Evolution, 50(3), 1996, pp. 1062–1073

A MARKER-BASED METHOD FOR INFERENCES ABOUT QUANTITATIVE INHERITANCE IN NATURAL POPULATIONS

KERMIT RITLAND[1]
Department of Botany, University of Toronto, Toronto, Ontario M5S 3B2, Canada

Abstract.—A marker-based method for studying quantitative genetic characters in natural populations is presented and evaluated. The method involves regressing quantitative trait similarity on marker-estimated relatedness between individuals. A procedure is first given for estimating the narrow sense heritability and additive genetic correlations among traits, incorporating shared environments. Estimation of the actual variance of relatedness is required for heritability, but not for genetic correlations. The approach is then extended to include isolation by distance of environments, dominance, and shared levels of inbreeding. Investigations of statistical properties show that good estimates do not require great marker polymorphism, but rather require significant variation of actual relatedness; optimal allocation generally favors sampling many individuals at the expense of assaying fewer marker loci; when relatedness declines with physical distance, it is optimal to restrict comparisons to within a certain distance; the power to estimate shared environments and inbreeding effects is reasonable, but estimates of dominance variance may be difficult under certain patterns of relationship; and any linkage of markers to quantitative trait loci does not cause significant problems. This marker-based method makes possible studies with long-lived organisms or with organisms difficult to culture, and opens the possibility that quantitative trait expression in natural environments can be analyzed in an unmanipulative way.

Key words.—Genetic markers, heritability, quantitative genetics, relatedness.

Received December 30, 1994. Accepted August 7, 1995.

The covariance between relatives for a quantitative trait is the basis for estimating the heritability of the trait (Falconer 1989). Classically, the level of relationship between relatives is calculated from known pedigrees (Cannings and Thompson 1981). In unmanipulated natural populations, pedigrees are usually unknown, which prevents inferences about heritability. However, genetic markers provide information about relatedness between individuals of unknown pedigree (Morton et al. 1971; Thompson 1975; Lynch 1988; Queller and Goodnight 1989; Ritland 1996). In principle, the joint distribution of markers and quantitative traits should provide information about heritability and other genetic components of traits expressed in the field.

Knowledge of heritable variation in natural populations is important in many contexts. While many studies have examined the intensity of selection on phenotypes S in the field (Lande and Arnold 1983), information about the heritability of phenotypes, which determines the genetic response to selection (Falconer 1989), is lacking in most field studies of phenotypic selection. Magnitudes of heritabilities and genetic correlations also provide information relevant to rates and directions of short-term evolution (Dickerson 1955), historical patterns of natural selection (Lande 1979), the targets of natural selection in a suite of correlated characters (Price et al. 1984), and strategies for the conservation of genetic variability underlying quantitative traits.

Because quantitative traits are highly dependent upon the environment for their expression, field-based measures of quantitative inheritance are most desirable for nondomesticated species. An excellent approach for estimating natural heritabilities is cross-fostering in birds (Boag and Grant 1978; Smith and Dhondt 1980; van Noordwijk et al. 1980; Dhondt 1982; Alatalo and Lundber 1986). However, other species rarely present such opportunities for non-disruptive manipulation. In *Drosophila,* studies are based on regressing lab-raised progeny on wild-caught parents (Prout 1958; Coyne and Beecham 1987; Riska et al. 1989), and these usually provide only a lower bound for heritability. Studies with plants predominately involve hand-planting sibships in the field (c.f. Mitchell-Olds 1986; Shaw 1986). While manipulative experiments of wild populations are powerful for estimation and hypothesis testing (Mitchell-Olds and Shaw 1987), even the most careful treatments may affect the expression of quantitative characters.

This paper presents a marker-based approach for estimating heritabilities, genetic correlations, and other components of quantitative genetic variation in natural populations. Barring the location of individuals, the only disturbance involved is the harvest of tissue for assay of genotypes at marker loci. Unlike the method of Shaw (1987), this procedure is based on inferred relatedness, as opposed to known relationship, and on a linear modeling procedure, as opposed to maximum likelihood. The linear approach does not make assumptions about the distribution of relatedness, which normally spans a continuum in natural populations. A maximum likelihood procedure, suited for populations with prespecified, discrete classes of relatives, is presented in Mousseau et al. (unpubl.).

In the wild, the expression of quantitative traits is more complex than under experimental conditions. Both the environment and the pattern of relationship are more variable, and consequently the phenotypic similarity between relatives is often not simply a function of relatedness and heritability. Other factors that may contribute include dominance effects (broad-sense heritability), sharing of environments, and sharing of inbreeding combined with inbreeding depression. This paper attempts to include these factors in progressively more complex estimation models.

Part of the procedure presented herein involves estimating "actual" variance for coefficients of relatedness, for which new estimators are developed. This is a previously unexplored

[1] Present address: Department of Forest Sciences, University of British Columbia, 193-2357 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada.

aspect of population structure that in itself is a feature worthy of study. For adequate statistical power to infer heritability, the population should exhibit substantial actual variation of relatedness. For inferences about quantitative inheritance under these complex models, variation must be detected for more complex modes of genetic relationship. Lack of variation for these modes of relationship prohibits inferences about these complex inheritance and actually bias simple heritability estimates if variation is present but undetectable. Thus, estimates should be interpreted with caution due to possible undetected violations of assumptions, and comparisons should be made with estimates obtained via classic procedures. At least, empirical studies are needed to document patterns of relationship in real populations and their feasibility for use in this procedure. To this end, a companion paper applies this procedure to two populations of the common monkeyflower, *Mimulus guttatus* (Ritland and Ritland 1996).

The Estimation Procedure and Its Properties

We consider a single population in which individuals are assayed for both genetic markers and quantitative traits. The markers are assumed codominant and at least two are required (dominant markers can be used if suitable measures of relatedness are defined). For the purposes of developing a regression-based estimator, the sample consists of pairs of individuals (the same individual may be present in more than one pair, as discussed later). Let the value of a quantitative trait for the ith pair of individuals be Y_i for the first individual and Y_i' for the second. In the following, our genetic inferences will rest upon a quantity termed the "phenotypic similarity":

$$Z_i = \frac{(Y_i - U)(Y_i' - U)}{V} \tag{1a}$$

where U and V are the sample mean and variance of Y, respectively, in the population. Among all pairs, the average Z_i equals the phenotypic correlation.

The following models for the genetic and environmental basis of phenotypic similarity are developed in a sequential manner, from the simple to the complex. At the simplest, relatives share additive effects of genes. At the next level of complexity, relatives may share environmental effects (including maternal effects), and this may be a function of distance between individuals. At the third level, relatives may share dominant effects of genes. If individuals are inbred, a fourth level involves shared phenotypes due to inbreeding depression. By specifying how these factors combine to determine phenotypic similarity, we can extract these genetic and environmental covariances from the observed phenotypic similarities, given the appropriate estimates of relatedness.

As Weir and Cockerham have emphasized (1984), one must distinguish parameters from their estimates. Thus, in the following, estimated values are given by capital letters, while the true parameter values are given in the corresponding lowercase letters.

A Regression Estimator for Heritability

With purely additive genetic variation and no sharing of environment, the phenotypic similarity between two relatives A and B for a quantitative trait is

$$Z_i = 2r_i h^2 + e_i \tag{1b}$$

(c.f. Jacquard 1974, eq. 23; Falconer 1989, eq. 9.13), where h^2 is the narrow-sense heritability, r_i is the coefficient of kinship between this pair of individuals i, and e_i is an error term, due to random environmental effects, with zero expectation. The coefficient of kinship is the probability that two homologous alleles, one sampled from each individual, are identical-by-descent (Jacquard 1974). Equation (1b) stems from the fact that alleles at quantitative trait loci (QTL) are identical between relatives with probability r_i, that there are two alleles in diploids to consider; and that given an identity, the expected product equals the heritability (otherwise it is zero). For our purposes below, we will specifically assume that variation of r_i is uncorrelated with variation at QTL, e.g., there is a random association between loci used to estimate r_i and QTL.

Equation (1b) is a linear model where h^2 is the parameter to be estimated and r_i is an observable parameter. Normally, regression requires that the predicting variable (relatedness) be known without error, but with the proper precautions, we can replace r_i with its estimate from markers, R_i. For N pairs of relatives, standard regression theory gives the estimator for heritability as

$$\hat{h}^2 = \frac{C_{ZR}}{2V_r}. \tag{1c}$$

where C_{ZR} is the sample covariance between phenotypic similarity (Z_i) and estimated relatedness (R_i), and V_r is the *actual* variance of relatedness among all pairs i. This actual variance of relatedness, whose estimation is treated below, is less than the variance among estimates of relatedness, the latter of which includes statistical variance.

It can be shown that in a haploid population where two markers are assayed, this regression estimator for heritability is identical to the marker-based estimator for heritability given by Ritland (1989). The current work greatly increases the power of heritability estimation by providing the most efficient extension to data from more than two marker loci, and to diploids.

Note that a method-of-moments estimator can be obtained from eq. (1b) as

$$\hat{h}^2 = \frac{\bar{Z}}{2\bar{R}}. \tag{1d}$$

The problem with this approach is the relativity of relatedness estimates: in the absence of knowing the drift variance incurred by the formation of relationships within populations, the expection of all pairwise relatedness estimates from markers is zero (Ritland 1996). The only practical estimators for heritability in this context are those based on variation of actual relationship. Alternatively, if one assumes a mixture of unrelated and related individuals of prespecified degree, a mixture model can be combined with maximum likelihood to give estimates without resorting to explicit computation of actual variance of relatedness (see Mousseau et al., unpubl.).

In natural populations, the clustering of relatives in similar environments inflates the phenotypic correlation between relatives, and must be taken into account for unbiased estimates

of heritability. If environments are shared between relatives, the phenotypic similarity can be rewritten as

$$Z_i = 2r_i\, h^2 + r_e + e_i \tag{2a}$$

where r_e is the correlation of environmental effects between the individuals being compared. The joint estimator for heritability and environmental correlation is

$$\hat{h}^2 = \frac{C_{ZR}}{2V_r} \tag{2b}$$

$$\hat{r}_e = Z - R\hat{h}^2 \tag{2c}$$

where Z and R are the means of the Z_i and R_i.

If $h^4 \cong 0$, the statistical variance of the marker-based heritability estimate is approximately

$$\mathrm{Var}(\hat{h}^2) \approx \frac{1}{4NV_R}\left(1 + \frac{E_R}{V_R}\right) \tag{3}$$

where E_R is the estimation variance of relatedness. Of note is the importance of actual variance of relatedness in the accuracy of heritability estimates.

When true relatedness is low, the estimation variance of relatedness (E_R) based on n marker loci, each with m alleles, is about $1/[4n(m - 1)]$ (Ritland 1996). The actual variance of relatedness (V_R) depends highly on the taxa studied, but in monkeyflowers lies within the range of 0.0025 to 0.01 (Ritland and Ritland 1996; and unpub. data). This lower bound corresponds to a combination of ⅛ half-sibs and ⅞ unrelated, and the upper bound to ¼ full-sibs and ¾ unrelated. These values provide ballpark figures for the statistical variances to expect when using markers to infer heritability via eq. (3). Statistical properties of eq. (3) are studied in greater detail in a following section.

Estimation of Pairwise Relationship

Pairwise relatedness is notoriously difficult to estimate, with the maximum likelihood method suffering from biases with few marker loci. To reduce bias, method-of-moments estimators can be used (Ritland 1996). This approach is based upon the fact that between a pair of individuals denoted i, each allele j at each locus k gives an estimate of relatedness, R_{ijk}. The estimate of relatedness is found by averaging estimates over alleles j and loci k using weights W_{jk},

$$R_i = \sum_{j,k} W_{jk} R_{ijk} \tag{4a}$$

Formulae for computation of optimal weights are given in Ritland (1996). A simplified estimator with weights efficient at low levels of relatedness is

$$R_i = c \sum_{j,k} \frac{S_{ijk} - P_{jk}^2}{P_{jk}} \tag{4b}$$

where S_{ijk} is the fraction of alleles of type j at locus k shared between the two relatives, P_{jk} the estimated population frequencies of allele j at locus k (with pair i excluded from the calculation of P_{jk}), and

$$c = \left(\sum_k (n_k - 1)\right)^{-1}, \tag{4c}$$

for n_k the number of alleles at locus k.

Estimation of Actual Variance of Relationship

Actual variance of relatedness occurs due to the presence of a mixture of full-sibs, half-sibs, first-cousins, unrelated individuals, and so on. There are no published methods for estimating actual variance of relatedness with genetic markers. In this section, I describe a random effects, weighted analysis of variance for estimating variance of relatedness. This ANOVA is based on the statistical independence of relatedness estimates among unlinked loci: estimates of relatedness from independent loci are subjected to a one-way weighted ANOVA wherein the intraclass covariance equals the actual variance of relatedness. While maximum likelihood can alternatively be used here, it would require undue assumptions about the distribution of actual relatedness.

In this approach, we first estimate the squared actual relatedness of a single pair of individuals i as follows. After eq. (4a), let $W_k = \Sigma_j W_{jk}$ be the locus-specific weights for estimating relatedness, such that relatedness is estimated as $R_i = \Sigma_k W_k R_{ik}$ where $R_{ik} = \Sigma_j W_{jk} R_{ijk}$. The expectation of R_i^2 is

$$r_i^2 + \sum_k w_k^2 e_k^2 \tag{5a}$$

for e_k^2 the error variance for locus k (this assumes errors are independent among loci). Now, the weighted sum of squares of locus-specific estimates of relatedness is

$$\sum_k W_k^2 R_{ik}^2,$$

whose expectation is

$$r_i^2 \sum_k w_k^2 + \sum_k w_k^2 e_k^2. \tag{5b}$$

By solving for r_i^2 in the above pair of equaitons, we obtain an estimator for squared relatedness for pair i. Then by averaging this over N pairs of relatives and subtracting the squared mean relatedness, we obtain the estimator for variance of actual relatedness:

$$\hat{V}_r = \left(\frac{1}{N}\right) \sum_{i=1}^{N} \left[\frac{\left(\sum_k W_k R_{ik}\right)^2 - \sum_k W_k^2 R_{ik}^2}{1 - \sum W_k^2} \right] - R^2 \tag{5c}$$

where $R = \Sigma_i R_i/N$ is the average pairwise relatedness in the population. If observations of marker genotypes are missing for some loci for some pairwise comparisons, weights should be normalized so they sum to unity. Note that since one computes a between-locus (within individual) covariance, at least two loci are required to estimate variance of actual relatedness (and heritability). This is in accord with the earlier work of Ritland (1989), wherein a two-locus model was used to develop a crude estimator for heritability. Simulations of marker data consisting of N full sib families indicate that eq. (5c) does recover reasonably unbiased estimates (proportionally within 2–5% of true values). By contrast, the expected value of eq. (4a) when averaged over all pairs of individuals is approximately zero regardless of the true levels of relatedness (Ritland 1996), suggesting that marker-based inferences about heritability can only be based on variance of relatedness, and not mean relatedness.

Genetic Correlations

The extension of this method to the estimation of genetic correlations among characters is straightforward. Interestingly, it does not require estimation of actual variance of relatedness. Hence, it is simplest to work with a linear model involving covariances, and not correlations as in eq. (1a). For pair i, let Y_{ki} be the value of trait k in the first individual, Y'_{ki} be the value of trait k in the second individual, and r_i be the relatedness between the two individuals. In the population, let U_1 and U_2 be the respective means of traits 1 and 2, let v_{A12} be the additive genetic covariance between the two traits and v_{Ak} the corresponding genetic variances, and let c_{e12} be the environmental covariance between traits between individuals and c_{ek} the corresponding covariances for single traits. The models for the phenotypic covariances between individuals for the same trait are

$$\begin{aligned}(Y_{1i} - U_1)(Y'_{1i} - U_1) &= Y_{11i} = 2r_i v_{A1} + c_{e1} + e_{i1}\\ (Y_{2i} - U_2)(Y'_{2i} - U_2) &= Y_{22i} = 2r_i v_{A2} + c_{e2} + e_{i2}\end{aligned} \quad (6a)$$

while that for different traits is

$$(Y_{1i} - U_1)(Y'_{2i} - U_2) = Y_{12i} = 2r_i v_{A12} + C_{e12} + e_{i12} \quad (6b)$$

wherein the e is random error with zero mean. The regression estimators for the additive genetic variances and covariances are then

$$\hat{V}_{A1} = \frac{C_{Y_{11}R}}{2V_r}$$

$$\hat{V}_{A2} = \frac{C_{Y_{22}R}}{2V_r}$$

$$\hat{V}_{A12} = \frac{C_{Y_{12}R}}{2V_r} \quad (6c)$$

where the C_{YR} are the covariances between the Ys of eq. (6b) and estimated relatedness, e.g., $C_{Y_{12}R} = \mathrm{Cov}(Y_{12}, R)$. Since what we seek is $(v_{A12}/\sqrt{(v_{A1}v_{A2})})$, an estimator for the genetic correlation is

$$\hat{r}_{A12} = \frac{C_{Y_{12}R}}{\sqrt{C_{Y_{11}R}C_{Y_{12}R}}} \quad (6d)$$

The actual variance of relatedness has canceled out. This implies that one marker locus (not two, as for heritability) is the minimum requirement. However the error of estimation of the genetic correlation as normally defined—the ratio of the covariance of additive effects to the geometric mean of the variance of additive effects—still has a large estimation variance, on the order of $1/h^2$ times that of heritability. Thus, one must have good estimates of additive genetic variances for both characters. Note that the *sign* of the genetic correlation, which is sufficient for tests of many hypotheses in evolutionary biology, is simply given by the sign of $C_{Y_{12}R}$.

Statistical Properties of Estimators

Heritability has always been difficult to estimate, suffering from large standard errors even in well-designed experiments.

TABLE 1. Estimated values of heritability in Monte-Carlo datasets (m = number of alleles per locus, n = number of loci). Case A: ¼ half-sibs, ¾ unrelated (r = 0.0313, V_r = 0.0029). Case B: ⅛ half-sibs, ⅛ full-sibs, ¾ unrelated (r = 0.0487, V_r = 0.0078).

m	n	h^2 (SE)
A. true $h^2 = 0.0$		
2	32	0.001 (0.005)
4	8	−0.005 (0.007)
B. true $h^2 = 0.0$		
2	8	0.005 (0.004)
4	4	0.000 (0.003)
16	2	−0.009 (0.002)
16	8	0.001 (0.001)
B. true $h^2 = 0.25$		
2	8	0.289 (0.005)
2	32	0.256 (0.003)
4	4	0.260 (0.004)
4	8	0.250 (0.003)

The inference of relatedness not only makes this problem more acute, but magnifies the potential for bias and increases the discrepency between predicted variance (eq. [3]) and true statistical variance. For part of what follows, Monte-Carlo simulations were performed. An artificial dataset of 1000 pairs was created with known values, then heritability estimated via the above procedure. This was repeated 10^5 times for each of several combinations of numbers of alleles m and number of loci n (allele frequencies were assumed even at each marker locus).

Potential Biases

Table 1 gives examples of bias that may occur. In case A, where ¼ of the pairs are half-sibs and the remaining unrelated, bias is less than 0.01 for zero true heritability. However, because actual variance of relatedness was low (0.003), estimates of heritability were very unreliable with few alleles or loci. In case B, where variance of actual relatedness was higher (0.0075), a pattern of bias is revealed in which at low true heritability, heritability is underestimated, while at higher heritability, it is overestimated. These biases are small, and disappear with greater numbers of marker loci or greater marker polymorphism. In general, it seems that when both the variance of relatedness and the degree of marker information are sufficiently large, the expected estimate of heritability is close to its true value.

Statistical Errors

The simulations show that eq. (3) with $E_R = 1/[4n(m - 1)]$ gives values within 10–20% of the true value, except when true variances are large, where predicted variance is too low (results not shown). The true variances drop sharply from n = 2 to 4 loci, and from m = 2 to 4 alleles (with one locus, variance of relatedness cannot be estimated). Beyond these values, the simulations demonstrated that alleles are roughly as informative as loci, e.g., doubling the number of alleles at all loci has same effect as doubling the number of loci in terms of decreasing the estimation variance of heritability.

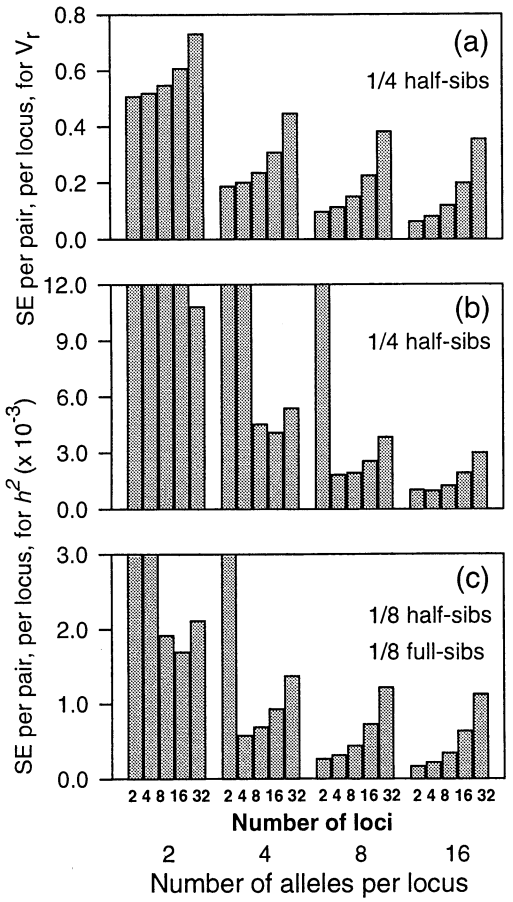

FIG. 1. Expected standard error (SE) per pair, per locus, of (a) the estimate of actual variance of relatedness V_r and (b,c) the estimate of heritability h^2, as a function of number of marker loci and number number of equifrequent alleles at the marker loci. The optimal allocation of experimental effort in terms of the numbers of marker loci assayed is given by the minumum SE for each allele number class (1000 pairs of individuals are assayed in each case, results based upon simulations; see text).

Experimental Design

Because inferences are based on nonmanipulated individuals of unknown pedigree, the main choices of experimental design involve altering the numbers of marker loci assayed, the number of individuals sampled, and the spatial pattern of sampling. The number of marker loci needed for adequate estimation of relatedness can be calculated roughly as that needed to ensure $E_R = V_R$ (the estimation variance of heritability is doubled by markers, c.f. eq. [3]). This is $n(m - 1)$ between 25 and 100 (e.g., 25–100 diallelic loci, or 12–50 triallelic loci, or 2–10 decallelic loci, etc.). Thus for cases of stronger population structure ($V_R = 0.01$) isozymes may be adequate, but with weaker structures, microsatellites or more variable markers may be needed.

This assumes no tradeoffs between assaying individuals versus marker loci. If total sample size is fixed (number of markers × number of individuals = constant), the question arises as to how to best partition effort into assaying more loci versus assaying more individuals. This is the "optimal allocation" problem for point estimation in a random-effects ANOVA (see Scheffe 1959). For estimating variance of relatedness, the optimum allocation formula of Scheffe (1959, p. 237, I_2 with $\theta = V_R/V_E = 4(m - 1)V_R$) is just 2 marker loci and as many individuals as possible, assuming low true V_R.

The optimal allocation for heritability estimation is not the same as that for variance of relatedness, but is difficult to obtain analytically. Based on the above simulations, Figure 1 shows the standard error of estimates, on a per-pair, per-locus basis (the variance of estimates obtained from 1000 pairs was divided by the number of loci used). For each case of allele number (m = 2, 4, 8, 16), the number of loci at which the SE is minimized (the number of loci ranges from n = 2, 4, 8, 16, 32) is the optimum allocation. Figure 1a shows that as predicted by Scheffe's formula, variance of actual relatedness is optimally estimated by the minimum number of markers (two).

By contrast, often more markers are optimal for estimating heritability (Figure 1b–c). Generally 4–16 loci are optimal, and more loci are needed with fewer alleles or when actual variance of relatedness is weaker; fewer loci are needed in the opposite cases (Figure 1b). Estimates with few markers and few alleles have very poor properties, reflecting cases where the actual variance of relatedness is often estimated as zero or negative. However, these calculations have assumed a constant per-locus cost of marker assay. Due to the effort of DNA isolation, the cost for the first locus is higher, and favors the assay of more marker loci than predicted by Figure 1.

The spatial pattern of sampling affects the magnitude of actual variance of relatedness in the sample (that is, if one cannot sample all individuals in a population). In our study (Ritland and Ritland 1996) we chose to sample a continuous population with a single transect spanning the population, within which adjacent pairs or triplets of plants were sampled, separated by a random distance. Sampling in both dimensions may increase the number of pairwise comparisons within a specified distance.

In any sampling plan involving a population with individuals genetically isolated by distance, the number of pairwise comparisons should be as large as possible (ca. 10^4 or even 10^5 if possible), owing to the innate large estimation variance of heritability, but not so large as to underrepresent larger scale patchiness of the population. In this light, an alternative to transect sampling would be to sample patches of individuals in a gridwork, where within each patch every individual is sampled. For total sample size N, if n were sampled in each patch, the total number of pairwise comparisons with this scheme is $N(n - 1)/2$. For example, if 500 individuals were sampled in 10 patches (50 in each patch), the number of pairwise comparisons is reasonable (12,250).

As a total alternative, one can choose those populations or even those taxa expected to show greater population substructure. One expects higher variance of relatedness in small populations consisting of a few family units, where randomly chosen pairs of individuals are likely to be related, or in large populations with limited dispersal of offspring, where individuals of close proximity are likely to be related.

Nonrandom Association of Markers with QTLs

Bias may occur from non-random association of markers with quantitative trait loci (QTLs). Simulations were run

TABLE 2. Bias of heritability estimates due to associations of QTLs with markers. True h^2 was 0.5, see text for other details. Bias is very small except with complete association.

No. markers	No. QTLs	Heritability estimate Mean	SE
QTLs randomly distributed			
2	2	0.5273	0.0058
2	8	0.5173	0.0063
8	2	0.4952	0.0032
8	8	0.5028	0.0031
QTLs completely disassociated with markers			
2	2	0.5185	0.0067
2	8	0.5319	0.0069
8	2	0.4993	0.0034
8	8	0.4989	0.0031
QTLs completely associated with markers			
2	2	0.7880	0.0077
4	4	0.6363	0.0044
8	8	0.5668	0.0033
16	16	0.5296	0.0029

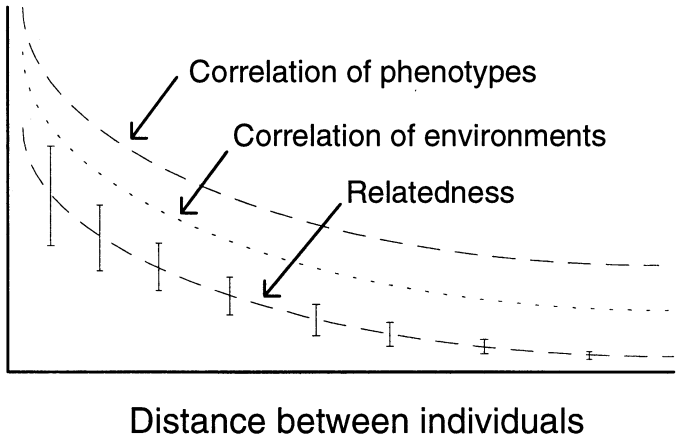

FIG. 2. Decrease of phenotypic, environmental, and genetic correlations (relatedness) as a function of physical distance between individuals in populations with restricted dispersal. The variance of relatedness has two components: that due to variation of distance between individuals, and that due to variation of pedigree at a constant distance (indicated by vertical bars).

where the genome consisted of 16 linkage blocks, with no recombination within blocks. The genetic component of the quantitative trait was determined by 2–16 QTLs that were, with respect to markers, either randomly distributed, completely associated with markers (no recombination), or completely disassociated with markers (always unlinked). All markers had four alleles.

Table 2 gives the main features of the simulation results. With random association of markers with QTLs, and with QTLs disassociated with markers, a slight upward bias was observed with low marker polymorphism, but no bias was observed with more marker polymorphism. The bias did not depend on the number of QTLs. If QTLs were completely associated with markers, significant positive bias occurred, particularly when markers were few. However, this is an extreme case, rarely expected to be encountered in real data.

Estimation of the Spatial Correlation of Environments

We now proceed to more complex versions of the linear model for estimating quantitative genetic parameters. For brevity the explorations of statistical properties are limited and the above conclusions are assumed to qualitatively apply. However, we emphasize that statistical problems only get worse from this point, that caution needs to be excercised with any treatment of real data, and that much further work remains to find improved statistical models and methods for marker-based inferences about quantitative inheritance in the field.

If resources have a patchy distribution within a population, individuals who reside near each other will share environments, causing an environmental correlation for a trait that decreases with the physical distance between individuals. If this is ignored, heritability estimates are not biased when the environmental correlation decreases independently of relatedness with distance. However, in populations with restricted gene flow, nearby individuals are also more related (e.g. Wright's "isolation-by-distance"), and both relatedness and the sharing of environments will decrease with distance (Fig. 2), confounding and biasing marker-based estimates of heritability.

One approach to incorporating distance-dependence of shared environments is to classify relatives into distance classes, such that pairwise comparisons are made between individuals of approximately the same distance, then to estimate heritability for each class and average estimates over classes. However, with this approach, small sample problems may be introduced. More importantly, limiting comparisons to within distance catagories omits variance of relatedness due to physical distance, reducing the overall statistical power. Under isolation-by-distance, variance of relatedness has two components: one due to variation of distance between individuals, and another due to variation of pedigree at a constant distance (c.f. Fig. 2).

A Model for Local Environments

A second approach is to directly incorporate environmental patchiness into the estimation model. However, this requires specifying a function for the environmental correlation. As a first approximation, the environmental correlation between a pair of individuals i decreases linearly with the physical distance d_i separating them as

$$a_e - b_e d_i \tag{7a}$$

where a_e is the environmental correlation between individuals sharing the same environment and b_e is the decrease of correlation per unit distance. Note that a_e is less than the total environmental variance, as it does not include variance specific to individuals such as accidents of development. Also, to avoid the inference of negative environmental correlations at greater distances, an exponential function or some other simple function can alternatively be used in place of eq. (7a), but this would require a nonlinear regression procedure. If comparison of relatives are restricted to close distances, this problem probably will not occur.

The linear model for phenotypic similarity becomes (after eq. [2a])

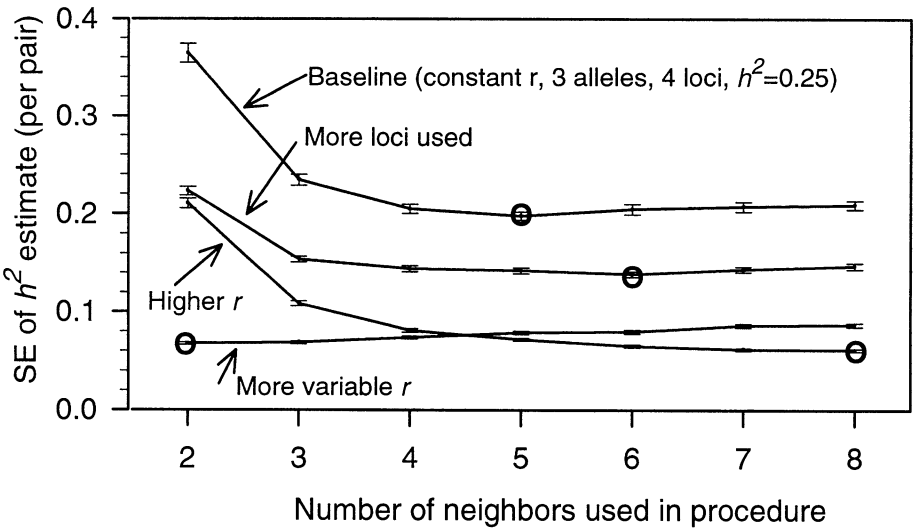

FIG. 3. Reliability of estimates as a function of number of adjacent neighbors used in the estimation procedure, under four cases of an isolation by distance model described in the text. The optimal number is circled in each case.

$$Z_i = 2r_i h^2 + a_e - b_e d_i + e_i \quad (7b)$$

with the joint estimates of heritability and environmental parameters being

$$\hat{h}^2 = \frac{C_{ZR}V_D - C_{DZ}C_{RZ}}{2(V_R V_D - C_{RD}^2)} \quad (8a)$$

$$\hat{b}_e = \frac{C_{DZ}V_R - C_{RZ}C_{RD}}{2(V_R V_D - C_{RD}^2)} \quad (8b)$$

$$\hat{a}_e = Z - R\hat{h}^2 - D\hat{b}_e \quad (8c)$$

wherein the V and C, respectively, denote sample variances and covariances, with indices signifying trait similarity (Z), genetic relatedness (R), and physical distance (D); Z, R and D are the means of the Z_i, R_i, and D_i, respectively. With no correlation of relatedness with distance ($C_{RZ} = 0$), the estimator for heritability reverts to eq. (1c) and the slope of the environmental correlation simplifies to $b_e = C_{DZ}V_D$.

Sampling Relatives under Isolation by Distance

In populations where relatedness declines with physical distance, the high relatedness between proximate individuals suggests that the heritability estimation procedure should restrict pairwise comparisons to individuals lying within a certain distance of each other (a "cutoff" distance). However, the power to infer heritability depends primarily upon variation of relatedness, which can be strongly associated with variation of physical distance between individuals (Fig. 2). This favors a larger distance cutoff that includes less related individuals. Now, if relatedness declines exponentially with distance, then as the cutoff distance is increased, variation of relatedness will eventually start to decrease as more unrelated pairs are added, suggesting the existance of an optimal cutoff of physical distance that confers the greatest power to infer heritability compared to other distance cutoffs.

To verify this optimal distance cutoff, a simulation was conducted where 10,000 individuals were placed along a transect at equal distances. In the baseline simulation, relatedness between successive individuals was randomly chosen, with equal probability, to be either zero or $r = 0.125$ (half-sibs), and data generated under the case of four triallelic marker loci and a true heritability of 0.25. This created a geometric decrease of relatedness with distance. Cases of greater numbers of loci, higher variability of relatedness ($r = 0$ or $r = 0.25$ with equal probability) and higher mean relatedness, were also considered. Figure 3 gives the results in terms of the SE of the heritability estimate as a function of the number of neighbors used (e.g., the distance cutoff).

Figure 3 shows that in two of four cases, an optimum distance cutoff does exist (the SE is minimized), but this optimum is shallow. With higher mean relatedness, the optimum increases, while with higher variance of relatedness, the optimum decreases. Also, increasing the number of markers slightly increases the optimum. These results suggest than in any analysis of real data where physical distance is a covariate, heritability should be estimated under several cutoffs, and that the cutoff giving lowest error of estimation be chosen as the optimum cutoff.

Dominance and Inbreeding Depression

At least two other factors may significantly affect the correlation between relatives in natural populations: dominance and inbreeding. Beyond this, the full description of the covariance between relatives in partially inbreeding populations is inordinately complex, requiring at least five genetic parameters (see Harris 1964; Jacquard 1974). While the appropriate coefficients of relationship for these can be estimated from marker data (Ritland 1987), and therefore these quantitative genetic parameters estimated in principle with a procedure analogous to above, the statistical power of such inferences would be extremely low. In this light, it has been noted if one assumes only additive and dominance effects (two of five parameters) together with a change of the mean phenotype with inbreeding, little bias of estimates occurs up

to an inbreeding coefficient of 0.35 (de Boer and van Arendonk 1992). Thus, with some sacrifice of realism, in this section we additionally consider just dominance and mean inbreeding depression, with the qualification that inbreeding not be very strong.

Dominance

Dominance variance inflates the correlation between relatives that may share both alleles at a locus identically by descent. These are generally stronger relatives, such as full-sibs, whose covariance is inflated by ¼ of the dominance variance (half-sibs are unaffected). In the absence of inbreeding and shared environments, the phenotypic similarity between relatives (eq. [1a]) becomes

$$Z_i = 2r_i h^2 + 2r_{2i}(H - h^2) + e_i \tag{9}$$

where H is the broad-sense heritability (additive plus dominance genetic variance), and r_{2i} is the probability that, at a randomly chosen polymorphic locus, the pair of relatives i share both alleles by descent (Δ_7, see Jacquard 1974, eq. 23). Estimators for this coefficient of relatedness for individuals are given in Ritland (1996).

Inbreeding

If the level of inbreeding varies among pairs of relatives, and inbreeding effects a systematic change on the phenotype (e.g. inbreeding depression), then relatives share phenotypes due to shared inbreeding. For the ith pair of relatives, let the inbreeding coefficient of the first individual be f_i, and that of the second individual be f_i'. To describe this correlation due to shared inbreeding, let inbreeding have a linear effect b_f upon the normalized phenotype value as

$$\frac{Y_i - U}{\sqrt{V}} = b_f(f_i - f) + e_i \tag{10a}$$

where b_f is the regression of the normalized phenotypic value on f_i, f is the average inbreeding coefficient in the population, and other terms are from eq. (1b). In the absence of other factors determining similarity, the expected phenotypic similarity between two individuals (eq. [1b]) equals the "inbreeding correlation,"

$$E[Z_i] = f_{2i}b_f^2 \tag{10b}$$

where

$$f_{2i} = (f_i - f)(f_i' - f) \tag{10c}$$

measures the sharing of inbreeding coefficients between the two individuals. Finding this quantity requires estimating f for each individual. This involves the same procedure as estimating r (eq. [4b]), with S_{ijk} defined as the observation of alleles identical-by-state for a single individual i; if homozygous, then $S_{ijk} = 1$; if heterozygous, then $S_{ijk} = 0$ (see Ritland 1996 for details).

This inbreeding correlation has two components, one due to shared inbreeding coefficients (f_i and f_i') and one due to inbreeding depression (b_f^2). Both must be present for its presence. Equation (10b) can also be obtained from Jacquard (1974, eq. 23) by redefining his V_H as $\Sigma p_i\delta_{ii}^2 - (\Sigma p_i\delta_{it})^2$, so that the coefficient multiplying his D_H^2 becomes f_2. Thus b_f^2 is closely related to the mean of dominance deviations in a corresponding homozygous population.

If one seeks estimates of heritability defined in terms of an outbred, reference population, one must account for the increase phenotypic variance due to inbreeding (by definition, the additive genetic variance is unaffected). In the absence of epistasis, it increases by the factor $(1 + f)$, so that inbreeding effects can be removed by multiplying heritability estimates by $(1 + f)$, using f estimated from markers.

Inbreeding has also been recognized to inflate the correlation between relatives and hence bias estimates of heritability, regardless of changes in phenotypic variance. For example, $r = ¼$ for full-sibs of outbred parents, while $r = ½$ for full-sibs of completely inbred parents, resulting in overestimation of heritability in the presence of any undetected inbreeding. However, a unique property of the marker-based estimate of heritability is that since r is estimated from genetic markers, the increase of r due to inbreeding is incorporated, resulting in no bias of heritability estimates due to inflated relatedness.

The General Model

A total of five quantitative genetic parameters enter the general model: narrow-(h^2) and broad-sense (H) heritabilities, the regression of fitness on f (b_f^2), and the intercept (a_e) and slope (b_e) of the environmental correlation. To jointly estimate all these, or any subset of these, a matrix formulation of the linear model quantitative genetic parameters is now presented. In many situations, a reduced set of parameters will be estimated. For example, inbreeding in animals, or dominance for certain characters, may be assumed absent. In the full model, the phenotypic similarity Z_i is related to the quantitative genetic parameters through the linear model

$$Z_i = a_e - d_i b_e + 2r_i h^2 + 2r_{2i}(H - h^2) + f_{2i}b_f^2 + e_i \tag{11}$$

The estimated parameters consists of the vector $\boldsymbol{\beta} = (a_e, b_e, h^2, H - h^2, b_f^2)$ and the independent variables lie in the vector $\mathbf{X} = (\{1, d_i, 2r_i, 2r_{2i}, f_{2i}\}, i = 1, N)$. The phenotypic similarities are placed in the vector $\mathbf{Y}$ (or matrix, if more than one character). The least-squares estimates of parameters are then $\boldsymbol{\beta} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}^T\mathbf{Y}$.

Estimating the Variance-Covariance Matrix of Actual Relationship ($\mathbf{X}\mathbf{X}^T$)

Finding the full set of estimates requires not only estimating three different types of relationship for each pair of individuals, but their actual variances and covariances in the population as well. This involves a straightforward extension of the ANOVA procedure developed earlier (eq. [5c]). To describe this, let the "coefficients of relationship R_ℓ and R_m" refer to either r, r_2 or f (e.g., ℓ and m index coefficients of relatedness and inbreeding). For a pair of individuals i, we define the weighted sum of squared relatedness ($\ell = m$) or cross-product of relatedness ($\ell \neq m$) among n loci (indexed by k) as

$$S_{i\ell m}^2 = \sum_{k=1}^{n} W_{k\ell}W_{km}R_{ik\ell}R_{ikm}. \tag{12a}$$

The individual coefficients of relatedness or inbreeding are estimated by the general form

$$R_{i\ell} = \sum_{k=1}^{n} W_{k\ell} R_{ik\ell}, \qquad (12b)$$

and are used for the total sum of squares. Taking advantage of the fact that estimates are statistically independent among loci, for N pairs of relatives, the estimator for actual variances and covariances of relatedness as

$$\hat{V}_{r\ell m} = \left(\frac{1}{N}\right) \sum_{i=1}^{N} \left(\frac{R_{i\ell} R_{im} - S^2_{i\ell m}}{1 - \sum W_{i\ell} W_{im}} \right). \qquad (12c)$$

There are two addendum to computing the variance-covariance matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$. First, covariances involving distance d_i are computed using relatedness estimated from all marker loci (there is no need for the ANOVA). Second, when computing f_{2i} on a locus-by-locus basis, I have found it best to use total f for one individual (regardless of locus) while using locus-specific estimates of f for the second individual. This eliminates a slight positive bias when estimating the actual variance of this quantity using the ANOVA formula (12b).

Statistical Properties of the General Model

The cost of more elaborate statistical models is often an increase of the statistical variance of parameters (such as heritability) formerly estimated in simpler models. This problem arises whenever the independent variables are correlated. Thus to adequately characterize properties, we need to know what are the natural levels of variation and covariation for the two- and four-gene coefficients of relatedness, the sharing of inbreeding coefficients, and the sharing of environments (the components of $\mathbf{X}$). Of course, this information is lacking, as these coefficients have never been estimated.

To examine statistical properties of the full model in the absence of such information about relevant parameter values, I nevertheless constructed artificial datasets under two scenarios of relationship: (a) a mixture of full-sibs, half-sibs, and unrelated, in proportions 0.2, 0.2 and 0.6 at a closer distance, and 0.1, 0.1 and 0.8 at a further distance (thus creating isolation-by-distance); (b) a mixture of full-sibs, selfed-sibs and unrelated in the same proportions at two distances. Markers all had four alleles with frequency distribution {0.4, 0.3, 0.2, 0.1}, the true values of quantitative genetic parameters were zero, and cases of 8, 16 and 32 markers were examined. Variances were computed by bootstrapping a dataset consisting of 1000 pairs of individuals.

When distance was added to the model for narrow-sense heritability (eq. [7b]), the error of the heritability estimate increased very slightly (< 5%). In addition, errors for the environmental factors (a, b_e) were several times less than that for heritability. This would seem to imply that the joint estimation of heritability and spatial environmental correlations has good power.

For inferring broad sense heritability and the inbreeding correlation, we need to first estimate the means and actual variances of the higher-order relatedness coefficients r_2 and f_2 from markers. These simulations indicated that their estimation properties were quite good, showing low bias and low variance. In fact, the errors in inferring actual variance of these quantities were generally less than that for the two-gene coefficient r (but other scenarios of relationship would result in greater errors for four-gene parameters). Increasing the number of marker loci always bestowed greater reductions in the statistical errors of four-gene quantities than for r. The estimation of r_2 showed greatest uncertainty but greatest benefit with increasing numbers of markers. For example, when 16 instead of 8 markers were used, the SE for the actual variance of r_2 fell by 77%, that for the actual variance of f_2 fell by 61%, while that for the actual variance of r fell by only 22%.

Under a model of narrow and broad-sense heritability, extreme variance of heritability estimates were encountered in both scenarios. This was due primarily to problems with jointly estimating r and r_2. Even though the estimation variance for the four-gene parameter, r_2, is normally less than that for the two-gene relatedness, r, (Ritland 1996), the correlation of r and r_2 among classes of relatives was strong. This colinearity of r and r_2 may be a general problem. Another problem is that actual variances of r_2 are smaller than that for r, increasing the likelihood that estimates will be zero or negative, disallowing joint estimates of narrow- and broad-sense heritability.

Under model of heritability and inbreeding correlations, the inbreeding correlation showed an error about equal to that for heritability. The increase of the SE for heritability ranged from 20% (32 markers) to 50% (8 markers). Thus, more markers seem to mitigate increases of variance due to model complexity. However, these increases depend greatly on the pattern of relatedness; with greater proportions of selfed sibs, the increase was not nearly as great.

A Computer Program

As part of the development of this procedure, a FORTRAN computer program was written that implements the full estimation model, including X − Y distance coordinates, as well as reduced versions of the model. A second program generates simulated data. For details of program features and availability, contact the author.

Discussion

This paper has presented an estimation procedure for quantitative genetic parameters using marker-inferred, natural levels of relatedness. Estimation of up to five quantitative genetic parameters has been described: narrow-(h^2) and broad-sense (H) heritabilities, the squared regression of fitness on the inbreeding coefficient (b_f^2), and the intercept (a_e) and slope (b_e) of the environmental correlation between individuals. This procedure is especially appropriate for the many organisms that either cannot be mated in any controlled fashion, or whose progeny are difficult to grow. This applies particularly to long-lived organisms, such as trees. In addition, for all species there is the appeal that inferences can be made with unmanipulated material in natural habitats, so that spurious environmental effects are eliminated.

However, the expense of molecular markers, the uncertainty of inferred relationships, and the limits of population structure place definite constaints on use of this method, particularly when more complex models involving several quan-

titiative genetic parameters are invoked. While simulations have demonstrated the ability of this method to recover information from real data, the statistical power of this procedure depends critically on the patterns of relatedness in real populations, specifically the extent of actual variance of relatedness. This is largely unknown, in large part because there has been no reason for population geneticists to measure actual variance of relatedness.

The development of complex models for quantitative trait similarity has also required a complex description of genetic relatedness, in that several different coefficients of relatedness are actually required for a full description of relationship. With the exception of the basic parameter of relatedness, the two-gene relatedness, r, statistical procedures for estimating these quantities and their actual variances have not been previously considered, and had to be developed for the purpose of this paper. The use of these estimators (eqs. [5c], [12a–c], and Ritland 1996) should allow workers to determine the ultimate feasibility of marker-based inference of quantitative genetic parameters. Also, these coefficients of relatedness are of interest in themselves as descriptors of population structure for other applications such as conservation genetics.

In all estimation formulae, the heritability has the pleasing interpretation as "the (partial) regression of character similarity Z on relatedness r." It is a partial regression if the additional quantitative genetic factors, such as the environmental correlation or shared inbreeding, are also taken into account. This interpretation provides a linkage of heritability to phenotypic selection theory, wherein the phenotypic selection intensity is "the regression of fitness on the character phenotype" (a partial regression if several characters are included in the analysis, c.f. Lande and Arnold 1983).

The Precision of Inference and Experimental Design

Heritability is a quantity already notorious for large errors of estimation, as standard errors of order 0.1 can result from the best optimized experimental designs involving known relationships. The inference of relatedness adds this uncertainty to the estimate of heritability. Beyond sample size, there are two major factors that determine the estimation variance (eq. [3]).

The first factor is the estimation variance of relatedness. In general, it does not take an enormous number of loci to get an adequate estimates of pairwise relatedness for the purpose of heritability estimation, for as discussed earlier, generally $n(m - 1)$ should be between 25 and 100, for n loci and m alleles per locus. One could do well with a dozen triallelic isozyme loci, with four microsatellite loci each with 12 alleles, or with other combinations. However, simulations indicated that the optimal division of experimental effort favors sampling more individuals than one might expect, at the expense of assaying more markers within those individuals (c.f. Fig. 1).

The second factor is the actual variance of relatedness. There are two basic ways the investigator can exert some control over this factor: optimized spatial sampling designs, as discussed earlier, or the choice of populations or species. Species with wide dispersal of progeny would be difficult to study, whereas species where progeny establish within a few neighbors, such as many conifers, would be feasible to study. In addition, if one wishes to jointly estimate additive and dominance variance, the levels of two- and four-gene relatedness (r and r_2) should not be greatly correlated among classes of relatives. However, this correlation of relatedness coefficients in natural populations is a great unknown.

If actual variation of r_2 is present, it may turn out that under a more continuous isolation-by-distance case, where local drift as well as recent coancestry via pedigree determines total relatedness, that r and r_2 are largely independent, allowing joint estimation. Barring this, one will probably have to exclude dominance variance, as in controlled experiments in mostly domesticated plants and animals, its quantity is quite small compared to additive genetic variance, particularly when a transformation is found that converts the phenotypic distribution to Gaussian (Powers 1950). However, theory predicts that dominance variance should be relatively large for fitness characters (Mousseau and Roff 1987), making its inference relevant for field studies of life history characters.

Unresolved Actual Variances

Lack of variation for certain parameters of actual relatedness or inbreeding may not bias estimates, but it does make certain quantitative genetic parameters nonestimable. In these cases, if mean levels of relatedness are still positive and significant, the overall phenotypic correlation would be affected through the constant term (a_e of eq. [7b]). For example, if both the means of r and r_2 are positive, but actual variance positive for only r, any dominance variance has a constant contribution to the phenotypic correlation, and does not bias estimates of h^2, since r_2 does not vary and therefore cannot affect the association between phenotypes (Z) and r via correlations between r and r_2. However, the broad-sense heritability is nonestimable.

Inbreeding depression combined with shared of inbreeding levels between compared individuals can also substantially contribute to the phenotypic correlation between relatives. Again, the mean level of shared inbreeding depression has no effect on the correlation; this shared inbreeding depression must vary significantly among pairs of individuals for it to bias estimates of other quantities such as heritability.

Inferences Involving Individuals Compared Several Times

When the data consists of groups of more than two related individuals, and all possible within-group pairwise relationships are considered, the estimates of relatedness and character similarity will be statistically correlated when groups of individuals are mutually interrelated. However, if the average level of relatedness is low, or the variance of relatedness within groups is low (or, if true, heritability is low), a sufficient approximation is to use all pairwise comparisons because the statistical correlation between pairs is low. For example, if groups consist of pure full-sib families, there is no within-group variance of relatedness, and the correlation computed from all pairwise combinations within families (the within-group correlation) is computationally equivalent to the

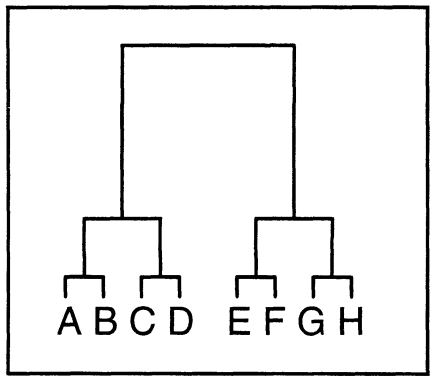

FIG. 4. A simple microphylogenetic structure that can bias estimates (see text).

statistically efficient estimate of the within-group correlation given by a one-way, random-effects ANOVA.

Previous approaches for estimating quantitative genetic parameters using the network of relationships, such as that of Shaw (1987), not only assume known relatedness, but also assume multivariate normality. In other words, all dependencies of the data higher than order two (variances and covariances) are negligible. With higher levels of relatedness and heritability, ideally, one should account for these higher-order dependencies of pairwise data. However, this requires that the estimation model include both higher-order coefficients of relationship (e.g., variance of relatedness among groups) as well as higher-order moments of quantitative variance. For example, in the simplest case of a group of three, the additional statistical dependencies are given by the third moment of relatedness and character similarity, $E[r_{123}(Y_1 - U_Y)(Y_2 - U_Y)(Y_3 - U_Y)] = M_3$, where r_{123} is the probability all three relatives share alleles identity-by-descent and M_3 is the third-moment of additive alleles (groups of four or more relatives would have terms of r^2_{ij} and V^2_A, where r_{ij} are the two-gene coefficients of relationship and V_A the variance of additive effects). This model would become prohibitively complex with groups of more than four individuals.

If one simply computes estimates using all pairwise combinations of relatives, regardless of their statistical interdependence, point estimates will likely be unbiased, but the validity of statistical tests is questionable, as the unit of independent observation is not easily determined. In the above simulations, and in our field study (Ritland and Ritland 1996), the bootstrap method was used to compute error variances, with individual the unit of resampling (identical comparisons were omitted). Bootstrapping of individuals would assume a homogenous "microphylogenetic" structure within the population. In Figure 4, a simple example of a nonhomogenous structure is given. The length of branches connecting individuals (labeled A–H) depict the level of relatedness: smaller branches denote closer relatedness, and individuals A–D are essentially unrelated to individuals E–H. In this figure, individuals within groups are *not* statistically independent from each other, and the effect of bootstrapping individuals would be to underestimate the true statistical variance. The extent of this bias due to microstructure depends on the extent of dichotomous phylogenetic structure, and is less of a problem when genetic isolation is gradual or continuous. Inspecting Figure 4, we see that bootstrapping should ideally be done at the group level (i.e., with groups A–D and E–H as the units of resampling). However, now this reduces the number of independent units such that we may not have enough for valid bootstrapping. The appropriate bootstrapping test under such structure is an area worthy of further investigation.

Conclusion and Prospects

In any estimation procedure, one must strike a compromise between model tractability, estimability of parameters, and biological realism. The procedure in this paper has been described for several genetic and environmental factors, such as dominance, inbreeding depression, and shared environments. The problem is that statistical power generally decreases with increased model complexity when the contributing factors are intercorrelated. While future studies should consider all relevant factors influencing the correlation between relatives, ultimately we must compromise between simplicity (idealism) and complexity (realism), which is the art of science.

This proposed method is but one of several new approaches for studying quantitative traits via molecular markers. In this area, the most frequent application is the mapping and characterization of individual quantitative trait loci, or QTL (Haseman and Elston 1972; Long et al. 1995), which requires controlled crosses and therefore known pedigrees. The mapping of QTL with unknown pedigrees, e.g., with material collected from wild populations, is an interesting prospect, and seems most possible using concepts borrowed from allele-sharing methods (Lander and Schork 1994). Another prospect for marker-aided inference is for those cases where inferences about heritability are made when maternity is known but paternity unknown (c.f. Konigsberg and Cheverud 1992).

ACKNOWLEDGMENTS

I thank Carol for encourgement, Laura for distractions, Kelvin for protractions, J. Felsenstein for discussion and computer time, and many others who have expressed interest in this idea over the years. This research was supported by a Natural Sciences and Engineering Research Council of Canada grant to K. R.

LITERATURE CITED

ALATALO, R. V., AND A. LUNDBER. 1986. Heritability and selection on tarsus length in the pied flycatcher (*Ficedula hypoleuca*). Evolution 40:574–583.

BOAG, P. T., AND P. R. GRANT. 1978. Heritability of external morphology in Darwin's finches. Nature 274:793–794.

CANNINGS, C., AND E. A. THOMPSON. 1981. Genealogical and genetic structure. Cambridge Univ. Press, Cambridge, MA.

COYNE, J. A., AND E. BEECHAM. 1987. Heritability of two morphological characters within and among natural populations of *Drosophila melanogaster*. Genetics 117:727–737.

DE BOER, I. J. M., AND J. A. M. VAN ARENDONK. 1992. Prediction of additive and dominance effects in selected or unselected populations with inbreeding. Theor. Appl. Genet. 84:451–459.

DHONDT, A. A. 1982. Heritability of blue tit tarsus length from normal and cross-fostered broods. Evolution 36:418–419.

DICKERSON, G. E. 1955. Genetic slippage in response to selection for multiple objectives. Cold Spring Harbor Symposia on Quant. Biol. 20:213–244.

FALCONER, D. S. 1989. An introduction to quantitative genetics. 3d ed. Longman and Wiley, New York.
HARRIS, D. L. 1964. Genotypic covariances between inbred relatives. Genetics 50:1319–1348.
HASEMAN, J. K., AND R. C. ELSTON. 1972. The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. 2:3–19.
JACQUARD, A. 1974. The genetic structure of populations. Springer, Berlin.
KONIGSBERG, L. W. AND J. M. CHEVERUD. 1992. Uncertain paternity in primate quantitative genetic studies. Am. J. Primatol. 27:133–143.
LANDE, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. Evolution 33:402–416.
LANDE, R., AND S. J. ARNOLD. 1983. The measurement of selection on correlated charactes. Evolution 37:1210–1227.
LANDER, E. S., AND N. J. SHORK. 1994. Genetic dissection of complex traits. Science 265:2037–2048.
LONG, A. D., S. L. MULLANEY, L. A. REID, J. D. FRY, C. H. LANGLEY, AND T. F. C. MACKAY. 1995. High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. Genetics 139:1273–1291.
LYNCH, M. 1988. Estimation of relatedness by DNA fingerprinting. Mol. Biol. Evol. 5:584–599.
MITCHELL-OLDS, T. 1986. Quantitative genetics of survival and growth in *Impatiens capensis*. Evolution 40:107–116.
MITCHELL-OLDS, T., AND R. G. SHAW. 1987. Regression analysis of natural selection: Statistical inference and biological interpretation. Evolution 41:1149–1161.
MORTON, N. E., S. YEE, D. E. HARRIS, AND R. LEW. 1971. Bioassay of kinship. Theor. Pop. Biol. 2:507–524.
MOUSSEAU, T. A., AND D. A. ROFF. 1987. Natural selection and the heritability of fitness components. Heredity 59:181–197.
POWERS, L. 1950. Determining scales and the use of transformations in studies of weight per locule of tomato fruit. Biometrics 6:145–163.
PRICE, T. D., P. R. GRANT, AND P. T. BOAG. 1984. Genetic changes in the morphological differentiation of Darwin's ground finches. Pp. 49–66 *in* K. Wohrmann and V. Loeschcke, eds. Population biology and evolution. Springer-Verlag, Berlin.
PROUT, T. 1958. A possible difference in genetic variance between wild and laboratory population. Drosophila Inf. Ser. 32:148–149.
QUELLER, D. C., AND K. F. GOODNIGHT. 1989. Estimating relatedness using genetic markers. Evolution 43:258–275.
RISKA, B., T. PROUT, AND M. TURELLI. 1989. Laboratory estimates of heritabilities and genetic correlations in nature. Genetics 123: 865–871.
RITLAND, K. 1987. Definition and estimation of higher-order gene fixation indices. Genetics 117:783–793.
———. 1989. Marker genes and the inference of quantitative genetic parameters in the field. Pp. 183–201 *in* A. H. D. Brown et al., eds. Population genetics, plant breeding and gene conservation. Sinauer Associates, Sunderland, MA.
———. 1996. Estimators for pairwise relatedness and inbreeding coefficients. Genet. Res. 67:175–186.
RITLAND, K., AND C. RITLAND. 1996. Inferences about quantitative inheritance based upon natural population structure in the common monkeyflower, *Mimulus guttatus*. Evolution 50:1074–1082.
SCHEFFE, H. 1959. The analysis of variance. John Wiley and Sons, Inc., New York.
SHAW, R. G. 1986. Response to density in a natural population of the perennial herb *Salvia lyrata*: Variation among families. Evolution 40:492–505.
———. 1987. Maximum likelihood approaches applied to quantitative genetics of natural populations. Evolution 41:812–826.
SMITH, J. N., AND A. A. DHONDT. 1980. Experimental confirmation of heritable morphological variation in a natural population of song sparrows. Evolution 34:1155–1158.
THOMPSON, E. A. 1975. The estimation of pairwise relationships. Ann. Hum. Genet. 39:173–188.
VAN NOORDWIJK, A. J., J. H. VAN BELAN, AND W. SCHARLOO. 1980. Heritability of body size in a natural population of the Great Tit (*Parus major*) and its relation to age and environmental conditions during growth. Genet. Res. (Cambridge) 51:149–162.
WEIR, B. S., AND C. C. COCKERHAM. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370.

Corresponding Editor: D. Waller