Behavior Genetics, Vol. 11, No. 2, 1981

Estimating Genetic Correlations from Inbred Strains

Joseph P. Hegmann[1] **and Bernard Possidente**[1]

Received 25 Apr. 1980—Final 29 Oct. 1980

Genetic correlations measure the extent of pleiotropic effects of polygenes on pairs of characters or the closeness of linkage between sets of loci influencing the traits and held in allelic (gametic) disequilibrium. Their importance for research lies primarily in predicting correlated responses of one trait to selection based on values for another, and secondarily in analyzing the complex organization of biological systems. Genetic correlations appear to limit the rate and set the direction of multivariate evolution. In view of this, efficient methods for estimating genetic correlations may be essential for understanding the role of behavior in adaptation and for predicting behavioral change in evolution. In this paper we present methods for the estimation of genetic correlations from inbred strain comparisons. Estimates from inbred strains are relatively easy to obtain and appear to be valid when compared to those derived from more demanding parent–offspring comparisons and to correlated responses to selection.

KEY WORDS: inbred lines; genetic correlations; heritability estimation; mice; prediction.

INTRODUCTION

Covariation between pairs of characters measured on the same individuals may have genetic and environmental bases. This fact, together with the genetic and environmental bases of variation in characters considered singly, allows the theoretical decomposition of observed correlations

$$r(\mathrm{p}) = h(x)\, h(y)\, r(\mathrm{a}) + e(x)\, e(y)\, r(\mathrm{e}),$$

where $r(\mathrm{p})$, $r(\mathrm{a})$, and $r(\mathrm{e})$ represent phenotypic (observed), genetic, and

[1] Zoology Department and Genetics Program, University of Iowa, Iowa City, Iowa 52242.

0001-8244/81/0300-0103$03.00/0

environmental correlations, and $h(x)$, $h(y)$, $e(x)$, and $e(y)$ are estimated by partitioning the variation for characters x and y considered singly. The full derivation of this decomposition of phenotypic correlations is given by Falconer (1960, p. 314). Note that in the absence of genetic variation for either character, phenotypic associations among traits reflect environmentally imposed covariances. Usually though, separation of covariance into genetic and environmental components is interesting because both variable genotypes and variable environments combine to produce the observed correlations.

Empirical estimates of genetic correlations can be achieved in a number of ways. In fact, Dickerson (1959) pointed out that "any method of estimating heritability also may be adapted to estimate genetic correlation among different traits. It is only necessary to isolate the covariances corresponding to each component of variance to permit estimates of both genetic and environmental correlations between traits" (p. 86). The purpose of this paper is to show a simple method for applying Dickerson's generality to the case of inbred strains for the estimation of genetic correlations. Inbred strain comparisons are probably the most rapid and the simplest method available for heritability estimation and they should provide the most rapid (one-generation) estimates of genetic correlations.

Rapid, straightforward methods for the estimation of genetic correlations are of considerable importance because of the analytical nature of genetic correlations and because of the growing awareness of the importance of these genetically imposed associations in populations under natural selection. The analytical importance of genetic correlations lies in their use in associating genetic variation in causal systems with genetically imposed covariation among effects. In this role, techniques for estimating genetic correlations among quantitative traits parallel the "genetic dissection" method of Benzer and others, but the techniques can be applied to organisms which are not ideal for dissection with induced mutations and to character differences which are more subtle than those imposed by major mutations. The recent, brilliant work of Lande (1976, 1979) on theoretical models of multivariate evolution illustrates the importance of genetic correlation estimates for the study of evolution. For example, Lande (1979) concludes that the behavior of a vector of means for characters under natural selection is "determined jointly by the selection gradient and the genetic covariance matrix" (p. 406).

Inbred strains present a unique biological circumstance. They must be used with exceptional care in the estimation of genetic variance and, we feel, with even more care for the estimation of genetic covariances and correlations. Nevertheless, their use should allow the estimation of genetic correlations with data from a single generation. Associations of

particular importance can be verified in subsequent, segregating generations derived from the set of inbred strains used to estimate the genetic correlations.

METHODS

Standard procedures for estimating heritability from inbred strain comparisons involve rearing a set of strains contemporaneously, taking care to avoid confounding (associating) environmental differences (like shelf effects in the case of mouse strains) with strain differences. If the sample of strains is large and if the strains are chosen without respect to their values for the characters to be analyzed, then differences among the strains should provide an unbiased estimate of gene-imposed differences affecting the characters. Notice that only homozygous genotypes are considered, so strain differences must be related to additive genetic variance and not to gene effects through dominance interactions within or between loci. The nature of the relationship of strain differences to additive genetic variance can be deduced from the discussion of Crow and Kimura (1970, p. 100) concerning the effects of inbreeding on character variance. The additive genetic variance among inbred lines [$V_a(i)$] increases relative to the additive genetic variance in a randomly mating population [$V_a(r)$] as the coefficient of inbreeding in the population (f) increases. The relationship is

$$V_a(i) = V_a(r)\,(1 + f).$$

Thus, under intense inbreeding the component of variance among inbred strains should provide a reasonable estimate of twice the additive genetic variance influencing a trait in a randomly mating population from which the strains were derived (or which crossing them would regenerate). The heritability of the trait [that is, the proportion of the variance due to $V(a)$] is then

$$\text{heritability} = \tfrac{1}{2}Com(AS)/[\tfrac{1}{2}Com(AS) + Com(WS)],$$

where $Com(AS)$ and $Com(WS)$ represent the components of variance among strains and within strains. Note that this is not the intraclass correlation frequently equated to "broad sense" heritability in such analyses (see Ehrman and Parsons, 1976). Using Dickerson's generality, the among-strain component of covariance of two traits should provide an analogous estimate of the additive genetic covariance for the traits and, thus, an estimate of their genetic correlation.

Standard analysis of variance procedures can be employed to partition the variance of measured characters within and between strains and

components of variance can be determined by solving two equations with two unknowns (see Sokal and Rohlf, 1969, p. 211).

The same standard procedures will yield a partitioning of the covariance for two traits if a synthetic variable, the sum of values for two characters, is formed for each animal in every strain and subjected to the same analysis. The component of variance among strains for this synthetic variable contains the component of variance among strains for each of the two characters singly plus twice their among-strains component of covariance. This component of covariance among strains is the sought-after bivariate analogue of the component of variance among strains. It can be isolated using results from single character analyses to solve the equation

$$\mathrm{Var}(x + y) = \mathrm{Var}(x) + \mathrm{Var}(y) + 2\,\mathrm{Cov}(xy)$$

derived by Sokal and Rohlf (1969, p. 651). Environmental correlations can be estimated using the component of variance within strains for two variables analyzed singly and a synthetic sum of the two to isolate the component of covariance imposed on two characters by sources of covariance other than additive genetic covariance.

It should be made clear that the components of covariance estimated from this analysis of synthetic variables are no different from those which would be obtained from a direct analysis of covariance (cf. Sokal and Rohlf, 1969. The advantage of synthetic variables lies solely in the clarity with which sources of covariance can be related to their genetic and environmental bases.

When inbred strains are available, strain comparisons are clearly the least demanding procedures for partitioning genetic variance. These comparisons are also far simpler than are other methods available for the estimation of genetic correlations (see Falconer, 1960). However, there is little to be gained unless the estimates obtained are reasonable in comparison to those from methods known to be extremely effective for estimating genetic correlations. That is, the method outlined above is of little use unless the estimates obtained are effective in predicting correlated responses to selection.

To investigate this aspect of the effectiveness of the method, we first reanalyzed scores for body weight, tail length, and caudal nerve conduction velocity for individuals of six inbred strains of mice (see Hegmann, 1972). The aim was to compare estimates from that reanalysis to those obtained for the same traits in the same laboratory, but from more demanding methods of analysis known for their predictive validity. Many of the comparisons presented are to estimates obtained from a segregating population (McClearn *et al.*, 1970) which was synthesized from eight

inbred mouse lines. Four of those strains were included as lines or sublines in the set of six strains reanalyzed here. Thus, the notion of comparing parameters estimated from inbred lines to parameters from a segregating population derived from them is only approximated in this reanalysis.

We subjected scores for body weight, tail length, and caudal nerve conduction velocity from mice in the two replicate experiments reported by Hegmann (1972) to an analysis of variance. The synthetic variables [body weight + tail length], [body weight + conduction velocity], and [tail length + conduction velocity] were generated from each individual's values and subjected to the same analysis.

RESULTS AND DISCUSSION

The components of variance among strains for the three measured variables were halved to estimate the additive genetic variances which were used, with the components of variance within strains, to estimate heritability for the traits. These heritabilities are displayed in Table I, which also shows estimates derived from parent–offspring regression and originally reported by Hegmann *et al.* (1973). Since only four of the strains from the eight-way-cross population were also included in the six-strain comparison, it is not reasonable to assume a priori that the heritabilities of the traits in the two populations are the same. However, the estimates from the inbred strain comparison are consistent with those from the parent–offspring regression in the sense that they suggest a low heritability for caudal nerve conduction velocity and only intermediate heritabilities for body weight and tail length. In fact, the response to selection for conduction velocity in the eight-way-cross population yielded a "realized heritability" estimate of 0.14 (0.09) and the response from selection

Table I. Heritabilities (SE) for Body Weight, Tail Length, and Caudal Nerve Conduction Velocity in *Mus musculus*

	Females	Males
From inbred strain analysis		
Body weight	0.41 (0.07)	0.20 (0.05)
Tail length	0.48 (0.07)	0.27 (0.06)
Velocity	0.12 (0.03)	0.11 (0.04)
From parent–offspring regression		
Body weight	0.35 (0.12)	0.29 (0.14)
Tail length	0.38 (0.10)	0.39 (0.11)
Velocity	0.11 (0.12)	0.25 (0.10)

imposed on segregating generations of a two-way cross indicated a conduction velocity heritability as low as 0.09 (0.06) (see Hegmann, 1975).

The standard errors for estimates from the inbred strain analysis appear low compared to those from regression (see Table I). We used the methods described by Osborne and Paterson (1952) to calculate standard errors. The increased precision may reflect the fact that standard errors in parent–offspring comparisons are constrained by the number of parent pairs.

The genetic correlations shown in Table II were estimated by subtracting the among-strains components of variance for each pair of single variables from the among-strains component for their synthetic sum to yield twice the among-strains component of covariance. This component of covariance estimates twice the additive genetic covariance among the pair of characters, so its division by 2 yields the numerator of the genetic correlation (see Falconer, 1960, p. 317). We used corresponding among-strain components of variance to solve for the additive genetic standard deviations required for the denominator of the genetic correlation. Estimates of environmental correlations, also presented in Table II, were obtained treating the within-strain components of variance in the same fashion but recognizing that those components are direct estimates of the environmentally imposed variance and covariance of characters. Standard errors of genetic correlations were estimated following Falconer (1960, p. 318), using current estimates of the sampling variance of character heritabilities.

The genetic correlation estimates shown in Table II are in fairly close

Table II. Genetic Correlations (Above the Diagonal) and Environmental Correlations (Below the Diagonal) Estimated from Analysis of Components of Covariance Among Strains and Within Strains[a]

	Body weight	Tail length	Velocity
	From females		
Body weight	*	0.83 (0.03)	0.42 (0.12)
Tail length	0.36	*	0.38 (0.12)
Velocity	0.24	0.14	*
	From males		
Body weight	*	0.57 (0.11)	0.30 (0.20)
Tail length	0.53	*	0.33 (0.18)
Velocity	0.25	0.13	*

[a] The analysis was easily accomplished using synthetic variables (see text). Values in parentheses are standard errors of estimates.

agreement with estimates available from other experiments. Most notably, the genetic correlations between conduction velocity and body weight and between conduction velocity and tail length can be estimated from the correlated responses of those traits to direct selection on conduction velocity. We used the correlated responses tabled by Hegmann (1975) to calculate the regression of the correlated responses of body weight and tail length on the direct response of conduction velocity and converted those regressions to genetic correlations following DeFries and Hegmann (1970, p. 47). From these calculations it appears that the genetic correlations among body weight and conduction velocity and tail length and conduction velocity were about 0.32 and 0.68, respectively, in the eight-way-cross population which had been subjected to selection. Inspection of the genetic correlations in Table II indicates agreement, at least in relative magnitude. However, bearing in mind that there are no exactly comparable estimates for the segregating populations derived from the six strains analyzed here, the estimates obtained are clearly quite in line with expectations based on other analyses.

A second, though not independent, reassurance of the effectiveness of the method is available from considering what might be referred to as the internal consistency of the estimates obtained. Recall that the initial motivation for seeking genetic correlations was the recognition that observed correlations had both genetic and environmental bases. Having now obtained estimates of genetic and environmental correlations as well as character heritabilities, all from the same set of inbred animals, it is of more than passing interest to use those estimates and the first equation presented above to generate phenotypic correlations expected in a randomly mating population derived by crossing the six strains employed. From previous experience with males and females from the eight-way cross (based on separate samples of more than 400 animals each), we expect phenotypic correlations for body weight and tail length of about 0.60. The predicted correlations from inbred strains are 0.54 and 0.57 for males and females, respectively. Predicted correlations for body weight and conduction velocity of 0.27 for both sexes can be compared to observed values in the eight-way-cross population of 0.28 and 0.22 for males and females. Finally, the correlation of tail length and conduction velocity predicted from the inbred analysis (0.16 for males and 0.19 for females) must be compared to observed values of 0.29 and 0.38 for males and females of the eight-way cross.

Because the method presented yields estimates of genetic correlation which agree in relative magnitude with those from more "expensive" procedures, because it seems to be internally consistent, and because it

rests on a firm theoretical and analytical base, we conclude that Dickerson's generality can easily be used to extend inbred strain comparisons to the estimation of genetic correlations.

There are four points that should be discussed regarding the procedures presented here. First, Blizard and Bailey (1979) estimated genetic correlations from recombinant-inbred strains by the simple expedient of calculating the correlation of strain means for the two characters. While we feel that their procedure is theoretically less sound than that presented here (because the correlation among strain means for two traits contains environmental sources of covariance as well as genetic and, thus, is not a "genetic" correlation), application of their method to the data presented here resulted in estimates differing only trivially from those shown in Table II. Employing group means in correlations reduces the numerator of the phenotypic correlation by $n - 1/n$ times the within-group covariance (where n is the number of animals measured for each strain) and has an analogous effect on the variances used to calculate the denominator of the correlation ratio. Under conditions of large sample sizes from strains or relatively low variance or covariance within strains (that is high heritabilities or genetic correlations), estimates from the two methods will be similar. However, correlations of strain means are never free from environmental sources of variation and covariation.

The second point which should be made is that extreme care should be exerted when methods like those presented here, or by Blizard and Bailey (1979), are used to estimate genetic correlations. Inbred lines are homozygous at all loci (theoretically) and may be fixed for different alleles at a number of loci quite by chance (see Wright, 1966). Genetic correlations imposed in this way would be detected by the method presented, but they would be transient in a randomly mating population and might not be seen at all. This detection of "transient" genetic correlations is more likely for traits influenced by gene differences at only a few loci than for those affected by genetic variance at many loci. Further, the larger the number of different strains used in analysis, the less likely it is that such chance associations will result in spuriously significant estimates of genetic correlation. While the precision of estimates depends on the number of individuals per strain, the generality of the estimate depends on the number of strains sampled.

Care should also be taken to choose strains without regard to their values for the traits under investigation. This is especially important if the intention is to generalize from analysis of the strains to the genetic bases of trait variation and covariation in segregating populations. If the strains analyzed provide a representative sample of the strain differences for the traits analyzed, the additive genetic variance estimates (and her-

itabilities) should also be representative. Strains chosen for extreme performance will yield heritabilities too high for a representative estimate. The same cautions apply regarding combinations of characters and the choice of strains. In this case, though, the pairwise character differences between strains must be representative in the sample to avoid misleading estimates.

The third point is that attributing genetic causes to strain differences can be potentially dangerous. For example, dietary deficiencies suffered by members of only one strain could produce an environmentally mediated effect on body length and contribute to among-strain variation in body length which could be interpreted, incorrectly, as genetic variation. The same dietary deficiency would likely also lower body weight and impose among-strain covariation for body length and body weight which could be misinterpreted as genetic covariation. The concern, though, is not so much with this kind of major environmental effect (which would show up even if members of the same strain were fed differently) as it is with more subtle maternal-mediated effects. When subtle maternal differences are consistent among strains (as they must be to be detected), it is likely that they trace to the same physiological cause and ultimately to the genetic differences among strains. Separating maternal effects from gene effects which are not mediated through maternal behavior or physiology requires experimental designs which involve at least cross-fostering and perhaps ova transplantation.

The final point we want to make is that careful use of inbred strains can allow analysis of the genetic basis of trait covariation in instances where that analysis would be impractical using more demanding methods. In general, this could occur whenever it is extremely laborious, expensive, or time-consuming to measure traits. Practical considerations like these should not preclude the genetic analysis of important aspects of biological systems but they tend to promote the intense analysis of traits that are easy to measure in the field or the laboratory, but which may be of limited biological relevance. Using the inbred strain procedures described here, genetic correlations could be examined among physiological assays of circulating hormone levels, various biophysical measures from excitable tissues, or parameters from time-series analyses (such as periods and phases of circadian rhythms). As an example, we present data (see Table III) indexing the period and phase of cyclical feeding displayed by four mice of each of five inbred strains. The data are a modified subset of observations analyzed by Possidente and Hegmann (1980). An expectation from current theory relating circadian pacemakers to behavioral rhythms is that phase and period (in constant light) for any single rhythm should display a negative genetic correlation (see Pittendrigh and Daan,

Table III. Observation of Phase and Period for Circadian Feeding Rhythms from Four Individuals of Five Inbred Mouse Strains[a]

Strain	Mouse	Phase	Period	Sum
C3H	1	2.67	12.00	14.67
	2	2.33	12.33	14.66
	3	1.33	12.71	14.04
	4	2.33	12.71	15.04
DBA	1	2.00	12.50	14.50
	2	1.50	12.86	14.36
	3	1.50	13.00	14.50
	4	3.00	12.00	15.00
C57	1	1.67	12.13	13.80
	2	3.00	11.00	14.00
	3	2.50	11.86	14.36
	4	3.00	11.56	14.56
C58	1	3.33	12.00	15.33
	2	2.00	11.77	13.77
	3	3.33	12.29	15.62
	4	3.00	12.00	15.00
BALB	1	1.33	13.40	14.73
	2	1.67	12.56	14.23
	3	2.67	11.84	14.51
	4	1.67	12.28	13.95

[a] The "Sum" column illustrates the synthetic variable employed in analysis.

1976, p. 296). The time-series analysis employed to test this expectation required 12 days of measurements at 2-hr intervals for each subject. Estimates of phase and period shown in Table III are from a total data set of 2880 observations. The "Sum" column is included for illustration of the analysis.

Using the synthetic variable method, the estimate for a genetic correlation between phase and period is -0.89. From the equation presented earlier, the among-strains component of covariance (see Table IV) is

Table IV. The Analysis of Variance for Feeding Phase, Feeding Period, and Their Sum from Four Mice of Five Strains[a]

Source	df	Variance components		
		Phase	Period	Sum
Among strains	4	0.09221	0.11196	0.02320
Within strains	15	0.39227	0.20568	0.22916

[a] The original data are given in Table III.

$$\text{Cov}(A) = (0.02320 - 0.09221 - 0.11196)/2.$$

Since this estimates twice the additive genetic covariance, the numerator of the genetic correlation is -0.0452. Direct analysis of covariance gives a component of covariance among strains of -0.090487, demonstrating the identity of the two procedures. The additive genetic variances for phase and period are estimated from half the among-strains components of variance for the variables. Reference to Table IV yields estimates of 0.046 and 0.056, respectively. Thus, the denominator of the genetic correlation (the product of the additive genetic standard deviations of the two variables) is 0.05075, completing the calculations required for the estimate. The variance component correlation taken directly from the analysis of covariance is also -0.89. The environmental correlation among these traits in this example is -0.64 and it can be calculated following the same procedures but using within-strains variance components. For completeness, we should indicate that the heritability (from Table III) of phase for feeding is 0.11 and that for period of feeding in constant light is 0.21. Even with the modest heritabilities (relatively large within-strain variation) found here, the estimate of genetic correlation using Blizard and Bailey's (1979) technique of correlating strain means is not substantially different (-0.77). The important point, though, is that the expectation from the circadian clock model was tested, using a one-generation technique for estimation of genetic correlation.

In view of the convenience of the method presented and the importance of empirical evidence of genetic correlations among characters, it is especially unfortunate that inbred lines are available for so few species. However, the procedure can be applied readily, and even retrospectively, to estimate genetic correlations in mice and *Drosophila*, two organisms whose genetic and ecological characteristics are of wide interest and of which inbred strains are commonly available.

REFERENCES

Blizard, D. W., and Bailey, D. W. (1979). Genetic correlation between open-field activity and defecation: Analysis with the CXB recombinant inbred strains. *Behav. Genet.* **9**:349–357.

Crow, J. F., and Kimura, M. (1970). *An Introduction to Population Genetics Theory*, Harper and Row, New York.

DeFries, J. C., and Hegmann, J. P. (1970). Genetic analysis of open-field behavior. In Lindzey, G., and Theissen, D. D. (eds.), *Contributions to Behavior-Genetic Analysis: The Mouse as a Prototype*, Appleton-Century-Crofts, New York, pp. 23–56.

Dickerson, G. E. (1959). Techniques for research in quantitative animal genetics. In *Techniques and Procedures in Animal Production Research*, American Society of Animal Production.

Ehrman, L., and Parsons, P. A. (1976). *The Genetics of Behavior*, Sinauer Associates, Sunderland, Mass.

Falconer, D. S. (1960). *Introduction to Quantitative Genetics,* Ronald Press, New York.
Hegmann, J. P. (1972). Physiological function and behavioral genetics. I. Genetic variance for peripheral conduction velocity in mice. *Behav. Genet.* **2**:55–67.
Hegmann, J. P. (1975). The response to selection for altered conduction velocity in mice. *Behav. Biol.* **13**:413–423.
Hegmann, J. P., White, J. E., and Kater, S. B. (1973). Physiological function and behavioral genetics. II. Quantitative genetic analysis of conduction velocity of caudal nerves of the mouse. *Mus musculus. Behav. Genet.* **3**:121–131.
Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**:314–334.
Lande, R. (1979). Quantitative genetic analysis of multivariate evolution applied to brain-body size allometry. *Evolution* **33**:402–416.
McClearn, G. E., Wilson, J. R., and Meredith, W. (1970). The use of isogenic and heterogenic mouse stocks in behavioral research. In Lindzey, G., and Theissen, D. D. (eds.), *Contributions to Behavior-Genetic Analysis: The Mouse as a Prototype,* Appleton-Century-Crofts, New York, pp. 3–22.
Osborne, R., and Paterson, W. S. B. (1952). On the sampling variance of heritability estimates derived from variance analyses. *Proc. Roy. Soc. Edinburgh S.* **64**:456–461.
Pittendrigh, C. S., and Daan, S. (1976). A functional analysis of circadian pacemakers in nocturnal rodents. IV. Pacemaker as clock. *J. Comp. Physiol.* **106**:291–331.
Possidente, B., and Hegmann, J. P. (1980). Circadian complexes: Circadian rhythms under common gene control. *J. Comp. Physiol.* **139**:121–125.
Sokal, R. R., and Rohlf, F. J. (1969). *Biometry,* W. H. Freeman, San Francisco.
Wright, S. (1966). Polyallelic random drift in relation in evolution. *Proc. Natl. Acad. Sci.* **55**:1074–1080.

Edited by Norman D. Henderson