

RESEARCH ARTICLE

Estimating genetic correlations based on phenotypic data: a simulation-based method

ELIAS ZINTZARAS*

*Department of Biomathematics, University of Thessaly School of Medicine,
2 Panepistimiou Str., Biopolis, Larissa 41110, Greece*

Abstract

Knowledge of genetic correlations is essential to understand the joint evolution of traits through correlated responses to selection, a difficult and seldom, very precise task even with easy-to-breed species. Here, a simulation-based method to estimate genetic correlations and genetic covariances that relies only on phenotypic measurements is proposed. The method does not require any degree of relatedness in the sampled individuals. Extensive numerical results suggest that the proposed method may provide relatively efficient estimates regardless of sample sizes and contributions from common environmental effects.

[Zintzaras E. 2011 Estimating genetic correlations based on phenotypic data: a simulation-based method. *J. Genet.* **90**, 51–58]

Introduction

The evolutionary response to selection is a function of the genetic covariance between characters as well as environmental associations between them (Young and Weiler 1960; Lande 1979). It is, therefore, important to ascertain how the genetic and phenotypic associations between traits jointly influence the evolutionary trajectory of a suite of characters. However, a fundamental dictum in quantitative genetics is that the degree of resemblance between relatives provides the means of estimating basic genetic parameters (Kempthorne 1957; Falconer and Mackay 1996; Lynch and Walsh 1998), a difficult or impossible endeavour with most species. Hence, some authors have proposed to use phenotypic correlations as surrogate estimates of genetic correlations (Cheverud 1988, 1995; Roff 1995, 1996), although this is still a very controversial issue (Willis *et al.* 1991; Lynch and Walsh 1998). Lynch (1999) has conceived a more promising avenue for research by showing that it is feasible to estimate genetic correlations provided our sample is relatively large and consists of at least 20% or so paired relatives. Nevertheless, in many circumstances it may be impossible to obtain samples that conform to this requisite,

and there seems to be no cheap alternative to the use of standard, time-consuming techniques of quantitative genetics. Even in this case, the estimation of genetic parameters in natural populations requires in most instances the regression of lab offspring on the midparent from nature, and the risk of obtaining unreliable results because of genotype \times environment interaction can be very high (Coyne and Beecham 1987; Riska *et al.* 1989).

Here, a method to estimate genetic correlation (correlation of genotypic values, symbolized as \mathfrak{R}_G) that does not require any information on pedigree structure or even the presence of relatives in our samples is suggested. First, the method and numerical results from extensive computer simulations are presented, then its application with two empirical data sets using *Drosophila buzzatii* samples are illustrated. One set comes from a natural population, and the other from a laboratory experiment that also allows estimating genetic correlations from standard techniques in quantitative genetics.

Methods

Procedure to estimate genetic correlations (\mathfrak{R}_G)

Let us assume a sample of N individuals measured for two metric traits x and y , and represent the phenotypes of individual i as $z_i(x)$ and $z_i(y)$. Then, randomly arrange them in two vectors and calculate the following correlation

*E-mail: zintza@med.uth.gr. Also affiliated to Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Tufts University School of Medicine, 800 Washington Street, Boston, MA 02111, USA.

Keywords. correlation; genetics; phenotype; environment; bias; simulation.

from the resulting paired individuals (Ritland 1996; Lynch 1999):

$$\rho^\bullet = \frac{\frac{1}{2} \{ \text{Cov}[z_i(x), z_j(y)] + \text{Cov}[z_i(y), z_j(x)] \}}{\sqrt{\text{Cov}[z_i(x), z_j(x)] \times \text{Cov}[z_i(y), z_j(y)]}}, \quad (1)$$

where Cov denotes an observed covariance and subscript letters refer now to the i th individual in the first vector and the j th individual in the second vector, both in the same row. Provided that (i) N is large (i.e., several hundreds); (ii) at least 20% or so of the paired individuals are relatives (e.g., full sibs); and (iii) shared environmental effects are absent, equation (1) can render an informative estimate of the genetic correlation (Lynch 1999). In most natural populations, however, it is difficult if not impossible to obtain samples that conform to requisites (ii) and/or (iii).

Here it is suggested that equation (1) can still be used to obtain a plug-in estimate of the genetic correlation (\mathfrak{R}_G) even if the random sample comes from the entire range of the population and does not contain relatives. In addition, sample sizes are not required to be very large. Simply stated, the genetic correlation is estimated as the average of ρ^\bullet from the ‘informative subsets’ after randomly assorting the N individuals in the sample in two column vectors R times (i.e., random permutations where each individual occurs only once in a vector). The informative subsets ($\sim 25\%$ of R) are defined as those that render both covariances in the denominator of (1) simultaneously positive.

The following expressions are not intended to be mathematically rigorous, but rather provide heuristic arguments sustained by computer simulations for why the method works (see below). Let us denote all covariances in (1) as:

$$\begin{aligned} \sigma^o &= \frac{1}{2} \{ \text{Cov}[z_i(x), z_j(y)] + \text{Cov}[z_i(y), z_j(x)] \}, \\ \sigma^* &= \text{Cov}[z_i(x), z_j(x)], \\ \sigma^{**} &= \text{Cov}[z_i(y), z_j(y)]. \end{aligned} \quad (2)$$

Assume that both traits follow a multivariate normal distribution with variance–covariance matrices:

$$\begin{aligned} Z &= \begin{bmatrix} \sigma_z^2(x) & \sigma_z(x, y) \\ \sigma_z(x, y) & \sigma_z^2(y) \end{bmatrix} = G + E = \begin{bmatrix} \sigma_g^2(x) & \sigma_g(x, y) \\ \sigma_g(x, y) & \sigma_g^2(y) \end{bmatrix} \\ &+ \begin{bmatrix} \sigma_e^2(x) & 0 \\ 0 & \sigma_e^2(y) \end{bmatrix}, \end{aligned} \quad (3)$$

where Z , G and E are variance–covariance matrices for phenotypic, genetic and environmental effects, respectively. For the time being, we will also assume the environmental correlation between traits within the same individual to be equal to zero. Both σ^* and σ^{**} are normally distributed with zero mean and variance that is a function of the phenotypic variance and that scales inversely with the number of paired individuals in the sample (actually, $\text{Var}(\sigma^*) = \text{Var}(\sigma^{**}) = \frac{1}{N/2}$ when both traits are standardized and N is even). For a truly

normal distribution with mean zero we know that (Kendall and Stuart 1951):

$$\begin{aligned} E(\sigma^* | \sigma^* \geq 0) &= \sqrt{\frac{\text{Var}(\sigma^*)}{\pi/2}} \approx \frac{\sigma_z^2(x) / \sqrt{N/2}}{\sqrt{\pi/2}}, \\ E(\sigma^{**} | \sigma^{**} \geq 0) &= \sqrt{\frac{\text{Var}(\sigma^{**})}{\pi/2}} \approx \frac{\sigma_z^2(y) / \sqrt{N/2}}{\sqrt{\pi/2}}, \end{aligned} \quad (4)$$

(E denotes ‘expectation’). Accordingly, the averages of σ^* , σ^{**} from the informative subsets, which are those coming from the intersection of the two-half-normal distributions in (4), inversely scale to $\sqrt{\frac{N}{4}}$. What is most pertinent is that the only source of covariance in the phenotypic distributions of traits x and y , namely the genetic covariance in (3), is retained as the average of σ^o . Thus, computer simulations (see below) show that a reliable estimate of the genetic covariance can be obtained as:

$$\widehat{\sigma_g(x, y)} = (\sigma^o | \sigma^* \geq 0, \sigma^{**} \geq 0) \sqrt{\frac{N}{4}}, \quad (5)$$

where the expression in brackets is the average of σ^o calculate from the informative subsets (notice that the factor $\sqrt{\frac{N}{4}}$ cancels out in formula (1)). It is obvious from (4) that $E(\sigma^* | \sigma^* \geq 0, \sigma^{**} \geq 0) \sqrt{\frac{N}{4}}$ and $E(\sigma^{**} | \sigma^* \geq 0, \sigma^{**} \geq 0) \sqrt{\frac{N}{4}}$ linearly fit with, but are not equal to, the phenotypic variances of traits x and y , and it is not feasible in general to obtain plug-in estimates of the genetic variances. However,

$$\begin{aligned} &\sqrt{\frac{E(\sigma^* | \sigma^* \geq 0, \sigma^{**} \geq 0) \times E(\sigma^{**} | \sigma^* \geq 0, \sigma^{**} \geq 0) \times N}{4}} \\ &\approx \sqrt{\frac{\sigma_z^2(x) \times \sigma_z^2(y)}{\pi}} \propto \sqrt{\sigma_g^2(x) \times \sigma_g^2(y)}, \end{aligned} \quad (6)$$

and a plug-in (but biased) estimate of the genetic correlation (\mathfrak{R}_G) can be obtained as the average of ρ^\bullet from the informative subsets. Expressions (5) and (6) clearly suggest that \mathfrak{R}_G will be upwards biased when the product of heritabilities is larger than $1/\pi$, and downwards biased otherwise. However, notice that

$$\begin{aligned} &E(\rho^\bullet | \sigma^* \geq 0, \sigma^{**} \geq 0) \\ &\neq \frac{E(\sigma^o | \sigma^* \geq 0, \sigma^{**} \geq 0)}{\sqrt{E(\sigma^* | \sigma^* \geq 0, \sigma^{**} \geq 0) \times E(\sigma^{**} | \sigma^* \geq 0, \sigma^{**} \geq 0)}}, \end{aligned}$$

and numerical results indicate that $E(\rho^\bullet | \sigma^* \geq 0, \sigma^{**} \geq 0)$ is more upwards biased than the ratio of expectations when heritabilities are high, whereas it is less downwards biased when heritabilities are low.

Computer simulations

MATLAB (v. 5, 1998 The language of technical computing, <http://www.mathworks.com>) was used in all simulations. First, it will be illustrated that reliable estimates of the genetic

covariance can be obtained from (5) by using the simple phenotypic model in (3). Samples of N (ranging from 100 to 1000) genetically independent individuals were generated from randomly defined G and E matrices (with zero means and zero environmental covariance). For a given joint distribution of the two traits and for a given N , 20,000 independent samples were obtained and each sample provided one set of the three covariances in (2). Numerical results are shown in figure 1 as the scatterplot between the estimate genetic covariance obtained from (5) and the parametric genetic covariance. It seems clear that a simple and reliable test to the null hypothesis of no genetic correlation between traits x and y could be performed from the empirical distribution of σ^o over the informative subsets.

First, it will be illustrated that \mathfrak{R}_G provides an informative plug-in estimate of the genetic correlation that is largely independent of sample size from numerical results with $N = 100$ randomly generated individuals (qualitatively identical results were obtained for $N = 500$). In these simulations, and with no loss of generality, it was assumed that environmental covariance was zero. The multivariate normal G matrix was randomly generated within the following bounds obtained from a uniform distribution:

$$\begin{aligned} 4 \geq [\sigma_g^2(x), \sigma_g^2(y)] \geq 0.25; \\ (+1) \times \min [\sigma_g^2(x), \sigma_g^2(y)] \geq \sigma_g(x, y) \geq (-1) \\ \times \min [\sigma_g^2(x), \sigma_g^2(y)], \end{aligned} \quad (7)$$

where min stands for ‘minimum of’ (heritabilities are therefore bounded within the interval 0.20–0.80). A total of 3000 random genetic variance–covariance matrices were generated, and numerical results are shown in figure 2. Each

point in the figure was obtained as the average of ρ_\bullet from the informative subsets within the limits ± 10 (to avoid bias from rare outliers) from the outcome of 12,000 random data sets. The overall correlation between the parametric genetic correlation (ρ_G) and \mathfrak{R}_G is 0.974. It is clear that \mathfrak{R}_G underestimates the true correlation when heritabilities are lower than 0.35, whereas the converse is true when heritabilities are higher than 0.65 (see above).

The root mean square error of \mathfrak{R}_G (i.e. $\sqrt{[\hat{\rho}_g(x, y) - \rho_g(x, y)]^2}$) can be used as a measure of bias (Efron and Tibshirani 1993), and the values are plotted in figure 3. Provided that some information on the heritabilities is available (which is more likely in laboratory samples), a more efficient estimate of the genetic correlation can be obtained from these plots regardless of sample size (see below).

Some suggestions as how to proceed with empirical data sets seem worthwhile here. As previously indicated, the average of ρ_\bullet from the informative subsets (within the limits ± 10) is more upwards biased than the ratio of expectations when heritabilities are high, whereas the converse is true when heritabilities are low (data not shown). Field heritabilities seem to be comparable in magnitude or even larger than laboratory estimates, ranging from 0.562 for morphology to 0.224 for behaviour (Weigensberg and Roff 1996). As a result, researches will likely deal with conditions as those plotted in the upper left part of figures 2 and 3 most of the time. Consequently, the best strategy is to take the average of ρ_\bullet from the informative subsets (within ± 10) as a plug-in estimate of the genetic correlation. Limited evidence suggests that statistical significance can be tested either way; namely, from the empirical distributions of σ^o or ρ_\bullet (within ± 10) over the informative subsets.

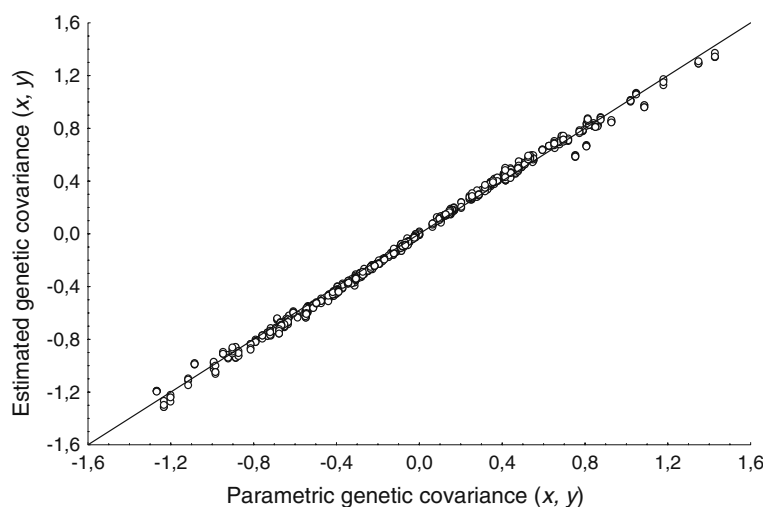


Figure 1. Relationship between the parametric and the estimated genetic covariance from equation (5). The values were obtained from numerical results after generating random samples of N (ranging from 100 to 1000) genetically independent individuals using expression (3). Parametric heritabilities were not bounded and range from 0.092 to 0.994 for trait x , and from 0.167 to 0.980 for trait y . Similarly, parametric genetic correlations range from -0.935 to 0.917 . Each point in the figure was obtained from the outcome of 20,000 random data sets. The solid line plots the ideal case of no bias.

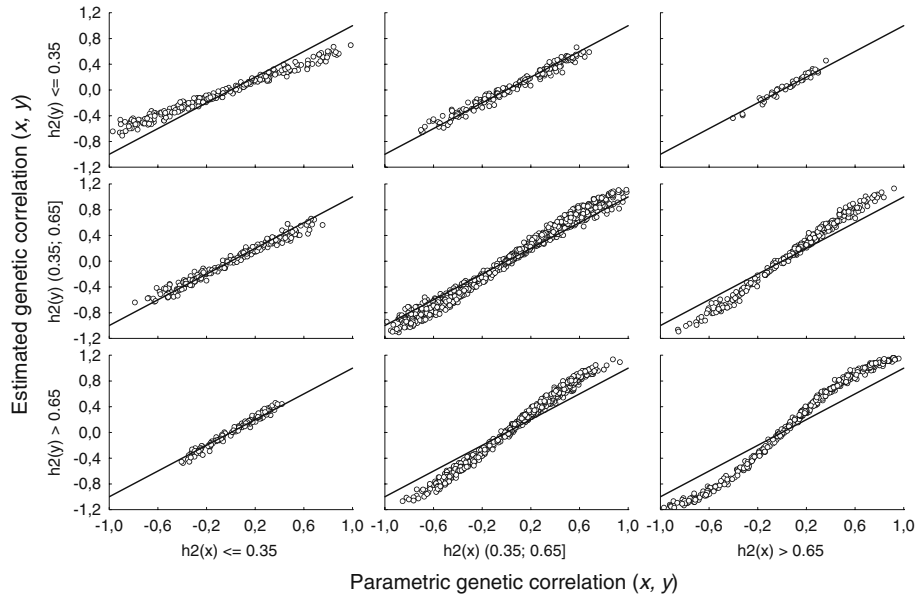


Figure 2. Scatterplots of the parametric ($\rho_g(x, y)$) and the estimated \mathfrak{R}_G genetic correlations for traits x and y , in relationship to their heritabilities (bounded within the limits 0.20–0.80), obtained from randomly generated samples of $N = 100$ genetically unrelated individuals. A total of 3000 variance–covariance genetic matrices (G) were generated, and each point in the figure was obtained as the average of ρ_{\bullet} from the informative subsets within the limits ± 10 from the outcome of 12,000 random data sets (see text for details). Solid lines stand for the ideal case when bias (i.e., $\rho_g(x, y) - \hat{\rho}_g(x, y)$) is zero (notice that the plots are arranged as 3×3 symmetrical matrices).

Potential problems with \mathfrak{R}_G

A major problem when estimating quantitative genetic parameters (Lynch 1999) is the contribution of common environmental effects and/or environmental covariance to

the phenotypic resemblance of relatives. \mathfrak{R}_G should be relatively insensitive to common environmental effects when samples are at random over the entire range of the population because no relatives are necessary to estimate the genetic correlation. However, environmental covariances can be a

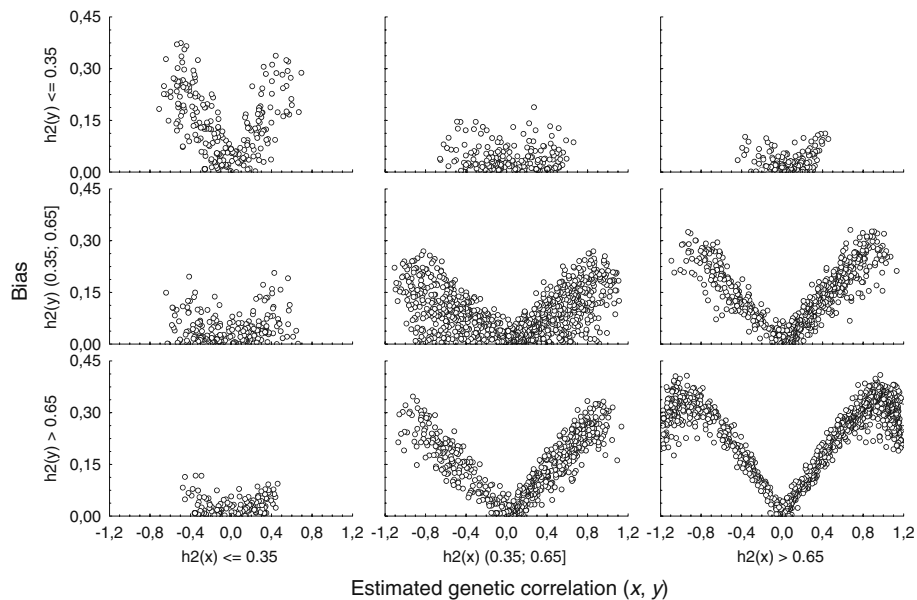


Figure 3. Scatterplots of the root mean square errors for $\mathfrak{R}_G \left(\sqrt{[\hat{\rho}_g(x, y) - \rho_g(x, y)]^2} \right)$ in relation to the heritabilities of traits x and y . Values were obtained from figure 1 with $N = 100$ but are about the same regardless the number of individuals included in our samples (a nice property if bias could be approximated in a particular data set; see text for details).

potentially important source of bias simply because the average of σ^o calculated from the informative subsets (5) will be a function of both the genetic and environmental covariances. There are, nevertheless, situations when \mathfrak{R}_G could also be relatively insensitive to this problem. To understand why this is so, an analogy with the random effect ANOVA model is useful (Searle *et al.* 1992).

Assume that individuals are sampled over the entire array of theoretically infinite microenvironments with environmental covariance in the k th microenvironment equal to ξ_k . Also assume that the standard assumptions in the ANOVA are met, namely:

$$\begin{aligned} E(\xi_k) &= 0, \\ \text{Cov}(\xi_k, \xi_l) &= 0, \quad \forall k \neq l, \\ \text{Cov}[(z_i(\cdot) | \xi_k), (z_j(\cdot) | \xi_k)] &= \text{Cov}[z_i(\cdot), (z_j(\cdot))] \\ &+ \text{Cov}(\xi_k, \xi_k), \quad i \neq j, \end{aligned} \quad (8)$$

where (\cdot) stands for trait x, y . The implications in (8) are clear. Thus, as far as individuals are sampled from the entire population and the probability of pairing two individuals coming from the same microhabitat is negligible, \mathfrak{R}_G would be relatively insensitive to environmental effects provided that the average environmental covariance across environments is zero. This conclusion is sustained by numerical results after assuming no genetic variation for traits x and y and microhabitat environmental covariances randomly ranging between ± 1 . In this case, the average value of \mathfrak{R}_G is zero when individuals are independently drawn regarding microenvironments, as expected (data not shown). The assumption that environmental effects are independent is likely to be realistic. The problem is, however, whether or not it is reasonable to assume that the average environmental covariance across environments is zero. If it were not, \mathfrak{R}_G would be biased towards the direction of the average environmental correlation, although the bias will likely to be lower than that from other estimation procedures demanding relatives that have grown in the same microhabitat. In any case, environmental differences might in principle be eliminated after standardizing phenotypic measures with respect to fixed environmental effects (Lynch 1999).

Results

An illustration of \mathfrak{R}_G using empirical data sets

In this section, we apply the method to estimate genetic correlations to two empirical data sets of *D. buzzatii*. The experimental data described next were kindly provided by Dr Mauro Santos (see Leibowitz *et al.* 1995; Laayouni *et al.* 2000), and I now briefly summarize those experiments.

Laboratory experiment: The *D. buzzatii* flies used in the experiment originated from 152 isofemale strains. A large outbreeding population was set up by dumping the flies into a plexiglas cage. The population was since kept in mass culture

on a 12:12 light/dark cycle at 23°C with uncontrolled humidity, and maintained by introducing three fresh bottles containing 50 mL of killed-yeast *Drosophila* medium (David 1962) once a week and removing them after five weeks. Throughout the experiments CO₂ was used when anaesthesia was necessary.

Approximately one year after the population cage was set up; a parental half-sib mating design was performed. Rearing conditions were standardized so that all flies experienced optimum growing environments. The final data set was a balanced design with 263 sires crossed to three dams each, and one random son per dam was measured for wing length (WL) and wing width (WW) on both wings (Leibowitz *et al.* 1995). After an entire round of measurements on all individuals was completed, an additional round was taken so that each wing was measured twice. Least-squares (ANOVA) estimates of the components of variance and covariance were obtained for the averages of WL and WW in each fly following Becker (1984).

Table 1 gives the estimates of heritabilities and genetic correlations. The genetic correlation between WL and WW estimated from the standard linear model (0.803) is comparable to the value of 0.833 previously obtained by Loeschcke *et al.* (1999) in *D. buzzatii*. According to the heritabilities given in table 1 and the numerical results in figure 2, \mathfrak{R}_G estimated as the average of ρ_\bullet from the informative subsets (within the limits ± 10) is expected to render an upwards-biased estimate. The corresponding result in table 1 suggests that this is indeed the case. However, the example also serves to illustrate how bias could be approximated from knowledge of the heritabilities (see lower triangular plots in figure 3). In this case, bias is ~ 0.2 and a better estimate of the genetic correlation is 0.788 (approximate 95% confidence interval: 0.734–0.855).

Field experiment: Thirty-six rotting *Opuntia ficus-indica* cladodes were collected from a disused plantation. Each rot was placed in a transparent plastic container on a bed of sand,

Table 1. Laboratory estimates of heritabilities (diagonal) and genetic correlations of wing length (WL) and wing width (WW) in *D. buzzatii*. Above the diagonal is the estimate of the genetic correlation obtained from the standard linear model with paternal half-sib mating designs, and below is that obtained from \mathfrak{R}_G after $R = 5000$ random permutations of paired individuals (in parenthesis are the approximate 95% confidence intervals based on the bootstrap percentile method over the remaining values of ρ_\bullet in the informative subsets within the limits ± 10 ; see Efron and Tibshirani 1993).

Trait	Trait	
	WL	WW
WL	0.494*	0.803*
WW	0.988 (0.934; 1.055)	0.746*

* $P < 0.05$.

closed with a fine-meshed fabric, and kept at room temperature (22–27°C) in the makeshift laboratory near the field site (Laayouni *et al.* 2000). A total of 5926 adult flies, all but one *D. buzzatii*, emerged from 28 rots. Both wings from a sample of 2951 flies coming out from 26 rots were measured once for WL and WW (figure 4). Large phenotypic differences are obvious among the flies raised from different rots, which is partially related to density (i.e., total number of flies raised) in each rot (see below). Let us focus first on rot BP3 (indicated by a two-headed arrow in figure 4).

A total of 132 flies were measured for rot BP3. For each sex, the data are plotted in figure 5 as a function of the day of emergence from the rot. Females are clearly bigger than males ($t = 3.59$, $P < 0.001$ for WL; $t = 2.56$, $P = 0.011$ for WW), and the average size of the flies decreases as a function of time ($\beta = -0.043$, $P < 0.001$ for WL; $\beta = -0.016$, $P < 0.001$ for WW), probably as a result of a drop in the quantity and/or quality of available resources in the rot. Therefore, the genetic correlation between WL and WW was estimated after differences in sex and day of emergence between flies were removed. This was achieved by using the residuals from the data fitted to a multiple linear regression including both effects as explanatory variables (figure 5c). After $R = 5000$ random permutations, \mathfrak{R}_G for rot BP3 was estimated to be 0.795 (approximate 95% confidence interval: 0.710–0.882). The resultant figure is remarkably similar to those obtained in the laboratory (see above).

The estimation of the genetic correlation from the whole data set ($N = 2951$) also serves to illustrate the potential biasing effects of environmental differences across rots. The average rot means are inversely related to the number of flies raised ($\beta = -1.43 \times 10^{-4}$, $P < 0.001$ for WL;

$\beta = -0.61 \times 10^{-4}$, $P < 0.001$ for WW), and the average environmental correlation affecting the expression of both size traits is likely to be positive across rots. Hence, we can expect that \mathfrak{R}_G would be biased towards the direction of the average environmental correlation if environmental effects are not removed, which seems to be the case ($\mathfrak{R}_G = 1.134$ after $R = 10,000$ random permutations). However, when the genetic correlation for each rot (with $N > 20$) is estimated from the residuals as explained above, the average estimate (\pm empirical s.d.) is 0.839 (± 0.199). Because field heritabilities for size traits are likely to be lower than laboratory ones in *D. buzzatii* (Prout and Barker 1989; Ruiz *et al.* 1991; Leibowitz *et al.* 1995), it seems reasonable to conclude that the field genetic correlation between WL and WW in this species is ~ 0.8 , roughly identical to that estimated in the laboratory.

The preceding example illustrates how differences in mean phenotypes due to sex or microhabitat variation (i.e., fixed genetic and environmental effects) within the sampling area could be dealt with adequately. Interestingly, the approach in Riska *et al.* (1989) was unable to render realistic estimates of heritabilities and genetic correlations because regressions of lab offspring on the midparent from nature (isofemale strains were set up by crossing flies emerging from each rot) were nonsignificantly different from 0 for WL and WW. Furthermore, because the genetic correlation between WL and WW has been estimated to be high, a lower bound for the heritabilities could tentatively be obtained as $\frac{(\sigma^2_{\text{gen}} | \sigma^2_{\text{gen}} \geq 0, \sigma^2_{\text{env}} \geq 0)}{V_P} \sqrt{\frac{N}{4}}$ (see equation 5), where V_P is the phenotypic variance. Care must of course be taken with the different scales of measurement, but a lower bound for the heritability of WL in the laboratory sample above is ~ 0.327 , and for the flies emerging from

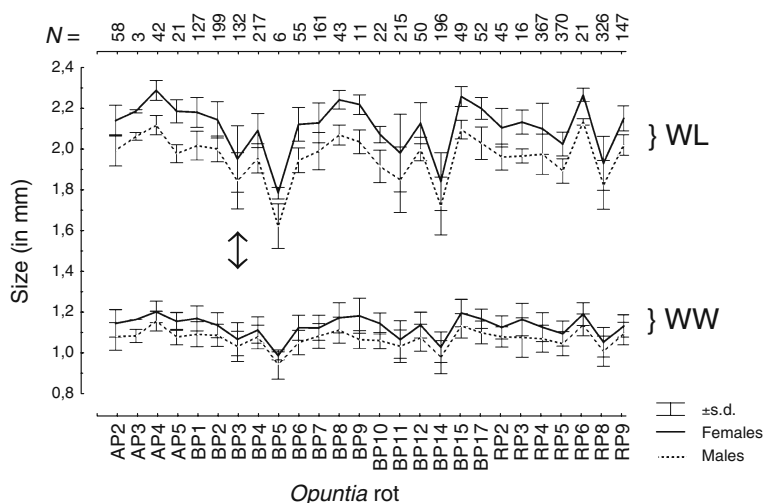


Figure 4. Profiles connecting the averages (\pm s.d.) of wing length (WL) and wing width (WW) of *Drosophila buzzatii* flies emerging from 26 *Opuntia ficus-indica* rots. The top axis indicates the corresponding number of flies measured for both traits and the two-headed arrow the rot BP3 used as an illustration (see text for details).

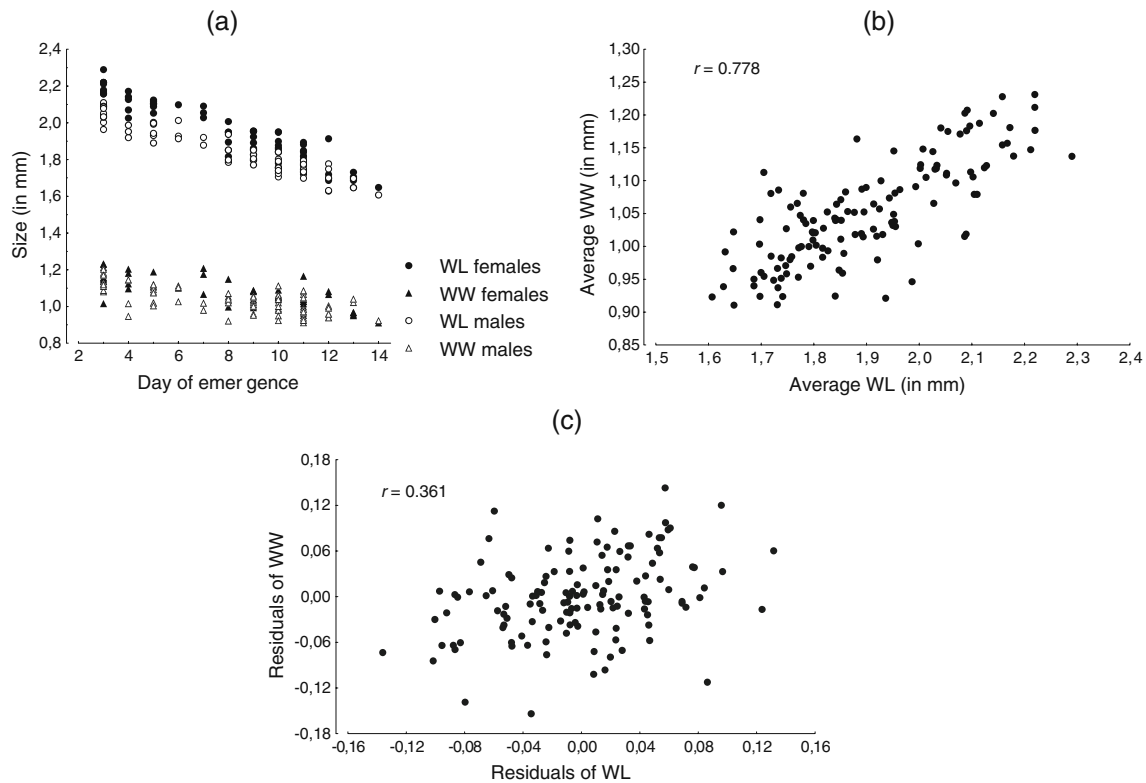


Figure 5. Relationships between wing length (WL) and wing width (WW) of *D. buzzatii* as a function of the day of emergence (counted from the day when the first fly emerged in the whole sample of *Opuntia ficus-indica* rots) from rot BP3 (a). The phenotypic correlation between WL and WW is plotted in (b), and the correlation of the residuals obtained after removing the ‘sex’ and ‘time’ fixed effects in (a) is plotted in (c).

rot BP3 is ~ 0.371 (however thought provoking these figures may be, they should obviously be taken with great caution). From methods described in Riska *et al.* (1989) a lower bound of ~ 0.04 was previously obtained for the natural heritability of WL in *D. buzzatii* (Leibowitz *et al.* 1995). The limited evidence provided clearly suggests that across environment estimates of genetic parameters in *Drosophila* or in other organisms could be very misleading because of the presence of genotype \times environment interactions, and \mathfrak{R}_G is a cheap and useful alternative.

Finally, although the main goal when applying \mathfrak{R}_G is likely to make reliable statements about the natural genetic correlations, its usefulness in laboratory or greenhouse samples should also be stressed. In these situations, growing conditions are usually standardized and estimates of genetic covariances and correlations using the proposed method only require a few hours of work.

Discussion

An obvious question is how many random permutations (R) should be carried out to obtain an estimate of the genetic correlation from \mathfrak{R}_G ? For a sample size N there are $\frac{N \times (N-1)}{2}$ different pair combinations and in general, only 25% of the random permutations are retrieved as ‘informative subsets’.

\mathfrak{R}_G quickly converges to roughly the same value and $R \approx 5000$ or less should be enough in most circumstances (data not shown). However, the estimated provided in equation (1) does a reasonable good job when the proportion of relatives in the data set is relatively large. With no relatives the expectation of (1) is zero, and the basic difference in the proposed method from Lynch (1999), is that we only use 25% of subsets after randomly assorting the N individuals. The transition is just from 0 to the plug-in estimate of the genetic correlation.

A number of issues remain to be more fully explored. Thus, some initial results suggest that the statistical power of \mathfrak{R}_G to detect genetic correlations when heritabilities are lower than 0.20 can be relatively high even though the estimates are downwards biased. In these circumstances, estimates of the genetic correlation from more standard techniques in quantitative genetics can easily be undefined when one of the genetic variance estimates is negative (Hill and Thompson 1978). \mathfrak{R}_G clearly provides a way round this particular limitation because it always renders an informative estimate.

Although the proposed methodology estimated the genetic correlation from phenotypic correlation quite efficiently, this estimation is valid within a given parameter space, i.e. when heritability is of intermediate value. However, for very low or high heritability, the method is likely to under-estimate

or over-estimate the genetic correlations, respectively. The present simulations made the assumption that environmental covariance was different from zero, and thus, a bias can be introduced in the estimated correlations. In the current work, the primary aim was to introduce the simulation methodology in estimating genetic correlation and to explore bias. However, the proposed methodology could be used as a base to develop a more advanced estimation model where the genetic correlation is adjusted for a parameter that indicates the degree of environmental covariance. Nevertheless, in standard experimental practice, the effect of environmental covariance may be avoided by randomizing individuals across environments, although it may not be possible to eliminate environmental covariance in natural settings. In the later case, phenotypic covariance may arise easily according to imposed sampling schemes of individuals.

Obtaining useful estimates of genetic correlations has traditionally been considered as a daunting problem, and only time-consuming standard experimental designs involving a few hundred individuals from easy-to-breed species typically allow for a rejection of the null hypothesis (Lynch and Walsh 1998). The method presented here, symbolized as \mathfrak{R}_G , seems to be relatively simple, far-reaching, and efficient under many realistic natural situations.

Acknowledgements

I thank Mauro Santos for comments on the method and the manuscript, and for providing experimental data. An anonymous reviewer provided useful comments on an earlier draft.

References

- Becker W. A. 1984 *Manual of quantitative genetics*, 4th edition. Academic Enterprises, Pullman, USA.
- Cheverud J. M. 1988 A comparison of genetic and phenotypic correlations. *Evolution* **42**, 958–968.
- Cheverud J. M. 1995 Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. *Am. Nat.* **145**, 63–89.
- Coyne J. A. and Beecham E. 1987 Heritability of two morphological characters within and among natural populations of *Drosophila melanogaster*. *Genetics* **117**, 727–737.
- David J. 1962 A new medium for rearing *Drosophila* in axenic conditions. *Drosophila Inform. Ser.* **36**, 128.
- Efron B. and Tibshirani R. J. 1993 *An introduction to the bootstrap*. Chapman and Hall, New York, USA.
- Falconer D. S. and Mackay T. F. C. 1996 *Introduction to quantitative genetics*, 4th edition. Longman, Harlow, UK.
- Hill W. G. and Thompson R. 1978 Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics* **34**, 429–439.
- Kempthorne O. 1957 *An introduction to genetic statistics*. John Wiley, New York, USA.
- Kendall M. G. and Stuart A. 1951 *The advanced theory of statistics*. Hafner, London, UK.
- Laayouni H., Santos M. and Fontdevila A. 2000 Toward a physical map of *Drosophila buzzatii*: use of randomly amplified polymorphic DNA polymorphisms and sequence-tagged-site landmarks. *Genetics* **156**, 1797–1816.
- Lande R. 1979 Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402–416.
- Leibowitz A., Santos M. and Fontdevila A. 1995 Heritability and selection on body size in a natural population of *Drosophila buzzatii*. *Genetics* **141**, 181–189.
- Loeschcke V., Bundgaard J. and Barker J. S. F. 1999 Reaction norms across and genetic parameters at different temperatures for thorax and wing size traits in *Drosophila aldrichi* and *D. buzzatii*. *J. Evol. Biol.* **12**, 605–623.
- Lynch M. 1999 Estimating genetic correlations in natural populations. *Genet. Res.* **74**, 255–264.
- Lynch M. and Walsh B. 1998 *Genetics and analyses of quantitative traits*. Sinauer, Sunderland, USA.
- Prout T. and Barker J. S. F. 1989 Ecological aspects of the heritability of body size in *Drosophila buzzatii*. *Genetics* **123**, 803–813.
- Riska B., Prout T. and Turelli M. 1989 Laboratory estimates of heritabilities and genetic correlations in nature. *Genetics* **123**, 865–871.
- Ritland K. 1996 A marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* **50**, 1062–1073.
- Roff D. A. 1995 The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture. *Heredity* **74**, 481–490.
- Roff D. A. 1996 The evolution of genetic correlations: an analysis of patterns. *Evolution* **50**, 1392–1403.
- Ruiz A., Santos M., Barbadilla A., Quezada-Díaz J. E., Hasson E. and Fontdevila A. 1991 Genetic variance for body size in a natural population of *Drosophila buzzatii*. *Genetics* **128**, 739–750.
- Searle S. R., Casella G. and McCulloch C. E. 1992 *Variance components*. John Wiley, New York, USA.
- Weigensberg I. and Roff D. A. 1996 Natural heritabilities: can they be reliably estimated in the laboratory? *Evolution* **50**, 2149–2157.
- Willis J. H., Coyne, J. A. and Kirkpatrick M. 1991 Can one predict the evolution of quantitative characters without genetics? *Evolution* **45**, 441–444.
- Young S. S. Y. and Weiler H. 1960 Selection for two correlated traits by independent culling levels. *J. Genet.* **57**, 329–233.

Received 19 December 2009, in revised form 30 April 2010; accepted 5 August 2010

Published on the Web: 19 May 2011