

Testing Natural Selection vs Genetic Drift in Phenotypic Evolution Using Quantitative Trait Locus Data

H. Allen Orr

Department of Biology, University of Rochester, Rochester, New York 14627

Manuscript received January 20, 1998
Accepted for publication April 27, 1998

ABSTRACT

Evolutionary biologists have long sought a way to determine whether a phenotypic difference between two taxa was caused by natural selection or random genetic drift. Here I argue that data from quantitative trait locus (QTL) analyses can be used to test the null hypothesis of neutral phenotypic evolution. I propose a sign test that compares the observed number of plus and minus alleles in the "high line" with that expected under neutrality, conditioning on the known phenotypic difference between the taxa. Rejection of the null hypothesis implies a role for directional natural selection. This test is applicable to any character in any organism in which QTL analysis can be performed.

TO determine if a genetic difference is adaptive, we typically consider its phenotypic consequences. We might ask, for instance, if different alleles at *PGI* in *Colias* affect flying time or if *Adh^F* survives better than *Adh^S* among *Drosophila* in wine cellars. Here I consider the reverse possibility: Can we show that a phenotypic difference is adaptive by looking at the genes involved?

The idea is simple. Imagine that two plant varieties differ in height by 50 mm and that (an unusually powerful) quantitative trait locus analysis reveals that this difference is due to 50 quantitative trait loci (QTL). QTL analysis also reveals the direction and magnitude of each factor's phenotypic effect. Although many QTL in the tall line are likely to be plus ("tall") factors, some may be minus ("short"). Indeed, QTL studies have shown that mixtures of plus and minus factors in the high line are common (Tanksey 1993). If our analysis revealed that the tall line carries 30 plus factors and 20 minus factors, we might entertain the possibility that the varieties arrived at different heights by chance: during divergence, the tall line happened to accumulate a few more plus than minus factors by genetic drift. But if *all* 50 QTL in the tall line are plus, it becomes very difficult to believe the height difference reflects chance. Instead, the phenotypic difference likely reflects a history of directional natural selection. Thus, as Coyne (1996) and Laurie and colleagues (Laurie *et al.* 1997; True *et al.* 1997) have hinted, an unusual concentration of plus alleles in the high line might serve as a footprint for directional natural selection.

The problem, however, is more subtle than it first appears. Because one variety is taller than the other, it is obvious that the tall line must contain *some* plus factors. But because the mere existence of the phenotypic

difference did not convince us of a role for natural selection, the necessary existence of some plus alleles in the high line cannot do so either. Instead, we must ask if the ratio of plus-to-minus alleles in the high line is more extreme than expected under neutrality *given* the known phenotypic difference.

The problem can be restated as follows. Under neutral evolution, we predict some distribution of phenotypic differences between taxa after a period of divergence (Lande 1976, 1977; Lynch and Hill 1986). But the cases chosen for QTL analysis are not a random sample of these neutral divergences: we perform QTL analysis only when differences are fairly pronounced, and such cases tend to involve a fair number of plus alleles in the high line. Thus, any unbiased test of neutrality must ask if the number of plus-to-minus alleles is more extreme than expected in that subset of neutral divergences showing a phenotypic difference as large as that seen.

I sketch such a test here. As we will see, given the phenotypic difference separating two taxa, the number of QTL, and the approximate distribution of QTL effects, we can find the probability that the observed number of plus factors would show up in the high line under the null hypothesis of neutrality.

THE NULL MODEL

Our null hypothesis is that the observed phenotypic difference between two taxa is neutral. In particular, we assume that all phenotypes in the neighborhood of the observed two populations have equal fitness (*e.g.*, Lande 1976) and thus that we can picture evolution as a random walk over a flat fitness surface.

The phenotypic difference between our two taxa is due to n loci. At any locus, genotypic values, G , are assigned as follows:

Author e-mail: haorr@urhlep.pas.rochester.edu

Genotype: *BB Bb bb*
G: *a d -a*

The quantities *a* and *d* can take different values at different loci. When *d* = 0, there is no dominance, and the heterozygote is intermediate between the homozygotes. The test described below makes no assumption about dominance.

Populations *P*₁ (the high line) and *P*₂ (the low line) are assumed to be homozygous at all *n* loci. The *i*th locus in *P*₁ might carry a plus allele (of effect *G*_{1*i*} = *a*) or a minus allele (*G*_{1*i*} = *-a*). With no epistasis, the mean phenotypes are $\bar{P}_1 = \sum_{i=1}^n G_{1i}$ and $\bar{P}_2 = \sum_{i=1}^n G_{2i}$ and the phenotypic difference between populations is $\bar{P}_1 - \bar{P}_2 = 2 \sum^n G_{1i}$.

Because QTL analysis is imperfect, detecting only a subset of the genes causing a phenotypic difference, in practice *n* will refer to the number of factors detected in an actual QTL analysis of two particular lines. Similarly, we will usually replace $\bar{P}_1 - \bar{P}_2$ with *R*, the phenotypic difference between two lines that are homozygous for the appropriate alleles at the *n* loci actually found in QTL analysis: $R = 2 \sum^n G_{1i}$.

Assume that the absolute value of QTL effects, *|a|*, for a trait are drawn from some probability density *f*(*|a|*). The choice of distribution does not matter for our purposes, as long as it can be written down. In practice, the distribution used will be decided by best fit to the data.

Under our null hypothesis of neutrality, note that:

- At each locus, each line has a 1/2 chance of fixing the plus allele by genetic drift.
- The distribution of fixed allelic effects reflects the distribution of mutations available. If alleles of small effect are more common than those of large effect, that is because small mutations arise more often than large ones.

These two facts let us model neutral phenotypic evolution as an unbiased random walk in which some step sizes are more common than others.

QTL SIGN TEST

General model: Our question is simple: Given all the ways of neutrally evolving a phenotypic difference of *R* or more when drawing QTL from our distribution of effects, how often does one see the observed number of plus factors or more in the high line? (We consider differences of *R* or more because, presumably, we would have performed QTL analysis in either case.) If cases of neutral evolution yielding differences of *R* or more usually involve the observed number of plus factors, we have no reason to reject the null hypothesis. But if differences of *R* or more rarely involve such an extreme number of plus factors, we reject the null hypothesis.

To build our test, we must find the probability that

n₊ = 1, 2, . . . , *n* plus factors will show up in the high line under neutrality conditioned on a line difference of *R*. Our critical probability, *P*, is just the total probability of observing at least *n*₊ obs plus factors in the high line by chance given *R*. In symbols,

$$P = \sum_{i=n_{+obs}}^n P(n_{+} = i | 2 \sum G_1 \geq R). \tag{1}$$

If *P* < 0.05, we reject the null hypothesis.

By Bayes' theorem, the probability of finding *n*₊ = *i* plus alleles in the high line is

$$P(n_{+} = i | 2 \sum G_1 \geq R) = \frac{P(2 \sum G_1 \geq R | n_{+} = i) P(n_{+} = i)}{\sum_{j=0}^n P(2 \sum G_1 \geq R | n_{+} = j) P(n_{+} = j)}, \tag{2}$$

where *P*(*n*₊ = *j*) is the probability that *j* plus factors appear in the high line by chance. This is given by the binomial $P(n_{+} = j) = \binom{n}{j} \frac{1}{2}^n$.

The probability *P*(2∑*G*₁ ≥ *R* | *n*₊ = *j*) is slightly more difficult to find and is derived in the appendix. It is

$$P(2 \sum G_1 \geq R | n_{+} = j) = 0 \quad (\text{for } j = 0) \\ = \int_{-\infty}^0 \left[1 - F_{(j)+} \left(\frac{R - 2(n-j)\bar{G}_-}{2j} \right) \right] f_{(n-j)-}(\bar{G}_-) d(\bar{G}_-) \quad (\text{for } 0 < j < n) \\ = 1 - F_{(j)+} \left(\frac{R}{2n} \right) \quad (\text{for } j = n). \tag{3}$$

\bar{G}_+ and \bar{G}_- give the mean effects of the *j* plus factors and *n* - *j* minus factors, respectively, residing in the high line. *f*_{(*j*)₊}(\bar{G}_+) and *f*_{(*n-j*)₋}(\bar{G}_-) are the sampling distributions of these means, and *F*_{(*j*)₊}(\bar{G}_+) is the cumulative distribution function of *f*_{(*j*)₊}(\bar{G}_+).

Substituting into Equation 1, the critical probability *P* is

$$P = \frac{\sum_{i=n_{+obs}}^n \binom{n}{i} \int_{-\infty}^0 \left\{ 1 - F_{(i)+} \left[\frac{R - 2(n-i)\bar{G}_-}{2i} \right] \right\} f_{(n-i)-}(\bar{G}_-) d\bar{G}_-}{\sum_{j=1}^n \binom{n}{j} \int_{-\infty}^0 \left\{ 1 - F_{(j)+} \left[\frac{R - 2(n-j)\bar{G}_-}{2j} \right] \right\} f_{(n-j)-}(\bar{G}_-) d\bar{G}_-}. \tag{4}$$

The sampling distributions in Equations 3 and 4 depend on the distribution of QTL effects. If, for instance, QTL effects are gamma distributed—as often assumed (Zeng 1992)—the sampling distributions are related to the χ^2 . In particular, with gamma distributed effects, that is, $f(|a|) = \alpha e^{-\alpha|a|} (\alpha|a|)^{\beta-1} / \Gamma(\beta)$, the density of sample means is

$$f_{(j)+}(\bar{G}_+) = \frac{(j\alpha)^\beta e^{-j\alpha\bar{G}_+} \bar{G}_+^{\beta-1}}{\Gamma(j\beta)}, \tag{5}$$

where *j* is the number of plus factors drawn from our gamma (Hendricks 1956, p. 100). (The density of sam-

ple means when drawing $n - j$ minus factors can be found by replacing j with $n - j$ and taking \bar{G}_- to refer to the absolute value of the mean of the minus factors.) With gamma distributed QTL effects, one can therefore calculate the critical probability P exactly by numerical integration.

The biologically important point is simple. Knowing only the phenotypic difference separating two taxa, the number of QTL, and the approximate distribution of QTL effects, we can find the probability that some number of plus factors would show up in the high line by chance. The ratio of plus-to-minus factors residing in the high line, in other words, *can* serve as a footprint for natural selection.

Although we assumed for illustration that QTL effects are gamma distributed, the QTL sign test can be performed for any distribution of QTL effects. Exact calculation of P is possible whenever the sampling distribution of means is known (Lindgren 1976). But even if the observed distribution of QTL effects is exotic—and thus the corresponding sampling distributions unknown— P can always be found by Monte Carlo simulation.

A C program that calculates P for gamma distributed QTL is available from the author (see appendix). This program also remedies a potential problem with the above simple approach: it allows one to set a threshold for QTL detection. Thus, although QTL are drawn from a distribution having estimated scale (α) and shape (β) parameters, factors having an effect smaller than T are assumed to be undetectable and so are ignored. Use of a truncated distribution surely better captures the realities of QTL mapping. Although I ignore all parameter estimation problems in this note, this program can also obviously be used to assess the effect on P of variation in α and β about their estimated values.

Equal effects: It is worth considering a variation on the above model that might be appropriate in certain cases. Imagine that evolution is constrained to build phenotypes from factors of equal effect (*cf.* Wright 1968). Thus, $\bar{G}_- = \bar{G}_+$. For any combination of j plus factors and $n - j$ minus factors, we can obtain the observed R only if $\bar{G}_j = R/[2(2j - n)]$. But $R > 0$ only when $j > n/2$. In other words, one cannot obtain the observed phenotypic difference of R unless most plus factors reside in the high line.

Thus, the probability of seeing n_{+obs} or more plus factors in the high line conditioned on $R > 0$ equals the probability of seeing n_{+obs} or more plus factors conditioned on the majority of plus factors residing in the high line, and

$$P(n_+ \geq n_{+obs} | n_+ \geq n/2) = \frac{\sum_{i=n_{+obs}}^n \binom{n}{i}}{\sum_{j>n/2}^n \binom{n}{j}}, \quad (6)$$

where it is understood that $j > n/2$ refers to the smallest integer $> n/2$.

In the fortunate case where n is large, we can use a normal approximation:

$$P \approx 2 \left[1 - \Phi \left(\frac{n_{+obs} - n/2}{\sqrt{n/4}} \right) \right]. \quad (7)$$

If $P < 0.05$, we reject the neutral null hypothesis. A minimum of $n = 6$ factors must be detected by QTL analysis to reject the null hypothesis.

Although this test ignores all information from QTL analysis about the actual sizes of the factors involved, generously allowing factors to assume whatever \bar{G} is required to explain R (for a given j), it has one obvious merit: it requires only that we know n and n_{+obs} . Although this test is not as biologically realistic as the one described above, it is preferable to—and more conservative than—a simple sign test, which fails to condition on the necessary existence of plus factors in the high line.

EXAMPLES

I now consider three QTL data sets to demonstrate the use of the QTL sign test.

Tomato fruit mass: In one of the best-known QTL analyses, Paterson *et al.* (1991) dissected several phenotypic differences distinguishing the domestic tomato *Lycopersicon esculentum* from its wild relative, *L. cheesmanii*. They detected 11 QTL contributing to the large difference in fruit weight between these varieties. All 11 acted in the expected (plus) direction. Lines homozygous for the appropriate alleles at these 11 QTL would phenotypically differ by $R = \Sigma 2a = 2.27$ [units are \log_{10} (grams)]. [I have averaged across the two California environments considered by Paterson *et al.* (1991) when estimating a as differences due to environment were typically small; see their Table 2.]

Looking across the several characters studied, Paterson *et al.* (1991) found that factors of large phenotypic effect were rarer than those of small effect. The distribution of QTL effects appears roughly exponential, that is, a gamma distribution with $\beta = 1$ (see their Figure 6). (Because of the fairly small number of QTL found in this and the following example, it seems best, for purposes of illustration, to simplify our estimation problem by letting $f(|a|)$ be exponential.) We will assume, then, that fruit mass QTL were drawn from an exponential distribution having a mean heterozygous effect of about $1/\alpha \approx 0.10$.

When evolution is neutral and involves factors of mean effect of $1/\alpha \approx 0.10$, there is a small probability of finding all 11 plus alleles in the high line by chance alone: $P = 0.02$. We thus reject the neutral null hypothesis. This probability remains essentially unchanged for similar values of mean QTL effect and for any realistic QTL effect threshold ($0 < T < 0.025$, where 0.025 is

close to the heterozygous effect of the smallest QTL actually seen).

This result is, of course, hardly surprising. Tomato fruit mass has obviously been subjected to strong directional artificial selection.

Maize grain weight: Factors affecting grain weight between two inbred lines of maize were studied by Edwards *et al.* (1987, 1992). These lines dramatically differ in many characters. Although this study suffers some problems (*e.g.*, true interval mapping was not performed), it enjoys one strength: a very large number of F_2 progeny were scored and genotyped. Edwards *et al.* (1987) thus had the power to detect a fair number of QTL, some of small effect. Perhaps most impressive, 13 QTL affecting grain weight were found (Edwards *et al.* 1992). Of these, 11 acted in the expected (plus) direction and two did not (minus). Unfortunately, Edwards *et al.* (1987) presented QTL effects only in units of percent of F_2 variance explained. Although these values are strictly proportional to a only in the absence of dominance, I will assume this popular measure of phenotypic effect is nearly proportional enough to a for purposes of illustration.

The magnitude of QTL effects is approximately exponentially distributed for most traits studied (see Figure 3 in Edwards *et al.* 1987), including grain weight. QTL affecting grain weight have a mean heterozygous effect of $1/\alpha \approx 3.33$. (This value is computed using only those QTL of significant effect; if all QTL are used, a slightly different value is obtained.)

Conditioning on the phenotypic difference explained by the QTL, $n_+ = 11$ or more factors would result by chance alone $P = 0.19$ of the time. This result does not qualitatively depend on the precise value of the mean of QTL effects nor on any realistic threshold value ($0 \leq T \leq 1.5$). We cannot, therefore, reject the neutral null hypothesis.

Posterior lobe area in *Drosophila*: True *et al.* (1997) recently analyzed several male secondary sexual traits that distinguish *Drosophila simulans* from its close relative *D. mauritiana*. Although this study was small (only 200 F_2 males were scored), True *et al.* (1997) were able to map eight QTL affecting the area of one of the genital structures distinguishing these taxa, the posterior lobe. All eight factors act in the same (plus) direction. Together, the eight detected QTL explain $R \approx 86\%$ of the species difference in the posterior lobe area. (I have assumed no dominance and have re-expressed True *et al.*'s (1997) data in units of the whole species difference; their data were originally expressed in units of half the species difference. I have also taken into account the fact that one locus is X-linked; this distinction matters because the trait is expressed only in males.)

All eight QTL have roughly similar effects, ranging between $|a| = 2.5\%$ and 8% of the species difference (see also Laurie *et al.* 1997). Given the large error on these measures, it does not seem unreasonable to use

the equal-effects test with these data. Equation 6 shows that, conditioning on a majority of plus factors residing in the high line, the chance of finding all eight plus factors in the high line (*i.e.*, *D. simulans*) by chance is only $P = 0.011$. We therefore reject the null hypothesis. True *et al.* (1997) similarly conclude that the concentration of plus factors in one species seems too striking to reflect random divergence. Instead, the male posterior lobe is probably subject to directional sexual selection. This conclusion is supported, of course, by the rapid evolution of male genitalia in insects generally (Eberhard 1985).

CONCLUSIONS

Evolutionary biologists have long desired some method for determining if a phenotypic difference between taxa reflects adaptation or neutrality. Ironically, quantitative trait locus analysis was never intended as such a method. It is clear, however, that QTL data do provide information on the roles of natural selection *vs.* genetic drift in phenotypic evolution. Here I have suggested a test to extract this information. Although the power of the test is obviously constrained by the present power of QTL analyses, there is every reason to believe that future analyses will uncover more factors as well as provide less biased estimates of their effects. In principle, then, QTL analyses may routinely provide the information required to test for the action of directional natural selection.

It is important to understand what the QTL sign test does *not* test. In particular, we do not ask if a phenotypic difference can or cannot be explained by natural selection. As Lande (1976) emphasized in a similar context, any pattern of morphological change can be explained by the right form of selection acting at the right time. Instead, we must ask a tractable question: can the observed phenotypic difference be plausibly explained by random change? If not, we infer a role for directional natural selection.

To see why failure to reject the null hypothesis does not imply no role for selection, imagine that natural selection "built" the high line by fixing a major plus factor (which overshoots the optimum somewhat), followed by several smaller compensatory minus factors. It is very unlikely that we can reject the neutral null hypothesis in this case. Such a pattern is simply too common under neutral evolution. Regardless of any intuitions we may have about how selection acts, it will always be easier to reject the null hypothesis when more, rather than fewer, plus factors reside in the high line. Intuitively, it might also seem that selection would sometimes fix minus factors in the high line: linked minus factors could hitchhike to fixation during strong selection for a major plus factor. Although this could well occur, it should not complicate the present test: if natural selection could not separate the undesirable minus

factor from the linked desired plus factor, it seems very unlikely that our F_2 QTL analysis could do so.

It is also worth noting that rejection of the null hypothesis does not, strictly speaking, allow us to conclude that the analyzed character was the direct target of selection. One can never completely exclude the possibility that the measured character changed as a correlated response to selection (although this seems less plausible for the larger, and sometimes dramatic, character differences often considered in QTL analysis). In any case, rejection of the null hypothesis does demonstrate that the character's evolution was not neutral.

Several previous tests of neutral phenotypic evolution have been proposed (Lande 1976, 1977; Lynch and Hill 1986). The most popular of these are "rate" tests in which the observed phenotypic differences between populations are compared with those expected under neutrality (Lande 1976; Lynch and Hill 1986). The present test differs from rate tests in two important ways. First, in rate tests, the distribution of phenotypic differences expected under neutrality is inferred from quantitative genetic parameters. The parameters required varies with the individual rate test. In Lande's (1976) test, for instance, one must know the effective population size N_e and the heritability h^2 of a character to predict the distribution of phenotypic differences. In Lynch and Hill's (1986) test, one must know N_e and the mutational variance V_m . In both cases, the number of generations separating populations must also be known, as larger phenotypic differences are expected, given more time.

The present QTL test, however, does not require information on any quantitative genetic parameters or on time since separation. It requires only that information provided in standard QTL studies: the size of the phenotypic difference distinguishing two taxa, the number of QTL detected, and the direction and magnitude of QTL effects.

Second, rate tests are *a priori* in the sense that they predict a distribution of phenotypes expected under neutrality. The present test, however, is *a posteriori*: we begin with an observed phenotypic difference but know nothing of the distribution of differences expected under neutrality. The distinction matters because, in practice, we perform QTL analysis only on characters that are fairly different between taxa. Thus, although a Lande-Lynch-Hill analysis might predict mostly slight differences between taxa separated for a given length of time, these will not be the differences analyzed in QTL studies, as explained earlier.

Last, it is worth noting that the QTL sign test does not depend on any assumptions about dominance or about the shape of the distribution of QTL effects. As long as the parental lines are homozygous for QTL (which will typically be true for the inbred lines used in such analyses), the QTL sign test makes no assumption about heterozygous effect *per se*. Similarly, although one

must specify some distribution giving reasonably good fit to observed QTL effects, the QTL sign test can be performed no matter what distribution is deemed appropriate.

The most important assumption underlying the QTL sign test involves epistasis. Because the crux of the test involves asking if a given combination of plus and minus factors "adds up" to explaining the observed phenotypic difference, we obviously assume no epistasis. It is difficult to see how this assumption could be relaxed. To the extent, then, that epistasis is common and strong, the QTL sign test is limited. Fortunately, QTL analyses often reveal little epistasis (Tanksley 1993). In any case, we need not blindly make any assumption about additivity: in any particular case, QTL analysis itself will reveal if epistasis between mapped factors is pronounced. If not, the test can be safely performed.

Of course, it may prove possible to build related tests of natural selection that are less sensitive to epistasis. And it will surely prove possible to refine the simple test sketched here, increasing its biological realism and addressing the complications of parameter estimation. But the underlying notion characterizing the test seems sound. QTL data must contain information on the role of natural selection in phenotypic evolution. Our task is to devise biologically realistic tests to extract this information.

I thank J. A. Coyne, P. D. Keightley, D. Presgraves, and two anonymous reviewers for very helpful comments. I especially thank M. Turelli for his careful reading of the manuscript (as well as for catching an error). This work was supported by National Institutes of Health grant GM-51932 and by the David and Lucile Packard Foundation.

LITERATURE CITED

- Coyne, J. A., 1996 Genetics of a difference in male cuticular hydrocarbons between two sibling species, *Drosophila simulans* and *D. sechellia*. *Genetics* **143**: 1689-1698.
- Eberhard, W. G., 1985 *Sexual Selection and Animal Genitalia*. Harvard University Press, Cambridge, MA.
- Edwards, M. D., C. W. Stuber and J. F. Wendel, 1987 Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* **116**: 113-125.
- Edwards, M. D., T. Helentjaris, S. Wright and C. W. Stuber, 1992 Molecular-marker-facilitated investigations of quantitative trait loci in maize. *Theor. Appl. Genet.* **83**: 765-774.
- Hendricks, W. A., 1956 *The Mathematical Theory of Sampling*. The Scarecrow Press, New Brunswick, NJ.
- Lande, R., 1976 Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**: 314-334.
- Lande, R., 1977 Statistical tests for natural selection on quantitative characters. *Evolution* **31**: 442-444.
- Laurie, C. C., J. R. True, J. Liu and J. M. Mercer, 1997 An introgression analysis of quantitative trait loci that contribute to a morphological difference between *Drosophila simulans* and *D. mauritiana*. *Genetics* **145**: 339-348.
- Lindgren, B. W., 1976 *Statistical Theory*. Macmillan Publishing Co., New York.
- Lynch, M., and W. G. Hill, 1986 Phenotypic evolution by neutral mutation. *Evolution* **40**: 915-935.
- Paterson, A. H., S. Damon, J. D. Hewitt, D. Zamir, H. D. Rabinowitch *et al.*, 1991 Mendelian factors underlying quantitative

- traits in the tomato: comparison across species, generations and environments. *Genetics* **127**: 181–197.
- Tanksley, S. D., 1993 Mapping polygenes. *Annu. Rev. Genet.* **27**: 205–233.
- True, J. R., J. Liu, L. F. Stam, Z.-B. Zeng and C. C. Laurie, 1997 Quantitative genetic analysis of divergence in male secondary sexual traits between *Drosophila simulans* and *D. mauritiana*. *Evolution* **51**: 816–832.
- Wright, S., 1968 *Evolution and the Genetics of Populations*. University of Chicago Press, Chicago.
- Zeng, Z.-B., 1992 Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics* **131**: 987–1001.

Communicating editor: P. D. Keightley

APPENDIX

Derivation of $P(2\Sigma G_1 \geq R \mid n_+ = j)$: We find this probability as follows: $f_{(j)+}(\bar{G}_+)$ is the sampling distribution of means when drawing j plus factors from our distribution of QTL effects. Similarly, $f_{(n-j)-}(\bar{G}_-)$ is the sampling distribution of means when drawing $n - j$ minus factors. Assume that, in some particular case, the $n - j$ minus factors in P_1 have a mean effect of \bar{G}_- . Then there is some probability that the mean of the j plus factors will be large enough to give a line difference of $2\Sigma G_1 \geq R$. This probability is

$$1 - F_{(j)+}\left[\frac{R - 2(n - j)\bar{G}_-}{2j}\right], \quad (\text{A1})$$

where $F_{(j)+}[\bar{G}_+]$ is the cumulative distribution function of $f_{(j)+}(\bar{G}_+)$. To find the total probability that j plus factors and $n - j$ minus factors will yield a phenotypic difference of at least R , we weight the probability in Equation A1 by the frequency with which \bar{G}_- takes different values. Thus, we get

$$\begin{aligned} &P(2\Sigma G_1 \geq R \mid n_+ = j) \\ &= \int_{-\infty}^0 \left[1 - F_{(j)+}\left(\frac{R - 2(n - j)\bar{G}_-}{2j}\right)\right] f_{(n-j)-}(\bar{G}_-) d(\bar{G}_-), \end{aligned} \quad (\text{A2})$$

as in the text.

Computer programs: A computer program that performs the QTL sign test is available from the author. This program assumes that QTL effects are drawn from a gamma distribution. Source code (in C) and a stand-alone program for the Power Macintosh are available and can be downloaded at <http://www.rochester.edu/College/bio/orrlab/orrhome.html>. The program is a Monte Carlo, that is, it does not calculate probabilities by numerical integration. Although it is easy to calculate the critical P numerically, for example, via *Mathematica*, the Monte Carlo approach has the advantage of allowing one to easily test the effects of different QTL detection thresholds.