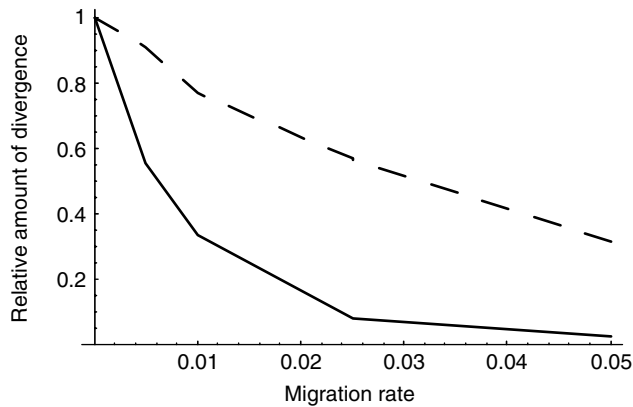

Figure 1 The relative amount of phenotypic divergence after 10^3 (solid line) and 10^4 (dashed line) generations of population divergence versus the migration rate between populations. Each population experienced pure directional selection, that is stabilizing selection was not initiated. The amount of phenotypic divergence is relative to when there is no gene flow for 10^3 and 10^4 generations of divergence, respectively. The average magnitude of a mutation was 0.01 and the strength of selection was 1.0; accordingly, the average magnitude of the selection coefficient of a mutation was 0.01. The absolute amount of divergence, on average, when populations diverged for 10^3 generations and there was no gene flow was 0.24 95% CI = {0.21,0.26} and when populations diverged for 10^4 generations and there was no gene flow, the absolute amount of divergence, on average, was 2.40 95% CI = {2.32,2.47}.

CI = {45,52}) of the phenotypic difference was caused by segregating alleles. When $m/s = 1.0$, segregating alleles caused all of the phenotypic difference. These results are very similar to the simulation results in which individuals in the two populations were genetically homogeneous at the beginning of divergence, suggesting that results based on a genetically homogeneous ancestor at the time of split are reasonably robust.

Statistical properties of fixed and segregating alleles

When both locally fixed and segregating alleles are present, the average magnitude of fixed alleles is larger than segregating alleles (Table 3). The greatest difference between fixed and segregating alleles ($\sim 300\%$) is seen when the divergence is relatively short (1000 generations) because for an allele to fix within that time it needs to have a large effect and be strongly selected. An increase in the migration rate tends to increase the average magnitude of locally fixed and segregating alleles, but interestingly the average magnitude of a random allele decreases (Table 3). There is a decrease because segregating alleles contribute more to the difference with an increase in migration and the average magnitude of their effects is smaller than locally fixed alleles.

Provided that the divergence time is not too long or the optima of the populations are not too far apart (relative to the average magnitude of a mutation), the majority of the time there was an allele that caused more than 20% of the phenotypic difference between populations (Table 4). If the divergence time is very long and populations have yet to reach their optimum, an allele that explains more than 20% of the difference becomes vanishingly rare.

Table 2 Fraction of the phenotypic difference caused by segregating alleles[a]

Divergence time	$m/s = 0.0$	$m/s = 0.01$	$m/s = 1.0$
10^3 (pure directional)[b]	0.522 (0.505, 0.540)[c]	0.890 (0.876, 0.904)	1.00 (1.00, 1.00)
10^4 (pure directional)[d]	0.051 (0.047, 0.055)	0.051 (0.047, 0.055)	0.802 (0.726, 0.872)
10^4 (stabilizing selection)[e]	0.000 (0.00, 0.010)	0.849 (0.808, 0.891)	1.0 (1.0, 1.0)

[a]In all cases, $s = E[|\delta|] = 0.01$ and migration was symmetric such that $m = m_{12} = m_{21}$.
[b]$o_1 = 1.0$, $o_2 = -1.0$.
[c]Values in parenthesis beside the mean are 95% confidence intervals of the mean.
[d]$o_1 = 10.0$, $o_2 = -10.0$.
[e]$o_1 = 0.1$, $o_2 = -0.1$.

Table 3 The average magnitude of locally fixed (upper), segregating (middle) and randomly chosen (lower) alleles in a local population[a]

Divergence time	$m/s = 0.0$	$m/s = 0.01$	$m/s = 1.0$
10^3 (pure directional)[b]	0.0316 (0.0307, 0.0327)[c]	0.0354 (0.0345, 0.0365)	—
	0.0108 (0.0102,0.0115)	0.0124 (0.0117, 0.0132)	0.0124 (0.0116, 0.0132)
	0.0207	0.0149	0.0124
10^4 (pure directional)[d]	0.0206 (0.0197, 0.0215)	0.0206 (0.0197, 0.0215)	0.0252 (0.0241, 0.0262)
	0.0100 (0.0094, 0.0106)	0.0097 (0.0091, 0.0103)	0.0166 (0.0158, 0.0175)
	0.0201	0.0200	0.0183
10^4 (stabilizing selection)[e]	0.0156 (0.0146, 0.0166)	0.0231 (0.0220, 0.0242)	—
	0.0082 (0.0076, 0.0087)	0.0096 (0.0090, 0.0103)	0.0117 (0.0110, 0.0124)
	0.0156	0.0116	0.0117

[a]In all cases, $s = E[|\delta|] = 0.01$ and migration was symmetric such that $m = m_{12} = m_{21}$.
[b]$o_1 = 1.0$, $o_2 = -1.0$.
[c]Values in parenthesis beside the mean are 95% confidence intervals of the mean.
[d]$o_1 = 10.0$, $o_2 = -10.0$.
[e]$o_1 = 0.1$, $o_2 = -0.1$.

Distribution of the effects of alleles causing phenotypic differences between populations

The distributions in Figure 2 consist of both segregating and locally fixed alleles whose statistical properties were summarized in the previous two sections. In part (a) of Figure 2, populations are isolated, whereas in part (b) populations are connected by gene flow. The distributions are sometimes quite similar despite very different histories. For instance, there is little difference between the distributions when populations experience gene flow

Table 4 The expected probability that there is an allele that causes at least 20% of the phenotypic difference between populations[a]

Divergence time	$m/s = 0.0$	$m/s = 1.0$
10^3 (pure directional)[b]	0.56 (0.51, 0.61)[c]	0.96 (0.94, 0.97)
10^4 (pure directional)[d]	0.0	0.0
10^4 (stabilizing selection)[e]	0.70 (0.60, 0.78)	0.67 (0.60, 0.73)

[a]In all cases, $s = E[|\delta|] = 0.01$ and migration was symmetric such that $m = m_{12} = m_{21}$.
[b]$o_1 = 1.0$, $o_2 = -1.0$.
[c]Values in parenthesis beside the mean are 95% confidence intervals of the mean.
[d]$o_1 = 10.0$, $o_2 = -10.0$.
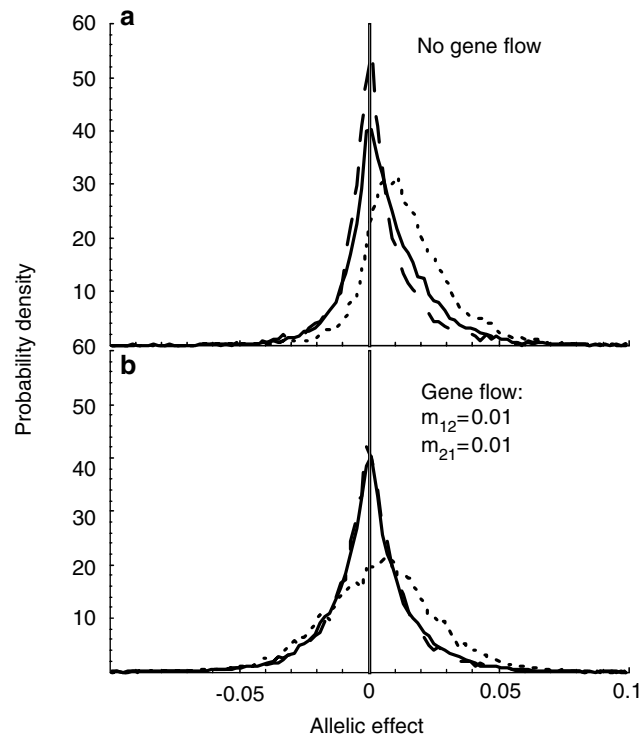[e]$o_1 = 0.1$, $o_2 = -0.1$.

Figure 2 The distribution of the phenotypic effects of alleles present in population one that cause adaptive differences between it and population two. The solid line corresponds to the directional selection case where divergence occurred for 10^3 generations and the fitness parameters were $o_1 = 1.0$, $o_2 = -1.0$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. The dotted line corresponds to the directional selection case where divergence occurred for 10^4 generations and the fitness parameters were $o_1 = 10.0$, $o_2 = -10.0$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. The dashed line corresponds to the case when the divergence time was 10^4 generations and stabilizing selection occurred. The fitness parameters in this case were $o_1 = 0.1$, $o_2 = -0.1$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. The average magnitude of the selection coefficient of a mutation was 0.01.

(Figure 2b) and have been diverging for 10^3 generations under pure divergent selection versus when the populations initially diverged and subsequently experienced stabilizing selection (over a period of 10^4 generations).

Average segregating frequency of alleles that cause differences between populations

The averages in Figure 3 consist of both segregating and locally fixed alleles. In the pure divergent selection case, migration causes the average frequency of locally beneficial alleles to be less than if populations were isolated and causes the average frequency of locally deleterious alleles to be higher (Figure 3a). When there is stabilizing selection, there is less of an effect of migration on the average frequency of alleles that were beneficial during the initial divergence (Figure 3b). Overall, alleles of smaller magnitude are at lower frequencies, on average, than alleles of larger magnitude. The reason alleles of smaller magnitude are at lower frequencies, on
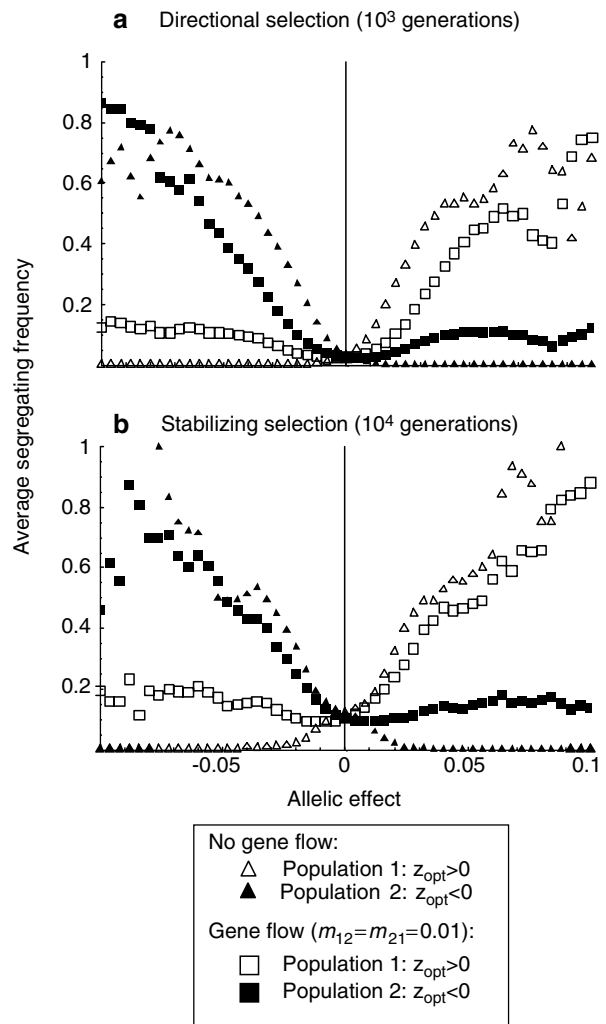

Figure 3 The average frequency of alleles that cause differences between populations. The values represent a moving average in which each average was taken in increments of 0.004 allelic units and consist of alleles within a bin range of ± 0.008 units about the central value. In the directional selection simulations: $o_1 = 1.0$, $o_2 = -1.0$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. In the stabilizing selection simulations: $o_1 = 0.1$, $o_2 = -0.1$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. The average magnitude of the selection coefficient of a mutation was 0.01.

average, is because they are typically young alleles and are weakly selected such that they do not increase in frequency quickly. Furthermore, with migration, any benefits of alleles with small effects are swamped by migration.

Proportion of the difference between populations caused by alleles of a given size

Despite alleles of smaller magnitude being present at higher densities (see Figure 2), alleles of moderate magnitude, generally, cause most of the difference between populations (Figure 4a). Alleles of small
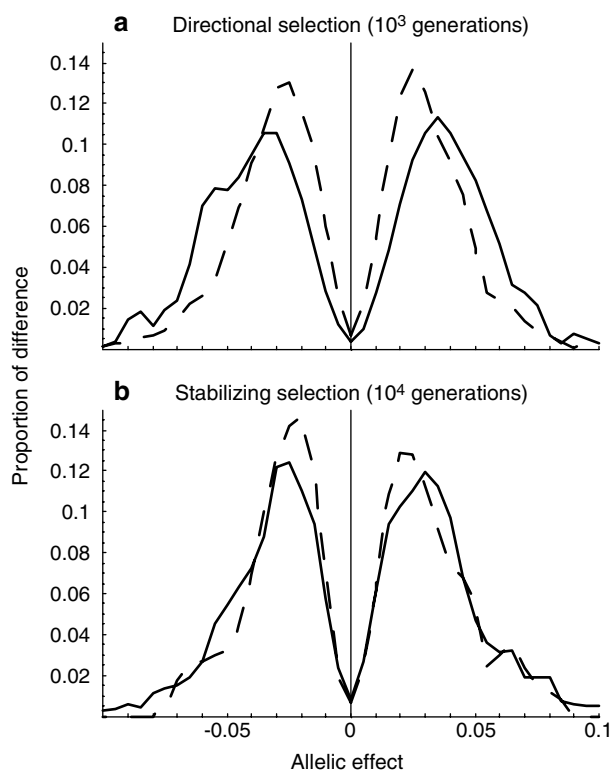

Figure 4 The proportion of the overall difference between populations caused by alleles of a given size. Dashed lines correspond to when the two populations are isolated and solid lines to when the populations exchange migrants ($m_{12} = 0.01$ and $m_{21} = 0.01$). The values represent the average proportion of the difference caused by alleles in increments of 0.005 allelic units and consist of alleles within a bin range of ± 0.005 units about the central value. In the directional selection simulations: $o_1 = 1.0$, $o_2 = -1.0$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. In the stabilizing selection simulations: $o_1 = 0.1$, $o_2 = -0.1$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. The average magnitude of the selection coefficient of a mutation was 0.01.

magnitude do not contribute much to the phenotypic difference between populations because they have small effects and their average frequencies are low and roughly equal in both populations, further eroding any difference they would cause. When populations are undergoing divergent selection alleles of larger effect cause a larger proportion of the difference when there is migration versus when populations are isolated. When populations have previously diverged and are experiencing stabilizing selection, results suggest that there is little to no effect of migration (Figure 4b).

Statistical properties of QTL effects in a mapping population

The expected magnitude of a QTL allele in a mapping population tends to increase with an increase in the migration rate between populations (Table 5). Contrast this with Table 3 that showed the true average magnitude, based on the alleles that are present in the populations used to form the mapping population, tends to decrease with an increase in the migration rate. Furthermore, the average magnitude of an allele in a mapping population is dependent on the design of a QTL experiment (ie, inbred versus outbred), with the inbred designs having larger average effects.

A stronger effect of migration appears to be on the PVE of alleles in a mapping population. Migration tends to increase the average PVE of an allele in the mapping population, all else being equal (Table 6). The increase is greatest when the divergence time between populations is relatively short (1000 generations) and the populations have not reached their respective optima. For the stabilizing selection case, the increase is marginal to nonexistent for outbred designs and present for an inbred line design. For modest divergence times and/or when the optima of the stabilizing selection functions are not too divergent relative to the average magnitude of a mutation, there is a very good chance that a mapping population will contain a QTL allele with an effect that explains $> 20\%$ of the variance (Table 7).

Fraction of the phenotypic difference explained by alleles present in a mapping population

A key issue for QTL studies that seek to determine the genetic basis of phenotypic differences between two populations is what fraction of the phenotypic difference is explained by the alleles that are present and segregating in the mapping population when populations have diverged under pure divergent directional selection for a short period of time (Figure 5a) and when populations have diverged for a short period of time and subse-

Table 5 The average magnitude of a QTL allele in a mapping population for outbred F_2 (upper) and inbred F_2 designs (lower)[a]

Divergence time	$m/s = 0.0$	$m/s = 0.01$	$m/s = 1.0$
10^3 (pure directional)[b]	0.0202 (0.0193, 0.0212)[c]	0.0207 (0.0198, 0.0216)	0.0210 (0.0200, 0.0220)
	0.0220 (0.0211, 0.0230)	0.0225 (0.0215, 0.0235)	0.0233 (0.0222, 0.0245)
10^4 (stabilizing selection)[d]	0.0122 (0.0114, 0.0130)	0.0130 (0.0122, 0.0139)	0.0155 (0.0146, 0.0164)
	0.0131 (0.0122, 0.0140)	0.0140 (0.0132, 0.0148)	0.0164 (0.0155, 0.0173)

[a]In all cases, $s = E[|\delta|] = 0.01$ and migration was symmetric such that $m = m_{12} = m_{21}$.
[b]$o_1 = 1.0$, $o_2 = -1.0$.
[c]Values in parenthesis beside the mean are 95% confidence intervals of the mean.
[d]$o_1 = 0.1$, $o_2 = -0.1$.

Table 6 The average percent variance explained by a QTL allele in a mapping population for outbred F_2 (upper) and inbred F_2 designs (lower)[a]

Divergence time	$m/s = 0.0$	$m/s = 0.01$	$m/s = 1.0$
10^3 (pure directional)[b]	0.0766 (0.0711, 0.0832)[c]	0.0781 (0.0718, 0.0844)	0.1180 (0.1081, 0.1279)
	0.0953 (0.0883, 0.1023)	0.0974 (0.0900, 0.1050)	0.2030 (0.1884, 0.2188)
10^4 (stabilizing selection)[d]	0.0478 (0.0421, 0.0537)	0.0523 (0.0463, 0.0586)	0.0492 (0.0439, 0.0545)
	0.0534 (0.0478, 0.0602)	0.0618 (0.0559, 0.0684)	0.0794 (0.0717, 0.0878)

[a]In all cases, $s = E[|\delta|] = 0.01$ and migration was symmetric such that $m = m_{12} = m_{21}$.
[b]$o_1 = 1.0$, $o_2 = -1.0$.
[c]Values in parenthesis beside the mean are 95% confidence intervals of the mean.
[d]$o_1 = 0.1$, $o_2 = -0.1$.

Table 7 The probability a QTL mapping population contains an allele that explains more than 20% of the variation for outbred F_2 (upper) and inbred F_2 designs (lower)[a]

Divergence time	$m/s = 0.0$	$m/s = 0.01$	$m/s = 1.0$
10^3 (pure directional)[b]	0.93 (0.91, 0.95)[c]	0.94 (0.92, 0.96)	0.98 (0.98, 1.0)
	0.95 (0.92, 0.96)	0.96 (0.94, 0.97)	0.98 (0.96, 0.99)
10^4 (stabilizing selection)[d]	0.93 (0.88, 0.98)	0.93 (0.88, 0.97)	0.87 (0.82, 0.91)
	0.94 (0.89, 0.98)	0.95 (0.90, 0.99)	0.95 (0.92, 0.98)

[a]In all cases, $s = E[|\delta|] = 0.01$ and migration was symmetric such that $m = m_{12} = m_{21}$.
[b]$o_1 = 1.0$, $o_2 = -1.0$.
[c]Values in parenthesis beside the mean are 95% confidence intervals of the mean.
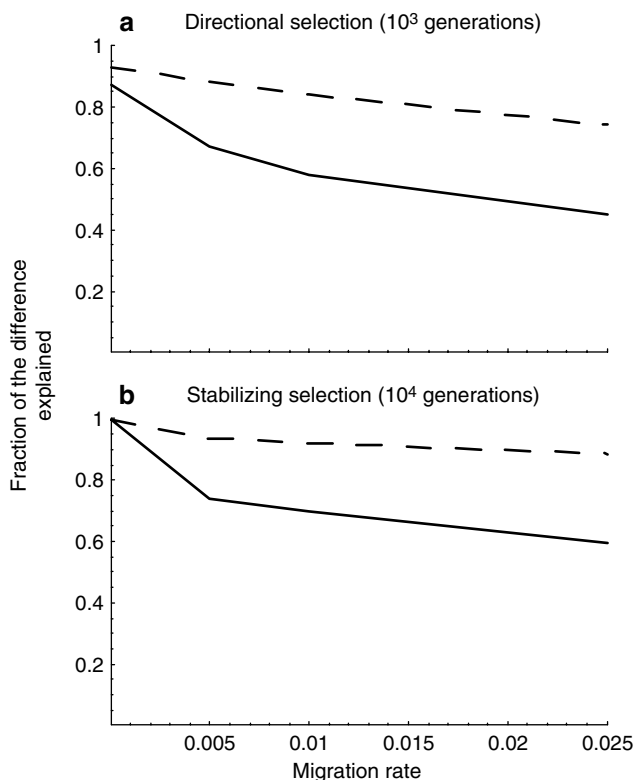[d]$o_1 = 0.1$, $o_2 = -0.1$.

Figure 5 The average fraction of the difference between two individuals randomly chosen from two populations explained by QTLs segregating in the mapping population as a function of migration rate. Solid lines correspond to an inbred line F_2 design and dashed lines to an outbred F_2 design. In the directional selection simulations: $o_1 = 1.0$, $o_2 = -1.0$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. In the stabilizing selection simulations: $o_1 = 0.1$, $o_2 = -0.1$, $\sigma_1 = 1.0$ and $\sigma_2 = 1.0$. The average magnitude of the selection coefficient of a mutation was 0.01. The populations evolved for 1000 generations in the directional selection simulations and for 10 000 generations in the stabilizing selection simulations.

quently experience stabilizing selection (Figure 5b) results show that modest levels of migration cause fairly substantial decreases in the phenotypic difference explained by the alleles present in a mapping population. Inbred line designs have the greatest decrease in the difference explained.

Discussion

This study has shown how migration alters the genetic architecture of a local adaptation. With an increase in migration and when populations are experiencing directional selection, the average magnitude of an allele that contributes to a local adaptation declines, but alleles of larger magnitude tend to cause more of the phenotypic difference. The reason alleles of larger magnitude cause more of the difference when there is migration is that they segregate, on average, at higher frequency than alleles of small magnitude, and there is a greater difference, on average, in the frequencies alleles of large magnitude segregate between populations than alleles of small magnitude. The effect that alleles of larger magnitude cause more of the difference with migration goes away if there has been an extended period of stabilizing selection because alleles of smaller effect are necessary for adaptation near a population's optimum (eg, Orr, 1998).

Predictions for empirical studies
The results of this paper lead to a few predictions that may be tested empirically or help interpret empirical results. One prediction is that along a phenotypic cline brought about by an environmental gradient, the alleles causing proportionally more of the difference between populations should be larger for populations that are geographically close versus far apart (assuming migration rate is inversely proportionally to geographic distance). Owing to the sampling properties of mapping

populations in a QTL study, this pattern should be revealed in QTL studies when the distance between mapping populations varies. The average PVE of an allele should be larger for mapping populations that are close together versus mapping populations that are far apart. If the cline has been stable for a long period of time then results suggest that there should be less of a difference in the average allelic size as the distance between populations varies. The results also lead to, perhaps, a counter-intuitive prediction regarding the genetic basis of differences in ring species. The results suggest that the average magnitude of alleles that cause most of the phenotypic differences between species at the 'ends' of a ring (the species that are reproductively isolated) may be smaller than the alleles bringing about differences between populations within the ring.

Relevance to QTL studies

QTL studies sometimes find alleles with large effects. For instance, Bradshaw *et al* (1998) found 26 alleles that explained more than 20% of the variance for various floral characters in *Mimulus*. In fact, they found effects that explained as much as 50 and 90% of the variation for different characters. Peichel *et al* (2001) found several QTL alleles that explain about 15% of the variation in sticklebacks. Are these effects unusually large?

This article presented theoretical predictions for the expected variance an allele will contribute to a mapping population. Based on the conditions that were simulated in this paper, a QTL allele with a large variance contribution is expected when the divergence time between populations is short or given that the divergence time is long, the local optima between populations are not too far apart such that the populations experience stabilizing selection for most of their history since divergence. If the divergence time has been long, but phenotypic optima are relatively close together, a few alleles of large effect can bring each population near its optimum and these alleles will typically not be subsequently replaced by alleles of smaller effect. Since the few alleles of large effect persist and subsequent adaptation toward the optimum consists of alleles of small effect, the few large alleles each contribute a high variance in a mapping population. It is unexpected to find alleles of large effect when populations have diverged for a long period of time and their phenotypic optima are far apart. Under these conditions, many mutations are necessary to cause the large phenotypic difference between populations. Although there may be large mutations that make up the difference between populations, there are many of these large mutations, such that individually they do not contribute (proportionally) a high variance to a mapping population.

It is important to note that the large variances presented in this article are not artifacts of the Beavis (1994) effect, which is a statistical bias in QTL studies with small mapping populations: the percentage of the variance that is explained by a particular QTL allele tends to be overestimated. This bias arises because of the difficulties of statistically detecting alleles of small effect. It was assumed in this study, however, that the mapping populations were of arbitrarily large size such that all QTL alleles could be detected and the estimates of their effects are unbiased.

When segregating alleles form the basis of differences between populations there is a chance that they will not be sampled when making the mapping population. This paper quantified how the fraction of the phenotypic difference that is explained by alleles that are segregating in a QTL mapping population goes down with increasing rates of migration and that QTL experiments of different design differ in the fraction of the phenotypic difference that is explained. Even with relatively low levels of migration (<0.025), there is a fairly substantial decrease in the fraction of the difference that is explained. Whether it is acceptable that a certain fraction of the difference will not be explained is a matter of choice. If the goal is a complete determination of the genetic basis of the difference between populations that exchange migrants, then outbreeding QTL designs should be used. Outbreeding QTL designs use more individuals to make up the mapping population, and there is consequently a better chance that an allele will be included in the mapping population. It should be highlighted that migration was modeled as a forward migration rate. The realized migration rate – the probability that an individual migrates to the other population and has offspring that survive and reproduce – is even lower because migrants tend not to be locally adapted.

The results of this paper suggest that results from a QTL study are a compromise between two processes that occur as the migration rates between populations increase. As the migration rates increase, the average magnitude of an allele that causes a difference between populations declines, yet alleles of larger magnitude explain more of the difference (under certain conditions). In a QTL mapping population, both the average magnitude of an allele and the average PVE tends to increase with migration. The increase for QTL studies is because alleles are randomly sampled from a population according to their frequency and alleles of larger magnitude segregate at higher frequency and are more likely to be sampled.

Conclusions

When two locally adapted populations exchanged migrants, the average effect of an allele that causes most of the phenotypic difference tends to increase compared to when populations are isolated. With migration, alleles that are beneficial and of larger effect tend to segregate, on average, at higher frequency in a population, accordingly they are more likely to be sampled when forming a QTL mapping population relative to alleles of smaller effect. Consequently, there is a bias, such that, the average magnitude of an allele in a mapping population is larger than the average in the natural populations used to form the mapping population.

Acknowledgements

References

Barton NH (1983). Multilocus clines. *Evolution* **37**: 454–471.

Beavis WD (1994). The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Wilkinson DB (ed.) *49th Annual Corn and Sorghum Research Conference*. American Seed Trade Association: Washington, DC. pp 250–266.

Bradshaw HD, Otto KG, Frewen BE, McKay JK, Schemske DW (1998). Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* **149**: 367–382.

Bulmer MG (1972). Multiple niche polymorphism. *Am Nat* **106**: 254–257.

Crow J, Kimura M (1970). *An Introduction to Population Genetics Theory*. Burgess Publisher Co.: Minneapolis, MN.

Felsenstein J (1976). The theoretical populations genetics of variable selection and migration. *Annu Rev Genet* **10**: 253–280.

Hawthorne DJ, Via S (2001). Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* **412**: 904–907.

Hendry AP, Day T, Taylor EB (2001). Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution* **55**: 459–466.

Karlin S (1982). Classification of selection-migration structures and conditions for a protected polymorphism. *Evol Biol* **3**: 186–209.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B *et al* (2002). A high-resolution recombination map of the human genome. *Nature Genetics* **31**: 241–247.

Levene H (1953). Genetic equilibrium when more than one ecological niche is available. *Am Nat* **87**: 331–333.

Lynch M, Walsh B (1998). *Genetic Analysis of Quantitative Traits*. Sinauer Assoc. Inc.: Sunderland, MA.

Orr HA (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**: 935–949.

Peichel CL, Nereng KS, Ohgi KA, Cole BLE, Colosimo PF, Buerkle CA *et al* (2001). The genetic architecture of divergence between threespine sticklebacks species. *Nature* **414**: 901–905.

Peters AD, Halligan DL, Whitlock MC, Keightley PD (2003). Dominance and overdominance of mildly deleterious induced mutations for fitness traits in *Caenorhabditis elegans*. *Genetics* **165**: 589–599.

Slatkin M (1978). Spatial patterns in the distributions of polygenic characters. *J Theor Biol* **70**: 213–228.

Wright S (1940). Breeding structure of populations in relation to speciation. *Am Nat* **74**: 232–248.