

# An analysis of **G** matrix variation in two closely related cricket species, *Gryllus firmus* and *G. pennsylvanicus*

M. BÉGIN & D. A. ROFF

Department of Biology, McGill University, Montreal, Québec, Canada

## Keywords:

crickets;  
Flury hierarchy;  
**G** matrix;  
genetic correlation;  
heritability;  
quantitative genetics.

## Abstract

An important issue in evolutionary biology is understanding the pattern of **G** matrix variation in natural populations. We estimated four **G** matrices based on the morphological traits of two cricket species, *Gryllus firmus* and *G. pennsylvanicus*, each reared in two environments. We used three matrix comparison approaches, including the Flury hierarchy, to improve our ability to perceive all aspects of matrix variation. Our results demonstrate that different methods perceive different aspects of the matrices, which suggests that, until more is known about these methods, future studies should use several different statistical approaches. We also found that the differences in **G** matrices within a species can be larger than the differences between species. We conclude that the expression of the genetic architecture can vary with the environment and that future studies should compare **G** matrices across several environments. We also conclude that **G** matrices can be conserved at the level of closely related species.

## Introduction

The application of quantitative genetic theory to natural populations has stressed the importance of understanding the evolution of genetic architecture, which can be represented as a matrix of additive genetic variances and covariances of quantitative traits (**G** matrix). In the past, most studies of genetic architecture have focused on trying to empirically validate the use of the multivariate response to selection equation  $\mathbf{R} = \mathbf{G}\boldsymbol{\beta}$ , where  $\mathbf{R}$  is the vector of phenotypic responses in the traits under study, **G** is the matrix of additive genetic (co)variances, and  $\boldsymbol{\beta}$  is the vector of selection gradients (Lande & Arnold, 1983). This equation can be used to predict long-term response to selection or to retrospectively estimate the selection gradient that gave rise to differences between populations or taxa (Lande, 1979). In either case a necessary assumption of the model is that **G** remains constant or that it is changed only proportionally throughout phenotypic evolution (Lande, 1979). This assumption was

originally justified by the proposition that selection intensities in nature are typically weak and thus variance that is eroded by selection will be replaced by pleiotropic mutation (Lande, 1976, 1980).

The problem with this proposition is that there is no theoretical reason to assume that **G** typically does not change with time and phylogenetic relationship (Lofsvold, 1986; Turelli, 1988; Camara & Pigliucci, 1999). Artificial selection experiments have clearly demonstrated that under strong selection genetic variances and covariances are not stable (Shaw *et al.*, 1995; see Table 4.3 and pp. 174–178 in Roff, 1997) and empirical estimates of selection intensities in the wild have shown that these can be as large as those used in artificial selection experiments (Endler, 1986). Furthermore, selection intensities and their effects on genetic architecture inevitably vary from case to case, making the assumption of a constant **G** difficult to justify. This is not to say that **G** matrices must change during population differentiation or that the predictive equation cannot be used, but rather that the evolution of genetic architecture is a continuous process which cannot be described as an equal/different dichotomy. An alternative approach is to assume that the **G** matrices of two or more populations may differ to any degree and investigate

Correspondence: Mattieu Bégin, Department of Biology, McGill University, 1205 Dr Penfield Avenue, Montréal, Québec, H3A 1B1, Canada.  
Tel.: +1 514 398 4949; fax: +1 514 398 5069;  
e-mail: mbegin1@po-box.mcgill.ca

relative differences between matrices instead of testing for statistical rejection of the null hypothesis of equality (i.e. interval estimation rather than hypothesis testing).

In addition to the direct information provided by the distance between matrices, the pattern of matrix differences is hypothesized to contain footprints of past evolutionary forces that shaped present-day genetic architecture. Theory predicts that selection should cause divergence in matrices (Lande, 1979), that random genetic drift alone should result in proportional changes (Lande, 1979; Lofsvold, 1988) and that low levels of selection and drift should not alter the structure of the matrices (Lande, 1979, 1980). Investigating **G** matrix variation between species might thus provide important insights into population evolution and might help to link changes in genetic constraints with phenotypic evolution.

Patterns of **G** matrix variation within species are also informative. Because genetic parameter estimates are theoretically only valid for the environment in which they are measured (Falconer & Mackay, 1996; Roff, 1997), it is reasonable to expect that the expression of the genetic architecture of a population will vary with the environment. In addition to providing information on that interaction, such investigations might help in the comparison of **G** matrices across species, because the optimal rearing conditions may not be the same and thus observed differences between species may instead reflect differences due to environment.

Analyses of **G** matrix evolution should therefore allow the investigation of two questions: what is the degree of similarity between matrices and what can this reveal about the evolutionary history of these matrices. Unfortunately multivariate data sets, such as the ones represented by **G** matrices, are extremely difficult to compare, and finding a satisfactory statistical method to do so is at present an unresolved problem. Several different techniques exist (reviewed in Roff, 1997, 2000) but it is not clear which approach, if any, is preferable. However, one method, the Flury hierarchy (Phillips & Arnold, 1999), currently receives strong support from investigators (Steppan, 1997a,b; Arnold & Phillips, 1999; Camara & Pigliucci, 1999; Merilä & Björklund, 1999). It is therefore important to evaluate the ability of the Flury hierarchy to answer the two previously stated objectives of **G** matrix analysis. In this paper we do so by comparing the Flury hierarchy with two other published methods; the element by element approach (Roff *et al.*, 1999) and the method of percentage reduction in mean square errors (Roff, 2000). Because all these use different statistical approaches, we expect the Flury hierarchy to yield similar but not necessarily identical results compared with the other methods.

In the present analysis we compare the **G** matrices of two species of wing dimorphic crickets; a *Gryllus firmus* population derived from Florida, USA and a *G. pennsylvanicus* population collected in Québec, Canada. These

two populations are isolated by over 1500 km and hence there has probably been no direct or indirect gene exchange for thousands of generations. *Gryllus firmus* occurs in coastal and lowland areas of Eastern North America from Florida to Connecticut (Harrison & Arnold, 1982). In contrast, *G. pennsylvanicus* is widely distributed throughout inland North America (Alexander, 1957a; Vickery & Kevan, 1983). The two species hybridize in a zone of overlap in the Appalachian and Blue Ridge mountains (Harrison & Arnold, 1982; Harrison, 1985). Although these species differ in a number of morphological characters (body size, ovipositor length, hind wing length, colour of tegmina, number of file teeth), none are diagnostic (Fulton, 1952; Alexander, 1957a; Harrison & Arnold, 1982). We therefore expect the **G** matrix of these two closely related cricket species to have remained relatively constant through species divergence.

This paper investigates three major questions. (1) Are the results produced by the Flury hierarchy corroborated by other statistical approaches and does using several methods improve the ability to perceive all aspects of matrix evolution? (2) What is the degree of similarity between the **G** matrices of two closely related cricket species and what type of evolutionary forces could have shaped the observed difference? (3) Within a species, does **G** vary with rearing condition?

## Materials and methods

### Experimental protocol

The data used in the present analysis come from two experiments already described (Simons & Roff, 1994; Roff, 1995) and thus we present here only an overview. Crickets were reared on rabbit chow and housed in 4-L buckets at a density of 25 (*G. pennsylvanicus*) or 60 (*G. firmus*) newly hatched nymphs. The *G. pennsylvanicus* nymphs were the offspring of nymphs collected from a field at Mont St-Hilaire, Québec, the previous year, whereas the *G. firmus* were taken from a stock originally collected from northern Florida and maintained in the lab for approximately 35 generations prior to the study.

Full-sib families of *G. pennsylvanicus* were each divided among four cages, two cages placed outside at the collecting location where they experienced ambient temperature and photoperiod (39 families, 705 individuals), and two cages maintained in a growth chamber at 24 °C and a photoperiod of 17 h of light (39 families, 505 individuals). These two treatments will be referred to as 'Pfield' and 'Plab'. Note that the temperature and photoperiod chosen for the lab environment are not significantly different from the typical natural conditions encountered by this species during the rearing period (Simons & Roff, 1994).

The full-sib families of *G. firmus* were raised in a growth chamber at 25 °C and 15 h of light (43 families, 382 individuals) and at 30 °C and 17 h of light (49 families,

587 individuals). For simplicity we shall refer to the two environments of *G. firmus* as 'F25' and 'F30'. Note that the number of *firmus* families used here slightly differs from Roff (1995) because only one sex was used (see below) and because the number of families for each environment were reversed in the previous paper.

Because there are differences in morphology between the sexes and the morphs, we only selected females of the most common wing morph within each treatment. For *G. pennsylvanicus*, only micropterous (short wing) females were included, whereas for *G. firmus*, short wing females were selected from the 25 °C treatment and macropterous (long wing) females from the 30 °C treatment. Because the wing morph is not the same in the two *G. firmus* treatments, it will not be possible to ascribe **G** matrix differences specifically to rearing conditions or morph. The term 'environment' will thus be interpreted to include both variables.

From each female, five morphological measurements were taken: femur length (FEMUR), head width (HEAD), prothorax length (PTHL), prothorax width (PTHW) and ovipositor length (OVIP). The normality of the data was tested using the one-sample-Kolmogorov-Smirnov test (Lilliefors option). Results (not shown) indicate that nine of 20 traits (two species with two treatments each, five traits per treatment) have a distribution that deviates significantly from normality. However, frequency histograms showed that none of the traits are highly skewed, which suggests that the significant deviation from normality is partly caused by the large sample sizes. Various transformations did not solve the problem. To test the effect of data distribution on **G** matrix comparison, we used original, log transformed and standardized data sets to construct matrices and compared them using the Flury and element by element methods (see description of these methods below). Results (not shown) indicate that matrix comparisons are not substantially affected. Therefore, we report the analyses for the untransformed data. However, tests using randomization (see below) involve standardization to a common mean.

### Quantitative genetic methodology

The vector of selection gradients,  $\beta$ , can be decomposed into  $\mathbf{P}^{-1}\mathbf{S}$ , where  $\mathbf{P}$  is the inverse of the matrix of phenotypic variance and covariances and  $\mathbf{S}$  is a vector of selection differentials. This decomposition illustrates that the  $\mathbf{P}$  matrix plays an important role in trait evolution. Thus in addition to considering **G** matrix variation we also analysed the  $\mathbf{P}$  matrices. A second reason for the latter analysis is to address the conjecture that this matrix can be used as a surrogate measure of **G** (Cheverud, 1988; Roff, 1995, 1996, 1997). Genetic correlation matrices were not compared as they are not useful to reconstruct evolutionary trajectories (Deng *et al.*, 1999).

Quantitative genetic parameters were estimated using the Bootstrap method (5000 reshuffling runs) carried out

by the package H2boot (Phillips, 1998a). These calculations are based on a one-way analysis of (co)variance among full-sib families. By definition, the genetic variances so estimated include additive and nonadditive genetic components and may be contaminated by maternal effects. A general review of the heritability of morphological traits indicated that heritabilities estimated from full-sib designs are generally similar to those estimated from parent-offspring regressions (Mousseau & Roff, 1987). Moreover, this was verified for the heritability of femur length in *G. firmus*, suggesting no significant effects because of nonadditive genetic or maternal effects (Roff, 1998). For this study, it is therefore probable that genetic (co)variances estimated using full-sib families represent additive genetic (co)variances. However, to be cautious, we will not make that assumption.

Because the H2boot and CPCrand programs (see description below) do not allow nesting of cages within families, estimation of genetic (co)variances was carried out by pooling all the individuals of a family. This procedure results in an overall inflation of these parameters, measured as a 27% (or 0.03) absolute average difference between genetic (co)variances arising from nested and non-nested ANOVA/ANCOVA models. For 13 out of 20 traits, cage effects are significant. The non-nested genetic parameters are thus generally contaminated by common environmental effects and should be considered as higher limit estimates. To test the effect of this inflation on **G** matrix comparison, genetic parameters were calculated using a nested design (cages nested within families) and matrices constructed from these were analysed using the element by element method (see description below). It was found (results not shown) that the omission of nesting does not substantially affect the comparison of matrices. Therefore, we report comparisons based on non-nested analyses of (co)variance.

### The Flury hierarchy

The Flury hierarchy is a principal components approach to the comparison of matrices (Flury, 1988) that has been applied to **G** matrix analysis by Phillips & Arnold (1999). This method based on maximum likelihood determines which model is the best descriptor of the structural differences between two or more matrices. The hierarchically nested models are (a) unrelated structure (matrices don't have a single principal component in common), (b) Partial Common Principal Components (matrices share 1, 2 or 3 principal components), (c) Common Principal Components (all principal components are shared), (d) proportionality (all principal components are shared and the eigenvalues all differ by the same constant between matrices) and (e) equality (identical principal components and eigenvalues). For each model, the Flury hierarchy calculates a log-likelihood statistic to quantify the fit of that model to the observed matrices.

We used the jump up procedure (Phillips & Arnold, 1999) to test the goodness of fit of each model against the model of unrelated structure, thus providing a significance test for each model.

To avoid the assumption of multivariate normality in hypothesis testing, randomization is used to determine the probability that the unrelated structure model fits the data significantly better than each other models. In this analysis, 4999 randomized data sets were created, each run randomly assigning whole families to a population. The best fitting model (referred to as the verdict) is determined as the model immediately under the first significant probability, going from the bottom (unrelated structure) to the top (equality) of the hierarchy (Phillips & Arnold, 1999). For simplicity, only the verdict is given in the results section. This analysis was performed using the program CPCrand (Phillips, 1998b).

### The element by element approach

This method (Roff *et al.*, 1999) is based on an element by element comparison of matrices and can be used to obtain three different types of information. The first use is to test the hypothesis of matrix equality by calculating  $T$ , defined as the sum of the absolute difference between each pair of elements:

$$T = \sum_{i=1}^c |\hat{\theta}_{i1} - \hat{\theta}_{i2}|$$

where  $\hat{\theta}_{ij}$  is the estimate of the  $i$ th element of the  $j$ th matrix and  $c$  is the number of elements in the matrix (sum of the number of diagonal elements plus the number of elements above the diagonal). The probability  $P_T$  that the two matrices come from the same statistical population is estimated by randomization (4999 runs):

$$P_T = \frac{1 + \text{number of randomizations in which } T_r \geq T_{\text{obs}}}{5000}$$

where  $T_r$  and  $T_{\text{obs}}$  are the comparison statistics for the randomized and observed data sets, respectively. For simplicity, only the verdict ('equal' or 'not equal') is given in the results section.

Second, the element by element method can be used for the comparison of individual pairs of (co)variances with the  $E$  statistic (Roff *et al.*, 1999):

$$E = |\hat{\theta}_{i1} - \hat{\theta}_{i2}|$$

The probability that the two matrix elements come from the same statistical population is obtained by randomization as in the previous method, using  $E$  instead of  $T$ . Because of the multiple estimations these values cannot be used individually but they do provide an indication of whether differences between the matrices arise because of a few strikingly variable elements or because of overall differences (the situation is analogous to the examination of individual cell values in a  $\chi^2$  test).

The third use of the element by element approach is to quantify the difference between matrices. The absolute average percentage difference statistic  $T\%$  is defined as:

$$T\% = \frac{T/c}{(\hat{\theta}_1 + \hat{\theta}_2)/2} 100$$

where  $\hat{\theta}_i$  is the average estimate of the (co)variances in matrix  $i$ . This statistic measures the absolute difference between the elements of two matrices as a percentage of the average size of the matrix elements.

### The method of percentage reduction in MSE

The following statistical procedure (Roff, 2000) can be used to partition the effects attributable to drift and selection, assuming that drift alone causes proportional changes in matrices and that selection causes any other type of change. This method is based on the calculation of the mean square errors for three regression models:

Model 1, equal matrices:

$$\text{MSE}_1 = \frac{1}{c} \sum (\hat{\theta}_{i1} - \hat{\theta}_{i2})^2$$

Model 2, proportional matrices:

$$\text{MSE}_2 = \frac{1}{2c} \sum (\hat{\theta}_{i1} - b_0 \hat{\theta}_{i2})^2 + \frac{1}{2c} \sum (\hat{\theta}_{i2} - b_0 \hat{\theta}_{i1})^2$$

Model 3, different matrices:

$$\text{MSE}_3 = \frac{1}{2c} \sum (\hat{\theta}_{i1} - [a + b\hat{\theta}_{i2}])^2 + \frac{1}{2c} \sum (\hat{\theta}_{i2} - [A + B\hat{\theta}_{i1}])^2$$

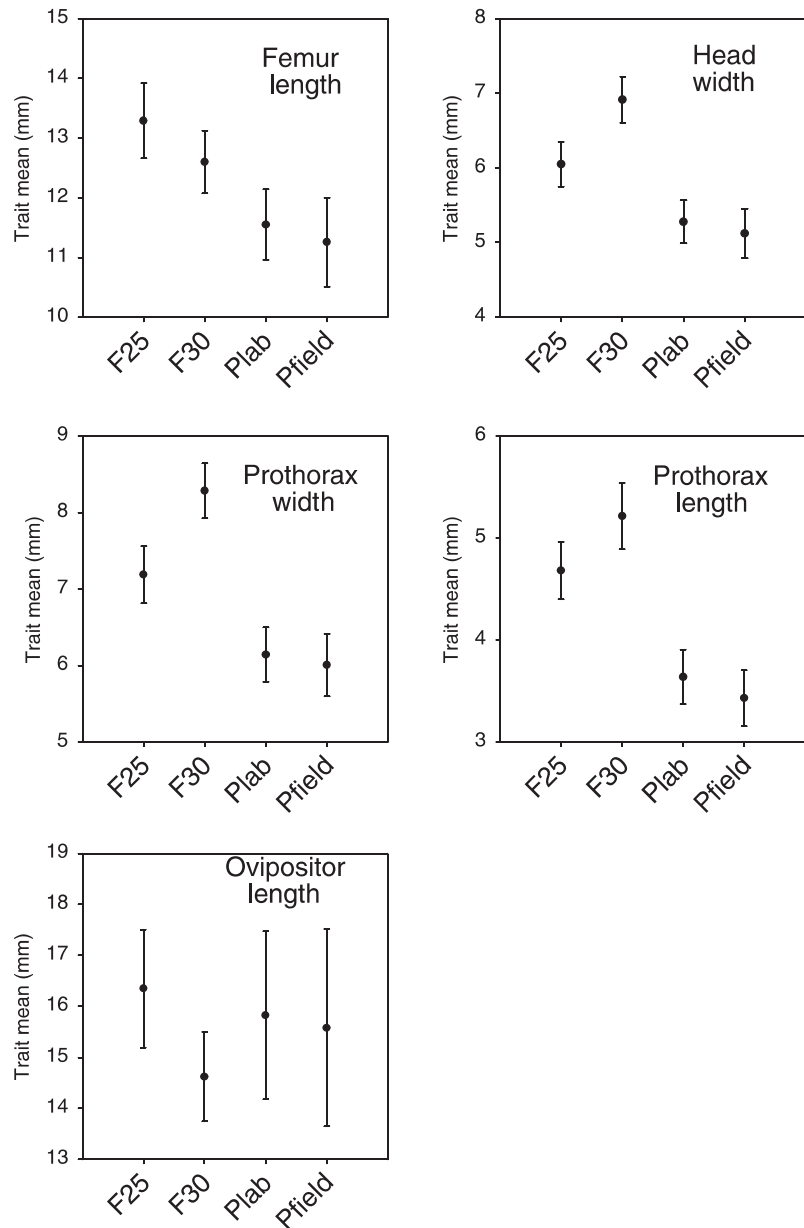
where  $b_0$  and  $B_0$  ( $B_0 = 1/b_0$ ) are the slopes of the reduced major axis regression forced through the origin, and  $a$ ,  $b$ ,  $A$  and  $B$  are parameters of the reduced major axis regression with the intercept included. Percentage reduction of the MSE from models 1 to 2 or 3 is then calculated as an estimator of the effect of drift alone (reduction from model 1 to 2) and of the effect of drift + selection (reduction from model 1 to 3).

## Results

### Overview of the data

In conformity with the published literature (Alexander, 1957a; Harrison & Arnold, 1982) the *G. pennsylvanicus* females are smaller than the *G. firmus* females (Fig. 1). An exception is ovipositor length for which both species are similar. More surprising is the extremely large differences in the variances of ovipositor length, the *G. pennsylvanicus* females being almost twice as variable (Fig. 1).

G matrix analyses performed in this paper are based on genetic (co)variances estimated from a one way ANOVA/ANCOVA (non-nested). To provide more accurate values we show heritabilities and genetic correlations estimated from a jack-knifed nested analysis of (co)variance (cages nested within full-sib families). Heritabilities (Table 1) are medium sized for morphological traits (mean values of 0.37 and 0.34 for F25 and F30, respectively; mean



**Fig. 1** Trait means (mm) plus or minus one standard deviation for the five traits used in the present analysis of *G. firmus* and *G. pennsylvanicus*.

values of 0.51 and 0.32 for Plab and Pfield, respectively). The genetic correlations between traits (Table 1) are positive and generally high (mean values of 0.78 and 0.49 for F25 and F30, respectively; mean values of 0.67 and 0.69 for Plab and Pfield, respectively). These data indicate that there is an abundance of genetic variance and covariance among the traits (**P** and **G** matrices are listed in Appendices 1 and 2).

### The Flury hierarchy

The results of the Flury hierarchy (and of the element by element method) for all six possible pairwise comparisons

of matrices are given in Table 2. The comparison of **P** matrices indicates that all but one comparison (Plab–Pfield) are best described by the unrelated structure model. This reveals that, given the degrees of freedom available, the Flury method ‘sees’ statistically large differences between most **P** matrices. The analysis of the **G** matrices yields a different pattern (Table 2). The verdicts indicate a general trend of conservation of matrix structure, as each comparison is described by either the CPC or equality models. The results also reveal that two comparisons (F30–Pfield and F25–F30) include pairs of matrices for which the equality and proportionality models are rejected.

**Table 1** Heritabilities (diagonal) and genetic correlations (off diagonal) followed by their standard error. These estimates are based on a jack-knifed nested ANOVA/ANCOVA (cages within full-sib families) and correspond to the five traits used in the present analysis.

	FEMUR	HEAD	PTHL	PTHW	OVIP
<i>Gryllus firmus</i> 25 °C					
FEMUR	0.33 (0.16)	0.82 (0.13)	0.72 (0.47)	0.89 (0.07)	0.59 (0.20)
HEAD		0.41 (0.13)	0.80 (0.15)	0.92 (0.05)	0.70 (0.14)
PTHL			0.23 (0.17)	0.85 (0.10)	0.68 (0.21)
PTHW				0.52 (0.15)	0.79 (0.12)
OVIP					0.38 (0.13)
<i>G. firmus</i> 30 °C					
FEMUR	0.30 (0.11)	0.51 (0.23)	0.63 (0.16)	0.62 (0.15)	0.42 (0.31)
HEAD		0.35 (0.10)	0.36 (0.21)	0.90 (0.05)	0.52 (0.23)
PTHL			0.35 (0.09)	0.35 (0.18)	0.14 (0.27)
PTHW				0.41 (0.09)	0.49 (0.22)
OVIP					0.31 (0.09)
<i>G. pennsylvanicus</i> lab					
FEMUR	0.62 (0.15)	0.80 (0.07)	0.67 (0.14)	0.84 (0.07)	0.62 (0.17)
HEAD		0.65 (0.12)	0.74 (0.11)	0.86 (0.06)	0.55 (0.17)
PTHL			0.39 (0.10)	0.66 (0.12)	0.40 (0.17)
PTHW				0.63 (0.14)	0.53 (0.20)
OVIP					0.24 (0.09)
<i>G. pennsylvanicus</i> field					
FEMUR	0.50 (0.14)	0.89 (0.06)	0.77 (0.13)	0.90 (0.07)	0.53 (0.12)
HEAD		0.29 (0.11)	0.72 (0.16)	0.91 (0.05)	0.52 (0.16)
PTHL			0.22 (0.12)	0.67 (0.19)	0.46 (0.25)
PTHW				0.38 (0.13)	0.49 (0.16)
OVIP					0.22 (0.07)

### The element by element approach

Hypothesis testing using the  $T$  statistic (Table 2) reveals rejection of equality for all pairwise comparisons of **P** matrices. It is interesting to note that the variation in  $T\%$  was not reflected in the verdicts, as the  $T$ -test had enough power to perceive statistical differences between all matrices. The comparison of **G** matrices produces a different pattern (Table 2). The hypothesis of equality is rejected in two cases (F30-Plab and F30-Pfield) whereas the verdict of the four other comparisons is 'equal'. Moreover, the  $T\%$  statistic suggests that two comparisons include relatively similar matrices (F25-Plab and Plab-Pfield), whereas the two comparisons yielding the verdict 'not equal' are associated with the two largest  $T\%$  values.

It can also be observed that the magnitude of  $T\%$  is similar for **P** and **G** comparisons (Table 2). A correlation analysis reveals a marginally significant positive slope between **P** and **G**-values of  $T\%$  ( $P = 0.04$ ,  $r = 0.83$ ). The  $T\%$  value for each comparison of **G** matrices was also regressed against the corresponding values of  $P_T$  (probabilities not shown). There is a highly significant negative regression ( $P = 0.005$ ,  $r = 0.94$ ) from which it was determined that a **G** matrix  $P_T$  of 0.05 corresponds to a  $T\%$  of 76.7%.

**Table 2** Results of the Flury hierarchy and element by element method for the pairwise comparisons of matrices from *G. firmus* reared in two environments (F25 and F30) and *G. pennsylvanicus* also reared in two environments (Plab and Pfield).  $T\%$  refers to the average absolute percentage difference statistic. A large  $T\%$  value corresponds to a large difference between two matrices. Bold characters indicate the two **G** matrix comparisons in which the Flury and element by element methods do not agree.

Comparison	Verdict		
	Flury hierarchy	Element by element	$T\%$
<b>P</b> matrices			
Across species			
F25-Plab	unrelated	not equal	37.0
F25-Pfield	unrelated	not equal	64.4
F30-Plab	unrelated	not equal	65.5
F30-Pfield	unrelated	not equal	94.7
Across environments*			
F25-F30	unrelated	not equal	36.5
Plab-Pfield	equality	not equal	38.1
<b>G</b> matrices			
Across species			
F25-Plab	equality	equal	27.6
F25-Pfield	equality	equal	45.8
<b>F30-Plab</b>	equality	not equal	78.4
F30-Pfield	CPC	not equal	96.0
Across environments*			
<b>F25-F30</b>	CPC	equal	58.5
Plab-Pfield	equality	equal	22.7

\* 'environment' is here interpreted to include rearing conditions and wing morph.

Investigation of the  $E$  statistic for all **G** matrix comparisons (results not shown) reveal that for each of the four pairs of equal matrices (according to the element by element method, Table 2), no individual element differs significantly between matrices (i.e.  $P_E$  is always greater than 0.05 with a majority of probabilities greater than 0.20). On the other hand, the two comparisons that rejected equality (F30-Plab and F30-Pfield) include at least one  $P_E$  value under 0.05. Concerning these two comparisons, it is interesting to note that in both cases the genetic variance of the ovipositor length (OVIP) has a  $P_E$  value under 0.05, and that several covariances involving ovipositor length also differ significantly in F30-Pfield. The genetic variance of this trait consistently yields one of the lowest  $P_E$  in each comparison (results not shown).

### The method of percent reduction in MSE

The calculation of differential reduction in MSE under the two models for **P** matrices (Table 3) reveals that the percentage difference between models 2 and 3 (maximum = 11.5%) is small compared with the difference between models 1 and 2 (range = 81.7–97.4%). These results indicate that including an additional parameter in

**Table 3** Percent reduction in mean square error (MSE) for comparisons of matrices from *G. firmus* reared in two environments (F25 and F30) and from *G. pennsylvanicus* also reared in two environments (Plab and Pfield). Results are given for reduction from model 1–2 (drift) and model 1–3 (drift + selection).

Comparison	Reduction of models (%)			
	P matrices		G matrices	
	2	3	2	3
Across species				
F25-Plab	81.7	93.2	85.4	91.7
F25-Pfield	93.0	97.9	91.2	95.7
F30-Plab	84.3	96.5	92.2	96.4
F30-Pfield	90.2	98.2	92.4	96.8
Across environments*				
F25-F30	92.2	99.0	89.4	91.1
Plab-Pfield	97.4	97.9	96.0	97.7

\*‘environment’ is here interpreted to include rearing conditions and wing morph.

model 3 (the intercept) does not produce a substantial reduction in MSE. Therefore, all six comparisons of **P** matrices are best represented by the proportionality model (model 2) although the percentage difference in MSE between models 2 and 3 cannot be neglected. The same pattern is observed for **G** matrices (Table 3). The percentage difference between models 2 and 3 (maximum = 6.3%) is once again small compared with the difference between models 1 and 2 (range = 85.4–96.0%). Proportionality thus seems to be the best fitting model.

### Comparative analysis of statistical methods

Because the results of the method of percentage reduction in MSE are constant for all types of comparisons, we will focus on the two other methods. The Flury and element by element methods (Table 2) agree that **P** matrices are statistically different. One exception is the Plab–Pfield comparison for which the equality model could not be rejected by the Flury method.

There also seems to be similarities between the Flury and element by element methods for the comparison of **G** matrices. These methods agree that the F25–Plab, F25–Pfield and Plab–Pfield comparisons contain similar matrices and that the F30–Pfield comparison includes the most divergent pair. However, the other two cases are problematic. For the F30–Plab comparison, the Flury method cannot reject the hypothesis of equality whereas the element by element method describes it as the second most divergent pair of **G** matrices. Conversely, the element by element method cannot reject equality for the F25–F30 comparison whereas the Flury method describes it as more divergent than four other comparisons. The reasons behind the discrepancy for these two

**Table 4** Principal component analysis of the three individual **G** matrices included in the two problematic comparisons. The eigenvalues and the component loadings corresponding to each of the five traits are given for the first two principal components.

	<i>G. pennsylvanicus</i> lab		<i>G. firmus</i> 30 °C		<i>G. firmus</i> 25 °C	
	PC 1	PC 2	PC 1	PC 2	PC 1	PC 2
Eigenvalue	0.99	0.17	0.36	0.11	0.73	0.11
FEMUR	0.36	0.70	0.40	0.62	0.36	0.78
HEAD	0.18	0.28	0.26	0.19	0.20	0.24
PTHL	0.10	0.21	0.15	0.41	0.16	0.07
PTHW	0.21	0.41	0.32	0.30	0.30	0.29
OVIP	0.89	–0.46	0.80	–0.57	0.85	–0.50

problematic comparisons should thus be examined for each method.

Starting with the Flury hierarchy, we computed and compared the principal components corresponding to each genetic matrix. This data exploration procedure provides a visual and simplified way of comparing matrices on a similar ground as the Flury hierarchy. The results were obtained by performing a principal component analysis on the genetic matrices directly. We expect this analysis to corroborate the results of the Flury hierarchy which indicate that F30 is closer to Plab than to F25. It can be observed from Table 4 that the structure of the first principal component is relatively similar in all matrices. However, the second component may differ more in F25–F30 than in F30–Plab, mostly because of the component loading of prothorax length. On the other hand, the eigenvalues seem to differ more in F30–Plab than in F25–F30. This ambiguous observation is hard to reconcile with the results of the Flury hierarchy. However, note that the Flury hierarchy does not compute and then compare individual principal component structures. Therefore, a lack of correspondence between the results from the Flury hierarchy and this data exploration procedure do not invalidate the Flury method.

Second, we investigated the results of the element by element method for these same two problematic comparisons. To do so, we used the individual element aspect of the method (the *E* statistic) to explore the differences one by one between each pair of (co)variances (Table 5). We expect that this analysis will corroborate the results from the *T*% statistic, that F30 is closer to F25 than to Plab. Observing the individual probabilities across the two comparisons reveals that 12 out of the 15 elements are more similar in the F25–F30 comparison than in the F30–Plab comparison. In addition to this general pattern, the ovipositor variance has a very low probability ( $P_E = 0.01$ , Table 5) in F30–Plab. Both observations support the results from the *T* statistic.

Despite these two problematic cases, we conclude that the results from the Flury and element by element

**Table 5** Results of the individual element tests ( $E$  statistic) for the two problematic **G** matrix comparisons. Probabilities corresponding to the null hypothesis of element equality ( $P_E$ ) are given for each trait (co)variance. A low probability indicates a large difference between the two matrices for that particular element.

	FEMUR	HEAD	PTHL	PTHW	OVIP
F30-Plab					
FEMUR	0.10	0.11	0.71	0.12	0.07
HEAD		0.48	0.38	0.53	0.17
PTHL			0.37	0.40	0.20
PTHW				0.33	0.18
OVIP					0.01
F25-F30					
FEMUR	0.45	0.38	0.89	0.22	0.22
HEAD		0.94	0.49	0.75	0.24
PTHL			0.43	0.25	0.07
PTHW				0.42	0.08
OVIP					0.08

methods yield some resemblance, on which the biological interpretation can be built. However, the results of the MSE method (Table 3) do not seem to relate to the pattern seen in Table 2. We will therefore deal with these results separately and concentrate on the Flury and element by element methods in the following paragraphs.

### Comparison of matrices across environments

The Flury and element by element analyses of **P** matrices (Table 2) clearly demonstrate that the *G. firmus* matrices (corresponding to two different environments) are different from each other. The **G** matrix analysis is not as obvious to interpret (Table 2) because the two methods are not in full agreement. The Flury hierarchy indicates differences, whereas the element by element method cannot reject the hypothesis of equality. However, the absolute average percentage difference statistic ( $T\% = 58.5\%$ ) suggests that the two matrices are not highly similar and approach the significance threshold of the test ( $T\%_{0.05} = 76.7\%$ ). Overall, there seems to be some variation in the **P** and **G** matrices corresponding to the two environments of *G. firmus*.

The situation is different for *G. pennsylvanicus* (Table 2). The Flury hierarchy cannot reject the hypothesis of equality of **P** matrices for the two environments. Conversely, the element by element method yields the verdict 'not equal'. However, the  $T\%$  statistic ( $T\% = 38.1$ ) is one of the lowest for **P** matrix comparisons and suggests that there are some similarities between the two matrices (Plab-Pfield). The **G** matrix analysis for these two treatments yields a unanimous verdict of matrix similarity. Overall, there is no evidence of large **P** and **G** matrix variation across the two environments of *G. pennsylvanicus*.

### Comparison of matrices across species

To analyse the variation between *G. firmus* and *G. pennsylvanicus*, we focus on F25-Plab because it compares individuals of the same wing morph (short wing) reared in similar laboratory conditions (25 °C/15 h light and 24 °C/17 h light, respectively). The Flury and element by element methods (Table 2) agree relatively well on the results from that comparison. Both methods are able to reject the equality model for **P** matrices, although the  $T\%$  statistic is relatively low ( $T\% = 37.0$ ). The **G** matrix analysis is unequivocal, as both methods suggest conservation of matrices.

It is interesting to note that the across species **G** matrix comparisons involving F30 do not lead to the same conclusions as those including F25 (Table 2). The Flury and element by element methods do not agree perfectly on this, but for both F30-Plab and F30-Pfield, at least one of the two methods indicates differences. Genetic matrices of *G. firmus* and *G. pennsylvanicus* thus seem to be similar if they correspond to similar environments, but can differ when rearing conditions and wing morph differ.

### Discussion

Analyses of **G** matrices have always been plagued by statistical problems arising from the complexity of multivariate data sets. Such studies would benefit greatly from a consensus on which method(s) is(are) the most reliable, thus allowing researchers to focus on the biological rather than statistical aspect of the problem. This paper attempts to comparatively evaluate three statistical approaches and to integrate information collected from all of them to get a general sense of **G** matrix variation within and between two species of crickets.

### Comparative analysis of statistical methods

As expected, the results from some of the statistical approaches share similarities but are not identical. The Flury and element by element methods are able to discern common patterns of matrix similarity but the approach of percentage reduction in MSE provides a very different view of the matrices. The important questions are therefore 'why do these methods give different results?' and 'do these differences reflect different aspects of matrix evolution?'

The Flury and element by element methods both have the ability to evaluate the distance between two matrices. The Flury hierarchy does it by assigning one of the seven possible models to the observed difference whereas the element by element method uses a numerical measure ( $T\%$ ), which is a more intuitive estimation of distance. However, the models of the Flury hierarchy are advantageous because they also

provide a way to investigate past evolutionary forces (e.g. proportionality/genetic drift), something that the element by element method cannot do. The approach of percentage reduction in MSE cannot evaluate distances but is designed to evaluate the type of differences between matrices.

The statistical approach of each method also differs. The Flury hierarchy compares the overall structure of the matrices whereas the other two methods look at the individual elements. This difference in approach is potentially responsible for the difference between the results of each method. For example the element by element method is probably more affected by a single very divergent element than is the Flury hierarchy. However, little is known about the statistical behaviour of these methods and simulation studies are needed.

The statistical approach also has implications for the biological interpretation of the data. The Flury hierarchy provides an overall view of the matrices which seems appropriate for studying a set of correlated traits. The element by element and MSE methods provide a less general view of the matrices, but compensate by being more transparent as the difference between matrices can be easily linked with particular elements (e.g. by visually inspecting a bivariate plot of two matrices). Information on individual elements is potentially useful because it provides a way to link **G** matrix evolution with the ecological function of important traits. Because each method provides a different biological perspective, there appears to be no single 'best' method.

Another potential problem for interpretation is the typically low power available in **G** matrix studies. No power analysis is currently available for these methods. It is possible that the observed differences between methods are caused by the large error associated with (co)variance components estimation. The only hint currently available to investigate that question is the comparison of parametric and nonparametric results of the Flury hierarchy (Phillips & Arnold, 1999). The parametric version (Phillips, 1998c), which computes the genetic (co)variances directly from the sample using family means, produces biased but statistically well behaved estimates. This version also appears to have more power than the nonparametric one (Phillips, personal communication). The results of the parametric analysis for **G** matrix comparisons (not shown) indicate much less shared structure across all matrices than does the nonparametric version. Our data does not allow us to know whether this result is because of the bias induced by the occasionally low number of individuals per family or to the problems associated with estimating (co)variance components, but it suggests that the results of the Flury hierarchy as presented in this paper may be affected by low power. This could also be the case for the two other methods and might be a cause of difference across methods.

Taken together, all the above uncertainties suggest that using one statistical method by itself might not be sufficient to obtain all necessary information on matrix variation. Until more is known about the properties of each method, we recommend the use of several different statistical approaches.

### Quantitative genetics issues

The proposition of using **P** as a surrogate for **G** was addressed using the average absolute percentage difference statistic. The range of *T*% values from Table 2 and the significant positive correlation between phenotypic and genetic *T*% values (see results) suggest that, in our analysis, these two types of matrices are characterized by approximately similar absolute differences. However, working only on phenotypic matrices would have been slightly misleading. The results thus loosely support the hypothesis that phenotypic covariation patterns reflect underlying genetic constraints (Cheverud, 1988). However, sample size might be an important confounding factor. Because phenotypic (co)variances are estimated more accurately than genetic parameters, the *T*% values associated with each type of matrix are, similarly, different in accuracy. Additional investigation is needed.

Sample sizes also cause problems in the interpretation of hypothesis testing results. In this study, **P** matrices have been found to generally differ significantly whereas **G** matrices tend to be much more similar. However, this conclusion is not necessarily biologically relevant because it could primarily reflect the degrees of freedom available for each type of analysis. This question could ideally be answered by linking the statistical difference between two matrices with the biological relevance of that difference. In the case of **G** matrix analysis, linking both aspects is not easy. The interpretation depends on the importance of the studied traits in the ecology of the organisms and on the time scale. If such analyses were to be applied in conservation biology programmes for example, only large differences in genetic architecture would be important whereas phylogenetic orientated studies should consider even small differences because these can have huge impacts over long periods of time.

The average absolute percentage difference (*T*%), a scale free index of similarity, is a suitable tool with which to assess the question of biological importance of observed differences, although it cannot replace an analysis of the role of the **G** matrix within its ecological context (e.g. Roff & Mousseau, 1999). Results from Table 2 suggest that several comparisons include matrices that differ by more than 50% and that the divergence can be as large 96% in one case. This represents a substantial variation which may have evolutionary implications. It was also determined that the **G** matrix significance threshold for the *T* statistic corresponds to a 76.7% average absolute difference, which could reasonably be assumed to represent real differences.

### G matrix variation within and across species

Since divergence from their ancestral population, *G. firmus* and *G. pennsylvanicus* can be presumed to have undergone some evolutionary changes. The phenotypic differences between the two species measured in a range of environments (Fig. 1) support this presumption. Phylogenetic studies have also demonstrated evolutionary divergence between the two species (Harrison, 1978; Huang *et al.*, 2000). Thus the evolutionary forces that acted on the two species could have modified their **G** matrix. However, changes in traits may or may not induce changes in the corresponding genetic architecture (Lande, 1979).

The comparative analysis of *G. firmus* and *G. pennsylvanicus* indicates that the differences in genetic architecture between these two species are relatively small. Note, however, that this conclusion only holds when rearing environments are similar. According to the theory (Lande, 1979), low intensities of selection on the measured traits could be responsible for this small amount of **G** matrix variation. The low levels of morphological divergence between *G. firmus* and *G. pennsylvanicus* (within a genus, cricket species tend to be difficult to distinguish on morphological grounds (Alexander, 1957b; Harrison, 1978)), and the fact that they can still interbreed suggest that this set of morphological traits as a whole have not been strongly selected during species divergence.

However, it appears that ovipositor length is an exception in that it may have been the target of stronger selection pressures. The ovipositor shows very high phenotypic and genetic variances, is the only trait with a similar average across the two species and yielded low probabilities for the *E* statistic in each comparison. These observations suggest that the evolution of the ovipositor has been important in the divergence of the two species. This is not surprising because the ovipositor length strongly influences the depth at which females lay their eggs in the soil, a reproductive behaviour which is known to be under selection (Masaki, 1979; Carrière *et al.*, 1997). To test the effect of ovipositor length on **G** matrix variation, we removed this trait from the analysis. Results for both the Flury and element by element methods (not shown) indicate that the omission of this trait does not affect substantially the comparison of matrices. Therefore, we conclude that the genetic basis of the ovipositor length has diverged more than that of the other traits but its effect is not strong enough to obscure general patterns of **G** matrix similarity.

Results from the method of percentage reduction in MSE suggest that proportionality is the principal source of observed variation between **G** matrices. This can be interpreted as meaning that random genetic drift has been the predominant force acting on **G** matrices whereas selection was weak enough to have its effect counteracted by mutation (Lande, 1979; Lofsvold, 1988). Both *G. firmus*

and *G. pennsylvanicus* usually live in temporally heterogeneous environments (Harrison, 1978) that probably produce frequent population bottlenecks. On these occasions, surviving individuals are the ones that are able to migrate and colonize a new habitat. This frequent reduction in numbers could cause random genetic drift to be strong enough to alter the genetic architecture of the cricket populations. On the other hand, the proportionality model also explains variation of matrices within each species. Obviously this variation cannot be interpreted as the drift between the two matrices because these were estimated from the same population. The observed variation across environments may be caused by environmental effects and is discussed in the following paragraphs. Without further investigation, it is difficult to assess the evolutionary importance of the observed proportionality in the comparisons between these species of crickets.

Another interesting result is that, according to the Flury and element by element methods, the matrix corresponding to *G. firmus* reared at 30 °C (F30) seems to differ from all other matrices, including F25. It is important to note that the F30 matrix comes from long wing individuals reared at a relatively high temperature. Both aspects of this environment are different from the other treatments. Because we cannot separate the effects of wing morph and rearing conditions with the present experimental design, we conclude that the environment (including both aspects) can have an effect on the **G** matrix.

The results also reveal that comparisons of **G** matrices across environments show larger differences in *G. firmus* than in *G. pennsylvanicus*. This observation (along with the preceding paragraph) stresses the importance of understanding the expression of the genetic architecture with respect to the environment. Different species may have different environmental patterns of genetic expression, and comparing the matrices of two species reared in one chosen environment might be misleading and only reveal part of the total difference. To avoid this confounding factor and to learn more about the expression of genetic architecture, comparisons of **G** matrices across species should ideally include several environments.

Similarly, it is important to know if laboratory estimates of genetic architecture are reliable predictors of genetic architecture in nature. Our results reveal that rearing condition (laboratory vs. natural environment) do not produce substantial variation at the **G** matrix level in *G. pennsylvanicus*. This result is complementary to two other studies that looked at the differences in genetic parameter expression between laboratory and natural environments in *G. pennsylvanicus* (Simons & Roff, 1994, 1996). These studies demonstrated that, for the same five morphological traits, heritabilities are generally higher in the lab than in the field, whereas genetic correlations are stable. The present results suggest that despite variation in genetic parameters, **G** matrices measured in a homogeneous laboratory environment can provide a good representation of natural genetic architecture.

The evolution of genetic architecture is at present poorly understood. Empirical studies of **G** matrix variation are greatly needed to shed light on theoretical models (Turelli, 1988). Our study demonstrated, using several statistical approaches, that the **G** matrix can be relatively conserved at the species level when rearing environments are similar. However, the expression of the genetic architecture can vary between environments, making comparisons across species difficult. Additional studies are required to test the generality of these results.

## Acknowledgments

This work was supported by a NSERC operating grant to D. A. Roff and by a FCAR graduate studentship to M. Bégin. We wish to thank A. M. Simons for making the data available to us. D. Réale and two anonymous reviewers provided very helpful comments on a previous draft of the manuscript.

## References

- Alexander, R.D. 1957a. The song relationships of four species of ground crickets. *Ohio J. Science* **153**: 153–163.
- Alexander, R.D. 1957b. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: acheta). *Ann. Entomol. Soc. Am.* **50**: 584–602.
- Arnold, S.J. & 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.
- Camara, M.D. & Pigliucci, M. 1999. Mutational contributions to variance-covariance matrices: an experimental approach using induced mutations in *Arabidopsis thaliana*. *Evolution* **53**: 1692–1703.
- Carrière, Y., Masaki, S. & Roff, D.A. 1997. The coadaptation of female morphology and offspring size: a comparative analysis in crickets. *Oecologia* **110**: 197–204.
- Cheverud, J.M. 1988. A comparison of genetic and phenotypic correlations. *Evolution* **42**: 958–968.
- Deng, H.-W., Haynatzka, V., Spitze, K. & Haynatzki, G. 1999. The determination of genetic covariances and prediction of evolutionary trajectories based on a genetic correlation matrix. *Evolution* **53**: 1592–1599.
- Endler, J.A. 1986. *Natural Selection in the Wild*. Princeton University Press, Princeton, New Jersey.
- Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*. Longman Harlow, England.
- Flury, B. 1988. *Common Principal Components and Related Multivariate Models*. Wiley, New York.
- Fulton, B.B. 1952. Speciation in the field cricket. *Evolution* **6**: 283–295.
- Harrison, R.G. 1978. Ecological parameters and speciation in field crickets. In: *Ecological Genetics: The Interface*. (Brussart, P., ed.), pp. 145–158. Springer-Verlag, Heidelberg.
- Harrison, R.G. 1985. Barriers to gene exchange between closely related cricket species. II. Life cycle variation and temporal isolation. *Evolution* **39**: 244–259.
- Harrison, R.G. & Arnold, J. 1982. A narrow hybrid zone between closely related cricket species. *Evolution* **36**: 535–552.
- Huang, Y., Orti, G., Sutherlin, M., Duhachek, A. & Zera, A. 2000. Phylogenetic relationships of North American field crickets inferred from mitochondrial DNA data. *Mol. Phylogenetics Evolution* **17**: 48–57.
- Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**: 314–334.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. *Evolution* **33**: 402–416.
- Lande, R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94**: 203–215.
- Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.
- Lofsvold, D. 1986. Quantitative genetics of morphological differentiation in *Peromyscus* I. Tests of the homogeneity of genetic covariance structure among species and subspecies. *Evolution* **40**: 559–573.
- Lofsvold, D. 1988. Quantitative genetics of morphological differentiation in *Peromyscus*. II. Analysis of selection and drift. *Evolution* **42**: 54–67.
- Masaki, S. 1979. Climatic adaptation and species status in the lawn ground cricket. III. Ovipositor length. *Oecologia* **43**: 207–219.
- Merilä, J. & Björklund, M. 1999. Population divergence and morphometric integration in the greenfinch (*Carduelis chloris*) – evolution against the trajectory of least resistance? *J. Evol. Biol.* **12**: 103–112.
- Mousseau, T.A. & Roff, D.A. 1987. Natural selection and the heritability of fitness components. *Heredity* **59**: 181–198.
- Phillips, P.C. 1998a. H2boot: bootstrap estimates and tests of quantitative genetic data. University of Texas at Arlington. Software available at [www.uta.edu/biology/phillips/software](http://www.uta.edu/biology/phillips/software)
- Phillips, P.C. 1998b. CPCrand: randomization test of the CPC hierarchy. University of Texas at Arlington. Software available at [www.uta.edu/biology/phillips/software](http://www.uta.edu/biology/phillips/software)
- Phillips, P.C. 1998c. CPC: Common principal components analysis. University of Texas at Arlington. Software available at [www.uta.edu/biology/phillips/software](http://www.uta.edu/biology/phillips/software)
- Phillips, P.C. & Arnold, S.J. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution* **53**: 1506–1515.
- 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.
- Roff, D.A. 1997. *Evolutionary Quantitative Genetics*. Chapman & Hall, New York.
- Roff, D.A. 1998. Effects of inbreeding on morphological and life history traits of the sand cricket, *Gryllus firmus*. *Heredity* **81**: 28–37.
- Roff, D.A. 2000. The evolution of the **G** matrix. *Heredity* **84**: 135–142.
- Roff, D.A. & Mousseau, T.A. 1999. Does natural selection alter genetic architecture? An evaluation of quantitative genetic variation among populations of *Allonemobius socius* and *A. fasciatus*. *J. Evol. Biol.* **12**: 361–369.
- Roff, D.A., Mousseau, T.A. & Howard, D.J. 1999. Variation in genetic architecture of calling song among populations of

- Allonemobius socius*. *A. fasciatus* and a hybrid population: drift or selection? *Evolution* **53**: 216–224.
- Shaw, F.H., Shaw, R.G., Wilkinson, G.S. & Turelli, M. 1995. Changes in genetic variances and covariances: G Whiz! *Evolution* **49**: 1260–1267.
- Simons, A.M. & Roff, D.A. 1994. The effect of environmental variability on the heritabilities of traits of a field cricket. *Evolution* **48**: 1637–1649.
- Simons, A.M. & Roff, D.A. 1996. The effect of a variable environment on the genetic correlation structure in a field cricket. *Evolution* **50**: 267–275.
- Steppan, S.J. 1997a. Phylogenetic analysis of phenotypic covariance structure. I. Contrasting results from matrix correlation and common principal component analyses. *Evolution* **51**: 571–586.
- Steppan, S.J. 1997b. Phylogenetic analysis of phenotypic covariance structure. II. Reconstructing matrix evolution. *Evolution* **51**: 587–594.
- Turelli, M. 1988. Phenotypic evolution, constant covariance, and the maintenance of additive variance. *Evolution* **42**: 1342–1347.
- Vickery, V.R. & Kevan, D.K. McE. 1983. *A Monograph Orthopteroid Insects Can Adjacent Regions*. Vols 1 and 2. Lyman Entomological Museum Research Library Memoirs **13**: 1–1467.

Received 18 September 2000; accepted 9 November 2000

## Appendix

	FEMUR	HEAD	PTHL	PTHW	OVIP
<i>G. firmus</i> 25 °C					
FEMUR	0.39 (0.04)	0.15 (0.02)	0.12 (0.02)	0.20 (0.02)	0.46 (0.05)
HEAD		0.09 (0.01)	0.06 (0.01)	0.10 (0.01)	0.23 (0.04)
PTHL			0.08 (0.01)	0.08 (0.01)	0.18 (0.02)
PTHW				0.14 (0.01)	0.28 (0.03)
OVIP					1.33 (0.12)
<i>G. firmus</i> 30 °C					
FEMUR	0.27 (0.02)	0.13 (0.01)	0.11 (0.01)	0.14 (0.01)	0.28 (0.03)
HEAD		0.10 (0.01)	0.06 (0.01)	0.09 (0.01)	0.17 (0.02)
PTHL			0.11 (0.01)	0.07 (0.01)	0.13 (0.02)
PTHW				0.13 (0.01)	0.19 (0.02)
OVIP					0.77 (0.06)
<i>G. pennsylvannicus</i> lab					
FEMUR	0.35 (0.03)	0.13 (0.01)	0.10 (0.01)	0.16 (0.02)	0.53 (0.06)
HEAD		0.08 (0.01)	0.05 (0.01)	0.08 (0.01)	0.24 (0.03)
PTHL			0.07 (0.01)	0.06 (0.01)	0.17 (0.02)
PTHW				0.12 (0.01)	0.28 (0.03)
OVIP					2.72 (0.35)
<i>G. pennsylvannicus</i> field					
FEMUR	0.55 (0.05)	0.21 (0.02)	0.14 (0.01)	0.26 (0.03)	0.91 (0.08)
HEAD		0.11 (0.01)	0.06 (0.01)	0.12 (0.01)	0.41 (0.03)
PTHL			0.08 (0.01)	0.08 (0.01)	0.25 (0.03)
PTHW				0.17 (0.01)	0.47 (0.04)
OVIP					3.75 (0.28)

**Appendix 1** Phenotypic (co)variance matrices (SE) for *G. firmus* and *G. pennsylvannicus*. These estimates correspond to the five traits used in the present analysis.

**Appendix 2** Genetic (co)variance matrices (SE) for *G. firmus* and *G. pennsylvanicus*.

These estimates correspond to the five traits used in the present analysis.

	FEMUR	HEAD	PTHL	PTHW	OVIP
<i>G. firmus</i> 25 °C					
FEMUR	0.16 (0.06)	0.07 (0.03)	0.04 (0.03)	0.10 (0.04)	0.18 (0.09)
HEAD		0.04 (0.01)	0.03 (0.01)	0.05 (0.02)	0.11 (0.04)
PTHL			0.03 (0.01)	0.04 (0.02)	0.09 (0.03)
PTHW				0.08 (0.02)	0.17 (0.06)
OVIP					0.55 (0.18)
<i>G. firmus</i> 30 °C					
FEMUR	0.11 (0.03)	0.04 (0.02)	0.05 (0.02)	0.06 (0.02)	0.08 (0.04)
HEAD		0.04 (0.01)	0.02 (0.01)	0.05 (0.01)	0.06 (0.02)
PTHL			0.04 (0.01)	0.02 (0.01)	0.02 (0.02)
PTHW				0.06 (0.01)	0.07 (0.03)
OVIP					0.27 (0.07)
<i>G. pennsylvanicus</i> lab					
FEMUR	0.22 (0.06)	0.09 (0.03)	0.06 (0.02)	0.12 (0.04)	0.26 (0.10)
HEAD		0.05 (0.01)	0.03 (0.01)	0.06 (0.02)	0.13 (0.04)
PTHL			0.03 (0.01)	0.03 (0.01)	0.07 (0.03)
PTHW				0.08 (0.02)	0.15 (0.06)
OVIP					0.81 (0.24)
<i>G. pennsylvanicus</i> field					
FEMUR	0.32 (0.08)	0.11 (0.03)	0.07 (0.02)	0.15 (0.04)	0.36 (0.08)
HEAD		0.04 (0.01)	0.03 (0.01)	0.06 (0.01)	0.13 (0.03)
PTHL			0.02 (0.01)	0.03 (0.01)	0.09 (0.03)
PTHW				0.08 (0.02)	0.17 (0.04)
OVIP					1.06 (0.27)