vol. 167, no. 4 the american naturalist april 2006

Geographical Variation in Selection, from Phenotypes to Molecules

John K. Kelly*

Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045

Submitted June 6, 2005; Accepted October 24, 2005; Electronically published March 8, 2006

Online enhancements: Excel workbook, tab-delineated ASCII data files.

ABSTRACT: Molecular technologies now allow researchers to isolate quantitative trait loci (QTLs) and measure patterns of gene sequence variation within chromosomal regions containing important polymorphisms. I develop a simulation model to investigate gene sequence evolution within genomic regions that harbor QTLs. The QTLs influence a trait experiencing geographical variation in selection, which is common in nature and produces obvious differentiation at the phenotypic level. Counter to expectations, the simulations suggest that selection can substantially affect quantitative genetic variation without altering the amount and pattern of molecular variation at sites closely linked to the QTLs. Even with large samples of gene sequences, the likelihood of rejecting neutrality is often low. The exception is situations where strong selection is combined with low migration among demes, conditions that may be common in many plant species. The results have implications for gene sequence surveys and, perhaps more generally, for interpreting the apparently weak connection between levels of molecular and quantitative trait variation within species.

Keywords: balancing selection, neutrality tests, quantitative trait loci, Tajima's D, Z_{nS}.

Natural populations exhibit a great deal of genetic variation at both molecular and phenotypic levels. The amount of variation reflects a balance of evolutionary forces: mutation, selection, migration, and genetic drift. Among these forces, natural selection is the most difficult to classify in terms of its net effect on the amount of genetic variation. Selection may be purifying, eliminating variation that is continually reintroduced by mutation or gene flow. It may be balancing, actively maintaining alternative alleles within the population. Finally, selection may have no effect on allele frequencies, a hypothesis commonly known as selective neutrality (Kimura 1983).

Neutrality frequently serves as the null hypothesis in studies of molecular variation (Tajima 1989; Kreitman 1996). Estimates of sequence polymorphism, linkage disequilibria, and divergence are evaluated relative to the likely range of outcomes under selective neutrality. Different forms of selection produce different sorts of deviation from neutral expectations. Purifying selection tends to reduce polymorphism and may skew the frequency spectrum toward an abundance of rare alleles (Charlesworth et al. 1993; Williamson and Orive 2002). The rapid fixation of an advantageous mutation can have similar effects on sequence variation at linked sites (Kaplan et al. 1989; Stephan et al. 1992; Aquadro 1997). In contrast, balancing selection should increase polymorphism and can generate a distinct "haplotype structure" within a population. Selectively maintained alleles will accumulate genetic differences at linked sites if those alleles are maintained for sufficient lengths of evolutionary time (Hudson and Kaplan 1988; Stadler and Delph 2002; Tian et al. 2002).

The neutrality hypothesis is less prominent in studies of variation at the phenotypic scale (but see Lande 1976, 1979; Lynch and Hill 1986). In part, this reflects uncertainty about the distribution of mutational effects on quantitative traits. However, neutrality is neglected more frequently because there is little doubt that selection is important at the phenotypic scale. Field studies routinely reveal significant associations between trait values and fitness components, even with small sample sizes (Endler 1986; Kingsolver et al. 2001). Genetic differences underpinning variation in characters with demonstrable effects on survival and reproduction are not likely to be selectively neutral.

A range of different selection regimes can maintain quantitative trait variation, for example, heterozygote advantage, frequency-dependent selection, temporal varia-

* E-mail: jkk@ku.edu.

Am. Nat. 2006. Vol. 167, pp. 481–495.

tion in selection, and genotype-environment interaction. However, the form of balancing selection most clearly demonstrated in experimental studies is spatial or geographic variation in selection. In a wide range of animal and plant species, phenotypic differences among populations reflect local adaptation and are therefore maintained by selection (Stebbins 1950; Mayr 1963; Endler 1977; Lewontin et al. 1981). A classic example is depicted in figure 1. Using a common garden experiment, Clausen et al. (1948) demonstrated the genetic basis of the striking geographical variation in morphology of *Achillea lanulosa*, which is likely a result of selection on variable climatic conditions. Scale is an important factor when considering the effect of selection in a spatially structured species. For a species like *A. lanulosa*, selection may be purifying within local populations (favoring multilocus genotypes closest to the local optimum). However, the persistence of distinct ecotypes that remain reproductively compatible generally implies balancing selection at the scale of the entire species. Spatial structure is obvious in some cases (e.g., fig. 1), but it may also be cryptic and able to exist on very small spatial scales. Local adaptation has been demonstrated at the scale of a few meters in plant populations (e.g., Schmitt and Gamble 1990).

This article examines the implications of selection at the phenotypic level for gene sequence evolution. What patterns of sequence variation do we expect in genomic regions that harbor quantitative trait loci (QTLs)? Imagine a QTL subject to geographical variation in selection, for example, a locus affecting growth form in *A. lanulosa* (fig. 1). We collect tissue from a large sample of individuals and sequence the genomic region containing the QTL. If we then apply standard molecular population genetic tests to the sequence data, are these tests likely to yield significant evidence of selection? I address this question by modeling neutral molecular evolution in sequences that flank QTLs. Patterns of variation are thus determined by mutation-drift balance, coupled with linkage to the QTLs under selection. Until recently, such a modeling exercise would have no direct application. However, advances in

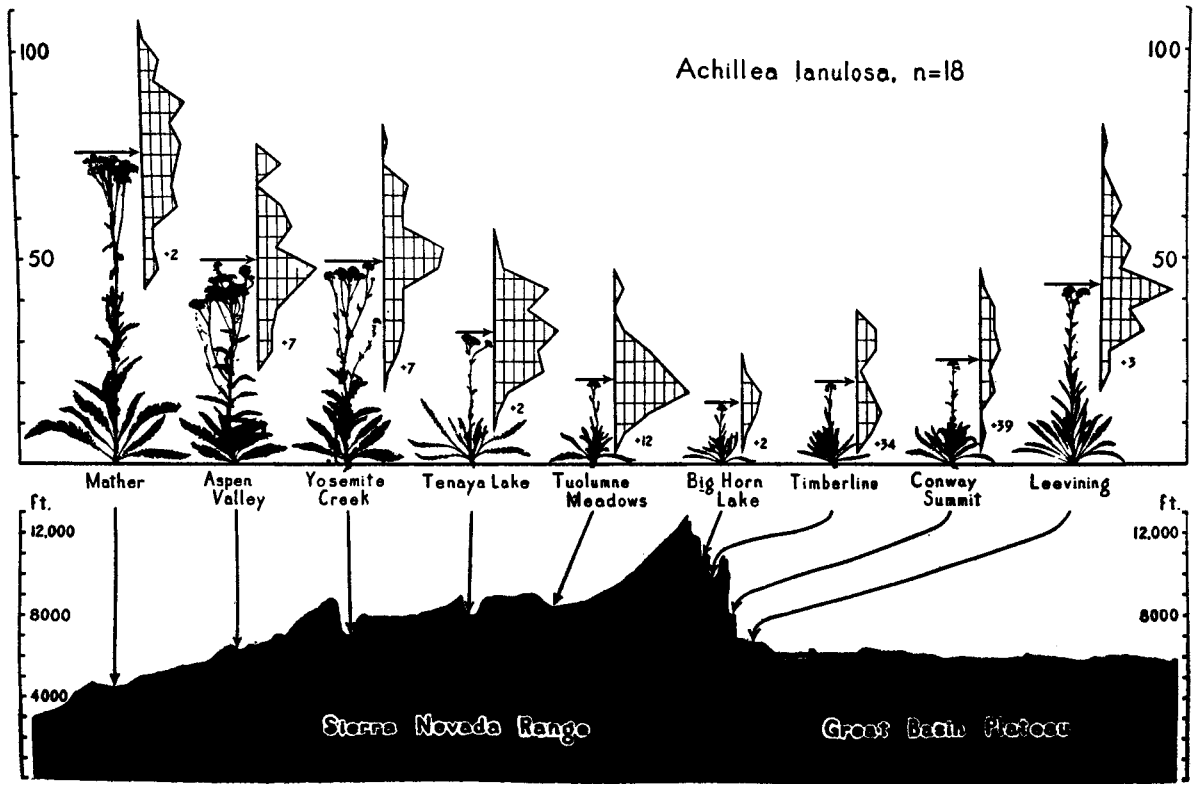

Figure 1: Morphological variation among populations of *Achillia lanulosa* as a function of geographical location. Each population in the upper panel is represented by a typical individual. The frequency distributions of plant height (in cm) within each population are based on measurements from approximately 60 individuals, with the arrow denoting the population mean. The numbers to the right of each distribution denote the number of nonflowering plants. The geographic location and elevation (in feet) are given in the lower panel. Figure is reprinted with permission from Clausen et al. (1948).

genomics now allow researchers to isolate QTLs and thus directly evaluate these predictions (Mackay 2004; see "Discussion").

The Model

The simulation consists of three basic parts. The demographic submodel describes the population size, its distribution, and the life cycle of the organism. The latter subsumes the evolutionary processes of migration, genetic drift, and natural selection. The quantitative genetic submodel describes the genetic underpinning of the character. The fitness of organism, which is assayed in the demographic submodel, is determined partly by this character value. The molecular submodel describes gene sequence evolution within flanking regions of 10 QTLs. The two genetic submodels contain the mutational processes at the QTL and sequence levels, respectively. Table 1 provides a summary of parameters.

The Demographic Submodel

The total population consists of N diploid individuals, which I initially assume to be distributed into two "demes" of equal size. There is symmetric migration between demes (characterized by the migration rate m), and each is subject to selection toward a deme-specific optimum. At the beginning of a generation, each zygote is formed from two gametes, either of which may be local (with probability $1 - m$) or migrant (with probability m). Given the ancestral deme, the parent is selected according to probabilities determined by the relative fitness of individuals. Each deme has a distinct phenotypic optimum, Z_i^*, and the fitness function (within demes) is Gaussian. For an individual with phenotypic value x in deme i, its fitness is

$$W_x = \exp\left[\frac{-(x - Z_i^*)^2}{V_S}\right], \quad (1)$$

where V_S is the variance of the fitness function within a deme (Lande 1975).

The selection regime is defined by the amount of difference between deme optima ($Z_1^* - Z_2^*$) and the rate at which fitness declines as individuals deviate from the (local) optimum. Selection becomes weaker as V_S increases. For $V_S \gg V_E$ (see "The Quantitative Genetic Submodel" for a definition of V_E), the quantitative trait is effectively neutral, and I use this special case to confirm the validity of the molecular evolution model (see below). This scheme represents a case of "soft selection" because individuals are competing against other deme members and the production of migrants is independent of deme mean fitnesses. The evolutionary processes of genetic drift and migration are essentially contained within the gametic selection regime of the simulation model.

Table 1: Summary of model parameters

Parameter	Definition
L	Number of QTLs
δ	Magnitude of allelic effects at QTL
μ	Mutation rate (per allele) at each QTL
u	Total mutation rate (summed across sites) in flanking region of each QTL
V_E	Environmental variance
Z_i^*	Optimum phenotypic value in deme i
V_S	Variance of selection surface around deme-specific optima
m	Migration rate among demes
N	Diploid population size (including all demes)

Note: QTL = quantitative trait locus.

The Quantitative Genetic Submodel

I assume that L loci contribute additively to variation in a quantitative trait. At each QTL, two alleles are possible (the high allele and the low allele). Each high allele increases the phenotypic value by an amount δ, while the low allele reduces the phenotype by an equivalent amount (heterozygotes make no net contribution). The genotypic value is the sum of allelic contributions across QTLs. The minimum genotypic value is $-2L\delta$ (for individuals that are homozygous low across QTLs), and the maximum is $2L\delta$ (for individuals that are fully homozygous for high alleles). The phenotypic value of an individual is equal to its genotypic value plus an environmental effect. The environmental effect is drawn from a normal distribution with mean 0 and variance V_E. A gamete is formed by randomly selecting one allele at each QTL from the parental genome. I assume that all QTLs are unlinked, so sampling is independent for each locus. Mutation may occur at this stage; alleles mutate to the alternative form with probability μ.

For each parameter combination, I compare the observed genetic variance in the quantitative trait (V_G) with that expected under neutrality. Let V_n denote the expected genetic variance under mutation-drift balance (without selection). I classify selection at the phenotypic scale as "purifying" if the average V_G from a simulation is less than V_n and as "balancing" if it is greater than V_n. An equation for V_n is obtained by noting that, under the assumptions outlined above, the genetic variance contributed by a QTL is $8p(1 - p)\delta^2$, where p is the frequency of the high allele (Falconer and Mackay 1996). Since QTLs are unlinked, it

is reasonable to assume that loci will be in linkage equilibrium if the trait is neutral (though not necessarily with selection). The expected genetic variance, $E(V_G)$, thus equals $8L\delta^2 E[p(1-p)]$. The probability density function for p under neutrality is a special case of Wright's formula (Wright 1931, 1937). Integrating over this distribution, we find that $E[p(1-p)] = 2N\mu/(1+8N\mu)$ and

$$V_n = L\delta^2\left(\frac{16N\mu}{1+8N\mu}\right). \quad (2)$$

With trait minima and maxima of -10 and 10, respectively, $\delta^2 = 25/L^2$ and $V_n = 400N\mu/[L(1+8N\mu)]$.

The Molecular Submodel

Molecular evolution is tracked within the flanking regions of 10 QTLs. Each flanking region is a nonrecombining sequence of nucleotides subject to infinite site mutation (Kimura 1983). Novel mutations are introduced at rate u. The quantity u is essentially the product of the neutral mutation rate per site times the number of sites in the flanking region. Thus, variation in the length of flanking sequences is absorbed into u. I assume that there is no recombination between the flanking region and QTLs (see "Discussion" for a comment on this assumption). The simulation tracks the identity of all mutations within the flanking region of each allele, as long as those mutations remain polymorphic within the population as a whole.

Parameter Values, Sampling Regime, and Molecular Statistics

A simulation run consists of 11 successive intervals, each 10^5 generations in duration. The optimum phenotype is the same in each deme during the first interval ($Z_1^* = Z_2^* = 0$). With sufficient migration among demes, this is effectively equivalent to stabilizing selection on a single deme (but see Goldstein and Holsinger 1992) and thus provides an opportunity to consider the form of purifying selection most frequently considered in theoretical studies. The optima of each niche progressively diverge in subsequent intervals, increasingly negative in niche 1 and increasingly positive in niche 2. The difference between optima increases by a single unit with each interval: Z_1^* and Z_2^* are $-1/2$ and $1/2$ in the second interval, -1 and 1 in the third interval, and eventually, -5 and 5 in the final interval. In generation 0, each individual is heterozygous at each QTL. Flanking regions of QTLs are free of neutral mutations and thus monomorphic within the population. Five independent simulation runs were performed for each parameter set.

In analyzing the model, I vary the number of QTLs while holding the minimum genotypic value to -10 and the maximum to 10. This implies that $\delta = 10/2L$. In addition, I set $V_E = 1$ throughout, as it is typical in quantitative genetics to standardize variation in both phenotype and fitness relative to V_E (e.g., Lynch 1988). The implication of these assumptions is that the maximum difference in phenotypic optima between demes is $10V_E$ and the maximum possible range of genotypic values is $20V_E$.

For all simulations described in this section, I assume $\mu = 0.01/(2L)$, $u = 10^{-3}$, and $N = 10^3$. The value for μ is based on empirical estimates of 10^{-2} for the total mutation rate affecting quantitative traits (Turelli 1984; see also Lynch and Walsh 1998, pp. 337–339). For this model, the total mutation rate for the trait is $2L\mu$. Regarding the other parameters, N was made as large as practical (given current technology) and u as small as practical, so that the product $Nu = 1$. While N is likely to be substantially greater than 10^3 for most natural species and u substantially less than 10^{-3} (unless a very long flanking sequence is considered), the amount of variation under neutrality is determined primarily by the product Nu (Kimura 1983; Hudson 1990). With $Nu = 1$, the simulations yield polymorphism levels comparable to those in empirical studies (see below).

Within a simulation run, the distribution of variation at QTLs equilibrates quite quickly after each change in fitness optima (the deme means and variances). However, molecular evolution requires more time to reach statistical equilibrium. For this reason, I do not begin sampling to estimate molecular variation until 10^4 generations into each interval (10 times the population size). After the "burn-in" within each interval, I sample 100 individuals randomly (without regard to deme) every 200 generations. The mutational composition of each flanking sequence is determined for a single haplotype of each sampled individual at 10 QTLs. For each QTL, I determine the number of polymorphic sites (S) and average pairwise difference (π) among sequences in the sample. I also calculate the neutrality test statistics D (Tajima 1989) and Z_{nS} (Kelly 1997) and determine whether these values are statistically significant. The critical values for the significance tests are conditioned on S for the sample (Hudson 1993; Simonsen et al. 1995; Kelly 1997). Concurrent with analysis of the flanking sequences, I calculate the mean and genetic variance of the full population and the means and variances (genotypic and phenotypic) within each deme. Finally, the program tracks the age of all neutral mutations within flanking regions. While these ages are not directly measurable in empirical studies, they are useful for interpreting the simulation results (see "Discussion").

The results from all 10 QTL flanking regions are distilled into the averages for S and π and the fraction of loci

yielding significant tests for D and Z_{nS}. Successive samples within a simulation run provide a series of unbiased but nonindependent estimates for the probability of a significant test result. They are based on distinct samples but from populations with a common evolutionary history. Thus, I average all estimates within each interval of a simulation run. Five independent simulation runs are conducted for each parameter set, and the set of run-specific averages provides a valid standard error for estimates. In most cases, estimates were very similar across replicates.

The values of molecular statistics are compared with their respective expectations under neutral evolution within an unstructured (panmictic) population. Given $Nu = 1.0$ and a sample size of 100 sequences, the expected value for π is 4.0, and the expected value for S is 20.7 (Watterson 1975; Tajima and Nei 1983; Hudson 1990). Simulations without selection ($Z_1^* = Z_2^* = 0$, $V_S = 10^8$) or population structure ($m = 0.5$) confirm these expectations. In these simulations, the average percentage of loci yielding significant tests for D and Z_{nS} is approximately 5%, which is expected, given that test-critical values are based on a Type I error rate of 5%.

Results

In exploring the parameter space, I consider all factorial combinations of three different QTL numbers ($L = 10$, 20, and 50), three different migration rates (low: $m = 0.01$; moderate: $m = 0.1$; and high: $m = 0.5$), three different selection intensities (weak: $V_S = 20$; moderate: $V_S = 10$; and strong: $V_S = 2$), and 11 different levels of divergence between deme optima ($Z_2^* - Z_1^* = 0, 1, \ldots, 10$). Biologically, the migration rate depends not only on the mobility of the organism (or its gametes) but also on the extent to which demes are spatially distinct. The high-migration cases include situations when the different habitats are spatially interspersed, for example, different host plants occurring within the range of a herbivorous insect. The simulations with $m = 0.5$ essentially describe disruptive selection within a single deme.

The fraction of significant tests for D and Z_{nS} and the average values for S, π, the age of neutral mutations, V_G, and the difference between the phenotypic means in demes 1 and 2 are given for each of these parameter combinations in appendix 1 in the online edition of the *American Naturalist*. Across parameter sets, there are strong positive correlations between the two measures of polymorphism (S and π) and between the frequencies of significant tests from D and Z_{nS} (fig. 2). The Z_{nS} statistic invariably yields a higher fraction of significant tests than D when $V_G/V_n > 1$. For this reason, figures 3–5 plot Z_{nS}, although the comparable graphs for D are similar in appearance (see online appendix).

Each trajectory in figure 3 represents the sequence of increasing values for divergence between deme optima ($Z_2^* - Z_1^*$) within a simulation run. The leftmost point corresponds to $Z_2^* - Z_1^* = 0$, whereas the rightmost point is for $Z_2^* - Z_1^* = 10$. The fraction of Z_{nS} values that are significant (rejecting neutrality) is given as a function of the average V_G/V_n for that parameter set. Selection is purifying for points left of the vertical line within each panel ($V_G < V_n$), as is typically the case with low divergence among deme optima. Where selection is purifying, molecular evolution in flanking regions is indistinguishable from neutrality: about 5% of tests for both D and Z_{nS} are significant, and average values for S and π are close to their neutral expectations (fig. 3; app. 1).

As $Z_2^* - Z_1^*$ increases, V_G increases and selection becomes balancing ($V_G > V_n$). With balancing selection, neutrality is rejected more frequently as V_G/V_n increases (fig. 3), although the rate of increase is rather slow and the relationship is often not monotonic. Substantially higher values for V_G/V_n obtain with $L = 50$ than with $L = 10$, primarily because V_n is greater with $L = 10$ (note the difference in the scale of the X-axis from fig. 3*A* and 3*C* to fig. 3*B* and 3*D*). The only cases where frequency of significant tests exceeds 50% involve high divergence between optima, intense selection ($V_S = 2$), and low migration ($m = 0.01$). In these cases, the low migration rate is still sufficiently high (averaging 10 migrants per generation) to prevent substantial differentiation by genetic drift alone.

Asymmetric Migration

A second set of simulations considers a "source-sink" migration scheme: individuals can migrate from deme 1 to deme 2 but not vice versa. The parameter m now denotes the fraction of gametes in deme 2 that are immigrants. This change in the pattern of gene flow affects both the maintenance of quantitative trait variation (the value of V_G/V_n, given the selection scheme) and molecular evolution in flanking regions (the power of D and Z_{nS}, given V_G/V_n). A full summary of results is given in appendix 2 in the online edition of the *American Naturalist*, excepting cases with $m = 0.5$.

As with symmetric migration, there are strong correlations between test results for D and Z_{nS} and between average values of S and π across parameter sets. In all cases with moderate to high divergence among optima, selection maintains substantial quantitative trait variation (fig. 4). However, for a given value of V_G/V_n, the power to detect selection in flanking regions is generally lower with asymmetric migration than in equivalent parameter combinations with symmetric migration (compare panels of figs. 3 and 4). As previously, the fraction of significant tests

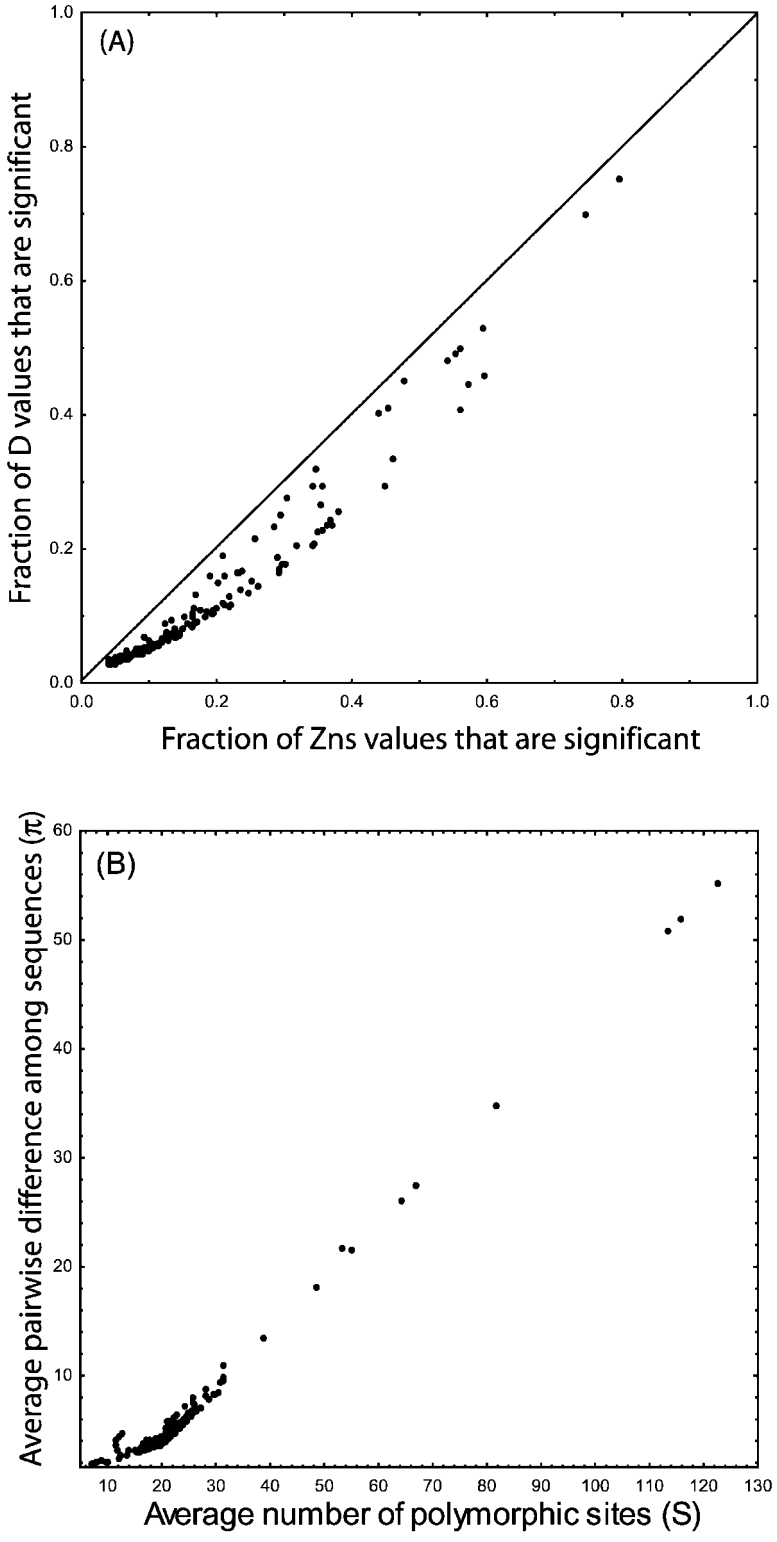

Figure 2: Relationship between molecular statistics across parameter sets of the two-deme, symmetric migration model. *A*, The fraction of values that are significant (reject neutrality) of Tajima's *D* versus Z_{ns} for the same parameter combinations; *B*, π versus *S* for the same parameter combinations.

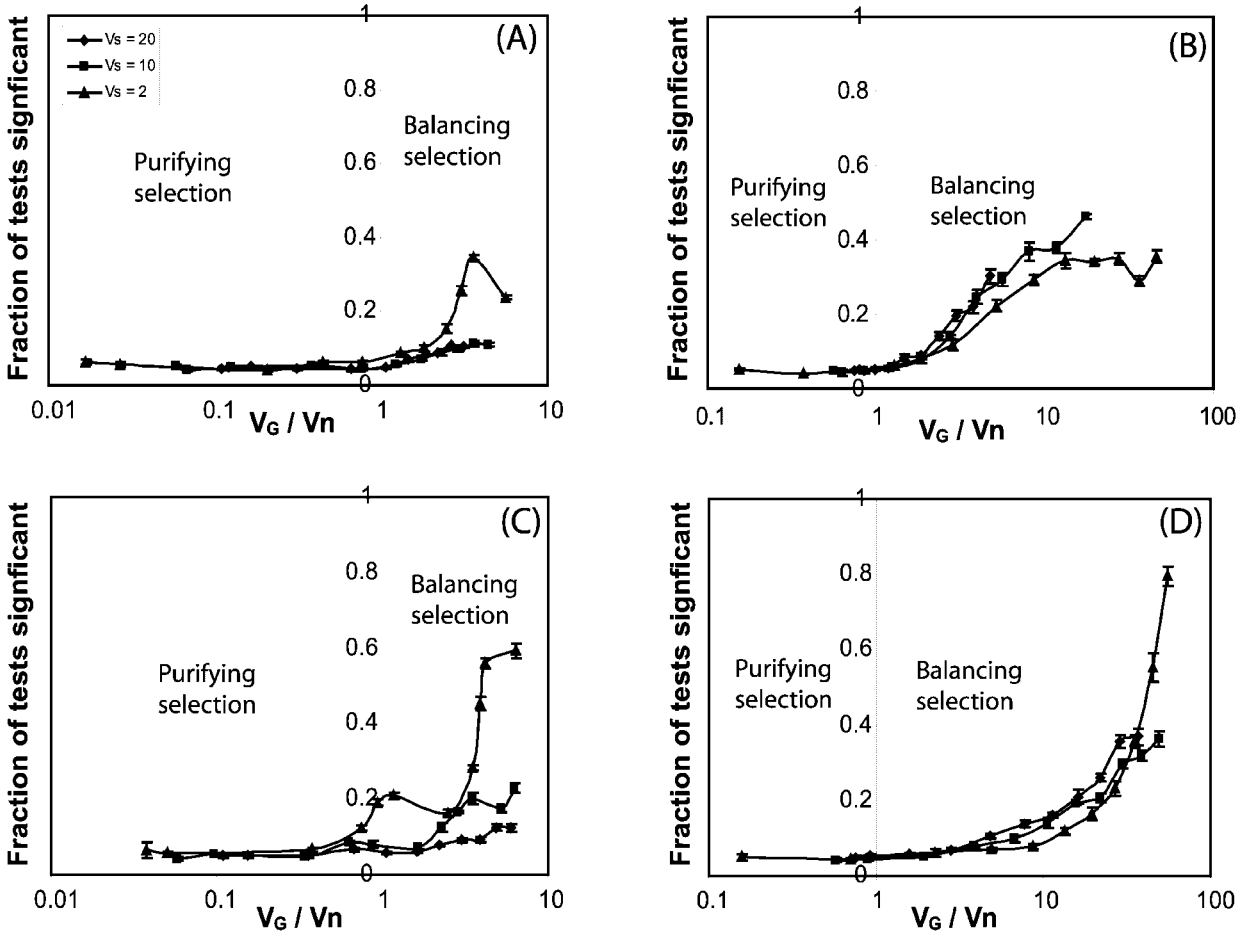

Figure 3: Fraction of Z_{nS} values that reject neutrality as a function of V_G/V_n for different parameter combinations of the two-deme, symmetric migration model. *A*, $L = 10$, $m = 0.1$; *B*, $L = 50$, $m = 0.1$; *C*, $L = 10$, $m = 0.01$; *D*, $L = 50$, $m = 0.01$. The three trajectories within each panel correspond to different intensities of selection: circles for $V_S = 20$, squares for $V_S = 10$, and triangles for $V_S = 2$. Each trajectory is composed of 11 points, each based on a different value for the difference between deme optima (see text). The error bars around each point denote ± 1 SE. In most cases, the confidence band is very small and obscured by the point.

exceeds 50% only with intense selection and low migration between demes.

Three Demes

A third set of simulations considers the same total population size split into three demes of equal size ($N/3$). As the optimum phenotype for demes 1 and 3 diverge, I assume that the optimum for deme 2 remains intermediate ($Z_2^* = 0$). Migration is symmetric among demes, and m denotes the probability that a gamete is an immigrant. Any immigrant is equally likely to come from either of the other two demes. As with asymmetric migration, increasing the number of demes does not prevent selection from maintaining high levels of quantitative trait variation (fig. 5). As the optima diverge, the phenotypic distributions of each deme typically become distinct (an example is given in fig. 6). Again, however, for a given V_G/V_n, the power to detect selection in flanking sequences is usually lower with three demes than with two (compare panels in figs. 3 and 5). The fraction of significant tests is below 50% even with intense selection and low migration among demes. As discussed in greater detail below, appendix 3 in the online edition of the *American Naturalist* also contains a set of simulation results with lower mutations rates at QTLs ($\mu = 10^{-6}$) and lower migration rates among demes ($m = 0.001$).

Discussion

This study investigates gene sequence evolution within genomic regions that harbor important loci, those respon-

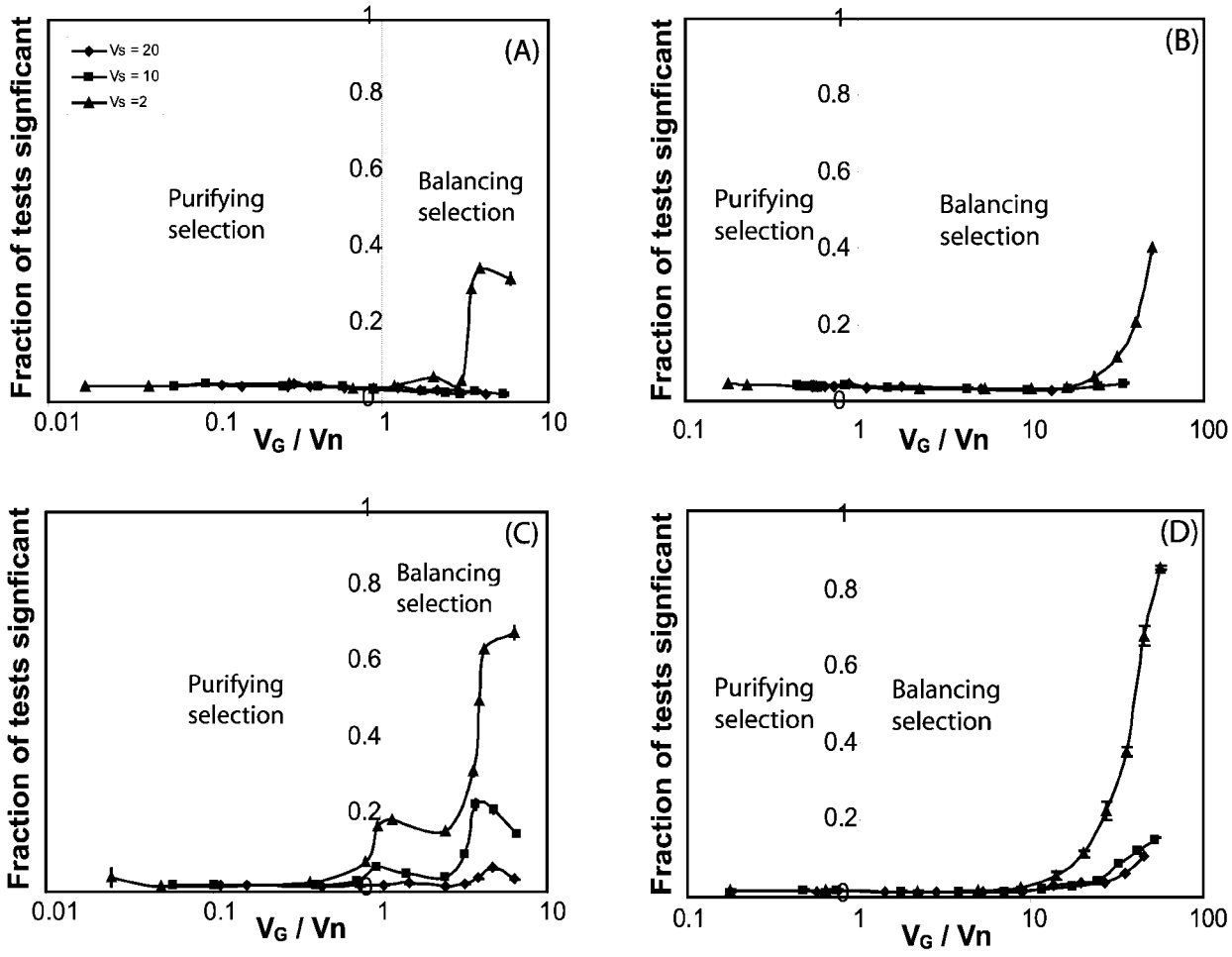

Figure 4: Fraction of Z_{nS} values that reject neutrality as a function of V_G/V_n for different parameter combinations of the two-deme model with asymmetric migration. Parameter combinations and symbols are the same as for the corresponding panels of figure 3.

sible for geographical variation in phenotype. The model has a large parameter space, and different outcomes obtain in different regions of this space. In most cases, however, selection substantially affects levels of quantitative genetic variation without leaving a pronounced signature in the patterns of molecular variation (app. 1; figs. 3–5). Gene sequences linked to QTLs may not be very different from sequences undergoing neutral evolution, at least in terms of polymorphism levels or the values for neutrality test statistics. A simple survey of gene sequences may thus greatly underestimate the frequency of selection at the phenotypic scale. The results may also bear on the noted lack of correspondence between levels of molecular and quantitative trait variation within species (Pfrender et al. 2000).

Selection is acting on QTLs across the full range of parameter sets. As a consequence, "fraction of significant tests" in figures 2–5 essentially estimates the statistical power of D and Z_{nS}. Power is the probability of rejecting the null hypothesis (neutrality) when it is false. However, in contrast to most statistical situations, the low power of D and Z_{nS} is not due to deficiencies of the experimental design, for example, inadequate sample sizes. Both statistics are strongly correlated with the average age of neutral mutations across parameter sets (fig. 7) and are thus measuring the intended signal (as described in the introduction to this article). Power is low because phenotypic selection does not substantially affect the age distribution of neutral mutations in most cases. In other words, the simulated selection regime simply did not produce the signal these statistics were devised to detect.

Power estimates for D and Z_{nS} that exceed 50% are observed only in simulations with intense selection ($V_S = 2$) and substantial separation of fitness optima between demes ($Z_2^* - Z_1^* \geq 8$; see app. 1). In terms of

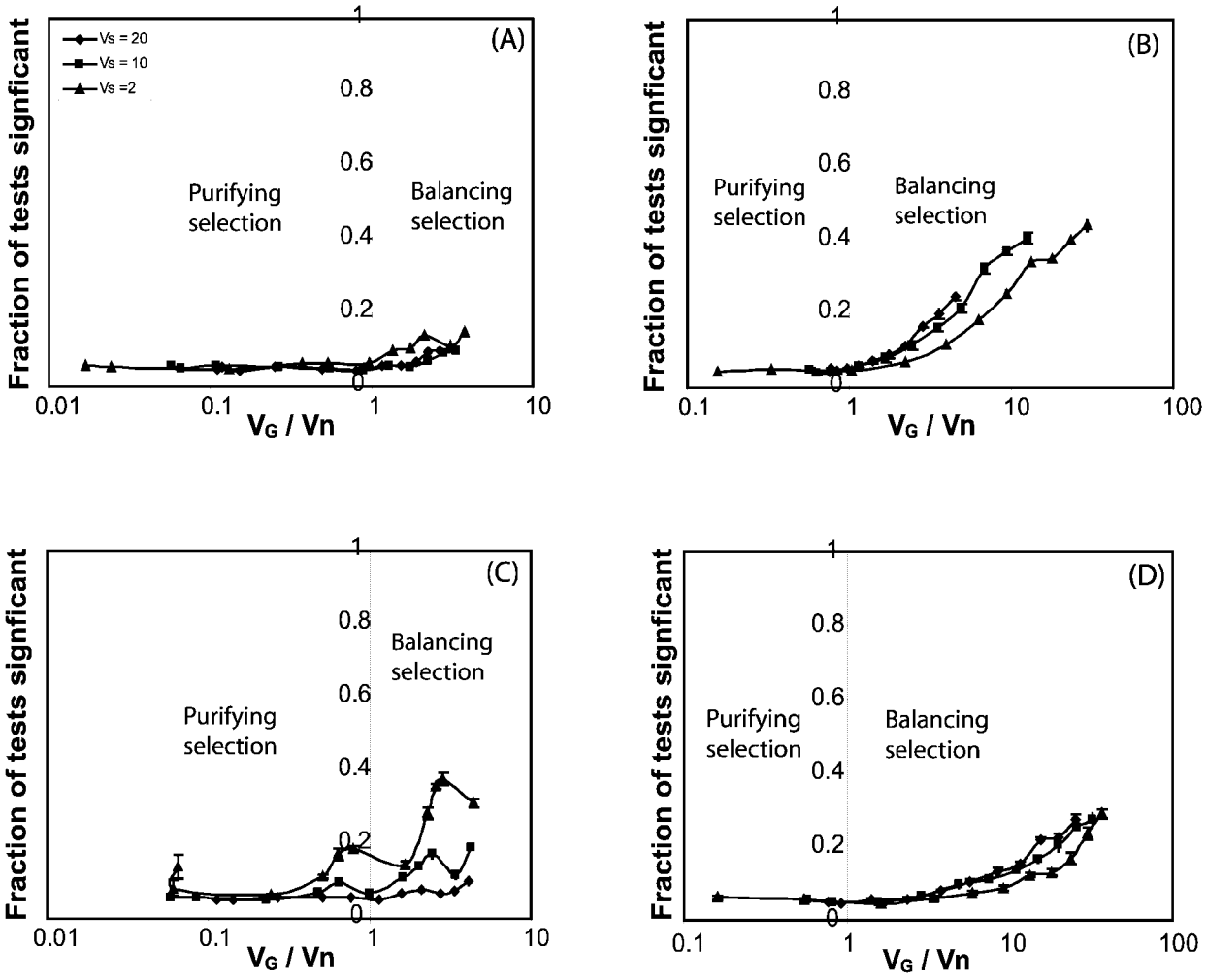

Figure 5: Fraction of Z_{nS} values that reject neutrality as a function of V_G/V_n for different parameter combinations of the three-deme model. Parameter combinations and symbols are the same as for the corresponding panels of figure 3.

Wright's adaptive topography, this represents a fitness surface with two (or more) steep peaks separated by deep, broad valleys. For example, in the two-deme model with $Z_2^* - Z_1^* = 10$ and $V_S = 2$, the fitness of individuals with phenotypic values between -2 and 2 is about 1% (or less) that of individuals close to the optima of -5 or 5, regardless of where they reside. Selection may be this strong in some cases, but it can be far weaker and still maintain large amounts of quantitative trait variation.

The low power of neutrality tests is surprising, given that, in most regards, the conditions of the simulations are favorable to detecting selection. First, the sample sizes prescribed in the simulations (100 sequences) are larger than those in most empirical studies. Second, I assume that there is no recombination within flanking regions or between QTLs and flanking regions. Either should reduce the signal of selection, although statistical tests can be adjusted to account for recombination within the flanking region (e.g., Filatov and Charlesworth 1999). However, recombination between QTL and flanking region decouples the evolutionary dynamics of selected and neutral variation, greatly reducing any effect of hitchhiking (Hudson and Kaplan 1988). The simulations also neglect gene conversion, which can further erase any signature of selection (Andolfatto and Nordborg 1998).

Perhaps most favorably to detecting selection, I assume that the selection regime is temporally stable over many thousands of generations. In nature, selection is likely to change, and over geological time, high or low trait values may periodically be favored over the entire species range. This will likely cause a rapid decay of variation at both molecular and quantitative trait levels. However, quanti-

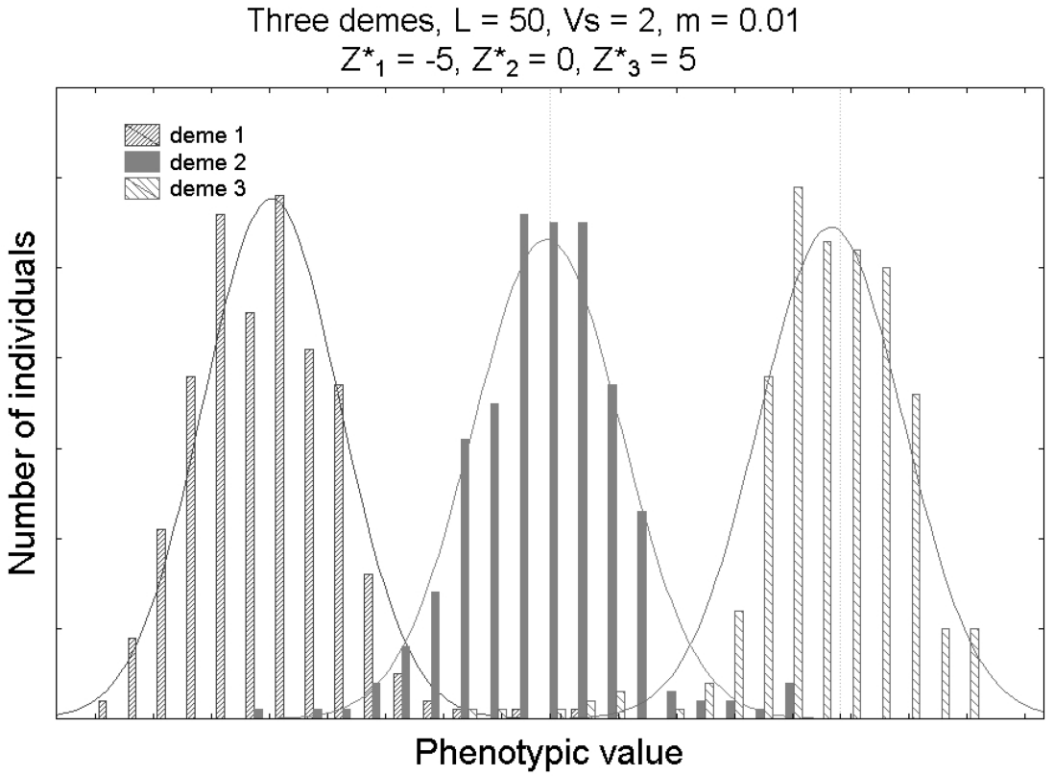

Figure 6: Typical distribution of phenotypic values is illustrated for simulations of the three-deme model. Parameter values are given in the figure.

tative genetic variation will recover much more rapidly than sequence variation once a multioptimum fitness regime is reestablished. This is a consequence of the higher cumulative mutation rate for quantitative trait variation than for nucleotide changes within flanking regions.

Three variants of the model population structure were considered, the simplest allowing a symmetrical exchange of migrants between two demes (app. 1). This model was generalized first to allow asymmetric migration (app. 2) and then to include a third deme with an intermediate optimum (app. 3). In detail, the effects of each generalization are complicated, but it is fair to say that neither asymmetric migration nor an additional deme is generally favorable to detecting selection. Across all three sets of simulations, the estimated power of Z_{nS} is generally greater than that of D. However, the importance of this result should not be overstated. With smaller sample sizes, or if recombination occurs within the flanking regions, the power of D (or alternative abstractions of the mutant frequency distribution; e.g., Fu and Li 1993) may be greater than that of Z_{nS}. The important point is that power of either statistic is likely to be lower than the estimates presented here.

QTL Mutation Rates and Genetic Redundancy

Balancing selection at the molecular level can maintain alleles within a population for long stretches of evolutionary time. There are at least two reasons why neutral mutations are unexpectedly "young" even when selection maintains large amounts of quantitative trait variation. The first is the relatively high rate of mutation between alternative alleles at QTLs. This allows "migration" of neutral alleles between genetic backgrounds, where backgrounds are defined by the identity of selectively maintained alleles (see Strobeck 1983; Hudson and Kaplan 1988; Kelly and Wade 2000). Like recombination, this at least partially decouples the fate of QTL alleles and the neutral variation in flanking regions. Neutral mutations can drift to fixation or loss in the population as a whole even if initially fixed within the collection of haplotypes that harbor a particular QTL allele. In the bulk of the simulations, I assume $\mu = 0.01/2L$, a value based on empirical estimates of 10^{-2} for the total mutation rate affecting quantitative traits (Turelli 1984). Unless L, the number of loci affecting the trait, is very large, this implies a rather high mutation rate per locus. With $L = 10$,

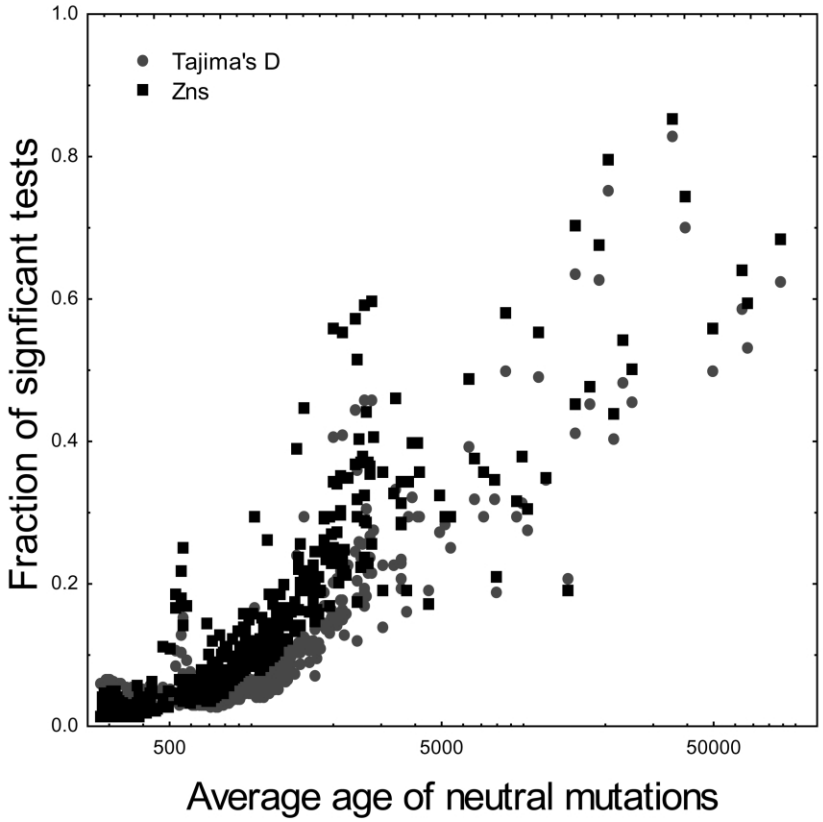

Figure 7: Estimated power of D (*circles*) and Z_{nS} (*squares*) as a function of the average age of neutral mutations for each parameter set.

$\mu = 10^{-3}$, whereas with $L = 50$, $\mu = 2 \times 10^{-4}$ (see Lynch and Walsh 1998, pp. 337–339, for a discussion of this issue). These values for μ allow substantial flow of neutral mutations between genetic backgrounds.

The high per-locus mutation rate of QTLs is not a sufficient explanation for the low power of D and Z_{nS}. This is illustrated by simulations of the three-deme model using a more conventional value for the per-locus mutation rate, $\mu = 10^{-6}$ (fig. 8; see last section of app. 3). Even with $L = 100$, the cumulative mutation rate for the quantitative trait is well below empirical estimates. Despite this, power estimates remain rather low, exceeding 50% only when the genetic variance in quantitative trait values is at least 100 times greater than the neutral expectation. Figure 8 also indicates that the effect of variation in the number of QTLs is rather modest, at least over the range of L considered here.

The second explanation for the lower power of neutrality tests is that the criteria for defining selection as purifying or balancing are different at the molecular and quantitative trait levels. For a quantitative trait, selection can maintain variation in multilocus genotypic values (and thus in trait values) without preserving particular alleles indefinitely. This is a natural consequence of "genetic redundancy" (Brookfield 1997). Mutations at multiple, perhaps many, different genes can have similar effects on a quantitative character. As a consequence, a mutation at one locus can effectively substitute for an allele with comparable effects at another locus. Such substitutions might occur because of selection or drift but in either case will likely reduce the life span of individual alleles.

A number of different observations suggest that genetic redundancy is the rule, rather than the exception, for quantitative traits. Perhaps most basic is the fact that artificial selection on most quantitative traits can rapidly move the mean value of a population outside the original range of variation (Falconer and Mackay 1996). Selection simply concentrates high or low alleles (across loci) into previously unrealized multilocus genotypes. This implies that individuals of intermediate phenotype within the original population were genetically heterogeneous, each harboring different combinations of high and low alleles across QTLs. Studies of transgressive segregation provide comparable evidence (Vega and Frey 1980; de Vicente and

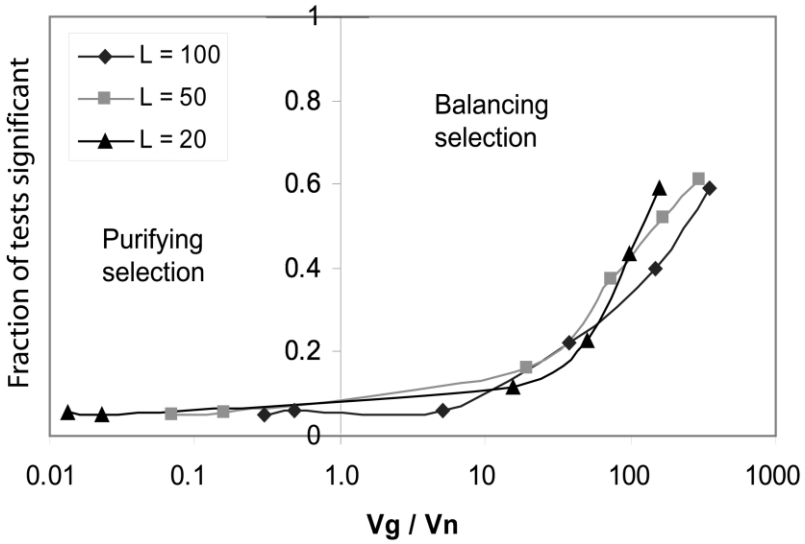

Figure 8: Fraction of Z_{nS} values that reject neutrality in simulations with $\mu = 10^{-6}$, $m = 0.1$, and $V_S = 2$. The trajectories represent power estimates from simulations with different numbers of QTL ($L = 20$, 50, or 100) as $Z_2^* - Z_0^*$ increases from 0 to 5. A smaller range for $Z_2^* - Z_0^*$ is considered because V_n is greatly reduced with $\mu = 10^{-6}$, and selection switches from purifying to balancing with little separation of deme fitness optima.

Tanksley 1993; Rieseberg et al. 1999). Finally, different populations of the same species often respond to the same selection pressure with similar changes in phenotype. However, the underlying genetic bases of phenotypic responses are often very different (Cohan 1984*a*, 1984*b*).

Genetic redundancy can itself facilitate the maintenance of genetic variation in quantitative characters. Goldstein and Holsinger (1992) model evolution of a quantitative trait in a structured population with uniform selection. In their model, different local populations reach the same mean phenotype (as dictated by the fitness optimum) via different genetic solutions (different combinations of high and low alleles across loci). This substantially increases V_G relative to an unstructured population experiencing uniform selection (see also Lande 1991).

Combining Molecular and Phenotypic Data

The increasing use of genomic techniques is bridging the gap between molecular and phenotypic studies. Mapping experiments identify genomic regions that contribute to quantitative trait variation (Tanksley 1993; Mackay 2004). While commonly denoted QTLs, these regions are generally quite large, containing many distinct genes. However, subsequent fine mapping (e.g., positional cloning) can resolve genetic differences to the scale of genes and even to the level of sequence variants. Paran and Zamir (2003) review studies that have successfully identified sequence-level differences contributing to quantitative trait variation in plants.

Mapping allows direct evaluation of QTL effects on ecology and fitness in nature (e.g., Schemske and Bradshaw 1999; Lexer et al. 2003; Weinig et al. 2003). When particular genes can be identified, molecular population genetic studies (collecting the sort of data that are simulated here) become possible. Interestingly, a number of gene sequence studies have found significant evidence for balancing selection in *Arabidopsis thaliana*. There is extensive gene sequence divergence between susceptible and resistant alleles of the *Rpm1* gene, which suggests an ancient, balanced polymorphism (Stahl et al. 1999; see also Tian et al. 2002; Mauricio et al. 2003). Substantial divergence among haplotypes has also been documented within the promoter region of *TFL1* (a gene involved in floral development; Olsen et al. 2002), within two genes of the phenylpropanoid pathway (Aguadé 2001), and between allozyme alleles of cytosolic *PGI* (Kawabe et al. 2000; see also Filatov and Charlesworth 1999). Statistically significant values for D and/or Z_{nS} are reported in each of these studies.

The population structure of *Arabidopsis* corresponds to the parameter region of this model that is most favorable to detecting selection at the molecular level. The species is structured into local races or types, and there is apparently low migration among populations. The estimated outcrossing rate of *A. thaliana* is less than 1% (Bergelson et al. 1998; Tian et al. 2002), and the migration rate among populations is likely to be substantially less unless seed

dispersal among population is common. Figures 3–5 indicate appreciable power for Z_{nS}, with strong selection and $m = 0.01$. If $m = 0.001$, reasonable power obtains even in the three-deme simulations (see app. 3).

Summary

My purpose has been to consider the effect of selection at the quantitative trait level for gene sequence evolution and, in particular, the consequences for molecular neutrality tests. The simulations are a hybrid of simple but widely used models from molecular and quantitative genetics. I assume that quantitative trait variation is due to the additive contributions of many loci, each with two possible alleles, that is, the Latter-Bulmer model (Latter 1960; Bulmer 1972). Molecular evolution within flanking regions is caused by neutral mutations introduced according to the "infinite-sites" model (Kimura 1983). The selection regime is based on a mixture of Gaussian functions with deme-specific optima. While simplified in many regards, this model effectively reproduces natural patterns of variation in morphology (compare figs. 1 and 6).

The assignment of fitness is a key difference between this and previous studies investigating the power of neutrality tests (e.g., Braverman et al. 1995; Williamson and Orive 2002). I assign fitness to phenotypes, whereas it is standard in molecular population genetics to assign fixed selection coefficients directly to alleles (or diploid genotypes). The latter approach follows a long tradition in population genetics (e.g., Fisher 1930; Wright 1931) and is reasonable when there is a simple mapping from genotype to phenotype to fitness. However, most interesting phenotypes are quantitative: variation is caused by both environmental and genetic differences, and the genetic contribution involves multiple (usually many) loci. Even the simplest models that explicitly characterize the distinct mappings from genotype to phenotype and from phenotype to fitness suggest that selection on alternative QTL alleles will not be constant.

The primary conclusion of the study is that neutrality tests applied to genomic regions containing QTLs are usually ineffective at revealing selection on those QTLs. This rather counterintuitive result is due, in large part, to the fact that selection on QTLs is context dependent. Within a particular multilocus genotype, an allele is advantageous if it brings the genotypic value closer to the local fitness optimum. This obviously depends on the alleles present at other QTLs within the genome. The second context is the environment. A multilocus genotype that performs well in one area may perform poorly elsewhere. Migration will move alleles throughout the range of a species, further complicating the mapping from genotype to fitness.

With low separation of deme optima in this model, selection is typically purifying. In other words, the genetic variance in quantitative trait values (V_G) is less than its expected value under neutrality (V_n). Consistent negative selection against particular mutations can substantially reduce sequence variation, both at sites subject to selection (Williamson and Orive 2002) and at linked sites undergoing neutral evolution (Charlesworth et al. 1993). However, for most parameter combinations of this model where $V_G < V_n$, there is no appreciable reduction in neutral variability (as measured by S and π; see app. 1). Under Gaussian stabilizing selection, selection against new mutations at QTLs is quite inconsistent. An allele that pushes the phenotype of its bearer away from the fitness optimum might be advantageous in that individual's progeny (depending on the genotypic value of mates and where the progeny reside). This, combined with the high mutation rate of QTLs, greatly reduces the signal of selection at linked sites subject to neutral evolution.

As the separation between the fitness optima of demes increases, selection becomes balancing ($V_G > V_n$). This usually increases variability within flanking regions. However, the effect is typically rather modest unless $V_G > 10V_n$ (app. 1), and even then, the power of neutrality tests often remains low (figs. 3–5). The exception is circumstances with strong selection combined with low migration and/or extensive differentiation in local optima. While only a small fraction of the parameter combinations that I considered, this set of conditions may not be uncommon in many natural plant species (Levin 1988).

The results presented here seem discouraging for gene sequence surveys as a method to identify QTLs under selection. However, I considered only a fraction of the possible ways that sequence data can be analyzed. Calculations of S, π, D, and Z_{nS} were conducted without regard to the phenotype or location of sampled individuals. Incorporating this information allows estimation of molecular differentiation between phenotypic or geographic classes (e.g., Akey et al. 2002; Emelianov et al. 2004). The viability of these more refined approaches will be explored elsewhere.

Acknowledgments

I would like to thank H. Alexander, J. Gleason, L. Holeski, S. Williamson, and an anonymous reviewer for comments on this article. Financial support was provided by National Institutes of Health grant R01 GM60792-01A1 and a National Science Foundation Epscor grant to the University of Kansas.

Literature Cited

Aguadé, M. 2001. Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. Molecular Biology and Evolution 18:1–9.

Akey, J. M., G. Zhang, K. Zhang, L. Jin, and M. D. Shriver. 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Research 12:1805–1814.

Andolfatto, P., and M. Nordborg. 1998. The effect of gene conversion on intralocus associations. Genetics 148:1397–1399.

Aquadro, C. F. 1997. Insights into the evolutionary process from patterns of DNA sequence variability. Current Opinion in Genetics and Development 7:835–840.

Bergelson, J., C. B. Purrington, and G. Wichmann. 1998. Promiscuity in transgenic plants. Nature 395:25.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140:783–796.

Brookfield, J. F. Y. 1997. Genetic redundancy. Advances in Genetics 36:137–155.

Bulmer, M. G. 1972. The genetic variability of polygenic characters under optimizing selection, mutation and drift. Genetical Research 19:17–25.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Clausen, J., D. D. Keck, and W. M. Hiesey. 1948. Experimental studies on the nature of species. III. Environmental responses of climatic races of *Achillea*. Carnegie Institution, Washington, DC.

Cohan, F. M. 1984*a*. Can uniform selection retard random genetic divergence between isolated conspecific populations? Evolution 38: 495–504.

———. 1984*b*. Genetic divergence under uniform selection. I. Similarity among populations of *Drosophila melanogaster* in their responses to artificial selection for modifiers of ci^{D}. Evolution 38: 55–71.

de Vicente, M. C., and S. D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134: 585–596.

Emelianov, I., F. Marec, and J. Mallet. 2004. Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proceedings of the Royal Society of London B 271:97–105.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton University Press, Princeton, NJ.

———. 1986. Natural selection in the wild. Princeton University Press, Princeton, NJ.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Prentice-Hall, London.

Filatov, D. A., and D. Charlesworth. 1999. DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia* PgiC locus. Genetics 153:1423–1434.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon, Oxford.

Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.

Goldstein, D. B., and K. E. Holsinger. 1992. Maintenance of polygenic variation in spatially structured populations: roles for local mating and genetic redundancy. Evolution 46:412–429.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology 7:1–44.

———. 1993. The how and why of generating gene genealogies. Pages 23–36 *in* N. Takahata and A. G. Clark, eds. Mechanisms of molecular evolution: introduction to molecular paleopopulation biology. Sinauer, Sunderland, MA.

Hudson, R. R., and N. L. Kaplan. 1988. The coalescent process in models with selection and recombination. Genetics 120:831–840.

Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. The hitchhiking effect revisited. Genetics 123:887–899.

Kawabe, A., K. Yamane, and N. T. Miyashita. 2000. DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. Genetics 156:1339–1347.

Kelly, J. K. 1997. A test of neutrality based on interlocus associations. Genetics 146:1197–1206.

Kelly, J. K., and M. J. Wade. 2000. Molecular evolution near a two-locus balanced polymorphism. Journal of Theoretical Biology 204: 83–101.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, New York.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. American Naturalist 157:245–261.

Kreitman, M. 1996. The neutral theory is dead: long live the neutral theory. Bioessays 18:678–683.

Lande, R. 1975. The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genetical Research 26: 221–235.

———. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

———. 1979. Quantitative genetic analysis of multivariate evolution applied to brain : body size allometry. Evolution 33:402–416.

———. 1991. Isolation by distance in a quantitative trait. Genetics 128:443–452.

Latter, B. D. H. 1960. Natural selection for an intermediate optimum. Australian Journal of Biological Sciences 13:30–35.

Levin, D. A. 1988. Local differentiation and the breeding structure of plant populations. Pages 305–329 *in* L. D. Gottlieb and S. K. Jain, eds. Plant evolutionary biology. Chapman & Hall, New York.

Lewontin, R. C., J. A. Moore, W. Provine, and B. Wallace, eds. 1981. Dobzhansky's genetics of natural populations I–XLIII. Columbia University Press, New York.

Lexer, C., Z. Lai, and L. H. Riesberg. 2003. Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. New Phytologist 161:225–233.

Lynch, M. 1988. The rate of polygenic mutation. Genetical Research 51:137–148.

Lynch, M., and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. Evolution 40:915–935.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative characters. Sinauer, Sunderland, MA.

Mackay, T. F. C. 2004. Genetic dissection of quantitative traits. Pages 51–73 *in* R. S. Singh and M. K. Uyenoyama, eds. The evolution of population biology. Cambridge University Press, Cambridge.

Mauricio, R., E. A. Stahl, T. Korves, D. Tian, M. Kreitman, and J. Bergelson. 2003. Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. Genetics 163:735–746.

Mayr, E. 1963. Animal species and evolution. Harvard University Press, Cambridge, MA.

Olsen, K. M., A. Womack, A. R. Garrett, J. I. Suddith, and M. D. Purugganan. 2002. Contrasting evolutionary forces in the *Arabidopsis thaliana* floral development pathway. Genetics 160:1641–1650.

Paran, I., and D. Zamir. 2003. Quantitative traits in plants: beyond the QTL. Trends in Genetics 19:303–306.

Pfrender, M. E., K. Spitze, J. Hicks, K. Morgan, L. Latta, and M. Lynch. 2000. Lack of concordance between genetic diversity estimates at the molecular and quantitative-trait levels. Conservation Genetics 1:263–269.

Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Trangressive segregation, adaptation and speciation. Heredity 83:363–372.

Schemske, D. W., and H. D. Bradshaw Jr. 1999. Pollinator preference and the evolution of floral traits in monkey flowers (*Mimulus*). Proceedings of the National Academy of Sciences of the USA 96: 11910–11915.

Schmitt, J., and S. E. Gamble. 1990. The effect of distance from parental site on offspring performance and inbreeding depression in *Impatiens capensis*: a test of the local adaptation hypothesis. Evolution 44:2022–2030.

Simonsen, K. L., G. A. Churchill, and C. F. Aquadro. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141:413–429.

Stadler, T., and L. F. Delph. 2002. Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. Proceedings of the National Academy of Sciences of the USA 99:11730–11735.

Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman, and J. Bergelson. 1999. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400:667–671.

Stebbins, G. L. 1950. Variation and evolution in plants. Columbia University Press, New York.

Stephan, W., T. H. E. Wiehe, and M. W. Lenz. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theoretical Population Biology 41:237–254.

Strobeck, C. 1983. Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics 103:545–555.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tajima, F., and M. Nei. 1983. Estimation of evolutionary distance between nucleotide-sequences. Japanese Journal of Genetics 58: 684–685.

Tanksley, S. D. 1993. Mapping polygenes. Annual Review of Genetics 27:205–233.

Tian, D., H. Araki, E. A. Stahl, J. Bergelson, and M. Kreitman. 2002. Signature of balancing selection in *Arabidopsis*. Proceedings of the National Academy of Sciences of the USA 99:11525–11530.

Turelli, M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. Theoretical Population Biology 25:138–193.

Vega, U., and K. J. Frey. 1980. Transgressive segregation in inter- and intraspecific crosses of barley. Euphytica 29:585–694.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7: 256–276.

Weinig, C., L. A. Dorn, N. C. Kane, Z. M. German, S. S. Halldorsdottir, M. C. Ungerer, Y. Toyonaga, T. F. C. Mackay, M. D. Purugganan, and J. Schmitt. 2003. Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. Genetics 165: 321–329.

Williamson, S., and M. E. Orive. 2002. The genealogy of a sequence subject to purify selection at multiple sites. Molecular Biology and Evolution 19:1376–1384.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 97–159.

———. 1937. The distribution of gene frequencies in populations. Proceedings of the National Academy of Sciences of the USA 23: 307–320.

Associate Editor: George W. Gilchrist
Editor: Michael C. Whitlock