

F_{ST} and Q_{ST} Under Neutrality

Judith R. Miller,^{*,1} Bryan P. Wood^{*,†} and Matthew B. Hamilton[†]

^{*}Department of Mathematics and [†]Department of Biology, Georgetown University, Washington, DC 20057

Manuscript received June 1, 2008

Accepted for publication July 30, 2008

ABSTRACT

A commonly used test for natural selection has been to compare population differentiation for neutral molecular loci estimated by F_{ST} and for the additive genetic component of quantitative traits estimated by Q_{ST} . Past analytical and empirical studies have led to the conclusion that when averaged over replicate evolutionary histories, $Q_{ST} = F_{ST}$ under neutrality. We used analytical and simulation techniques to study the impact of stochastic fluctuation among replicate outcomes of an evolutionary process, or the evolutionary variance, of Q_{ST} and F_{ST} for a neutral quantitative trait determined by n unlinked diallelic loci with additive gene action. We studied analytical models of two scenarios. In one, a pair of demes has recently been formed through subdivision of a panmictic population; in the other, a pair of demes has been evolving in allopatry for a long time. A rigorous analysis of these two models showed that in general, it is not necessarily true that mean $Q_{ST} = F_{ST}$ (across evolutionary replicates) for a neutral, additive quantitative trait. In addition, we used finite-island model simulations to show there is a strong positive correlation between Q_{ST} and the difference $Q_{ST} - F_{ST}$ because the evolutionary variance of Q_{ST} is much larger than that of F_{ST} . If traits with relatively large Q_{ST} values are preferentially sampled for study, the difference between Q_{ST} and F_{ST} will also be large and positive because of this correlation. Many recent studies have used tests of the null hypothesis $Q_{ST} = F_{ST}$ to identify diversifying or uniform selection among subpopulations for quantitative traits. Our findings suggest that the distributions of Q_{ST} and F_{ST} under the null hypothesis of neutrality will depend on species-specific biology such as the number of subpopulations and the history of subpopulation divergence. In addition, the manner in which researchers select quantitative traits for study may introduce bias into the tests. As a result, researchers must be cautious before concluding that selection is occurring when $Q_{ST} \neq F_{ST}$.

IDENTIFYING signatures of natural selection based on patterns of genetic variation within and among populations is a long-standing goal in population genetics. Distinguishing between natural selection and neutral processes such as genetic drift and gene flow is central to testing the hypothesis that selection is primarily responsible for patterns of phenotypic variation in natural populations. In the past 15 years, comparisons of two statistics, F_{ST} and Q_{ST} , have been used to test for the action of natural selection on quantitative traits in subdivided populations. The fixation index F_{ST} is a measure of population differentiation defined as the ratio of among-deme to total variance at the allelic level (WRIGHT 1951), and Q_{ST} is the analogous ratio of among-deme to total additive genetic variation for quantitative phenotypes (precise definitions of both these quantities are given below).

Theoretical analyses and simulations have indicated that under neutrality, F_{ST} and Q_{ST} should be equal. In particular, LANDE (1992) showed that in a finite-island model and in a metapopulation model with extinction and recolonization, F_{ST} described the differentiation of

a neutral quantitative trait among subpopulations, *i.e.*, that $F_{ST} = Q_{ST}$. SPITZE (1993) used Lande's result as a null hypothesis to argue that diversifying selection was acting in populations of *Daphnia obtusa*; in this work Spitze coined the name Q_{ST} . WHITLOCK (1999) used a coalescent argument to generalize LANDE's (1992) results to additional population structures, including stepping-stone models, under drift-mutation equilibrium. LE CORRE and KREMER (2003) related Q_{ST} to F_{ST} for traits under selection, arguing that Q_{ST} should be greater than F_{ST} under diversifying selection (*i.e.*, selection for different trait optima in different demes in a subdivided population) and less than F_{ST} under uniform selection (*i.e.*, selection for the same optimum in all demes). LE CORRE and KREMER (2003) paid particular attention to covariances between loci, pointing out that when linkage disequilibrium contributed equally to within- and between-deme trait variances, their formula predicted equal values of Q_{ST} and F_{ST} even for a trait under selection.

Further studies have examined the influence of additional evolutionary factors, such as gene action, on Q_{ST} and its relationship to F_{ST} . LÓPEZ-FANJUL *et al.* (2003) showed that with neutral evolution of a quantitative trait, dominance causes Q_{ST} to be $<F_{ST}$ for low to moderate recessive allele frequencies and Q_{ST} to be $>F_{ST}$ otherwise. Epistasis also broadens the range of

¹Corresponding author: Department of Mathematics, Georgetown University, 37th and O Sts. NW, Washington, DC 20057.
E-mail: jrm32@georgetown.edu

allele frequencies that cause Q_{ST} to be $<F_{ST}$. More recently, GOUDET and BÜCHI (2006) confirmed that dominance tends to depress Q_{ST} for a neutral quantitative trait and further showed that the effect of dominance disappears with consanguineous mating due to the resulting deficit of heterozygosity.

A number of empirical studies have accordingly been performed using the difference between Q_{ST} and F_{ST} as an indicator of the presence, and type, of selection acting on traits in a variety of organisms (STEINGER *et al.* 2002; PALO *et al.* 2003; SAINT-LAURENT *et al.* 2003; BARUCH *et al.* 2004; LE CORRE 2005; CONOVER *et al.* 2006; JOHANSSON *et al.* 2007; KNOPP *et al.* 2007; ROBERGE *et al.* 2007; reviewed in MERILÄ and CRNOKRAK 2001; MCKAY and LATTA 2002; LEINONEN *et al.* 2008). Meta-analyses of numerous studies have found that Q_{ST} tends to be $>F_{ST}$ (MERILÄ and CRNOKRAK 2001; MCKAY and LATTA 2002; LEINONEN *et al.* 2008). These results from empirical studies have been interpreted widely as evidence that quantitative traits are commonly influenced by natural selection and that diversifying selection is more common than balancing selection in natural populations (MERILÄ and CRNOKRAK 2001; MCKAY and LATTA 2002; LEINONEN *et al.* 2008). A direct approach that exposed experimental *Arabidopsis thaliana* populations to various levels of diversifying and balancing selection found that mean Q_{ST} for seven traits did increase with heterogeneous selection pressures among demes at the largest effective population size used in the study (PORCHER *et al.* 2004, 2006). Similarly, MORGAN *et al.* (2005) compared F_{ST} and Q_{ST} in mice with a known history of selection for wheel-running activity and the genetically correlated trait of body mass. They found that $Q_{ST} > F_{ST}$ among groups that experienced divergent selection pressures, but that this conclusion was contingent on confidence intervals calculated with a nonparametric bootstrap procedure (confidence intervals calculated with a parametric bootstrap procedure were overlapping).

Despite the growing number of empirical studies comparing Q_{ST} and F_{ST} , some theoretical questions germane to their interpretation remain unaddressed. In particular, the original theoretical arguments that led to the conclusion that Q_{ST} should equal F_{ST} under neutrality bear revisiting in a way that explicitly incorporates the effects of random fluctuations in allele frequencies and across loci for independent replicate populations that evolve neutrally. Such random fluctuation among replicate outcomes of an evolutionary process is called “evolutionary variance” in the context of coalescent models. It is necessary to estimate (*e.g.*, through simulations) the evolutionary variance of Q_{ST} and of the difference $Q_{ST} - F_{ST}$ under the null hypothesis of neutrality to understand the relationship between Q_{ST} and F_{ST} when a quantitative trait and marker allele frequencies are both evolving neutrally. A formal test of the hypothesis that selection is acting requires knowledge of the “evolutionary sampling distribution” of

$Q_{ST} - F_{ST}$. Only by comparing measured values of this difference with such a distribution can one determine that the difference is large enough to reject the null hypothesis of neutrality.

We stress here and below the distinction between evolutionary variance and sampling variance. Throughout this article we use the terms Q_{ST} and F_{ST} to denote parameters, that is, exact quantities describing entire populations. Others have used the terms Q_{ST} and F_{ST} to denote statistics, that is, estimates of underlying parameters that are calculated from sample data. To create confidence intervals, or otherwise understand the error inherent in estimation from a sample, one must ascertain the sampling variance of the statistic used. This has been done for estimators of F_{ST} (*e.g.*, PONS and CHAUCHE 1995; WEIR 1996), as well as for estimators of Q_{ST} in recent work (O’HARA and MERILÄ 2005; GOUDET and BÜCHI 2006). Even with complete census data (*i.e.*, zero sampling variance), however, evolutionary variance would still result in a range of values for Q_{ST} , F_{ST} , and the difference $Q_{ST} - F_{ST}$. Only by understanding the evolutionary sampling distributions of the underlying parameters Q_{ST} and F_{ST} , over all possible outcomes of a stochastic evolutionary process, can we determine what values of Q_{ST} and F_{ST} are not expected for neutral traits and therefore can serve as evidence of selection.

In this article we analyze the effect of evolutionary variance on F_{ST} and Q_{ST} through analysis and simulation of a neutral quantitative trait determined by n unlinked diallelic loci acting additively in a population composed of two demes. The additive effects of the loci on the quantitative trait are not necessarily equal. In the second and third sections (DEFINITIONS AND THEORY and ANALYTICAL EXAMPLES) we rederive expressions for Q_{ST} and for F_{ST} at the loci contributing to the neutral trait, following LE CORRE and KREMER (2003). We obtain conditions under which the expectations (across replicate evolutionary histories) of Q_{ST} and F_{ST} will be equal under neutrality and provide examples showing that equality need not hold even for an additive quantitative trait. In the fourth section (SIMULATIONS) we explore the joint sampling distribution of Q_{ST} and F_{ST} under neutrality through simulations. In the final section we discuss the implications of our findings for hypothesis tests constructed around the difference between Q_{ST} and F_{ST} in subdivided populations.

We emphasize again that throughout this article the terms Q_{ST} and F_{ST} refer to population parameters, not to statistics that estimate these parameters from samples of populations and individuals.

DEFINITIONS AND THEORY

We consider d demes of N diploid individuals each and n diallelic loci (with alleles denoted by + and -) contributing additively to a quantitative trait.

We first define some necessary quantities. When discussing F_{ST} and F_{IS} , the underlying variable is an indicator variable \tilde{x}_i for the + allele at locus i ; thus, $\tilde{x}_i = 1$ if an allele is + and $\tilde{x}_i = 0$ if the allele is -. We let p_{ik} denote the mean of \tilde{x}_i in deme k , i.e., the fraction of + alleles at locus i in deme k ; we let p_i denote the overall fraction of + alleles at locus i in the population (composed of two or more demes). Thus

$$p_i = \frac{1}{d} \sum_{k=1}^d p_{ik}. \tag{1}$$

We let $q = 1 - p$ with any subscript. We let

$$\tilde{\sigma}_i^2 = p_i q_i \tag{2}$$

denote the total variance of \tilde{x}_i in the population,

$$\tilde{\sigma}_{B_i}^2 = \frac{1}{d} \sum_{k=1}^d (p_{ik} - p_i)^2 \tag{3}$$

the between-deme variance, and

$$\begin{aligned} \tilde{\sigma}_{0W_i}^2 &= \frac{1}{d} \sum_{k=1}^d \left[p_{ik}^2 (1 - p_{ik})^2 + 2p_{ik} q_{ik} \left(\frac{1}{2} - p_{ik} \right)^2 + q_{ik}^2 (0 - p_{ik})^2 \right] \\ &= \frac{1}{2d} \sum_{k=1}^d p_{ik} q_{ik} \end{aligned} \tag{4}$$

the average within-deme variance of \tilde{x}_i expected in a set of random-mating demes with + allele frequencies p_{ik} . (In Equation 4, the middle term is obtained as the sum of squared deviations of \tilde{x}_i from the mean for each haplotype, weighted by the frequencies of the haplotypes.) The fixation index F_{ST_i} at locus i and the overall multilocus fixation index F_{ST} are then defined as

$$\begin{aligned} F_{ST_i} &= \frac{\tilde{\sigma}_{B_i}^2}{\tilde{\sigma}_i^2} = \frac{\tilde{\sigma}_{B_i}^2}{\tilde{\sigma}_{B_i}^2 + 2\tilde{\sigma}_{0W_i}^2}, \\ F_{ST} &= \frac{\sum_{i=1}^n \tilde{\sigma}_{B_i}^2}{\sum_{i=1}^n (\tilde{\sigma}_{B_i}^2 + 2\tilde{\sigma}_{0W_i}^2)} = \frac{\sum_{i=1}^n \tilde{\sigma}_{B_i}^2}{\sum_{i=1}^n p_i q_i}. \end{aligned} \tag{5}$$

We note that our definition of F_{ST} , which averages variances over all loci before taking a ratio, is analogous to the preferred estimator of F_{ST} proposed by WEIR and COCKERHAM (1984).

We let p_{++ik} denote the frequency of individuals in deme k with two + alleles at locus i , with p_{+-ik} and p_{--ik} defined analogously. We define the inbreeding coefficient $F_{IS_{ik}}$ as the correlation in deme k between the indicator variables for homologous alleles at locus i within an individual:

$$\begin{aligned} F_{IS_{ik}} &= \frac{p_{++ik}(1 - p_{ik})^2 + p_{+-ik}(1 - p_{ik})(0 - p_{ik}) + p_{--ik}(0 - p_{ik})^2}{p_{ik} q_{ik}} \\ &= \frac{2p_{ik} q_{ik} - p_{+-ik}}{2p_{ik} q_{ik}}. \end{aligned} \tag{6}$$

To discuss Q_{ST} , we must define trait means and variances. We denote the additive effect of locus i by a_i and the phenotypic value at the i th locus in the j th individual in deme k by x_{ijk} . We let p_{ijk} denote the fraction of + alleles at locus i in that individual, so that p_{ijk} can take on the genotypic values 0, $\frac{1}{2}$, or 1. Under additivity, therefore, we find that the phenotypic value of the j th individual in deme k is given by

$$x_{jk} = \sum_{i=1}^n x_{ijk} = \sum_{i=1}^n 2a_i \left(p_{ijk} - \frac{1}{2} \right). \tag{7}$$

We let

$$\sigma_{GB_i}^2 = \frac{4a_i^2}{d} \sum_{k=1}^d (p_{ik} - p_i)^2 \tag{8}$$

denote the between-deme component of the (additive) genetic variance contributed by locus i . The between-deme component of the total trait variance is given by

$$\sigma_B^2 = \sum_{i=1}^n \left[\sigma_{GB_i}^2 + \sum_{i' \neq i} \text{Cov}_{ii'} \right], \tag{9}$$

where

$$\text{Cov}_{ii'} = \frac{1}{d} \sum_{k=1}^d 4a_i a_{i'} (p_{ik} - p_i)(p_{i'k} - p_{i'}) \tag{10}$$

is the covariance between loci i and i' .

The between-deme trait variance can be partitioned into two components: one comprising covariances between loci contributing to the trait (quantitative trait loci, QTL) and one comprising variances at individual loci. We write the ratio of these two components as

$$\theta_B = \left[\sum_{i=1}^n \sum_{i' \neq i} \text{Cov}_{ii'} \right] / \left[\sum_{i=1}^n \sigma_{GB_i}^2 \right], \tag{11}$$

so that

$$\sigma_B^2 = (1 + \theta_B) \sum_{i=1}^n \sigma_{GB_i}^2. \tag{12}$$

As noted by LE CORRE and KREMER (2003), the quantity θ_B is a measure of gametic disequilibrium among QTL contributing to the trait. If the trait is neutral, then it should average to zero across replicate evolutionary histories (ROGERS and HARPENDING 1983). We see below that writing Q_{ST} in terms of θ_B and its within-deme analog θ_W facilitates comparison of Q_{ST} and F_{ST} .

Analogously to (4), we define the within-deme genetic variance that would be expected in a random-mating population with + allele frequencies p_{ik} ,

$$\sigma_{0W}^2 = (1 + \theta_W) \sum_{i=1}^n \sum_{k=1}^d \frac{2a_i^2}{d} p_{ik} q_{ik}, \tag{13}$$

where θ_W is the ratio of average within-deme covariances between indicator variables for different loci to average within-deme additive genetic variance. In other words,

$$\theta_W = \left[\frac{1}{d} \sum_{k=1}^d \sum_{i=1}^n \sum_{i' \neq i} \text{Cov}_{ii'k} \right] / \sigma_{GW}^2, \quad (14)$$

where

$$\text{Cov}_{ii'k} = \frac{1}{N} \sum_{j=1}^N 4a_i a_{i'} (p_{ijk} - p_{ik})(p_{i'jk} - p_{i'k}) \quad (15)$$

and

$$\begin{aligned} \sigma_{GW}^2 &= \frac{1}{Nd} \sum_{k=1}^d \sum_{i=1}^n \sum_{j=1}^N (x_{ijk} - x_{ik})^2 \\ &= \frac{4}{Nd} \sum_{k=1}^d \sum_{i=1}^n a_i^2 \sum_{j=1}^N (p_{ijk} - p_{ik})^2. \end{aligned} \quad (16)$$

As with θ_B (defined above), θ_W is the ratio of the component of within-deme trait variance that is due to covariances between loci to the component due to individual loci and is expected to be zero for a neutral trait (ROGERS and HARPENDING 1983). Also, it is straightforward to check that $\theta_W = 0$ within a single population if Hardy–Weinberg equilibrium holds.

In the notation we have now established, the within-deme component of total trait variance is $\sigma_{GW}^2(1 + \theta_W)$. Also,

$$\sigma_{GW}^2 = \sum_{i=1}^n \sum_{k=1}^d \frac{2a_i^2}{d} p_{ik} q_{ik} (1 + F_{IS_{ik}}), \quad (17)$$

so if $F_{IS_{ik}} = F_{IS}$ for all i and k , then

$$\sigma_{GW}^2 = (1 + F_{IS})\sigma_{0W}^2. \quad (18)$$

Analogously to (5), and in keeping with the usage of LE CORRE and KREMER (2003), we finally define

$$Q_{ST} = \frac{\sigma_B^2}{\sigma_B^2 + 2\sigma_{0W}^2}. \quad (19)$$

This is the same as $\sigma_B^2 / (\sigma_B^2 + 2\sigma_{GW}^2)$ if mating is random (so $F_{IS} = 0$).

We now consider the relationship between F_{ST} and Q_{ST} . First, from (8), (12), (13), and (19) we obtain

$$Q_{ST} = \frac{(1 + \theta_B) \sum_{i=1}^n \sum_{k=1}^d a_i^2 (p_{ik} - p_i)^2}{(1 + \theta_B) \sum_{i=1}^n \sum_{k=1}^d a_i^2 (p_{ik} - p_i)^2 + (1 + \theta_W) \sum_{i=1}^n \sum_{k=1}^d a_i^2 p_{ik} q_{ik}}, \quad (20)$$

while from (3)–(5) we obtain

$$F_{ST} = \frac{\sum_{i=1}^n \sum_{k=1}^d (p_{ik} - p_i)^2}{\sum_{i=1}^n \sum_{k=1}^d (p_{ik} - p_i)^2 + \sum_{i=1}^n \sum_{k=1}^d p_{ik} q_{ik}}. \quad (21)$$

Comparison of (20) and (21) shows that if $\theta_B = \theta_W$, and if in addition all QTL have equal effects on the trait (*i.e.*, $a_i = a$ for all i), then $Q_{ST} = F_{ST}$. This was observed by LE CORRE and KREMER (2003), although they did not examine the case of unequal a_i . As LE CORRE and KREMER (2003, p. 1207) noted, the condition that $\theta_B = \theta_W$ means that “linkage disequilibrium among QTL contributes equally to the within- and between-deme variances for the trait.” This condition does not seem to have an intuitive biological interpretation, except when both θ_B and θ_W are zero, as would be expected (on average across evolutionary replicates with random mating) for a neutral trait (ROGERS and HARPENDING 1983).

Alternatively, using (3)–(5) one can show that

$$F_{ST_i} = \frac{(1/d) \sum_{k=1}^d (p_{ik} - p_i)^2}{p_i q_i}; \quad (22)$$

then from (22) and (20) it eventually follows that

$$Q_{ST} = \frac{(1 + \theta_B) \sum_{i=1}^n a_i^2 p_i q_i F_{ST_i}}{\sum_{i=1}^n a_i^2 p_i q_i [(\theta_B - \theta_W) F_{ST_i} + (1 + \theta_W)]}. \quad (23)$$

Thus if $\theta_B = \theta_W$ and if

$$F_{ST_i} = F_{ST} \text{ for all } i, \quad (24)$$

we have $Q_{ST} = F_{ST}$. However, we do not have $Q_{ST} = F_{ST}$ in general.

If testing a null hypothesis of neutral evolution is the goal, then we must ascertain whether the expectations $E[Q_{ST}]$ and $E[F_{ST}]$, taken over replicate populations (*i.e.*, replicate evolutionary histories), are equal. If $\theta_B = \theta_W$ (as would be true if both are equal to zero), we have

$$Q_{ST} = \frac{\sum_{i=1}^n a_i^2 p_i q_i F_{ST_i}}{\sum_{i=1}^n a_i^2 p_i q_i}. \quad (25)$$

We see from (22) and (25) [using (1)] that both Q_{ST} and F_{ST} are nonlinear functions of the random variables p_{ik} . Thus even if the ratio of the expected values of the numerator and denominator of Q_{ST} does equal the expected value of F_{ST} , there is no reason to anticipate that $E[Q_{ST}]$ will also equal $E[F_{ST}]$.

If θ_B and θ_W are not constrained to be equal, then Q_{ST} and F_{ST} will be nonlinear functions of these two quantities as well as of the allele frequencies p_{ik} . It is conceivable that in this case θ_B and θ_W could vary across evolutionary replicates in such a way as to make $E[Q_{ST}] = E[F_{ST}]$. However, we are not aware of any biological mechanism that could plausibly produce such a phenomenon, except as a rare coincidence.

ANALYTICAL EXAMPLES

We now provide several concrete examples to show that it is indeed possible for the expected values of

$E[Q_{ST}]$ and $E[F_{ST}]$ to differ, even for a neutral, additive trait. We remind the reader that we are not concerned with detecting bias in estimators of Q_{ST} and F_{ST} , which would involve calculating the expectations of the estimators over repeated samples from the same populations and comparing them with the value of the parameters Q_{ST} and F_{ST} in those populations. Rather, we are concerned with the values of Q_{ST} and F_{ST} themselves (known exactly for each evolutionary replicate), averaged over all possible outcomes of a stochastic evolutionary process.

Our examples concern populations composed of two isolated demes and a neutral trait determined by two diallelic loci. They compare Q_{ST} with F_{ST} calculated for the loci contributing to the trait. Before presenting our examples in detail, we note that they hinge on a simple principle: genetic drift produces variance in neutral allele and haplotype frequencies across evolutionarily replicate lineages. This produces evolutionary variance in Q_{ST} and F_{ST} as well, because F_{ST} and Q_{ST} for an additive trait are simply nonlinear functions of allele and haplotype frequencies (*viz.* Equations 20 and 21).

In other words, Q_{ST} and F_{ST} can be viewed as random variables, which take on different values in different replicate populations. These values depend only on allele and haplotype frequencies, together with the effect sizes of the relevant QTL. Indeed, if Hardy–Weinberg equilibrium holds, then only allele frequencies and effect sizes are required to compute Q_{ST} and F_{ST} . Therefore, if we postulate a distribution of allele frequencies across evolutionary replicates, this in turn yields distributions of Q_{ST} and F_{ST} . In what follows, we approximate the means of such distributions. For populations in Hardy–Weinberg equilibrium, we find that these means are functions simply of allelic effect sizes and of parameters describing distributions of allele frequencies across replicate populations.

Numerous calculations in this section were performed with the computer algebra software Maple (version 9); the Maple worksheets are available upon request.

We recall that we consider populations composed of two isolated demes, denoted 1 and 2, and two diallelic loci, denoted A and B . The allelic effects of the loci are a_A and a_B ; the frequency of $+$ alleles at locus A in population 1 is denoted p_{A1} , with p_{B1} , p_{A2} , and p_{B2} defined similarly. We assume throughout this section that $\theta_W = 0$, which will be the case if Hardy–Weinberg equilibrium holds within demes. As for θ_B , it follows from Equation 11 that in the present case

$$\theta_B = \frac{2a_A a_B (p_{A1} - p_{A2})(p_{B1} - p_{B2})}{a_A^2 (p_{A1} - p_{A2}) + a_B^2 (p_{B1} - p_{B2})}. \quad (26)$$

Thus Equations 20 and 21 for Q_{ST} and F_{ST} become

$$Q_{ST} = \frac{(1/2)(1 + \theta_B)(a_A^2(p_{A1} - p_{A2})^2 + a_B^2(p_{B1} - p_{B2})^2)}{\left((1/2)(1 + \theta_B)(a_A^2(p_{A1} - p_{A2})^2 + a_B^2(p_{B1} - p_{B2})^2) + a_A^2(p_{A1}(1 - p_{A1}) + p_{A2}(1 - p_{A2})) + a_B^2(p_{B1}(1 - p_{B1}) + p_{B2}(1 - p_{B2})) \right)} \quad (27)$$

and

$$F_{ST} = \frac{(p_{A1} - p_{A2})^2 + (p_{B1} - p_{B2})^2}{(p_{A1} + p_{A2})(2 - p_{A1} - p_{A2}) + (p_{B1} + p_{B2})(2 - p_{B1} - p_{B2})}. \quad (28)$$

For the population model we have specified, we can therefore express both Q_{ST} and F_{ST} in the case of Hardy–Weinberg equilibrium as functions of allele frequencies and effect sizes alone [by combining (27) with (26)]. This will allow us to draw conclusions about the evolutionary sampling distributions of Q_{ST} and F_{ST} once we have specified distributions for the allele frequencies. We choose the allele-frequency distributions to model two scenarios. In the first scenario, two demes have recently arisen through subdivision of a previously panmictic population. In the second scenario, two demes have been evolving in isolation for a long time. By analyzing Equations 26–28 for each scenario, we obtain concrete examples that provide insight into how the distributions of Q_{ST} and F_{ST} can behave. Among other findings, we show that it is not generally true that mean Q_{ST} for a neutral, additive trait equals mean F_{ST} for a neutral marker. Although the relevant formulas are somewhat complicated, the principle behind the calculations is not: in the simple demographic and genetic model that leads to (26)–(28), Q_{ST} and F_{ST} depend only on allele frequencies and allelic effect sizes.

Recent subdivision: Our aim is to calculate the expectation $E[Q_{ST} - F_{ST}]$ over a set of replicate (structured) populations, for specific examples. To do so, we let $f_{A1}(p)$ denote the probability distribution of the allele frequency p_{A1} across replicates. For a first example, we specify a probability distribution whose graph is a tall, narrow rectangle. Thus

$$f_{A1}(p) = \begin{cases} \frac{1}{2\varepsilon} & p_{A1}^0 - \varepsilon \leq p \leq p_{A1}^0 + \varepsilon \\ 0 & |p - p_{A1}^0| > \varepsilon \end{cases} \quad (29)$$

for some small positive number ε ; we specify f_{A2} , f_{B1} , and f_{B2} similarly. When $p_{A1}^0 = p_{A2}^0$ and $p_{B1}^0 = p_{B2}^0$, the allele-frequency distributions described by (29) can be viewed as approximations of the distributions that would be seen across replicates in which a panmictic ancestral population had recently been divided into two subpopulations. [We have not used more exact formulas for these distributions (KIMURA 1964) because to do so would render the example analytically intractable.] In this case p_{A1}^0 and p_{B1}^0 represent the ancestral allele frequencies. At the moment of subdivision ε (the width of the distribution, correlated with variance) would

equal zero; as time went on ε would gradually increase as genetic drift caused the subpopulations to diverge.

We assume all four allele frequencies to be independent of each other (as should be the case in the absence of selection and linkage), so that

$$E[Q_{ST} - F_{ST}] = \iiint \iiint (Q_{ST} - F_{ST}) f_{A1}(p_{A1}) f_{A2}(p_{A2}) f_{B1}(p_{B1}) f_{B2}(p_{B2}) \times dp_{A1} dp_{A2} dp_{B1} dp_{B2}. \quad (30)$$

The integral obtained by substituting (27)–(29) into (30) is analytically intractable. However, it can be approximated by integrating a Taylor approximation to the function $Q_{ST} - F_{ST}$ around the point $(p_{A1}^0, p_{A2}^0, p_{B1}^0, p_{B2}^0)$. The resulting integral is still too complicated to be displayed in full generality, but can be examined in particular cases. For example, if $p_{A1}^0 = p_{A2}^0 = \frac{1}{2}$ and $p_{B1}^0 = p_{B2}^0 = \frac{1}{4}$, then as ε tends to zero we find that $E[Q_{ST} - F_{ST}] \approx -8\varepsilon^2(a_A^2 - a_B^2)/21(3a_B^2 + 4a_A^2)$. From this expression we conclude that mean $Q_{ST} - F_{ST}$ may be positive, negative, or zero for a neutral additive trait, depending on mean allele frequencies and on the relative effects of the loci contributing to the trait.

More generally, if we take

$$p_{A1}^0 = p_{A2}^0 = p_A^0 \quad \text{and} \quad p_{B1}^0 = p_{B2}^0 = p_B^0, \quad (31)$$

then

$$E[Q_{ST} - F_{ST}] = \frac{\varepsilon^2(a_A^2 - a_B^2)(p_A^0 - p_B^0)(p_A^0 + p_B^0 - 1)}{6[a_A^2 p_A^0(1 - p_A^0) + a_B^2 p_B^0(1 - p_B^0)][p_A^0(1 - p_A^0) + p_B^0(1 - p_B^0)]} + O(\varepsilon^4). \quad (32)$$

We note that when (31) holds, mean Q_{ST} and F_{ST} are equal when $\varepsilon = 0$, that is, when there is no variance in allele frequencies across replicate populations. The key conclusion from (31) and (32) is that the evolutionary variance in allele frequencies that results from genetic drift can cause mean $Q_{ST} - F_{ST}$ to differ from zero even when mean allele frequencies in the two demes (across replicate populations) are equal.

We also note that in (32), the mean difference $E[Q_{ST} - F_{ST}]$ depends on the ratio of allelic effects $\rho = a_A/a_B$. Mean Q_{ST} and F_{ST} are equal to at least the second order in ε [*i.e.*, to $O(\varepsilon^2)$] when a_A and a_B are equal. This is not always the case when (31) does not hold (calculations not shown). However, since (31) should hold in scenarios where two demes descend from a common ancestral population, simulations of such scenarios that use equal allelic effects may not detect a difference between mean Q_{ST} and F_{ST} . When the allelic effects a_A and a_B are not equal, some additional calculation (not shown) reveals that the $O(\varepsilon^2)$ term in $E[Q_{ST} - F_{ST}]$ grows toward finite limits as ρ tends either to infinity or to zero. The precise values of the limits depend on the mean

allele frequencies p_A^0 and p_B^0 . Also, since (32) contains a factor of ε^2 that corresponds roughly to the variability of allele frequencies across demes, it suggests that the difference between mean Q_{ST} and F_{ST} will be negligible immediately after subdivision (when ε is very small) and will then grow until the “narrow rectangle” model (29) is no longer applicable.

We have just concluded that in our simple model of a neutral trait in two demes that have recently emerged through subdivision of a panmictic ancestral population, the difference $Q_{ST} - F_{ST}$ can be of any sign. However, we can also ask what happens when we examine a large number of independent neutral traits, perhaps in different organisms, in such pairs of demes. Thus, we now examine the mean value E^* of $Q_{ST} - F_{ST}$, where the average is taken over all possible traits determined additively by two unlinked loci with a distribution of allele frequencies as in (29).

Since the names of loci A and B are arbitrary, we may assume that the ratio of effect sizes $r = a_B/a_A$ for each trait is chosen from a distribution $g(r)$ satisfying $g(r) = g(1/r)$ for all $r > 0$, *i.e.*, that the ratio a_B/a_A has the same distribution as a_A/a_B . Since we are modeling two recently isolated demes, we also assume that $p_{A1}^0 = p_{A2}^0 = p_A^0$ and $p_{B1}^0 = p_{B2}^0 = p_B^0$, as in (31) above. It turns out that these two assumptions alone suffice to constrain the across-trait mean $E^* = E[Q_{ST} - F_{ST}]$ to be positive when ε is small [recall that ε is the radius of the rectangular distribution of allele frequencies in (29)]. To see this, we first calculate the difference $Q_{ST} - F_{ST}$, using (26)–(28) and (31), and then exchange the allelic effects a_A and a_B to create a new expression. Finally we average the new expression with the original difference, which yields an approximate expression for the mean value E^* of $Q_{ST} - F_{ST}$:

$$E^* = \frac{\varepsilon^2(p_{A0} + p_{B0} - 1)^2(p_{A0} - p_{B0})^2(a_A^2 - a_B^2)^2}{(a_B^2 + a_A^2)(p_{B0}(1 - p_{B0}) + p_{A0}(1 - p_{A0}))(p_{A0}(1 - p_{A0}) + p_{B0}(1 - p_{B0}))} + O(\varepsilon^2). \quad (33)$$

Since the mean allele frequencies p_{A0} and p_{B0} lie in the interval $[0, 1]$, we conclude from (33) that E^* is always positive for small ε whenever p_{A0} and p_{B0} differ from each other and at least one of these frequencies lies strictly between 0 and 1. Biologically, this suggests that if many independent neutral, additive traits are measured in populations consisting of two demes that have recently arisen through subdivision of a panmictic ancestral population, the mean difference $Q_{ST} - F_{ST}$ for these traits is likely to be positive. However, the methods used to arrive at this finding do not allow us to estimate the magnitude of the difference.

Long evolution in allopatry: It is also possible to investigate the mean difference between Q_{ST} and F_{ST} over replicate evolutions in which two demes have evolved in allopatry for a long time after population subdivision and are therefore more diverged in allele

TABLE 1

Possible states of pairs of demes after long evolution in allopatry without mutation, migration, or selection

Group	Locus A		Locus B		$E[Q_{ST} - F_{ST}]$	Frequency
	Deme 1	Deme 2	Deme 1	Deme 2		
1	Seg	Seg	Seg	Seg	<0	A
2	0	Seg	Seg	Seg	Depends	B
3	1	Seg	Seg	Seg	Depends	B
4	0	0	Seg	Seg	0	C
5	1	1	Seg	Seg	0	C
6	0	1	Seg	Seg	Depends	D
7	0	Seg	0	Seg	Depends	E
8	1	Seg	1	Seg	Depends	E
9	0	Seg	1	Seg	<0	2E
10	0	1	0	Seg	Depends	F
11	0	1	1	Seg	Depends	F
12	0	0	0 or 1	Seg	0	G
13	1	1	0 or 1	Seg	0	G
14	0 or 1	0 or 1	0 or 1	0 or 1	0	H

In the locus/deme columns, 1 means that the + allele is fixed, 0 means that the - allele is fixed, and “Seg” means that segregation persists. The $E[Q_{ST} - F_{ST}]$ column contains the sign of the mean value of $Q_{ST} - F_{ST}$ for pairs of demes in each group; “depends” means that the sign can be positive, negative, or zero, depending on a_A/a_B . In the frequency column, groups with the same letter occur in replicate scenarios in equal proportions, if loci are unlinked and if each scenario is initialized by choosing + allele frequencies for each locus and deme from independent distributions symmetric about $\frac{1}{2}$. The notation “2E” means that group 9 occurs twice as often as groups 7 and 8.

frequency than in the first example. We consider a neutral trait determined additively by two loci, as in the previous scenario, and we retain notation (a_A, p_{A1} , etc.) from that scenario. The first task is to specify probability distributions, analogous to (29), for the allele frequencies at the two loci. These are obtained from KIMURA (1964), who showed that if mutation is negligible, in this scenario we can expect that fixation or loss at an originally segregating locus will occur in more and more replicate demes as time goes on. At loci and in populations where segregation persists, allele frequencies will become approximately uniformly distributed.

If we consider numerous replicate pairs of demes evolving in allopatry, we can separate the replicate pairs into groups according to KIMURA’s (1964) result. For example, one group consists of pairs in which segregation persists at both loci in both demes, another group consists of pairs in which fixation of the + allele at locus A has occurred in deme 1 but segregation persists at locus A in deme 2 and at locus B in both demes, and so on. All possible groups are enumerated in Table 1.

It is possible, using the integration formula (30), to calculate numerically (and sometimes analytically) the sign of the difference $Q_{ST} - F_{ST}$ averaged over all replicate evolutions in a group for a specified genetic architecture. As with the previous scenario (short evolution in allopatry), in many cases it turns out that $Q_{ST} - F_{ST}$ can be positive, negative, or zero, depending on the ratio of allelic effect sizes $r = a_B/a_A$ (calculations not shown).

To calculate the grand mean E^* of $Q_{ST} - F_{ST}$ over all replicate evolutions for all two-locus additive traits, we should first find the mean value of $Q_{ST} - F_{ST}$ for each group and then take a weighted average of the within-group means. This would require knowledge of the relative frequencies of all the groups, which in turn depend on the time since subdivision of the ancestral population as well as on the initial distribution of allele frequencies. It is therefore impossible to compute a “one size fits all” numerical value for E^* .

However, it turns out that it is possible to determine the sign of E^* , assuming only that evolution in allopatry without significant mutation has been going on for a long time and that the initial distributions of allele frequencies were symmetrical with respect to perturbations of the labels A and B (for loci) and 1 and 2 (for demes). To find the sign of E^* , we first note that although we cannot find numerical values for the frequencies of the various groups, certain groups in Table 1 must have equal frequencies. For example, group 2 and group 3 must be equally frequent, because within the class of deme pairs in which segregation persists at locus A in deme 2 and at locus B in both demes, fixation and loss at locus A in deme 1 should be equally likely. The “frequency” column in Table 1 shows these relationships; we note that group 9 occurs twice as frequently as group 7.

Next, we observe that any combination of allele frequencies and effect sizes, which we may express as a vector $C = [a_A, a_B, p_{A1}, p_{A2}, p_{B1}, p_{B2}]$, occurs with the

same frequency as the combination $C^* = [a_B, a_A, p_{A1}, p_{A2}, p_{B1}, p_{B2}]$ obtained by exchanging the values of a_A and a_B in C . To determine the sign of E^* , therefore, we can proceed as follows. First, we numerically calculate the value of mean $Q_{ST} - F_{ST}$ for each group listed in Table 1, assuming uniform distributions of allele frequencies at loci where segregation persists (we write the mean for group i as E_i^*). We also calculate \tilde{E}_i^* , the mean values of $Q_{ST} - F_{ST}$ for each group i , but with the effect sizes a_A and a_B exchanged. Next, we average the values of E_i^* and \tilde{E}_i^* for all groups in each frequency class (given in the rightmost column of Table 1) to obtain the mean values of $Q_{ST} - F_{ST}$ for each frequency class.

The mean values obtained from this procedure depend on the ratio of allelic effects a_A/a_B . If some of these mean values are negative for all such ratios and none are positive for any ratio, then it must be the case that the overall mean E^* of $Q_{ST} - F_{ST}$ is negative. In fact, this is precisely what occurs (details may be viewed in a Maple worksheet available from the authors upon request); an example is plotted in Figure 1. Biologically, this result means that for traits with genetic architecture like that modeled here, Q_{ST} is expected to be $<F_{ST}$ in populations where two demes were created long ago by subdivision of a panmictic ancestral population and mutation is weak. The precise magnitude of mean $Q_{ST} - F_{ST}$ cannot be determined without making additional assumptions. However, the case of group 14 in Table 1 shows that it will be zero if *no* replicate populations maintain segregating loci.

SIMULATIONS

To explore the relationship between Q_{ST} and F_{ST} in a more general finite-island model, we carried out two types of stochastic simulations. One set of simulations explicitly modeled individual organisms with multiple unlinked loci contributing additively to a single quantitative trait, as well as a second collection of loci with no phenotypic effects. The other set of simulations incorporated the same population structure and loci, but allele frequencies were summarized at the level of demes rather than being calculated from individual genotypes. In all simulations, Q_{ST} and F_{ST} were computed from complete information on all loci, individuals, and demes. Therefore, the calculated values of Q_{ST} and F_{ST} were unaffected by sampling variance. Rather, all variance among replicates arose from stochastic variation among replicate outcomes of the same evolutionary processes.

All simulations were written using the mathematical software and programming language Matlab 7.0.1. The deme-level simulations were executed using Matlab as well, while the individual-based simulations were then translated into C, compiled, and executed on a Beowulf PC cluster.

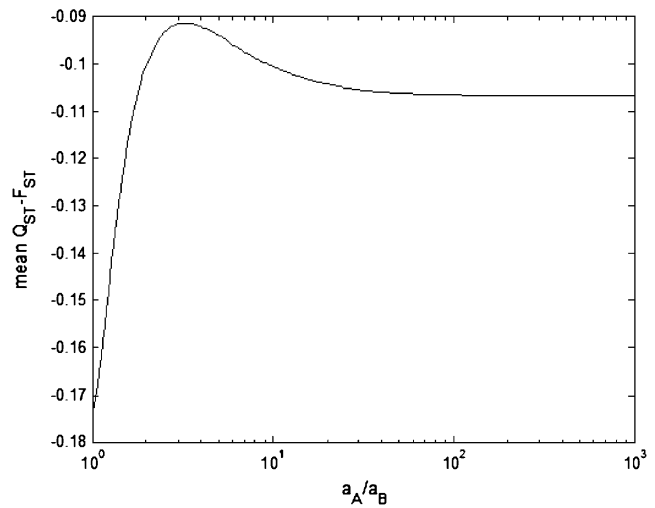


FIGURE 1.—Mean $Q_{ST} - F_{ST}$ across replicate populations in groups 10 and 11 from Table 1, together with their equivalents after exchanging the effect sizes a_A and a_B . These values assume uniform distributions of allele frequencies (across replicates) at loci where segregation persists. The value of mean $Q_{ST} - F_{ST}$ depends only on the ratio a_A/a_B ; because of the symmetry induced by exchanging a_A and a_B , including ratios <1 would simply reflect the graph across the vertical line $a_A/a_B = 1$.

Individual-based simulations: The individual-based simulations modeled a monoecious diploid population inhabiting two islands, with two unlinked loci (“QTL”) contributing to a single neutral trait and four unlinked loci (“markers”) with no phenotypic effect (except for one set of replicate simulations that used 20 islands, 10 QTL, and 10 markers). Mutation was absent. Thus the model used for individual-based simulations resembled that used in the ANALYTICAL EXAMPLES section above, with the main difference being the inclusion of migration in some of the simulation runs. The effective population size of individual demes (N) was set to 100 or 500, and the rate of gene flow among demes (m) was set to yield four values of Nm ranging from near panmixia ($Nm = 10$) to virtually complete isolation ($Nm = 10^{-10}$).

To initialize the simulations, first two allelic effect sizes (A and a) for each locus were chosen randomly from independent, identical Laplace, or reflected exponential, distributions. Next, initial frequencies of the A allele for each locus were chosen randomly from independent uniform distributions on $[0, 1]$. Finally, each individual was randomly assigned alleles using these frequencies. The two demes were initialized identically, to represent recent subdivision of a panmictic ancestral population. For each of the eight combinations of N and Nm used, two initializations were generated. One initialization was used in 100 replicate simulations for 100 generations each, and the other initialization was used in 100 replicate simulations for 1000 generations each. In each generation, migration

according to an island model preceded random mating within demes. Each locus, whether QTL or marker, was assumed to lie on a separate chromosome, giving a recombination rate of 0.5 for all pairs of loci.

The neutral trait was strictly additive. Thus, trait values were calculated by summing the allelic effects at all QTL for each individual. Q_{ST} was calculated according to the formula of LE CORRE and KREMER (2003). This formula incorporates the inbreeding coefficient F_{IS} (WRIGHT 1951); F_{IS} was calculated as a ratio of variances using data from all markers, summing over loci for the numerator and denominator separately before dividing as recommended by WEIR and COCKERHAM (1984). F_{ST} at the markers was calculated using NEI's (1973) formula for G_{ST} , a multiallelic analog that equals F_{ST} when only two alleles are present. Both Q_{ST} and G_{ST} for each marker were calculated in every generation.

For each of the 17 sets of replicate runs, the final values of Q_{ST} and mean (over markers) G_{ST} were averaged over the 100 replicates, yielding 17 values of the means $E[Q_{ST}]$ and $E[G_{ST}]$. The difference $E^* = E[Q_{ST}] - E[G_{ST}]$ was negative in 14 of the 17 sets. In addition, the absolute value of the difference E^* exceeded 20% of $E[G_{ST}]$ in 8 of 16 sets (excluding one set for which all but 4 runs led to near-complete fixation and thus undefined Q_{ST} and/or G_{ST}). E^* differed significantly from zero (at the 0.05 significance level) in only three cases after Bonferroni correction; however, in many cases the power of the t -test was relatively low because both Q_{ST} and F_{ST} had large variance across replicate runs. In all three significant cases, E^* was negative. Thus the simulations provided some confirmation for the analytical conclusion that mean Q_{ST} for an additive trait can differ from mean G_{ST} under neutrality. Increasing the number of replicate simulations, and thus the power of the t -tests used, might lead to stronger conclusions.

As expected, both Q_{ST} and G_{ST} tended to be higher in more isolated pairs of populations (*i.e.*, for lower values of Nm). The difference E^* showed no particular dependence on Nm .

Analysis of E^* computed after different numbers of generations (100 *vs.* 1000) also yielded results consistent with the analytical conclusion that mean Q_{ST} is expected to be less than mean G_{ST} after a long period of evolution in allopatry. In particular, the mean value of E^* after 1000 generations was -0.0970 , compared with -0.0177 after 100 generations. Thus Q_{ST} decreased, relative to G_{ST} , over time. However, the differences were not statistically significant: 95% confidence intervals for E^* were $(-0.243, 0.049)$ after 1000 generations and $(-0.053, 0.017)$ after 100 generations.

The analytical prediction that mean Q_{ST} is expected to be less than mean G_{ST} in populations that have recently become isolated was neither strongly supported nor contradicted by the simulation results. As noted above, the 95% confidence interval for E^* after 100 generations

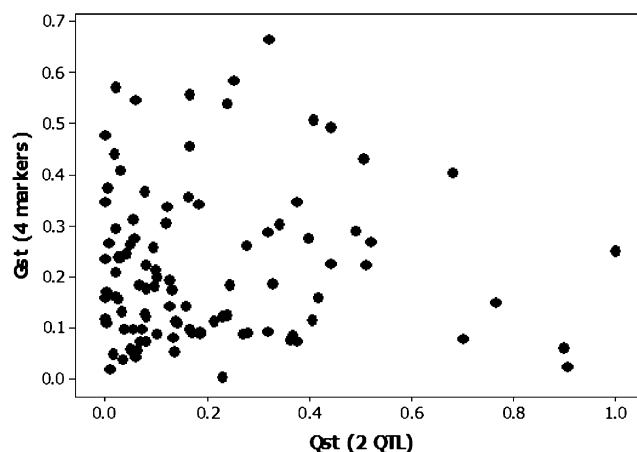


FIGURE 2.— G_{ST} (calculated using four markers) *vs.* Q_{ST} (for a neutral trait determined additively by two QTL) after 100 generations for 100 runs of an individual-based simulation with $d = 2$ initially identical demes of $N = 100$ individuals each, with a mean of $Nm = 0.1$ migrants per deme per generation.

contained 0, although its center was negative. Grouping the 100-generation runs by subpopulation size N , we found that E^* for $N = 100$ was -0.0379 , while E^* for $N = 500$ was 0.00243 . This suggests that even 100 generations may have been too long to yield positive values of E^* for very small populations; we recall that the relevant timescale for many population-genetic processes is in units of N generations (CROW and KIMURA 1970). However, 95% confidence intervals for E^* in both the $N = 100$ and the $N = 500$ cases included 0, which is not surprising given the small number of cases (4 for each).

Besides comparing the means of Q_{ST} and G_{ST} via the difference E^* , it is useful to examine the joint distribution of these two quantities across replicate runs with the same parameters and initial data, and in particular to compare their standard deviations. In Figures 2 and 3, we plot G_{ST} against Q_{ST} for all replicate runs with two different initializations and parameter sets. In both plots, though Q_{ST} and G_{ST} have similar means, Q_{ST} is much more variable than G_{ST} . This was typical for the conditions simulated. Indeed, the ratio of standard deviations $\sigma(Q_{ST})/\sigma(G_{ST})$ was 3.78 for the set of runs that used $d = 10$ demes of $N = 500$ individuals each, with 10 QTL and 10 markers (depicted in Figure 3), and had a median of 1.59 for the other sets of runs (excluding one set for which all but 4 runs led to near-complete fixation and thus undefined Q_{ST} and/or G_{ST} ; range $[0.0823, 2.08]$, with 12 of 15 sets yielding values > 1). A two-sided sign test comparing $\sigma(Q_{ST})$ with $\sigma(G_{ST})$ for the 16 sets of runs with meaningful output rejected the null hypothesis that the probability of finding $\sigma(Q_{ST}) > \sigma(G_{ST})$ was 50% ($P = 0.0213$). This finding is of practical importance in that the relatively large variability of Q_{ST} implies a strong correlation between Q_{ST} and the difference $Q_{ST} - G_{ST}$ that can aggravate bias in $Q_{ST} - G_{ST}$

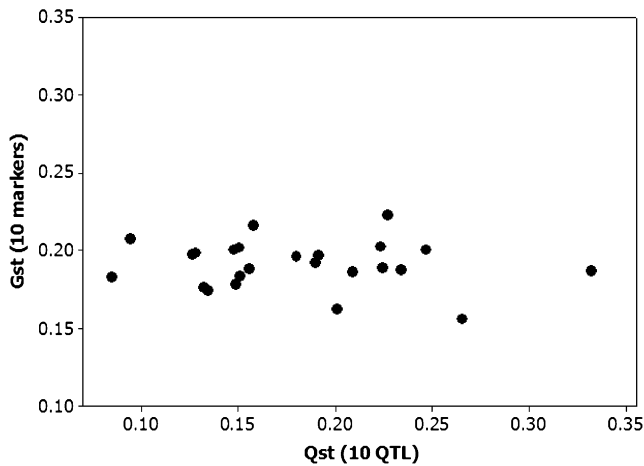


FIGURE 3.— G_{ST} (using 10 markers) *vs.* Q_{ST} (for a trait determined by 10 QTL) after 1000 generations for 24 runs of a simulation as in Figure 2 but with $d = 20$, $N = 500$, and $Nm = 1.0$.

comparisons (see the description of the deme-based simulations below, as well as the DISCUSSION). It should be noted that the difference between $\sigma(Q_{ST})$ and $\sigma(G_{ST})$ is sensitive to the number of independent markers used to calculate G_{ST} . Indeed, when G_{ST} was recalculated using a single marker, the hypothesis that $\Pr(\sigma(Q_{ST}) > \sigma(G_{ST})) = 0.5$ could not be rejected ($P = 0.2101$).

Finally, we used the individual-based simulations to assess the importance of θ_w , the ratio of within-deme trait variance attributable to covariances between loci to within-deme trait variance due to individual loci, in Equation 20 for Q_{ST} . To do so, we recalculated Q_{ST} for each run, using a formula that was the equivalent of (20) with θ_w set to 0, as would be expected for a neutral trait under random mating (ROGERS and HARPENDING 1983). The resulting mean value of Q_{ST} (over all runs for a given set of parameters) changed by $\leq 1.4\%$ and the standard deviation of Q_{ST} changed by $\leq 2.3\%$, with no clear bias toward positive or negative changes. The mean difference $Q_{ST} - G_{ST}$ changed by $\leq 6.6\%$, again with no clear directional bias. Furthermore, in no case was the sign of mean $Q_{ST} - G_{ST}$ changed from positive to negative, or vice versa, by setting $\theta_w = 0$. Thus, the main findings discussed above would remain unchanged if θ_w were ignored. This justifies the use of (computationally much less expensive) deme-based simulations in which θ_w is assumed to equal 0. We now outline the results of such simulations.

Deme-based simulations: For the deme-based simulations, a single additive quantitative trait was simulated with contributions from 10 diallelic loci, each with a phenotypic effect that could be specified individually. Simulations used equal phenotypic effects at all QTL, equal phenotypic effects at all QTL except for one or two loci of major effect (*e.g.*, 10 times larger than the next largest effect), or QTL effects drawn from a gamma

distribution. Ten additional independent diallelic loci not contributing to the quantitative trait were used to estimate F_{ST} . This is approximately the mean number of loci used to estimate F_{ST} in the studies summarized by MERILÄ and CRNOKRAK (2001). All loci, whether QTL or marker, were assumed to be independent, with a recombination rate of 0.5 for all pairs of loci.

The effective population size of individual demes (N) was set to either 50 or 20, and the rate of gene flow among demes (m) was set to 0.1, 0.025, 0.01, 0.001, or 0.0001 to yield six values of Nm ranging from 5 to 0.002. Allele frequencies were initialized at 0.5 for all QTL and marker loci in the results presented (although several initial allele-frequency schemes including random frequencies at all loci were examined). Simulations for 100 replicates of each set of conditions were run for 1000 generations for Nm between 5 and 0.02 and 10,000 generations for Nm of 0.002 in populations made up of 20 or 200 total demes.

In this section, we focus primarily on the results from simulations with 20 demes and equal effects. (We remind readers that while sampling from 200 demes is virtually unheard of, we are not simulating sample data but rather entire populations, which may more plausibly contain large numbers of demes.) Simulation code was written and executed in Matlab 7.0.1.

Since these simulations were carried out at the level of allele frequencies within each deme as opposed to the level of genotypes for each individual, large numbers of demes, individuals, and QTL could be simulated in a reasonable amount of computational time. As a result of the deme-level simulation, however, the allelic covariance among loci within individuals (θ_w in Equation 14) could not be estimated. Since θ_w is expected to approach zero under neutrality as the number of loci and demes grows large (ROGERS and HARPENDING 1983), and since analysis of individual-based simulations suggested that θ_w could be ignored without meaningful changes in the simulation results (see above), θ_w was set to zero in calculations of Q_{ST} based on Equation 20. This assumption would not be warranted in simulations of a quantitative trait under natural selection, since selection will cause correlations in allele frequencies between loci both within individuals and between demes (LATTA 1998, 2003; LE CORRE and KREMER 2003). Estimates of Q_{ST} for simulated populations did explicitly incorporate the allelic covariance over all pairs of loci between demes by calculating θ_B using Equation 11.

Our analysis above assumed a 2-deme model; these simulations used at least 20 demes (5-deme simulations often reached global fixation and loss for $Nm = 0.002$). The genetic basis of the modeled quantitative trait was 2 loci in the analysis and 10 loci in the deme-based simulations. Nevertheless, the results of the deme-based simulations were consistent with those of the individual-based simulations and thus indicated that those conclusions should apply even in situations involving large

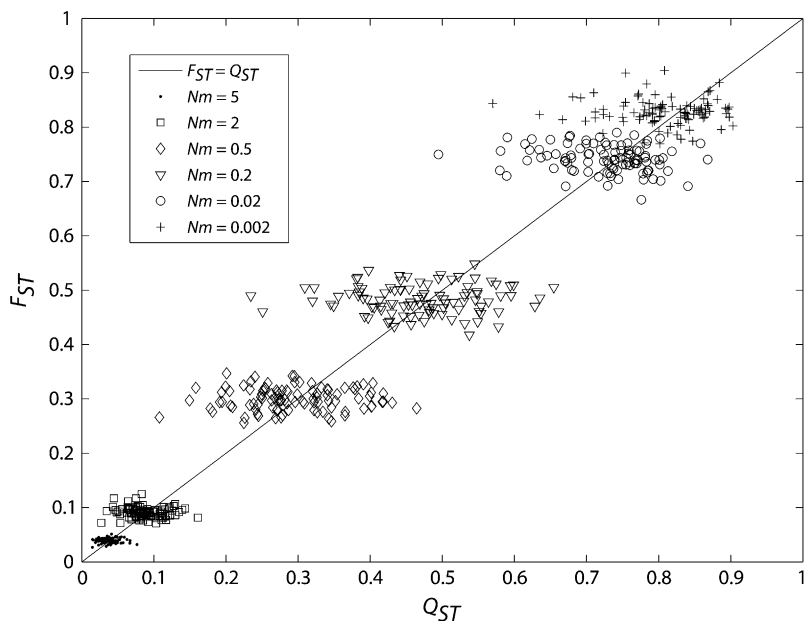


FIGURE 4.—The joint distribution of differentiation at the quantitative trait (Q_{ST}) and individual loci (F_{ST}) from finite-island model simulations carried out over a range of values for the product of the effective population size of each deme (N) and the rate of gene flow into each deme per generation (m). Q_{ST} shows more variation than F_{ST} for a given value of Nm . Q_{ST} and F_{ST} are each based on 10 independent diallelic loci, the quantitative trait loci had equal phenotypic effects, 100 replicate simulations were carried out for each Nm value, and the total population was 20 demes. Values at the 10,000th generation are shown for $Nm = 0.002$ and at the 1000th generation for all other Nm values.

populations, large numbers of demes, numerous QTL, and/or large numbers of markers. Here we focus on both the distribution of $F_{ST} - Q_{ST}$ and the joint distribution of F_{ST} and Q_{ST} .

The joint distribution of F_{ST} and Q_{ST} from the 100 replicate runs for each value of Nm is shown in Figure 4. The mean difference $E[Q_{ST} - F_{ST}]$ is negative (on average across all runs for a given Nm), though not significantly so due to the large variance. Furthermore, $E[Q_{ST} - F_{ST}]$ is lower (*i.e.*, more negative) for lower values of Nm , which may be thought of as modeling longer evolution in allopatry if time is measured in units of Nm generations. This is consistent with the analysis in the previous section that predicted negative values of $E[Q_{ST} - F_{ST}]$ after long evolution with $Nm = 0$.

In relative terms, the absolute value of $E[Q_{ST} - F_{ST}]$ in 20-deme simulations ranged from $\sim 25\%$ of $E[F_{ST}]$ for $Nm = 5$ to $\sim 6\%$ of $E[F_{ST}]$ for $Nm = 0.002$. In 200-deme simulations, the range was from $\sim 10\%$ for $Nm = 5$ to $\sim 2\%$ for $Nm = 0.002$. These values suggest the relative error that would be incurred in estimating Q_{ST} from F_{ST} .

Figure 4 also illustrates that the evolutionary variance of both F_{ST} and Q_{ST} among replicate simulation runs for a given parameter set is greatest for intermediate values of Nm , which result in intermediate values of F_{ST} and Q_{ST} , and least when F_{ST} and Q_{ST} are near zero or one. At all levels of Nm , the evolutionary variance for Q_{ST} is markedly greater than for F_{ST} .

The joint distribution of $Q_{ST} - F_{ST}$ and Q_{ST} from the same set of 100 replicate runs for each value of Nm is shown in Figure 5. We include this figure to emphasize that there is a strong positive correlation between the magnitude of the difference between Q_{ST} and F_{ST} and the value of Q_{ST} itself. This correlation arises because Q_{ST} has much greater evolutionary variance and there-

fore makes a much greater contribution to the evolutionary variance of the difference than does F_{ST} . In effect, subtracting F_{ST} from Q_{ST} simply adds “noise” to the (obviously perfect) positive correlation between Q_{ST} and itself.

The mean of $Q_{ST} - F_{ST}$ is shown as a function of time in Figure 6, with the standard deviation of $Q_{ST} - F_{ST}$ shown extending above and below the mean. The mean of $Q_{ST} - F_{ST}$ for the replicate simulations hovers near zero for all values of Nm . Unless Nm is large, there is substantial variability in $Q_{ST} - F_{ST}$ across replicate evolutions of the quantitative trait. Any temporal trend in $Q_{ST} - F_{ST}$ is weak compared to the amount of evolutionary variance among replicates. The confidence intervals for smaller values of Nm are somewhat asymmetric around zero, with a larger number of replicates having negative values of $Q_{ST} - F_{ST}$.

Figures 4–6 show values for $Nm = 0.002$ at 10,000 rather than 1000 generations. With such a low rate of gene flow, elimination of short-term transient effects on F_{ST} and Q_{ST} takes many more generations than for higher values of Nm . (We note that F_{ST} in a finite-island model with no mutation eventually goes to zero as the entire population approaches fixation or loss for all alleles. Hence the point at which the simulations were stopped is somewhat arbitrary, since there is no nonzero equilibrium value of F_{ST} as there is in the infinite-island model.)

As expected in a finite-island model, variance in $Q_{ST} - F_{ST}$ among replicates was markedly greater in simulations with 20 demes than in simulations with 200 demes. Specifically, the standard deviation of $Q_{ST} - F_{ST}$ was approximately three times greater in simulations with 20 demes than with 200 demes (~ 0.05 – 0.10 with 20 demes and ~ 0.02 – 0.03 with 200 demes for Nm values

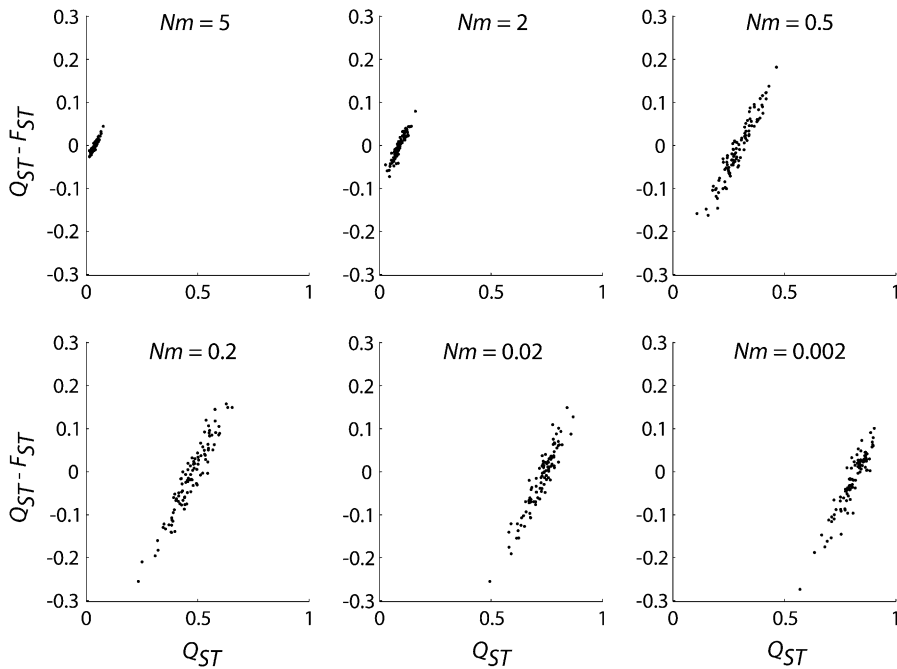


FIGURE 5.—The joint distribution of differentiation at the quantitative trait (Q_{ST}) by the difference between differentiation at the quantitative trait and individual loci ($Q_{ST} - F_{ST}$). The difference between Q_{ST} and F_{ST} is positively correlated with Q_{ST} . Q_{ST} and F_{ST} are each based on 10 independent diallelic loci, the quantitative trait loci had equal phenotypic effects, 100 replicate simulations were carried out for each Nm value, and the total population was 20 demes. Values at the 10,000th generation are shown for $Nm = 0.002$ and at the 1000th generation for all other Nm values. These graphs use the same simulation data sets as Figure 4.

between 0.5 and 0.002). Simulations run with 20 and 200 demes also exhibited roughly the same large positive correlation between Q_{ST} and $Q_{ST} - F_{ST}$ (correlation coefficient ≥ 0.92 with both 20 and 200 demes). On the other hand, the mean values over replicate simulations of the difference $Q_{ST} - F_{ST}$ were similarly slightly negative in simulations with 20 and 200 total demes, although with 200 demes there was less random

fluctuation. Simulations using a quantitative trait architecture with a single major gene yielded similar results in 20 and 200 demes as well (mean $Q_{ST} - F_{ST}$ over replicates slightly negative; standard deviation of $Q_{ST} - F_{ST}$ over replicates ~ 0.05 – 0.10 with 20 demes and ~ 0.02 – 0.03 with 200 demes for Nm values between 0.5 and 0.002; correlation between Q_{ST} and $Q_{ST} - F_{ST} \geq 0.94$).

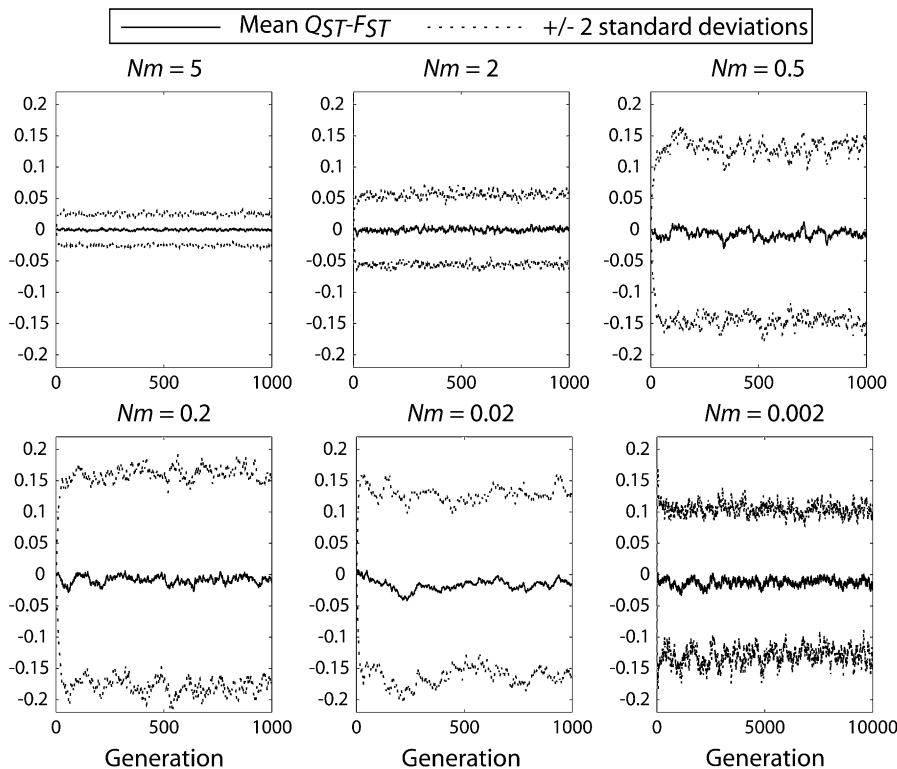


FIGURE 6.—Mean difference between differentiation at the quantitative trait and individual loci ($Q_{ST} - F_{ST}$) over 100 replicate simulations along with two standard deviations above and below the mean. Q_{ST} and F_{ST} are each based on 10 independent diallelic loci, the quantitative trait loci had equal phenotypic effects, 100 replicate simulations were carried out for each Nm value, and the total population was 20 demes. These graphs use the same simulation data sets as Figures 4 and 5.

DISCUSSION

Our rigorous analysis of the distributions of Q_{ST} and F_{ST} among evolutionarily replicate populations of two demes each shows that mean Q_{ST} need not equal mean F_{ST} even for a neutral quantitative trait with additive gene action. Indeed, mean Q_{ST} for such traits is expected to exceed mean F_{ST} when a panmictic population has recently been subdivided into separate demes, while mean Q_{ST} is expected to be less than mean F_{ST} in demes that have evolved in isolation for a long time. Nonlinear effects such as dominance and epistasis played no role in our models. Thus, all that is required for F_{ST} and Q_{ST} to differ in expectation is that not all loci contribute equally to the trait. Our simulations confirm that the mean difference between Q_{ST} and F_{ST} need not be zero in the model under study, although the large variance of both Q_{ST} and F_{ST} resulted in relatively low statistical power to detect this. The deviation of the mean difference $E[Q_{ST} - F_{ST}]$ from zero may not always be large enough to cause practical difficulties (see, for example, WHITLOCK 2008), but our simulations show that it can sometimes be substantial, *e.g.*, as large (in absolute value) as 25% of $E[F_{ST}]$.

An obvious question is why the analytical findings presented here differ from those of previous studies (LANDE 1992; WHITLOCK 1999), which found that $Q_{ST} = F_{ST}$ in expectation for an additive trait under neutrality. The answer appears to be that previous work (LANDE 1992; WHITLOCK 1999) calculated the expected values of $Q_{ST} = F_{ST}$ by assuming that the mean of a ratio of two quantities is equal to the ratio of their means. This assumption is strictly correct only when the variance of the denominator is zero (or in certain more specialized, arcane examples). For example, WHITLOCK (1999) noted that expressions obtained for components of additive genetic variance are expectations across (evolutionarily) replicate populations and that variance components in any given population may differ from these expectations. Q_{ST} is computed as a quotient of variance components. However, since Q_{ST} is a nonlinear function of the variance components [call it $f(V)$], it is not necessarily the case that $E[Q_{ST}] = f(E[V])$, where E denotes an expectation. The same generalization holds for F_{ST} . Thus, the result obtained by WHITLOCK (1999) is strictly applicable only in the case where variance components are identically equal to their expectations and not when evolutionary variance, *i.e.*, variance among replicate outcomes of the evolutionary process, is taken into account. Similar considerations hold true for the calculations of LE CORRE and KREMER (2003). In addition, the results of LE CORRE and KREMER (2003) do not account for stochastic variability in the values of F_{ST} and F_{IS} among different loci and further assume additive effects to be equal for all loci.

We also note that while sampling from a larger number of subpopulations will improve the precision

of individual Q_{ST} estimates, the impact of evolutionary variance can be evaluated only by sampling multiple independent quantitative traits. This means that testing the hypothesis that observed Q_{ST} is consistent with neutral evolution will be difficult for an individual quantitative trait. In short, because the evolutionary variance of $Q_{ST} - F_{ST}$ is large, tests for the presence of selection on an individual trait will have little power. Details of the population biology of a specific study species will also affect the degree to which Q_{ST} for one or a few quantitative traits may be perceived as larger than F_{ST} . Because the evolutionary variance of both F_{ST} and Q_{ST} is larger in populations with fewer subpopulations, differences between estimated Q_{ST} and F_{ST} for different species could simply be a function of differences in population structure rather than of differences in selection pressures.

Our simulations also indicate that Q_{ST} has a larger evolutionary variance than F_{ST} . This was noted by ROGERS and HARPENDING (1983), but seems to have escaped notice more recently. (However, it has been observed that the sampling variance of estimates of Q_{ST} is greater than that of estimates of F_{ST} ; see, for example, KOSKINEN *et al.* 2002; PALO *et al.* 2003; O'HARA and MERILÄ 2005.) The difference in evolutionary variance may be at least partially due to the fact that F_{ST} can be calculated using multiple independent markers, thus reducing "evolutionary sampling variability." This finding is reminiscent of a result due to ROGERS and HARPENDING (1983, p. 992), who showed theoretically that "the variance of the [between-group] variance is the same for polygenic characters as for single loci."

As a consequence, there is a strong positive correlation between Q_{ST} and the difference $Q_{ST} - F_{ST}$. This correlation has noteworthy implications for empirical comparisons of Q_{ST} and F_{ST} , since quantitative traits are not often sampled at random for comparisons of Q_{ST} and F_{ST} . In fact, traits exhibiting large (or less commonly small) phenotypic differences among subpopulations are often the focus of detailed investigation with a working hypothesis that natural selection is responsible for their high level of variation among subpopulations. If traits varying widely among subpopulations are preferentially chosen for study *because of their degree of variation*, then in statistical jargon a "selection bias" (having nothing to do with natural selection) exists in favor of studying such traits. Large phenotypic differences among subpopulations should often correspond to high Q_{ST} values and therefore to high values of $Q_{ST} - F_{ST}$ for neutral traits. Thus selection bias, if present, will result in a preponderance of empirical studies that show $Q_{ST} > F_{ST}$ even if all the traits studied are in fact evolving neutrally.

The results presented here pertain only to neutral traits and do not contradict earlier results that predict differences in Q_{ST} and F_{ST} for traits under selection as in LE CORRE and KREMER (2003). However, our results do

raise the question of what constitutes the proper test statistic and sampling distribution to use in tests to identify diversifying or stabilizing selection among subpopulations. Indeed, if $Q_{ST} > F_{ST}$ can be observed in neutrally evolving traits, what values of $Q_{ST} - F_{ST}$ should lead us to reject the null hypothesis of neutrality and conclude that natural selection is acting?

The gloomiest possible conclusion would be that $Q_{ST} - F_{ST}$ comparisons cannot readily be used to test for the action of selection on a particular trait, because (a) the proper null hypothesis for the test is not clear, (b) determining the proper sampling distribution from which to calculate the P -value for the test would require knowledge of Q_{ST} for many independent neutral traits, and (c) the large evolutionary variance of the difference $Q_{ST} - F_{ST}$ will limit the power of the test. However, the basic idea behind the use of $Q_{ST} - F_{ST}$ comparisons is surely valid: atypically high or low between-population trait variance (as measured by Q_{ST}) is evidence of diversifying or uniform selection, and since such variance depends on demographic parameters like migration, marker data (e.g., marker F_{ST}) should help to establish the ranges of “typical” and “atypical” values for Q_{ST} . We suspect that it will be possible to improve upon a $Q_{ST} - F_{ST}$ comparison test by comparing Q_{ST} with some function of data obtained from numerous markers distributed throughout the genome. However, this function might not correspond in a straightforward way to marker F_{ST} . Further work will be needed to find an optimal way to combine trait and marker data in a hypothesis test for the action of natural selection on quantitative traits in subdivided populations.

We thank S. Rottenstreich for discussion and comment, M. Whitlock for comments on a preliminary version of the DEFINITIONS AND THEORY and the ANALYTICAL EXAMPLES sections, and A. Miles and J. Cannata of Georgetown University’s Advanced Research Computing unit for technical assistance. This research was supported by National Science Foundation grant DMS-0201173 to J.R.M. and M.B.H.

LITERATURE CITED

- BARUCH, Z., J. M. NASSAR and J. BUBIS, 2004 Quantitative trait, genetic, environmental, and geographic distances among populations of the C_4 grass *Trachypogon plumosus* in neotropical savannas. *Div. Dist.* **10**: 283–292.
- CONOVER, D. O., L. M. CLARKE, S. B. MUNCH and G. N. WAGNER, 2006 Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *J. Fish Biol.* **69**: 21–47.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- GOUDET, J., and L. BÜCHI, 2006 The effects of dominance, regular inbreeding and sampling design on Q_{ST} , an estimator of population differentiation for quantitative traits. *Genetics* **172**: 1337–1347.
- JOHANSSON, M., C. R. PRIMMER and J. MERILÄ, 2007 Does habitat fragmentation reduce fitness and adaptability? A case study of the common frog (*Rana temporaria*). *Mol. Ecol.* **16**: 2693–2700.
- KIMURA, M., 1964 Diffusion models in population genetics. *J. Appl. Probab.* **1**: 177–232.
- KNOPP, T., J. M. CANO, P. A. CROCHET and J. MERILÄ, 2007 Contrasting levels of variation in neutral and quantitative genetic loci on island populations of moor frogs (*Rana arvalis*). *Conserv. Genet.* **8**: 45–56.
- KOSKINEN, M. T., T. O. HAUGEN and C. R. PRIMMER, 2002 Contemporary Fisherian life-history evolution in small salmonid populations. *Nature* **419**: 826–830.
- LANDE, R., 1992 Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution* **46**: 381–389.
- LATTA, R. G., 1998 Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *Am. Nat.* **151**: 283–292.
- LATTA, R. G., 2003 Gene flow, adaptive population divergence and comparative populations structure across loci. *New Phytol.* **161**: 51–58.
- LE CORRE, V., 2005 Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Mol. Ecol.* **14**: 4181–4192.
- LE CORRE, V., and A. KREMER, 2003 Genetic variability at neutral markers, quantitative trait loci and traits in a subdivided population under selection. *Genetics* **164**: 1205–1219.
- LEINONEN, T., R. B. O’HARA, J. M. CANO and J. MERILÄ, 2008 Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *J. Evol. Biol.* **21**: 1–17.
- LÓPEZ-FANJUL, C., A. FERNÁNDEZ and M. A. TORO, 2003 The effect of neutral nonadditive gene action on the quantitative index of population divergence. *Genetics* **164**: 1627–1633.
- McKAY, J. K., and R. G. LATTA, 2002 Adaptive population divergence: markers, QTL and traits. *Trends Ecol. Evol.* **17**: 285–291.
- MERILÄ, J., and P. CRNOKRAK, 2001 Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* **14**: 892–903.
- MORGAN, T. J., M. A. EVANS, T. GARLAND, J. G. SWALLOW and P. A. CARTER, 2005 Molecular and quantitative genetic divergence among populations of house mice with known evolutionary histories. *Heredity* **94**: 518–525.
- NEI, M., 1973 Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**: 3321–3323.
- O’HARA, R. B., and J. MERILÄ, 2005 Bias and precision in Q_{ST} estimates: problems and some solutions. *Genetics* **171**: 1331–1339.
- PALO, J. U., R. B. O’HARA, A. T. LAUGEN, A. LAURILA, C. R. PRIMMER *et al.*, 2003 Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Mol. Ecol.* **12**: 1963–1978.
- PONS, O., and K. CHAOUACHE, 1995 Estimation, variance and optimal sampling of gene diversity. II. Diploid locus. *Theor. Appl. Genet.* **91**: 122–130.
- PORCHER, E., T. GIRAUD, I. GOLDRINGER and C. LAVIGNE, 2004 Experimental demonstration of a causal relationship between heterogeneity of selection and genetic differentiation in quantitative traits. *Evolution* **58**: 1434–1445.
- PORCHER, E., T. GIRAUD and C. LAVIGNE, 2006 Genetic differentiation of neutral markers and quantitative traits in predominantly selfing metapopulations: confronting theory and experiments with *Arabidopsis thaliana*. *Genet. Res.* **87**: 1–12.
- ROBERGE, C., H. GUDERLEY and L. BERNATCHEZ, 2007 Genomewide identification of genes under directional selection: gene transcription Q_{ST} scan in diverging Atlantic salmon subpopulations. *Genetics* **177**: 1011–1022.
- ROGERS, A., and H. HARPENDING, 1983 Population structure and quantitative characters. *Genetics* **105**: 985–1002.
- SAINT-LAURENT, R., M. LEGAULT and L. BERNATCHEZ, 2003 Divergent selection maintains adaptive differentiation despite high levels of gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchell). *Mol. Ecol.* **12**: 315–330.
- SPITZE, K., 1993 Population structure in *Daphnia obtusa*: quantitative genetic and allozyme variation. *Genetics* **135**: 367–374.
- STEINGER, T., P. HALDIMANN, K. A. LEISS and H. MÜLLER-SCHÄRER, 2002 Does natural selection promote population divergence? A comparative analysis of population structure using amplified

- fragment length polymorphism markers and quantitative traits. *Mol. Ecol.* **11**: 2583–2590.
- WEIR, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
- WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- WHITLOCK, M. C., 1999 Neutral additive genetic variance in a meta-population. *Genet. Res.* **74**: 215–221.
- WHITLOCK, M. C., 2008 Evolutionary inference from Q_{ST} . *Mol. Ecol.* **17**: 1885–1896.
- WRIGHT, S., 1951 The genetical structure of populations. *Ann. Eugen.* **15**: 323–354.

Communicating editor: J. WAKELEY