Biology and Philosophy (2006) 21:1–23
DOI 10.1007/s10539-005-0399-z
© Springer 2006

Genetic variance–covariance matrices: a critique of the evolutionary quantitative genetics research program

MASSIMO PIGLIUCCI
Departments of Ecology & Evolution and of Philosophy, SUNY-Stony Brook, NY, USA; (e-mail: pigliucci@genotypebyenvironment.org; phone: +1-631-632-1097)

Received 23 January 2004; accepted in revised form 10 January 2005

Key words: Evolutionary theory, Genetic covariances, Heritability, Pattern vs. process, Quantitative genetics

Abstract. This paper outlines a critique of the use of the genetic variance–covariance matrix (**G**), one of the central concepts in the modern study of natural selection and evolution. Specifically, I argue that for both conceptual and empirical reasons, studies of **G** cannot be used to elucidate so-called constraints on natural selection, nor can they be employed to detect or to measure past selection in natural populations – contrary to what assumed by most practicing biologists. I suggest that the search for a general solution to the difficult problem of identifying causal structures given observed correlation's has led evolutionary quantitative geneticists to substitute statistical modeling for the more difficult, but much more valuable, job of teasing apart the many possible causes underlying the action of natural selection. Hence, the entire evolutionary quantitative genetics research program may be in need of a fundamental reconsideration of its goals and how they correspond to the array of mathematical and experimental techniques normally employed by its practitioners.

Introduction: the emergence of a new field?

Very little philosophy of biology has yet been focused on evolutionary quantitative genetics, despite the fact that the field has seen a renaissance during the past decade, with several new textbooks (e.g., Falconer and Mackay 1996; Roff 1997; Pigliucci and Preston 2004) and a plethora of empirical papers in major biological journals such as *Evolution, Genetics*, *Journal of Evolutionary Biology*, etc. One concept in particular has emerged as crucial for the quantitative genetic study of complex (multivariate) phenotypes: the so-called genetic variance–covariance matrix, **G**. This is thought of and calculated as an actual matrix summarizing the genetic portion of the phenotypic variance of a series of morphological, life history, or behavioral characteristics, as well as all possible pairwise (genetic) covariances between said characteristics (Figure 1).

The basic idea is that **G** describes the degree to which the 'genetic architecture' (i.e., how traits are genetically connected to each other) determines the response of a population to natural selection. Suppose, for example, that there happens to be a positive genetic covariance between phenotypic traits X and Y (Figure 2a). Theory (and intuition) then predict that, other things being equal, if selection favors, say, an increase in trait X, trait Y will also be 'lifted' upwards as an indirect result of its covariance (or correlation, which is a

	Trait 1	Trait 2	Trait 3
Trait 1	genetic variance (or heritability) of trait 1		
Trait 2	genetic covariance (or correlation) between traits 1 & 2	genetic variance (or heritability) of trait 2	
Trait 3	genetic covariance (or correlation) between traits 1 & 3	genetic covariance (or correlation) between traits 2 & 3	genetic variance (or heritability) of trait 3

Figure 1. The idea of a genetic variance–covariance matrix.

standardized covariance) with X. On the contrary, should there be a negative genetic covariance between X and Y (Figure 2c), any selection favoring an increase in X's mean would automatically decrease Y's mean in the same population. This, of course, may speed up or greatly hinder adaptation, since the correlated response may or may not increase the average fitness of the organisms in question. Finally, if the two traits happen to be weakly correlated (a near-zero covariance, Figure 2b), then they are free to respond to selective pressures independently of each other.[1]

What makes all of this interesting to philosophers is that many evolutionary biologists have embraced the study of **G** with unabashed enthusiasm, despite major conceptual problems lurking in the wings, potentially underlying the whole research program. Steppan et al. (2002), for example, have declared the study of **G** to be 'a new emerging field' which will 'provide one of the most promising frameworks with which to unify the fields of macroevolution and microevolution.' Similarly, Shaw et al. (1995) have claimed that 'one motivation for estimating **G** matrices is that they will reveal the most likely paths of evolution,' and Roff (2000) has called the multivariate 'breeder's equation' (see below) – which features **G** – the 'central equation of evolutionary quantitative genetics.' Nor are these isolated instances; the biological literature on **G** has been copious for the last decade or so, and the pace is, if anything, quickening,

[1] Of course, things are not really that simple even under the best circumstances, since any two traits are then influenced by their joint covariance with any third trait; the trio is then influenced by how it is related to any fourth character, and so on. However, mathematical analysis and conceptual discussions of **G** rarely if ever step beyond the bivariate setting. I thank Lev Ginzburg (SUNY-Stony Brook) for pointing this out to me.

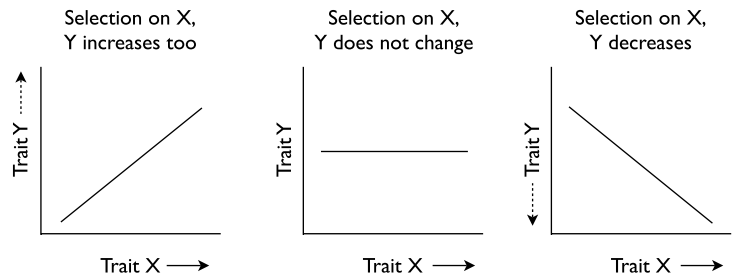

Figure 2. How genetic covariation between two traits, X and Y, influences the potential response to natural selection. From left to right, the value of the genetic covariance between the two traits changes from positive to zero to negative. Arrows indicate direct selection on X (solid) and the indirect (i.e., mediated by the covariance) response of Y (dotted).

with major reviews on empirical and statistical issues surrounding this topic being published on top journals (e.g., Turelli 1988; Cowley and Atchley 1992; Shaw et al. 1995; Pigliucci and Schlichting 1997; Phillips and Arnold 1999; Roff 2000; Johnson and Porter 2001; Steppan et al. 2002).

A philosophical analysis and critique of the uses and interpretations of evolutionary quantitative genetics and **G** is important because not only are the conceptual underpinnings of these ideas fundamental to modern biological theory, but **G** is an especially problematic concept in many ways. It is, I will argue, difficult to study empirically, and of limited value for conceptual reasons, and I will here present the beginnings of a critique that will hopefully stimulate future discussion. I will first attempt to explicate what evolutionary biologists who study and make use of **G** mean by the concept. Next, I will explore some of the practical problems faced in studying **G**; these problems will point towards some of the conceptual limitations of **G**. Finally, I will address what I consider to be perhaps the most important problem of all, namely the ways in which concepts like **G** can be invoked to obscure more fundamental issues in evolutionary quantitative genetics, and indeed, in evolutionary biology at large. Claiming that an emerging field will permit us to address long-standing problems, and interpreting the results of studies in ways that make it *seem* that such problems are being address, is not, I will suggest, the same thing as actually addressing such problems.

In a sense, the chief issue here is a particularly subtle and insidious one, sometimes referred to as 'reification.' Once that biologists have given a name to a certain concept (the genetic variance–covariance matrix, in this case), they proceed as if the label actually reflects a solid biological reality that can be meaningfully measured and dissected. Instead, I think that **G** is much too similar to other problematic concepts, such as the 'general intelligence' (also referred to as *g*) allegedly underlying responses to IQ tests. Clearly, statistics are measuring *something*, but the relationship between what they are measuring and most biologists' interpretation of it is far more complex than generally thought, leading to much less informative conclusions than are presented in the relevant literature.

Background: what exactly is G?

G is a crucial component of the multivariate extension of the classical 'breeder's equation' that has been the centerpiece of most analyses of phenotypic responses to selection in both natural and artificial systems throughout the 20th century (Falconer and Mackay 1996). The univariate (i.e., applied to one trait at a time) breeder's equation reads as follows:

$$R = h^2 S \tag{1}$$

where R is the response of a trait to selection applied with strength S, and h^2 is the heritability (i.e., the ratio between genetic and phenotypic variances) of that trait. The derivation of this equation is straightforward from simple assumptions about Mendelian genetics and multilocus genetics (e.g., see Falconer and Mackay 1996). What is important here is its conceptual underpinning: essentially, the equation says that heritability is the 'fuel' for any response to selection; if the heritability is close to one (100% of the phenotypic variance in the population is statistically associated with the genetic variance existing in the same population), then the population's mean for the trait under selection will shift by the quantity measuring the intensity of selection. In other words, with a heritability close to one, if selection favors, say, an increase in the height of the organism by 10 mm, the next generation will in fact have an average height of about 10 mm higher than the previous generation. At the other extreme, if the trait under selection has a heritability close to zero (i.e., there is no genetic variance for that trait in the population), then there will be no response to selection, regardless of how intense the selection pressure is: no fuel, no go. In many cases, of course, the heritability value will be intermediate, and the breeder's equation predicts a proportionality between selection pressure and response, with the exact value depending on the heritability. (Indeed, notice that simple algebra shows that the heritability can be estimated simply as the ratio between the selective response and the selective pressure – this method is actually common in biological practice, and the resulting quantity is often referred to as 'realized' heritability, because it is a direct measure of how much genetic 'fuel' the population really had during the selective episode.)

As is well known (Layzer 1974; Lewontin 1974; Kempthorne 1978), there are a number of problems with the idea of heritability (discussed for a philosophical audience by Sarkar 1998 and Downes 2004). As these problems translate to the multivariate level of **G**, it is worth briefly reiterating the primary difficulties. First, let us distinguish between the so-called 'broad' (often symbolized as H^2) and 'narrow' (often symbolized as h^2) senses of heritability. Mathematically, they are defined as:

$$H^2 = \frac{V_g}{V_p} \tag{2}$$

$$h^2 = \frac{V_a}{V_p} \tag{3}$$

where V_p is the total phenotypic variance for a given trait in a particular population (in a specific environment), V_g is the fraction of that variance that is statistically (not necessarily causally: see Kempthorne 1978) attributable to differences among genotypes, and V_a (referred to as the 'additive' genetic variance) is a subset of V_g which depends only on those allelic effects that actually contribute to a response to selection.[2]

Criticism of H^2 are more severe than those of h^2, because the latter measure is statistically more refined and derived from more sophisticated experimental designs (H^2 is the only measure of heritability that can be obtained for human populations, where controlled breeding is obviously unethical; which implies that any claim as to the heritability of human traits is in fact highly dubious at best, as remarked by all the authors cited above). Nonetheless, the following limits apply to all measures of heritability, and *a fortiori* to its multivariate extension, **G**:

(1) Heritability is a *local measure*, meaning that it can, and often does, change with changes in the population's gene frequencies and environments encountered;
(2) heritabilities do not reveal the *causal pathways* acting through development, nor, *a fortiori*, do they tell us anything about the extent of genotype-environment interactions; and, therefore,
(3) heritabilities do not provide a useful measure of the *long-term capability of traits to respond to selection*, nor are they of much use in determining the likely long-term evolutionary trajectories of phenotypic traits.

All of these limitations emerge from the simple fact that heritabilities measure a co-variation (which is a statistical attribute) between phenotypes and genotypes; as it is well known to any student of statistics, covariation does not necessarily imply causation, and – worse – often several distinct causal mechanisms can yield the same observable statistical pattern (Shipley 2000), making any inference from pattern to process doubtful at best.[3]

[2] V_g includes other statistical terms such as 'dominance' or 'epistatic' variance; these are portions of the total phenotypic variance that can be attributed to non-additive genetic effects, and that therefore do not contribute to the response to selection in sexually outbreeding organisms. As a subtle caveat, it should be noted that in some organisms that reproduce asexually, or where outbreeding is not complete, non-additive portions of V_g can also be expected to contribute to the response to selection.

[3] This, of course, does not imply that it is useless to obtain statistical summaries of the patterns of phenotypic variation found in natural populations. As Shipley (2000) very clearly explains, the question is what to do with those statistical descriptions, and the problem is that many biologists seem to have a tendency to do the wrong thing with them (i.e., to use them directly to infer underlying causal mechanisms).

Be that as it may, evolutionary quantitative genetic theory has been built upon a multivariate generalization of the breeder's equation, first proposed by Lande and Arnold (1983), whose intent was to provide a general (empirically applicable and theoretically sound) method to quantify natural selection on complex ensembles of phenotypic traits. The interested reader is referred to Lande and Arnold's landmark paper for the relatively simple mathematical derivation of the crucial formulas (which relate to the general statistical technique of multiple regression analysis). For our purposes here, it suffices to examine one version of the multivariate breeder's equation that has striking formal and conceptual similarities with the univariate version introduce above:[4]

$$\Delta Z = G\beta \tag{4}$$

where ΔZ is a vector of phenotypic responses to selection for a set of phenotypic traits (the delta sign indicates that the vector is made of the differences between each phenotypic trait's mean after and before the selection episode); β is a similar vector (often called of selection 'gradients') that specifies the intensity of selection on each phenotypic trait; and **G** is the above-mentioned genetic variance–covariance matrix. The similarity with the univariate version of the same equation is striking: ΔZ is a vector that is in fact made of a column of individual R's; **G** includes (unstandardized) measures of h^2 (plus the off-diagonal covariances, see Figure 1), and β is a vector made of individual values of S. Even at the multivariate level the intuition is that the response of a population to selection is proportional to the product of the intensity of selection on the traits in question and the 'fuel' available at the level of genetic architecture, this time in the extended sense of genetic variance–covariances.

Lande and Arnold (1983) proposed the use of the multivariate version of the breeding equation to approach two long-standing problems in evolutionary biology: predicting long-term multivariate responses to selection, and estimating past selection. Their idea was that given **G** and a multivariate selection regime, one could read off the likely response to selection from ΔZ, the phenotypic vector. Conversely, they reasoned, one could invert the equation to derive estimates of past selection from knowledge of current phenotypic values in the population and, again, the genetic variance–covariance matrix.[5]

Another conceptually important interpretation of **G** was then articulated by Cheverud (1984, 1988), who pointed out that the set of variance–covariances

[4] It is to be noted that several assumptions go into the derivation of both the uni- and the multivariate versions of the breeder's equation, perhaps the most crucial of which is that the phenotypic values in the population are normally distributed. These assumptions are summarized and discussed by Lande and Arnold (1983) among several others. However, what is of concern to me here is the decoupling between the statistical representation of the data (even if accurate) and the biological inferences that such representation allows. Accordingly, I will treat the equations as sound from a purely statistical perspective and concentrate on the biological problems they pose.

[5] It should be noted that Lande and Arnold are clearly pluralists when it comes to methods of measuring selection, but their version of the breeder's equation has in fact become the standard, though not the exclusive, way of proceeding for evolutionary quantitative geneticists.

can be thought of as a measure of genetic *constraints* on future evolution. In this sense, in the same way that the diagonal elements of **G** measure the short-term readiness of a character to respond to selection (just as heritability does in the univariate breeder's equation), the off-diagonal elements measure how much the evolution of a given trait is slowed down (or accelerated) by the co-evolution of another one (again, see Figure 2). In the same way that Lande and Arnold suggested that the equation could be inverted to detect selection in the past, on Cheverud's interpretation one can invert the equation in order to detect the constraints that have limited the effectiveness of past episodes of selection.

The many problems with G

The study of **G** and its related equations is often thought to provide insight into long-standing important problems in evolutionary theory, and has accordingly been actively embraced by practitioners of the field. But the rhetoric surrounding **G** ought not blind us to the limitations intrinsic in the idea, both practical and, especially, conceptual. Indeed, I will argue that if these problems are taken as seriously as they deserve to be, many of the promises of **G** turn out to be seriously overstated. In this central section, I will gradually build a case that, when considered in its entirety, undermines the whole **G**-based research program. I begin with a brief discussion of technical issues, such as the choice of traits to characterize **G** and the artificiality of the experimental designs used. I then move to more conceptual and philosophical issues, such as the consequences of the locality and instability inherent in the definition of **G**, problems related to the allegedly crucial role of **G** in inferring the existence of constraints, the near-impossibility to distinguish between selection and drift acting on covariance matrices, and the difficult issues related to the inferential move from population-level patterns to individual-level processes.

While most of these problems could be individually discounted as only partially undermining the usefulness of **G** in evolutionary quantitative genetics, I maintain that – when taken seriously as a whole – they shake the foundations of an entire research program. In the last section of this paper, then, I suggest some alternative approaches that have in fact at times been discussed by biologists, but that have curiously taken a back seat to the use of **G**.

Which traits make up G?

The first difficulty faced in studying **G** is finding a non-arbitrary way of choosing the traits to be used to construct the variance–covariance matrices themselves. Asking how a *generic* **G** matrix evolves (something that seems to be on people's minds, if one considers the titles and tone of most papers published in this field) is close to meaningless, since empirically we will always be dealing with *specific* matrices measured on a small subset of all possible or even all

biologically meaningful traits; generally, in fact, we will be dealing with matrices comprised of just a few traits because of logistic limitations. The answer to how stable **G** is over time, in other words, can depend critically on the traits actually used to empirically estimate the matrix, as well as on the phenotypic plasticity (the responsiveness of the traits to changes in environmental conditions: Pigliucci 2001) displayed by those traits in ecologically relevant environments. These facts are obvious, but they are rarely discussed in the biological literature on **G**.

To make the problem concrete, notice that *some* pair wise combinations of traits will tend to show stable covariances not only within, but across species, or even across higher taxonomic/phylogenetic levels of analysis; in these cases, **G** will seem quite stable. For example, most of the species of the plant family Brassicaceae have flowers characterized by two sets of stamens (the male sexual organs), in which four stamens are longer than the remaining two. While the reason for this is far from being clear (Karoly and Conner 2000; Conner 2002), the size of the two sets of stamens will show covariation at the taxonomically high level of an entire *family* of plants. On the other hand, it is well known that many traits can be uncoupled not merely within a population or species, but even within an individual organism! For example, many semi-aquatic plants produce two distinct kinds of leaves in response to the particular environment in which they find themselves in (below or above water, see Wells and Pigliucci 2000 for a review); parameters describing the shape of these two sets of leaves are completely uncoupled by the phenotypic plasticity that allows the plant to produce the two kinds of structures in an environment-dependent fashion; a **G** matrix constructed using parameters describing leaf shape as some of the traits would, therefore, be unstable not just over evolutionarily meaningful time frames, but even within an individual's developmental life-cycle.

The upshot of this is that quantitative geneticists who focus on stamens in Brassicaceae are likely to reach completely different conclusions about the stability of *the* genetic variance–covariance matrix from their colleagues who instead study leaf evolution in semi-aquatic plants. Indeed, research by Waldmann and Andersson (2000) in populations of *Scabiosa columbaria* and *S. canescens* found, among other patterns, that 'the magnitude of (co)variances was more variable among characters than among populations,' i.e., the results of a given study of **G** depend more strongly on which traits the investigators choose to focus on then it does on the species selected! Again, this ought to be obvious, but much of the literature is written in a way that implies that the evolution of **G** is a single kind of thing, and that it makes sense to think of **G** itself, somewhat independently of the particular traits used to calculate it.

Artificiality of estimates

The way in which quantitative geneticists study **G** empirically often involves the creation of an artificial set of offspring derived from carefully designed

crosses among individuals sampled at random from a natural population (Falconer and Mackay 1996). The most effective breeding design to estimate components of **G** is the half-sib approach, in which one compares the phenotypes of half sibs, i.e., of individuals that share only one of the two parents. This design is better than alternative ones (such as the full sib design, or the comparison of clones) because it allows a finer partitioning of the phenotypic variance. For example, while comparing clones means that it is possible only to estimate a gross genetic variance (as distinct from environmental, interaction, and error variances), a half-sib design permits the experimenter to further distinguish additive and non-additive genetic variances, and even maternal effects.[6]

There are two major problems with this approach: first, one can reasonably wonder if, in carrying out such experiments, one is really studying the genetic (co)variance matrix of a natural population. What I am questioning here is whether the **G** matrix derived from approaches using controlled crosses should be thought of as the matrix of the natural population from which the crosses were derived. The researchers in these cases are estimating genetic parameters in an artificially created set of genotypes, one that was not actually found in nature, and that had little chance of ever being there even if the population is mostly outbreeding. This is because it is vanishingly unlikely that the individuals in the population in question would ever cross in even approximately the same pattern as required by statistical tests and laboratory experiments. This problem seems to be completely ignored by practicing biologists, who happily go about discussing the relative merits of various breeding designs. But breeding designs are experimental devices that make sense, as the name clearly implies, when the goal is to *breed* certain characteristics in a population. The problem that evolutionary quantitative geneticists are trying to solve, on the other hand, is very different: it is to *estimate* a quantity allegedly characteristic of natural populations. But this cannot be what they actually do, and the problem of the resemblance, if any, between estimates of **G** obtained with breeding designs vs. whatever **G** would be obtained by actually sampling a natural population directly is rarely, if ever, addressed.[7]

---

[6] In practice, though, this can be achieved only by the use of very large sample sizes, which still do not provide much power for the complex statistical analyses necessary to make sense of these breeding designs. Furthermore, these 'variance components' often turn out to be of difficult or dubious biological interpretation.

[7] A similar problem is well known to arise (and often is also merrily ignored) in the case of estimates of heritability. This quantity can be estimated by using artificial breeding designs, or it can be obtained by comparing the generation after selection to the one before selection (as mentioned above, the second method yields what is often referred to as the 'realized' heritability). It is very often the case (Falconer and Mackay 1996) that realized heritabilities are much lower than those estimated by breeding designs, precisely because these designs create artificially structured populations with high level of outbreeding (and hence higher genetic variation), not generally found in nature.

One way to understand the problem is this: imagine that we wish to know the average height of individuals in a population of plants. What we do is to sample individuals from the population, measure their height, and then calculate the mean and some measure of dispersion (the variance, for example). We do *not* pick some individuals, cross them in specific combinations, and then measure the height of their progeny. The resemblance between the results obtained with the first and the latter methods is a matter of empirical determination, but the further the breeding design is from the actual mating pattern in nature, the less reliable the experimental approach will likely be.

Which brings us to a related problem: in what sense, if any, can one measure **G** in non-outbreeding organisms? The fact is that for the many species of organisms that are more or less inbred (which, outside of mammals and some other groups of vertebrates, is by no means the exception) the effective population size may be closer to a dozen, sometime only two or three individuals. But the statistical methods usually employed to estimate **G** require large sample sizes, often of the order of 90 or more 'families' of genetically related individuals. This is *more* than the actual number of genotypes in many natural populations of inbreeders (and of some outbreeders as well). What then? Does that mean that **G** does not exist in those populations? Or, if it does, is it possible to 'sample' it only by creating a highly artificial outbreed population from the few naturally available (inbred) genotypes? If the latter is the case, **G** is beginning to sound like something that may materialize with great effort and expense in the biologist's laboratory, but that has little to do with the natural populations that were the supposed objects of study.

Locality and instability

Moving toward more conceptual difficulties, we come to the fact that **G**, like heritability (h^2), is a *local* measure; that is, any estimate of **G** applies only to the particular population (with the particular genes and allelic frequencies actually extant), in the particular environment, in which **G** is calculated. As with h^2, if the available genes or the gene frequencies in the population change, so too might **G**; similarly, in a different range of environments (or if the population becomes differently distributed among the environments encountered), **G** might well change. Given this, if one wants to make use of **G** in simulations of long-term evolution (e.g., Via 1987), one must assume that the matrix stays constant (or at least proportional) to the ancestral state over the time period one is investigating (generally thousands of generations). But several authors (e.g., Turelli 1988; Pigliucci and Schlichting 1997) have pointed out that this is highly unlikely on first principles, because evolution *de facto* changes gene frequencies, and therefore **G** itself; nor is it unreasonable to suppose that over such time periods the environment encountered by the population may change as well. Therefore, while it is sensible to assume near-constancy of **G** for short-term applications (e.g., crop or animal breeding, artificial selection experi-

ments, or perhaps even evolution in wild populations over few generations), the hypothesis of approximate constancy becomes less and less likely the more widely separated the relevant populations are in time and/or space.

Ultimately, the question of the stability of **G** over long time periods and across environments is of course empirical, and it has accordingly generated a significant number of studies aimed at settling it. While a few authors have found constancy of the elements of **G** (e.g., Brodie III 1993 for anti-predator traits in two populations of garter snakes; Arnold and Phillips 1999, also in garter snakes), several others have demonstrated evolutionary divergence between populations (Waldmann and Andersson 2000 in species of the plant genus *Scabiosa*; Phillips et al. 2001 in *Drosophila melanogaster*), or species (Paulsen 1996 in buckeye butterflies; Steppan 1997 in leaf-eared mice; Waldmann and Andersson 2000 between species of *Scabiosa*).

The same locality issue, of course, applies to the environments experienced by the populations whose traits we are measuring in order to estimate **G**. Before the relatively recent resurgence of interest in the concept of phenotypic plasticity, the genotype-specific property of producing distinct phenotypes in different environments (e.g., Bradshaw 1965; Schlichting 1986; Sultan 1987; Scheiner 1993; Pigliucci 2001; West-Eberhard 2003), it was customary for quantitative geneticists to ignore environmental effects and assume that their results would hold more or less regardless of what environments were used in their experiments. There were always very good theoretical reasons not to do so (Lewontin 1974), but there is now overwhelming empirical evidence that the parameters important in statistical genetics can be highly sensitive (plastic) to environmental conditions. I listed above some references concerning the special case of heritability, but the literature is becoming clear also in the more general situation of components of **G** matrices. For example, Begin and Roff (2001) studied genetic (co)variances in two species of crickets, each reared in two environments, concluding that 'the expression of the genetic architecture can vary with the environment' and admonishing that 'future studies should compare **G** matrices across several environments.'

Laudable as the recommendation that **G** matrices be compared across multiple environments might be, for a variety of reasons actually doing so will not prove easy (and, in fact, is rarely done). First, it should be noted that quantitative genetic experiments, by their very nature, already tend to be very large (and, hence, expensive and time consuming). Since one is studying (often subtle) differences among many traits, and because the statistical power of even the best crossing designs is fairly limited (Mitchell-Olds and Shaw 1987), large samples are needed if statistically significant results are to be obtained. In order to study genetic (co)variances in multiple environments, the size of the experiment increases by at least as many times as the number of distinct environments one wishes to consider; the situation is even worse if more than one kind of environmental factor (e.g., temperature *and* light quantity) is to be explored. Expanding the already large-scale experiments designed to estimate

G to include multiple environments presents very serious logistical and funding difficulties.

Even if such difficulties could be solved in particular cases, however, it is generally not at all clear what set of environments should be chosen for such experiments. It goes without saying that one cannot test all possible environments, nor even all those environments a species is likely to have encountered in its recent evolutionary history (of which we are often ignorant anyway). If one wishes to carry out research in the field, the logistical challenges become quickly very severe (e.g., Mitchell-Olds and Bergelson 1990); if one opts for controlled conditions in the laboratory, growth chambers or greenhouse, then a whole different set of problems is raised by the fact that these conditions actually represent somewhat (although not entirely) 'novel' environments that may alter the genetic parameters of interest in unpredictable fashions (e.g., Weigensberg and Roff 1996; Sgró and Partridge 2000; Hoffmann et al. 2001). All of this may sound excessively negative, but these are serious difficulties that ought to be dealt with explicitly. That ignoring these problems makes doing quantitative evolutionary genetics easier is, in the end, no excuse at all.

*Does **G** indicate constraints imposed by the genetic architecture?*

More philosophically interesting are problems concerning **G** that cut to the core of *why* biologists use the concept to begin with. Let us start with the notion that the predictability of future phenotypic evolution is predicated on **G** revealing 'constraints' on evolutionary change imposed by the 'genetic architecture' underlying complex phenotypes. The idea is that trade-offs between traits to which an organism can allocate available resources (for example, between survival and reproduction) should manifest themselves as observable (negative) genetic covariances between the traits in question. If this were true, studies of **G** matrices could reveal features of the underlying trade-offs that influence the direction of phenotypic evolution, a major goal of evolutionary biology. Unfortunately, work by Houle (1991) and Gromko (1995) has dealt what should have been devastating blows to these uses of **G** in evolutionary theory; oddly, despite these key papers being published in *Evolution*, the premier journal in the field, their arguments have scarcely made a dent in the literature.

Let us begin by considering Houle's contribution. He examined a simple model of genetic architecture underlying a trade-off between two phenotypic traits (Figure 3): the observed (genetic) covariance between traits z_1 and z_2 is generated by the sum of effects due to two basic classes of genes. The **A** loci influence the acquisition of resources: the more resources there are, the more *both* z_1 and z_2 will increase in value; the **R** loci influence the partition of the available resources, thereby determining a negative relationship between z_1 and z_2. The standard quantitative genetics theory simply says that if there are **R** loci, then one should observe a negative genetic correlation between z_1 and z_2,

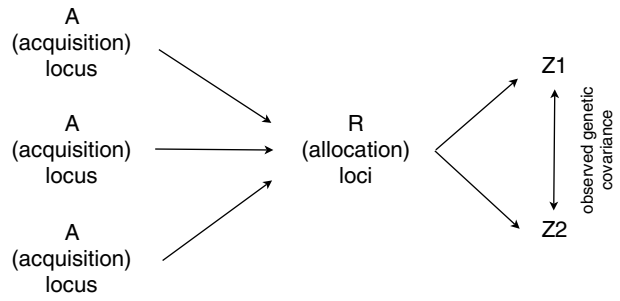

Figure 3. A rendition of Houle's model of trade-offs and genetic covariances (adapted from Houle, 1991).

corresponding to the underlying trade-off imposed by the particular genetic architecture. But Houle (1991) clearly showed that this is not necessarily the case: if the variance generated by the **A** loci is large (e.g., because there are many more **A** than **R** loci), then the observed genetic correlation between z_1 and z_2 will be *positive*, even though there is, in fact, an underlying trade-off. Houle's model implies that the goal of inferring trade-offs from observable genetic correlations is problematic at best.

Can we at least conclude that there are no underlying constraints if there is no observable genetic covariation between traits? Gromko's (1995) model crushed that expectation as well: even where there is no genetic covariation between two traits, these can in fact severely constrain each other's phenotypic evolution! Gromko (1995) was interested in pleiotropy, the genetic phenomenon by which the same gene (or set of genes) can affect multiple characters, some of which may not necessarily be related in any 'logical' fashion (i.e., developmental or selective) to each other. Genetic covariances are often interpreted as the result (mechanistically) of one of two phenomena: linkage (i.e., two genes are physically close to each other on the same chromosome, so their effects are statistically difficult to decouple because recombination is rare), or pleiotropy. Most authors think that only the latter is interesting biologically, because – in outbreeding organisms – linkage is eventually eliminated by continuous recombination. If indeed linkage is insignificant on evolutionary time scales, genetic covariances become important indicators of pleiotropy, which is itself an entry point into the developmental genetics of organisms that are not yet (nor likely to be any time soon) amenable to detailed molecular studies.

The goal of Gromko's work was to investigate the correlated responses to selection of those characters tied to each other pleiotropically (i.e., through the effect of the same genes on more than one character). He ran a series of simulations in which virtual populations were characterized by different combinations of pleiotropic effects. One of the results was that many different patterns of pleiotropy produced the same genetic correlations – in other words, the same **G** matrices could be produced by many different genetic mechanisms.

To put it as Gromko (1995) did, therefore, 'the genetic correlation does not uniquely determine a set of pleiotropic effects.' Further, the sign and magnitude of the genetic (co)variances could not be used to predict the way that the pleiotropic effects would influence actual evolutionary outcomes. Some sets of pleiotropic relationships allow correlated responses to selection that, if observed in nature or the laboratory, would lead to the conclusion that there is little or no pleiotropy connecting the traits in question. Indeed, in some cases, the simulated correlated responses to selection were in the *opposite* direction to that predicted on the basis of the observable genetic covariance. Again Gromko (1995): 'Whereas it has been previously established that genetic correlations are not necessarily constraints (see Houle's work discussed above), the alternative is also true: correlated response can be strictly constrained despite a genetic correlation of zero.'

In short, genetic (co)variance matrices do not reflect in a simple manner the underlying genetic or developmental structures, and therefore cannot be used to predict long-term evolutionary change, which depends on changes in the genetic architecture underlying the traits of interest. Where the genetic (co)variance matrix implies that the traits in question are independent (i.e., where there is no observable covariation), the traits can in fact be functionally connected; conversely, where the genetic (co)variance matrix registers a strong correlation between two characters, the latter can in fact be biologically independent, or even functionally coupled in the *opposite* direction of that implied by the genetic covariance. Genetic covariances, then, are not indicators of trade-offs or constraints, and they cannot be reliably used to elucidate the genetic architecture underlying a set of phenotypic characters.

There is an important note to add here. Whenever I have confronted some of my colleagues with the implications of the work by Houle and Gromko, a frequent response takes the form of an interesting conceptual maneuver: people *redefine* 'genetic architecture' precisely as the observable pattern of genetic (co)variances. This, by fiat, solves all problems, because one can verify the presence or absence of constraints *by definition* simply by examining the structure of **G**. But this is a very odd move indeed, for the real interest is in using **G** to *infer* glimpses of the true underlying genetic architecture. To say that one studies genetic covariances in order to study genetic covariances is obviously circular and rather empty. The move, seems to me, underscores the fragility of the whole **G**-based research program.

Selection vs. drift

Another major reason for the interest in studying **G** reflects an old problem in quantitative evolutionary theory: distinguishing between the effects of selection and drift in the phenotypic evolution of natural populations (a problem that has been analyzed in depth from a philosophical perspective; for example see Millstein 2002; Skipper 2002; Walsh et al. 2002 and references therein). This is

part of what Shaw et al. (1995) were getting at when they wrote that 'one motivation for estimating **G** matrices is that they will reveal the most likely paths of evolution.' But there actually is no reason to think that this problem is in general tractable, and good reasons to think that it may not be. The difficulty, of course, is that evolutionary biology is by its very nature a historical science (Cleland 2002), and hence one in which the problem of multiple processes (e.g., selection and drift) yielding similar (and often empirically indistinguishable) patterns is difficult to address because we often lack sufficient information about actual historical paths (Shipley 2000). No statistical technique, no matter how sophisticated, is going to provide a silver bullet, despite the hopes expressed by practitioners of the field, as we shall see soon.

The hope of distinguishing selection from drift in the case of multivariate phenotypic evolution hangs on the idea that selection is expected to generate patterns of change in (co)variance matrices that are qualitatively different from those typically generated by drift. More precisely, the standard prediction (see Roff 2000 for a review) is that drift should affect all elements of a (co)variance matrix in the same way, thereby causing proportional changes in the matrix over time, when compared to the ancestral **G**. Selection, on the other hand, is expected to act on subsets of characters in different manners; this would thereby alter the structure of **G** so to make the (co)variance matrix of descendant populations (after selection) qualitatively different from (i.e., not simply proportional to) that of their ancestors. If all this were true, one could distinguish selection from drift by comparing (co)variance matrices over time, or between different populations. Unfortunately, in a survey of available studies, Roff (2000) concluded that 'the null hypothesis that most of the variation can be attributed to drift rather than selection cannot be rejected,' even when the action of selection was suspected. In other words, in many cases the expected results of drift on a (co)variance matrix and the predicted effect of selection on the same (co)variance matrix cannot be distinguished. Roff (2000) concluded his review with the hope that 'further development of statistical tests' would help 'distinguish these two forces.'[8]

But the problem is not statistical in nature, and neither larger samples nor fancier math will help. Work by Phillips et al. (2001) shows that, except in very special situations, it will be impossible to use **G** matrices to distinguish selection from drift, since in any particular case the patterns that emerge as changes in the co(variance) matrix are consistent with both the action of selection and of neutral drift. Phillips and coworkers demonstrated this very elegantly by establishing several populations of the fruit fly *Drosophila melanogaster* and subjecting them to genetic drift by severe bottlenecks in a selectively benign (quasi-neutral) laboratory environment. They then compared the **G** matrices of

[8] Incidentally, I do not subscribe to the commonplace idea among biologists (and some philosophers: Sober 1984) that drift is an evolutionary 'force.' Rather, I tend to agree with analyses by Matthen and Ariew (2002), Walsh et al. (2002), Walsh (2003) and Pigliucci and Kaplan (forthcoming) that call into question the whole metaphor of 'forces' in biology.

the resulting descendant populations with that of the founding stock. When these authors considered the *average* **G** across populations (a statistically useful, but biologically meaningless, construct), this did indeed follow Roff's expectations: the populations that had undergone drift had an average **G** that was proportional to that of the founding population, as predicted by the theory *for a set of replicated populations* (i.e., when one actually knows the historical path of evolution). If these results had held for the individual populations that underwent drift, rather than for the statistical construct created by averaging them, this would have been good news indeed. But alas, when Phillips and collaborators examined the **G**s of individual populations that had undergone drift, they found that most were not proportional to their control at all: i.e., these matrices *appeared* to have been produced by selection, not drift, even though there was no significant selection going on![9]

The problem can be summarized in this fashion: *if* evolutionary biologists had access to neatly replicated historical events (as in the artificially constructed case of Phillips et al. 2001), *then* they could use the observed variation among **G** matrices to distinguish between selection and drift. The problem is that – except in very special cases – all biologists have access to is a series of populations that evolved naturally, generally from a set of ancestors whose phenotypic and genetic features are known only vaguely, if at all. Under these conditions, it is entirely possible that what looks like the result of selection is in fact the outcome of purely random processes. Once again, we run straight into the problem that many possible causal paths may converge to the same observable pattern, so that backward inference from pattern to process cannot be carried out in any straightforward way.

Individual vs. population levels

The last question I wish to raise in this section involves the way evolutionary quantitative geneticists talk about the mechanics of the evolution of **G**. Genetic variance–covariance matrices are often interpreted as the result of past

[9] A posteriori, this result should not have been surprising at all. Consider the much simpler, and better understood (see any population genetics textbook, such as Hartl and Clark 1989, e.g., chapter 2, and in particular Figure 1 on p. 63) case of drift affecting the changes in gene frequency of two alleles at a single locus. Elementary population genetic theory (and empirical evidence) shows that the two allele's frequencies will fluctuate over time in each small population (if there is no selection), following a random walk which will end up, eventually, in the fixation of one or the other allele. Across populations, each allele is expected to be fixed by drift at a frequency proportional to that allele's initial frequency in the founding population. Now, *on average*, this is exactly what happens. However, if one examines the actual evolutionary trajectory of either allele in any particular sub-population, one will see a definite trend toward fixation, which one could reasonably interpret as the result of directional selection! It is *only* because the investigator has several replicated populations available that it is clear that it is drift, and not selection, that is changing the allelic frequencies. Alas, such epistemic luxury is usually not available for studies of actual natural populations.

evolutionary 'forces,' such as selection and drift (see footnote 8 on the concept of 'forces' in evolution). But surely biologists do not mean to say that selection acts directly on the variance components; variance components are statistical constructs that summarize individual attributes at the *population* level, and are therefore not the sort of thing that discriminate physical processes can interact with. The physical processes that *are* natural selection act on particular entities, usually the individual organisms involved. While some physical processes are discriminate with respect to groups, group selection has never – to my knowledge – been invoked in discussions of **G**. I am forced to conclude that what those biologists who refer to the evolution of **G** *mean* is that **G** provides a summary indicator of the more complex and multifarious processes that occur at the level of individuals.

The problem with the latter interpretation, of course, is that the inference from observed patterns at the population level to underlying causes at the individual level is anything but straightforward. One reason is provided by the combined works of Houle (1991) and Gromko (1995) referred to above, but there has also been recent and extensive debate bearing on this issue (in the case of natural selection) in the philosophical literature (e.g., Matthen and Ariew 2002; Millstein 2002; Walsh et al. 2002; Pigliucci and Kaplan forthcoming). When considering natural selection, it is clear that physical interactions at the individual level may result in predictable statistical patterns at the population level, and yet this does not imply that the reverse move (from population to individual) is just as straightforward. The point has been made more generally by Shipley (2000), who – in the context of discussing the relationship between causation and correlation in biology – concluded that biologists can test hypothesized causal models by comparing them with their predicted statistical 'shadows,' but cannot reasonably go from the latter to the former. Alas, that is exactly what a great part of the research project in evolutionary quantitative genetics[10] is all about! To put it into another fashion, we can calculate the statistics, but what sort of biological questions are they answering, if any?

*What then? Getting over **G***

If the idea of **G** is so fraught with both conceptual and empirical problems, why do evolutionary quantitative geneticists use **G** as the centerpiece of their research program? Furthermore, assuming that evolutionary quantitative geneticists would be better off if they gave up on **G**, are there alternative approaches to addressing the questions that **G** was supposed to, but cannot, answer?

The two issues are in effect somewhat linked. Quantitative geneticists may be continuing to use **G** for the same reason that many of them keep using the concept of heritability: on the one hand many practitioners (but by no means

[10] As opposed to quantitative genetics for the purposes of plant and animal breeding, which has much more modest practical goals.

all) do not often pause to seriously consider the limitations inherent in these tools, a professional hazard originated by the drive to 'do' things, rather than 'speculate' about them; on the other hand, it may not be clear to most practitioners what alternatives, if any, are available at the moment. One of the roles that philosophy of biology ought to pursue is of course that of providing thorough criticism of the currently used tools, as I have began to do in this paper in the case of genetic variance–covariance matrices. My hope is that if these difficulties are clearly presented they will be taken seriously, and that the biological community may become more interested in investigating what alternatives may be available. Indeed, some possibilities can be sketched even now, as they have been considered by some practicing biologists, on and off, for the past two decades.

As I showed above, a careful reading of the literature reveals that even staunch supporters of the current methods of evolutionary quantitative genetics feel uncomfortable about the situation they now find themselves in: remember Roff's (2000) call for entirely new analytical tools to distinguish the effects of drift from selection on **G**. Even earlier, in 1989, Barton and Turelli, two of the most prominent theoreticians in the field, published a rather skeptical article entitled: 'Evolutionary quantitative genetics: how little do we know?' The answer they arrived at – that in fact, we know very little about what we would like to know – has not changed much in the intervening 15 years. Furthermore, Shaw et al. (1995) state: 'Short-term predictions based on sound estimates of genetic parameters are likely to be *qualitatively* informative (my emphasis). In contrast, both empirical results … and theoretical considerations suggest that quantitative predictions for long-term selection and retrospective analyses of selection should be interpreted with caution.' This is rather an understatement of the problem, though probably as frank an admission as one can expect from scientists that have invested their career in the field. Moreover, this position is hardly adequate if we have no guidelines for how cautious we should be in any particular case, nor for what a 'cautious interpretation' would involve to begin with.

The success of plant and animal breeding reveals the adequacy of quantitative genetic theory for making short-term predictions. But, as Shaw and colleagues note, even within that restricted (and evolutionarily rather insignificant) time frame, the predictions are mostly 'qualitative' – fortunately, for plant and animal breeding, that is all that is ever necessary. However, over evolutionary time spans even the qualitative predictions are likely to fail because of the problems of locality and environmental-sensitivity of components of **G**. We do not have a theory of how **G** changes during long-term evolution, though we know that it does change. Developing such theory has been an elusive goal of theoretical evolutionary biology for a long time, but at the moment at least there is no sign of anybody coming even close to it (S. Gavrilets, University of Tennessee, pers. comm.).

Be that as it may, by now it should be clear that quantitative evolutionary biologists ought not to think of these statistical constructs as 'first

approximations' to be refined by further research; the difficulty with these constructs is not that they are imprecise (and, therefore, amenable to 'refinement'), but that they do not answer the questions we wish answered. For example, recall the above-mentioned peculiar shift in focus from using **G** matrices to *infer* the genetic architecture underlying certain traits to the much less interesting 'solution' of *defining* genetic architecture as the observed pattern of genetic variance–covariances.[11] In any event, 'further research' has been conducted in quantitative genetics for more than a century, and we still have not gotten much past the first steps. While the 1980s saw the formulation of the multivariate extension of the breeders' equation and some related refinements, the fundamental problems have remained the same. In fact, many of the difficulties have worsened now that we have empirical demonstration of the environmental liability and phylogenetic variation of genetic (co)variances.

Alternative approaches and ingenuous solutions to the problems outlined in this paper have, in fact, been proposed and occasionally pursued by biologists.[12] Even Lande and Arnold (1983), as well as Mitchell-Olds and Shaw (1987) pointed out that the broader goal is really to understand the mechanics and causality of natural selection and response to it in natural populations. They frankly admitted that the multivariate breeder's equation should truly be used only as a preliminary step to gain some insight into a particular system. That insight should then be combined with as much knowledge of the biology of the species in question to propose specific hypotheses about the causes of the observed patterns (for example, as done by Johnson 2002 in his study of the selective factors affecting life history in populations of a Poecilid fish). These hypotheses can then be tested by a variety of approaches, ranging from more sophisticated statistical techniques like path analysis and structural equation modeling (Shipley 2000), to experimental approaches using a variety of manipulations of the phenotype and environment of the organisms under study (Schmitt et al. 1999). Neither of these suggestions has gained widespread favor among biologists (though they have been used occasionally throughout the past two decades), and I suspect this is chiefly for two reasons, one practical the other conceptual.

The practical reason is less of interest to philosophers, though it may be worth documenting and pursuing by sociologists of science: it is simply *much* easier to measure a bunch of phenotypic traits in a single (or few) field season in a single (or few) location, run the appropriate multiple regression analyses,

[11] This is somewhat similar to avoiding to deal with what exactly is the relationship between IQ scores and intelligence by defining intelligence as whatever property is measured by IQ tests. One would seem to be justified in smelling question begging.

[12] I will not discuss here molecularly based approaches such as Quantitative Trait Loci mapping, or the whole field of evolutionary developmental biology. These are areas of research large enough to deserve their own treatment. Also, while these other approaches can in some way interact with evolutionary quantitative genetics, I do maintain that the latter is characterized by a sufficient degree of intellectual and methodological independence to warrant a more narrow focus at this stage of analysis.

and claim that one has 'studied natural selection.' To move from there (the supposed 'preliminary' step) to an extended series of field and laboratory studies that include repeated observations, manipulations of environmental circumstances, as well as manipulation of crucial aspects of the phenotype (by mutation, or cleverly exploiting the organism's phenotypic plasticity) requires much more time, sophistication, and money. The conceptual reason, of more philosophical interest, can be traced back to the rationale that went into the publication of Lande and Arnold's (1983) paper: the main goal there was to provide not just a way to statistically quantify natural selection in action, but to do so while obtaining coefficients of selection that could be directly plugged into the standard quantitative genetics equations for the prediction (or postdiction) of phenotypic evolution. It turns out that, until now, nobody has figured out a way to use path coefficients for the same purpose (but see Scheiner et al. 2000 for the beginning of such an attempt). This implies that a theoretical goal has been for all effective purposes overriding serious conceptual and methodological limitations of the techniques used. What makes this a possible dead end for the entire field is that there are good reasons to believe that the theoretical goal in question – the long-term prediction of evolutionary trajectories – is simply not achievable because of the problems of locality and liability of **G** discussed above.

What, then, is an evolutionary quantitative geneticist to do? I suggest that practitioners should take seriously the limitations outlined here and re-evaluate their goals in light of what not only the available techniques, but nature itself, will allow them to do. If it turns out that the problem of long-term prediction of evolution is not tractable, people should not be too bothered by it. Plenty of good science can be done without having to engage at that level of predictability. As Sober (1984, pp. 137–138) put it when referring to studies of natural selection: 'The world we live in may simply be such that certain questions are very hard to answer … It is not the scientist's fault that nature has made some of its secrets relatively opaque to human scrutiny.' But the scientist can surely be faulted for insisting in banging his head against nature's staunchest walls when it ought to be clear that other strategies circumventing such walls can be pursued and be much more productive. The alternative, at least equally interesting, research program of developing a mechanistic understanding of natural selection and how it acts in populations of animals and plants, has been lagging behind *despite the availability of sound empirical and analytical approaches* to pursue it. This implies that much could be gained by a re-assessment and re-alignment of evolutionary quantitative genetics' goals and practices.

Acknowledgments

I wish to thank Jonathan Kaplan for discussions and suggestions on a previous version of this manuscript. I also benefited from discussing these topics with

Josh Banta, Kent Holsinger, David Houle, Norris Muth, Patrick Phillips, James Rohlf, and Carl Schlichting. This work was partially supported by National Science Foundation grants DEB-0089493 and NSF-032446.

References

Arnold S.J. and Phillips P.C. 1999. Hierarchical comparison of genetic variance–covariance matrices. II. Coastal-inland divergence in the garter snake, *Thamnophis elegans*. Evolution 53: 1516–1527.

Barton N.H. and Turelli M. 1989. Evolutionary quantitative genetics: how little do we know? Annu. Rev. Genet. 23: 337–370.

Begin M. and Roff D.A. 2001. An analysis of *G* matrix variation in two closely related cricket species, *Gryllus firmus* and *G. pennsylvanicus*. J. Evol. Biol. 14: 1–13.

Bradshaw A.D. 1965. Evolutionary significance of phenotypic plasticity in plants. Adv. Genet. 13: 115–155.

Brodie E.D. III 1993. Homogeneity of the genetic variance–covariance matrix for antipredator traits in two natural populations of the garter snake *Thamnophis ordinoides*. Evolution 47: 844–854.

Cheverud J.M. 1984. Quantitative genetics and developmental constraints on evolution by selection. J. Theor. Biol. 110: 155–171.

Cheverud J.M. 1988. The evolution of genetic correlation and developmental constraints. In: Jong G.D. (ed.), Population Genetics and Evolution. Springer-Verlag, Berlin, pp. 94–101.

Cleland C.E. 2002. Methodological and epistemic differences between historical science and experimental science. Philos. Sci. 69: 474–496.

Conner J.K. 2002. Genetic mechanisms of floral trait correlations in a natural population. Nature 420: 407–410.

Cowley D.E. and Atchley W. 1992. Comparison of quantitative genetic parameters. Evolution 46: 1965–1966.

Downes S.M. 2004. Heredity and heritability. In: Stanford Encyclopedia of Philosophy, plato.stanford.edu. Stanford University, Stanford, CA, pp. plato.stanford.edu/entries/heredity.

Falconer D.S. and Mackay T.F.C. 1996. Introduction to Quantitative Genetics. Prentice Hall, Harlow, England.

Gromko M.H. 1995. Unpredictability of correlated response to selection: pleiotropy and sampling interact. Evolution 49: 685–693.

Hartl D.L. and Clark A.G. 1989. Principles of Population Genetics. Sinauer, Sunderland, MA.

Hoffmann A.A., Hallas R., Sinclair C. and Patridge L. 2001. Rapid loss of stress resistance in *Drosophila melanogaster* under adaptation to laboratory culture. Evolution 55: 436–438.

Houle D. 1991. Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters. Evolution 45: 630–648.

Johnson J.B. 2002. Divergent life histories among populations of the fish *Brachyrhaphis rhabdophora*: detecting putative agents of selection by candidate model analysis. Oikos 96: 82–91.

Johnson N.A. and Porter A.H. 2001. Toward a new synthesis: population genetics and evolutionary developmental biology. Genetica 112–113: 45–58.

Karoly K. and Conner J.K. 2000. Heritable variation in a family-diagnostic trait. Evolution 54: 1433–1438.

Kempthorne O. 1978. Logical, epistemological and statistical aspects of nature-nurture data interpretation. Biometrics 34: 1–23.

Lande R. and Arnold S.J. 1983. The measurement of selection on correlated characters. Evolution 37: 1210–1226.

Layzer D. 1974. Heritability analyses of IQ scores: science or numerology? Science 183: 1259–1266.

Lewontin R.C. 1974. The analysis of variance and the analysis of causes. Am. J. Hum. Genet. 26: 400–411.

Matthen M. and Ariew A. 2002. Two ways of thinking about fitness and natural selection. J. Philos. 49: 55–83.

Millstein R.L. 2002. Are random drift and natural selection conceptually distinct? Biol. Philos. 17: 33–53.

Mitchell-Olds T. and Bergelson J. 1990. Statistical genetics of an annual plant, *Impatiens capensis*. I. Genetic basis of quantitative variation. Genetics 124: 407–415.

Mitchell-Olds T. and Shaw R.G. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. Evolution 41: 1149–1161.

Paulsen S.M. 1996. Quantitative genetics of the wing color pattern in the buckeye butterfly (*Precis coenia* and *Precis evarete*). Evolution 50: 1585–1597.

Phillips P.C. and Arnold S.J. 1999. Hierarchical comparison of genetic variance–covariance matrices. I. Using the Flury hierarchy. Evolution 53: 1506–1515.

Phillips P.C., Whitlock M.C. and Fowler K. 2001. Inbreeding changes the shape of the genetic covariance matrix in *Drosophila melanogaster*. Genetics 158: 1137–1145.

Pigliucci M. 2001. Phenotypic Plasticity: Beyond Nature and Nurture. Johns Hopkins University Press, Baltimore, MD.

Pigliucci M. and Kaplan J. forthcoming. Making Sense of Evolution: Toward a Coherent Picture of Evolutionary Theory. Chicago University Press, Chicago, IL.

Pigliucci M. and Preston K. (eds) 2004. Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes. Oxford University Press, Oxford.

Pigliucci M. and Schlichting C.D. 1997. On the limits of quantitative genetics for the study of phenotypic evolution. Acta Biotheoret. 45: 143–160.

Roff D.A. 1997. Evolutionary Quantitative Genetics. Chapman & Hall, New York.

Roff D.A. 2000. The evolution of the *G* matrix: selection or drift? Heredity 84: 135–142.

Roff D.A. 2002. Comparing *G* matrices: a MANOVA approach. Evolution 56: 1286–1291.

Sarkar S. 1998. Genetics and Reductionism. Cambridge University Press, Cambridge, England.

Scheiner S.M. 1993. Genetics and evolution of phenotypic plasticity. Annu. Rev. Ecol. Syst. 24: 35–68.

Scheiner S.M., Mitchell R.J. and Callahan H.S. 2000. Using path analysis to measure natural selection. J. Evol. Biol. 13: 423–433.

Schlichting C.D. 1986. The evolution of phenotypic plasticity in plants. Annu. Rev. Ecol. Syst. 17: 667–693.

Schmitt J., Dudley S. and Pigliucci M. 1999. Manipulative approaches to testing adaptive plasticity: phytochrome-mediated shade avoidance responses in plants. Am. Nat. 154: S43–S54.

Sgró C.M. and Partridge L. 2000. Evolutionary responses of the life history of wild-caught *Drosophila melanogaster* to two standard methods of laboratory culture. Am. Nat. 156: 341–353.

Shaw R.G. 1992. Comparison of quantitative genetic parameters: reply to Cowley and Atchley. Evolution 46: 1967–1968.

Shaw F.H., Shaw R.G., Wilkinson G.S. and Turelli M. 1995. Changes in genetic variances and covariances: **G** whiz!. Evolution 49: 1260–1267.

Shipley B. 2000. Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference. Cambridge University Press, Cambridge, England.

Skipper R.A.J. 2002. The persistence of the R.A. Fisher–Sewall Wright controversy. Biol. Philos. 17: 341–367.

Sober E. 1984. The Nature of Selection: Evolutionary Theory in Philosophical Focus. MIT Press, Cambridge, MA.

Steppan S.J. 1997. Phylogenetic analysis of phenotypic covariance structure. I. Contrasting results from matrix correlation and common principal component analyses. Evolution 51: 571–586.

Steppan S.J., Phillips P.C. and Houle D. 2002. Comparative quantitative genetics: evolution of the *G* matrix. Trends Ecol. Evol. 17: 320–327.

Sultan S.E. 1987. Evolutionary implications of phenotypic plasticity in plants. Evol. Biol. 21: 127–178.

Turelli M. 1988. Phenotypic evolution, constant covariances, and the maintenance of additive variance. Evolution 42: 1342–1347.

Via S. 1987. Genetic constraints on the evolution of phenotypic plasticity. In: Loeschcke V. (ed.), Genetic Constraints on Adaptive Evolution. Springer-Verlag, Berlin, pp. 47–71.

Waldmann P. and Andersson S. 2000. Comparison of genetic (co)variance matrices within and between *Scabiosa canescens* and *S. columbaria*. J. Evol. Biol. 13: 826–835.

Walsh D.M. 2003. Fit and diversity: explaining adaptive evolution. Philos. Sci. 70: 280–301.

Walsh D.M., Lewens T. and Ariew A. 2002. The trials of life: natural selection and random drift. Philos. Sci. 69: 452–473.

Weigensberg I. and Roff D.A. 1996. Natural heritabilities: can they be reliably estimated in the laboratory? Evolution 50: 2149–2157.

Wells C. and Pigliucci M. 2000. Heterophylly in aquatic plants: considering the evidence for adaptive plasticity. Perspect. Plant Ecol. Evol. Syst. 3: 1–18.

West-Eberhard M.J. 2003. Developmental Plasticity and Evolution. Oxford University Press, Oxford, England.