Genet. Res., Camb. (1998), **71**, *pp.* 193–212. *With* 1 *figure*. Printed in the United Kingdom © 1998 Cambridge University Press

Evolution of animal genitalia: patterns of phenotypic and genotypic variation and condition dependence of genital and non-genital morphology in water strider (Heteroptera: Gerridae: Insecta)

GÖRAN ARNQVIST[1]* AND RANDY THORNHILL[2]
[1] *Department of Animal Ecology, University of Umeå, S-901 87 Umeå, Sweden*
[2] *Department of Biology, University of New Mexico, Albuquerque, NM 87131-1091, USA*

(*Received 29 September 1997 and in revised form 27 January 1998*)

Summary

Rapid and divergent evolution of male genitalia represents one of the most general evolutionary patterns in animals with internal fertilization, but the causes of this evolutionary trend are poorly understood. Several hypotheses have been proposed to account for genitalic evolution, most prominent of which are the lock-and-key, sexual selection and pleiotropy hypotheses. However, insights into the evolutionary mechanisms of genitalic evolution are hindered by a lack of relevant in-depth studies of genital morphology. We used a biparental progenies breeding design to study the effects of food stress during ontogeny on phenotypic expression of a suite of genital and non-genital morphological traits, both linear traits and multivariate shape indices, in a natural population of the water strider *Gerris incognitus*. In general, genitalic traits were as variable as non-genital traits, both phenotypically and genotypically. Average narrow-sense heritability of genital traits was 0·47 (SE = 0·05). Further, while food stress during development had a large impact on adult morphology, and expression of genitalic traits exhibited significant levels of condition dependence, different genotypes did not significantly differ in their ability to cope with food stress. Genitalic conformation was also both phenotypically and genetically correlated with general morphological traits. These patterns are in disagreement with certain predictions generated by the long-standing lock-and-key hypothesis, but are in general agreement with several other hypotheses of genital evolution. We failed to find any additive genetic components in fluctuating asymmetry of any bilaterally symmetrical traits and the effects on fluctuating asymmetry of food stress during development were very low and insignificant. Some methodological implications of our study are discussed, such as the bias introduced by the non-negativity constraint in restricted maximum likelihood estimation of variance components.

1. Introduction

In animals with internal fertilization, diversification of genitalia represents one of the most general and striking forms of evolutionary trends (Eberhard, 1985). Animal genitalia evolve rapidly and divergently, and genitalic morphology typically differs even between closely related species. Though one of the most general evolutionary trends, it remains one of the most poorly understood (Scudder, 1971; Eberhard, 1985, 1990, 1993, 1996; Shapiro & Porter, 1989; Andersson, 1994; Alexander *et al*., 1997; Arnqvist, 1997*b*). A number of different hypotheses have been proposed to account for the evolution of animal genitalia, but empirical data are very scarce. In particular, in-depth studies of quantitative genetics and patterns of selection on genitalic traits are largely lacking, rendering discussions of genitalic evolution speculative (Eberhard, 1993, 1996; Andersson, 1994; Arnqvist, 1997*b*).

Arnqvist (1997*b*) suggested that relevant data from single species may be used to illuminate the patterns and processes of genitalic evolution, in much the same

* Corresponding author. Fax: +46-90-167665. e-mail: goran.arnqvist@animecol.umu.se.

way that data from single species studies are key in other adaptational domains. Studies of genitalia should address patterns of morphological variation and inheritance of genitalic traits on the one hand, and patterns of phenotypic selection on, and fitness consequences of, genitalic variation on the other. The current study represents the former line of investigation.

The three main contending hypotheses for genitalic evolution – the lock-and-key, the sexual selection and the pleiotropy hypotheses – all make a number of predictions of the pattern of genitalic variation expected within species (Eberhard, 1985, 1990, 1993; Shapiro & Porter, 1989; Alexander *et al*., 1997), many of which are relative and non-rigorous (Arnqvist, 1997*b*). The lock-and-key hypothesis states that genitalia evolve via selection for pre-insemination mechanical reproductive isolation, so that male genitalia evolve to be specific and unique (the key) in order to fit appropriately in female genitalia (the lock). Under this hypothesis, development of male genitalia is predicted to be canalized and under stabilizing selection, leading to relatively low degrees of phenotypic and genotypic variation in genital morphology (cf. Pomiankowski & Møller, 1995) and relatively little condition dependence in expression of genitalic traits (Arnqvist, 1997*b*).

Intromittent male genitalia may also evolve via post-mating sexual selection (Eberhard, 1985, 1993, 1996; Arnqvist, 1997*b*). Non-random fertilization success among males, based on genital morphology, may be brought about by differences in stimulatory/titillating ability (Thornhill, 1983; Eberhard, 1985, 1990, 1993, 1996), by differences in ability to coerce/control fertilization events (Lloyd, 1979; Arnqvist & Rowe, 1995; Rice, 1996; Alexander *et al*., 1997) or by differences in ability to displace/dislocate sperm of competing males (Smith, 1984; Waage, 1984; Birkhead & Hunter, 1990). Traits under sexual selection tend to exhibit relatively high levels of phenotypic and genotypic variation (Pomiankowski & Møller, 1995), and genitalia should be no exception (Arnqvist, 1997*b*). Phenotypic expression of genitalia would also be expected to evolve to be condition dependent (Andersson, 1994; Johnstone, 1995; Rowe & Houle, 1996), typically leading to significant phenotypic correlations between genital and general morphological traits.

The pleiotropy hypothesis holds that genitalic evolution is an indirect result of evolution of genetically correlated characters, via pleiotropic effects of genes that code for both genitalic and general morphology (Mayr, 1963; Eberhard, 1985, 1990). Genitalic variation *per se* is essentially neutral, and pleiotropic effects on genitalia are not directly selected against and can thus accumulate. A key assumption of this hypothesis, which is very difficult to refute with empirical data, is that genital and non-genital morphological traits are genetically correlated (Arnqvist, 1997*b*).

The current contribution represents the first extensive study to explicitly address questions about the degree and nature of genetic and environmental control of intraspecific variability in genital morphology, and to relate this to other components of morphology. As a model system, we use a natural population of water striders (see below). First, we quantify the extent of phenotypic variation in genitalic traits relative to other traits. Secondly, we characterize the patterns of inheritance of genitalic versus general morphological traits. Thirdly, we experimentally assess the degree of condition dependence in phenotypic expression of genital and general morphological traits, by inducing varying levels of food stress during ontogeny.

Recent research has indicated that differences in morphology between the left and right sides of bilaterally symmetrical traits may be revealing of the phenotypic and genotypic condition of individuals (Møller & Pomiankowski, 1993; Swaddle *et al*., 1994; Watson & Thornhill, 1994). In particular, fluctuating asymmetry (FA) results from the inability of individuals to undergo identical development on both sides of the body, and is thus believed to reflect the ability of individuals to cope with environmental stress (Mather, 1953; Van Valen, 1962; Palmer & Strobeck, 1986). The degree of FA of traits has thus been put forth as a potentially integrative measure of individual phenotypic and genotypic quality (Møller & Pomiankowski, 1993; Watson & Thornhill, 1994). In the current study, we use measures of FA as an additional tool for studying patterns of condition dependence and phenotypic quality across traits, by experimentally assessing how food stress affects the degree of FA in a suite of genital and general morphological traits.

The study of small-scale complex morphological variation, such as that of genitalic variation in arthropods, has long been hindered by methodological problems. Recent development in methods of capturing data (digital techniques) have reduced the problem of measuring small-scale traits accurately. More importantly, new statistical methods designed to describe two- or three-dimensional variation in shape among specimens have opened up novel possibilities to work with complex morphologies (Rohlf & Marcus, 1993; Marcus *et al*., 1996). These multivariate methods potentially provide much more synthetic and integrative measures of morphological variation than traditional linear measures (Liu *et al*., 1996). We use both standard linear traits and multivariate measures of shape variation, and are thus able to compare directly the relative merits and problems of the two approaches in this context.

2. Materials and methods

(i) *The organisms*

Water striders (Insecta; Heteroptera: Gerridae) form an ecologically rather homogeneous family of true bugs. They inhabit water surfaces of various aquatic habitats both as juveniles and as adults, and are predators/scavengers feeding mainly on arthropods trapped at the water surface (Andersen, 1982; Spence & Andersen, 1994; Rowe *et al*., 1994; Arnqvist, 1997*a*). In terms of morphological divergence within the group, water striders are typical insects in the sense that relatively rapid evolutionary radiation of male genitalia is a particularly striking and consistent pattern. Hence, characteristics of male genitalia is very important both in grouping higher-order taxa and in distinguishing among closely related congeneric species (Andersen, 1982, 1993; Arnqvist, 1992, 1997*a*). The genitalia of water striders consists of a proximal cylindrical segment (i.e. 1st genital segment) containing a boat-shaped structure representing the pygophore and the proctiger. These structures hold the intromittent phallic organ, which is inflated/extended and inserted into the female genital tract during copulation. Apart from membranous tissue, the phallus consists of a proximal sclerotized phallotheca, and an apical capsule, the vesica, which carries an armature of sclerites (for illustrations see Andersen, 1982, 1993, Fig. 1).

This study was performed on *Gerris incognitus* (Drake and Hottes), a member of the primarily Holarctic and relatively species-rich genus *Gerris* (42 species). In this genus, characters of male genitalia are of great taxonomic and phylogenetic importance. In particular, the set of sclerites in the apical part of the intromittent phallus (the vesical armature) have evolved rapidly and divergently, and their morphology is a critical species character within the genus (Andersen, 1993). Though little is known of the functional morphology of these structures in *Gerris*, the vesical sclerites could play an important part during insertion and/or positioning of the male phallus in the female genital tract during copulation (Andersen, 1982; Heming-van Battum & Heming, 1989).

(ii) *Experimental design*

Parental adult water striders (*G. incognitus*) were captured from a monomorphic apterous natural population at the Bosque del Apache Wildlife Refuge, New Mexico, USA, 6 March 1994. Males and females were isolated in unisexual tanks in the laboratory, and fed frozen crickets *ad libitum*. Since some females laid fertile eggs at the time of capture, the following procedure was applied to enable determination of paternity of offspring. Females were kept separated from males and the fertilization rate of their eggs was monitored continuously. Average fertilization rates among females were 5% after 12 d, 2% after 16 d and 0 after 19 d of isolation. Sperm longevity in congeneric species is known to range between 12 and 20 d (Kaitala, 1987; Arnqvist, 1988, 1989). The laboratory rearing experiment described below was initiated after 21 d of isolation of the sexes in the laboratory, when it can safely be assumed that females did not carry viable sperm. Additional confidence in ascribing paternity to assigned sires is provided by the fact that last male sperm precedence has been established in this group of insects (Arnqvist, 1988; Rubenstein, 1989).

For the breeding programme, we used a biparental progenies design (BIPs), which has the advantage of using a relatively large sample of parental individuals, and thus is powerful when one wishes to characterize the genetic structure of a natural population rather than a set of inbred lines (Kearsey, 1965). Males and females were introduced in pairs ($n = 61$) into containers provided with a Styrox floater to serve as resting and oviposition substrate. The parental individuals were fed field crickets *ad libitum* for 10 days, after which they were frozen for subsequent morphometric analysis.

Twelve randomly selected (upon hatching) full-sibling offspring from each family were reared individually in circular 9 cm diameter containers, each provided with a Styrox floater. All offspring were fed frozen *Daphnia* zooplankton *ad libitum* during the first 2 d. To assess the effects of food stress during ontogeny on morphology (the degree of condition dependence of phenotypic traits), offspring were thereafter randomly assigned to each of two food treatments within families ($n = 6$ offspring per family and food treatment). Offspring were fed frozen *Drosophila* fruitflies, larval *Gryllus* field crickets and *Tenebrio* beetle larvae, according to the scheme in Table 1. Offspring in food treatment I (aimed at providing a sufficient food supply) were never completely starved (deprived of all food), and received a larger and more varied food supply compared with offspring in food treatment II (aimed at creating food stress). Water was changed and the rearing containers cleaned once every 4 d. Offspring were individually preserved by freezing when sclerotized, 24–48 h after moulting into the adult stage, for subsequent morphometric analysis.

(iii) *Morphometric analysis*

A landmark-based approach was used to acquire morphometric measurements. Two-dimensional digital morphometric landmark maps were attained by placing a digitizing tablet (Summasketch III) under a

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

Table 1. *Summary of the feeding regimes for the two offspring food treatments*

Offspring larval stage	Treatment I (food per individual)	Treatment II (food per individual)
I–II	1 *Drosophila*[a] every 1 d	1 *Drosophila* every 2 d
III	1 *Drosophila* every 1 d	1 *Drosophila* every 2 d
	1 *Gryllus* (1 week old)[b] every 2 d	1 *Gryllus* (1 week old) every 3 d
	No starvation	Starved for 24 h in late III instar
IV–V	1 *Drosophila* every 1 d	1 *Drosophila* every 2 d
	1 *Gryllus* (2 weeks old)[c] every 2 d	1 *Gryllus* (1 week old) every 3 d
	1 *Tenebrio*[d] every 3 d	—
	No starvation	Starved for 24 h in IV instar

Mean weight per food item ($n = 12$): [a] 0·86 mg (SD = 0·42), [b] 3·90 mg (SD = 0·34), [c] 18·38 mg (SD = 6·48), [d] 15·36 mg (SD = 4·03).

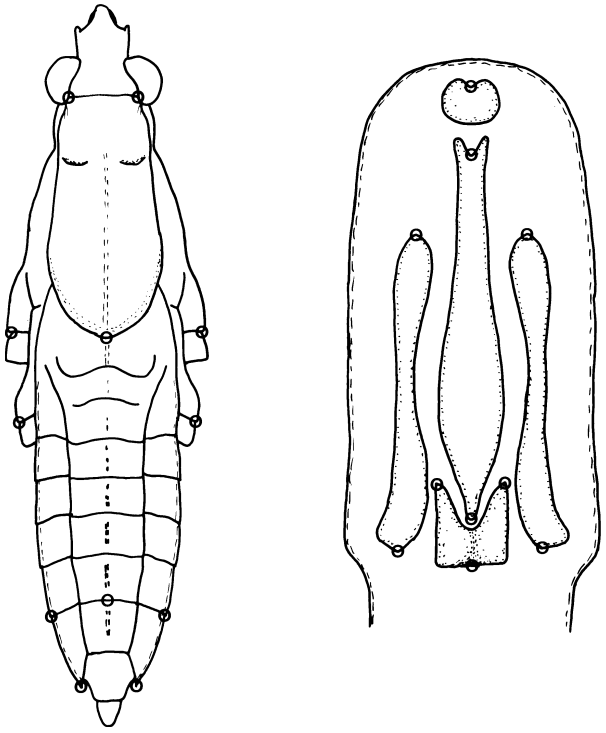

Fig. 1. Locations of landmarks on the body (dorsal view) and the intromittent genitalic capsule (the vesica) with its armature of internal sclerites (dorsal view). Landmarks used are indicated with circles.

side-mounted camera lucida attached to a dissecting microscope (Wild M5). For each individual water strider, several separate landmark maps were digitized (two for females and four for males): one of the body in dorsal view (12 landmarks, Fig. 1), one of the appendages (28 landmarks), one of the proximal parts of the genitalia (8 landmarks) and one of the armature of sclerites in the intromittent genitalic capsule (the vesica) in dorsal view (10 landmarks, Fig. 1). From each of these landmark maps, we used a number of distances between pairs of landmarks, selected *a priori* because of the very large number of potential distances that could be extracted from such maps, to acquire linear measures of size of a number of traits.

We also used multivariate geometric shape analysis to acquire a more synthetic quantification of morphological variation in shape of genitalia and body among individuals (Rohlf & Marcus, 1993). Landmark maps were subjected to thin-plate spline relative warp analysis, a flexible, powerful and interpretable multivariate technique for analysis of morphometric shape variation (Bookstein, 1991; Rohlf, 1993; Marcus *et al*., 1996). In short, the method involves fitting a function to the landmark coordinates of each specimen, where shape variation within the population is manifested as variance in the parameters of the fitted function. Relative warps represent principal component vectors in the multivariate shape space, and each relative warp can be thought of as representing a unique multivariate shape dimension, orthogonal to all other such dimensions. The dimensionality of each relative warp can be visualized as a displacement of landmarks relative to the average configuration of landmarks. For a more detailed description and discussion of thin-plate spline relative warp analysis see Bookstein (1991), Rohlf (1993) and Marcus *et al*. (1996).

Landmark maps of both the dorsal view of the body and the sclerites of the genital capsule (Fig. 1) were analysed with thin-plate spline relative warp analysis. We used the following procedure. Each landmark map from each specimen was divided into two separate images prior to analysis by duplicating all landmarks along the mid-line, one image representing the left and one the right side of the specimen. The left and right sides of all specimens were treated as separate landmark maps, and both sides of all individuals (parental as well as offspring) were pooled into one

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

single data set (i.e. the total number of images included in the analysis equalled the total number of individuals times 2). This analytic strategy has the benefit of producing a number of unique orthogonal shape dimensions, the scores of which are directly comparable not only between the left and right sides of a given individual, but also across all specimens. First, all images were optimally translated, rotated and scaled to the same position and size, using the generalized least-squares superimposition in the morphometric computer program GRF-ND (Slice, 1994). This procedure removes isometric size variation, and generates an average consensus configuration of landmarks as well as an integrative measure of the relative size of each landmark map (the centroid size). Secondly, pure shape variation among the scaled and aligned images was analysed with thin-plate spline relative warp analysis, using the consensus configuration as reference in the analysis. The computer software TPSRW (Rohlf, 1993) was used for estimating relative warps, as proposed by Bookstein (1991) (i.e. using a value of $\alpha = 1$). The six first relative warps were retained and used in subsequent analysis. These were concluded to collectively account for virtually all appreciable shape variation among specimens. However, the relative warps summarize only non-uniform (localizable) shape variation. To quantify uniform shape variation (non-localizable), the two uniform shape components of Bookstein (1996) were estimated separately. Using the analogy of a grid, the relative warps measure shape change which causes certain lines to be displaced or non-linearized relative to other lines, whereas the pair of uniform shape components parameterizes shape variation that leaves parallel lines parallel throughout the grid after the shape change. Thus, three groups of variables were extracted from the shape analysis to describe multivariate variation in shape among specimens: (a) centroid size, representing an integrative measure of overall size, (b) relative warp scores and (c) uniform shape component scores.

(iv) *Fluctuating asymmetry*

We estimated FA for both linear distances and multivariate shape scores (cf. Auffray *et al.*, 1996; Smith *et al.*, 1997). The average size/shape of the right and left side $(R+L/2)$ was used to characterize a particular individual for all morphometric variables that were potentially bilaterally symmetrical (e.g. length of appendages, shape scores of left and right sides of the body or genitalia). For these variables, we also calculated the absolute degree of asymmetry, by subtracting the size/shape of the left side from that of the right side $(R-L)$. FA, as opposed to directional or antisymmetry, is characterized by a normal distribution centred around zero within populations (Palmer & Strobeck, 1986; Swaddle *et al.*, 1994). All our measures of morphological asymmetry were thus tested for normality and a mean of zero (data on parental individuals; see Appendix A), and variables failing to comply with these basic assumptions of FA were excluded from further analysis.

For traits exhibiting FA, the absolute value of FA scores $(|R-L|)$ was used to characterize the degree of asymmetry of individuals (i.e. the directionality of the bilateral asymmetry was ignored). These variables will in theory be approximately half-normally distributed. To stabilize the variances of the models used for subsequent statistical analysis, all measures of FA were transformed prior to analysis as (cf. Sokal & Rohlf, 1981; Swaddle *et al.*, 1994): $x' = \log(1+10^{c}x)$, where c is a factor (>1) specific for each variable, chosen to minimize the sum of the absolute values of skewness and kurtosis of the distribution of the variable (Berry, 1987). These transformations also tended to optimize linearity in normal probability plots of the variables describing FA (Swaddle *et al.*, 1994).

(v) *Repeatability*

Taking repeated measures and estimating the repeatability of morphological variation is key in any thorough quantitative genetics morphometric study, for two related reasons (Arnold, 1994). First, the repeatability provides an approximate upper limit on the heritability of any given character (Falconer & Mackay, 1996). Secondly, the repeatability provides a reliability assessment of the quantification of morphological variation, since it represents a measure of the relative magnitudes of measurement error and true between-individual phenotypic variance (Lessels & Boag, 1987; Arnqvist & Mårtensson, 1998). This is particularly important for measures of FA, since measurement error alone can cause apparent FA (Palmer & Strobeck, 1986; Swaddle *et al.*, 1994; Merilä & Björklund, 1995). To enable estimation of the repeatability of our morphological variables, three repeated measures of all measurements were taken from each of 40 separate individuals (20 individuals of each sex), collected simultaneously with the parental individuals from the field (see above). We followed the procedure of Lessels & Boag (1987), by which data are analysed in one-way analyses of variance. The repeatability, or the intraclass correlation coefficient (r_I), of a given variable equals the ratio between the between-individuals variance component and the sum of between- and within-individuals components of variance. For bilaterally symmetrical traits, we based our repeatability estimates of size/shape on the average between sides $(R+L/2)$, and of asymmetry in

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

size/shape on the signed difference between sides (R − L), for each repeated measure respectively, since these were the metrics used in other parts of the study (cf. Palmer & Strobeck, 1986; Swaddle *et al*., 1994; Merilä & Björklund, 1995).

(vi) *Full-sib analyses*

Due to differential mortality and uneven sex ratios within families, the full-sibling data for estimation of quantitative genetic parameters were unbalanced. Thus, variance component estimations were made using restricted maximum likelihood (REML) fitting. This iterative method is already widely used for estimating the parameters of the genetic variance–covariance matrix (**G**) in the domain of animal breeding (Henderson, 1985; Searle *et al*., 1992), and its merits have more recently been appreciated in evolutionary genetics (Shaw, 1987; Shaw *et al*., 1995; Falconer and Mackay, 1996). In contrast to other methods, REML offers methods for estimating relatively unbiased variance components under a range of circumstances even when the data are unbalanced (which is typically the case with biological data on sets of relatives), as well as appropriate estimates of fixed effects. Variance components were estimated by fitting the mixed model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ represents the fixed effects model matrix (sex, food treatment and their interaction), $\mathbf{Z}$ represents the random effects model matrix (family and family × food treatment interaction) and $\boldsymbol{\varepsilon}$ represents a random error vector. In our case, the variance component associated with $\boldsymbol{\varepsilon}$ will include all components of unattributable environmental variation, but also higher-order interactions. REML estimation uses a regular least-squares based approach to fit fixed effects only, and then estimates relevant variance components of random factors by maximizing the likelihood for the remaining residuals (for discussions of the merits and problems of REML estimation in quantitative genetics see Shaw, 1987; Searle *et al*., 1992). We restricted the above analyses to include only families with four or more surviving offspring, limiting the number of families to $n = 47$.

The degree of condition dependence in expression of phenotypic traits was assessed directly of the *F*-value of the fixed effect of the food treatment. Since the degrees of freedom are identical across traits, this *F*-statistic provides an unbiased measure of the magnitude (effect size) of difference in morphology under food-stressed and non-food-stressed conditions that is directly comparable across traits.

Estimates of broad-sense heritabilities were derived from the analysis of variance components, as $V_G/V_P = 2\sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2 + \sigma_e^2)$, where σ^2 represent estimated variance components due to family (σ_g^2), family × food treatment interaction (σ_{ge}^2) and residual error (σ_e^2), respectively (Falconer & Mackay, 1996). The family × food treatment interaction term provides a direct estimate of the genotype × environment interaction (V_{GE}). In our case, this component (σ_{ge}^2) can be said to test the heritability of condition dependence, i.e. the extent to which different genotypes respond differently to environmental stress during ontogeny (Falconer & Mackay, 1996).

Estimations and significance tests of variance components assume constant variances and normality, irrespective of which method of fitting is used. While it has been suggested that parameter estimates produced by REML are relatively robust to deviations from the normality assumption (Harville, 1977; Smith, 1980; Searle *et al*., 1992), this is not necessarily the case in tests of significance (Shaw, 1987). We examined all fitted mixed models for compliance with the assumptions of statistical inference by visual inspection of plots of residuals versus fitted values (McCullagh & Nelder, 1983; Searle *et al*., 1992). Unsatisfactory residual variance distribution was indicated in one case only, for the distance between tips of abdominal spines, in which we were able to stabilize the variance by logarithmic transformations of the original data. Any cells identified as having extreme residuals were removed (three cases in total), and the parameters re-estimated with reduced data.

(vii) *Parent–offspring analyses*

In general, estimates of genetic parameters based on parent–offspring resemblance are more accurate and reliable than those derived from full-sib analyses (Falconer, 1973; Falconer & Mackay, 1996). Further, while our full-sib analyses allow estimation of fixed effects and genotype–environment interactions, estimates of heritability may be biased (Shaw, 1987) and the analysis does not allow separation of additive genetic variance and other components of V_G (e.g. dominance effects). Thus, we used conventional midparent–midoffspring regression to estimate narrow-sense heritabilities for the traits included in the analysis.

Since offspring were subjected to different food regimes, and since the sexes typically differ in size, we fitted a standard fixed effects linear model (including sex, food treatment and their interaction) to all offspring values. By analogy with the REML method of estimating variance components, these models were used to generate residual offspring values to be related to parental values. The logic behind this procedure is to examine the influence of parental phenotypic values on deviations in offspring values from their expectation based on the food regime and sex of each offspring, assuming that the heritability is equal in

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

both laboratory environments. This assumption (that the response to food stress is equal across genotypes) is directly tested by the genotype × environment interaction in the full-sib analysis (see above).

For each family and morphological trait, we calculated average offspring residual value and mid-parental value. Prior to the regression analyses, all variables were standardized to a mean of zero and unit variance. Standardizing the variables is necessary in order to attain offspring and parental distributions that are comparable, i.e. are of equal scales. While this exercise does not affect the precision and significance testing of narrow-sense heritability estimates, it may introduce bias in their magnitude. This will, however, only be the case if the phenotypic and additive genotypic variances are not equal in the parental and offspring generations (Coyne & Beecham, 1987; Lande, 1987; Falconer & Mackay, 1996). Phenotypic variances may be higher in natural populations compared with those reared in the laboratory (Coyne & Beecham, 1987). Giving both generations equal phenotypic variance (standardizing the phenotypic distributions) may thus increase offspring variance relative to parental variance, which would cause a general inflation of the magnitude of the estimates of narrow-sense heritabilities. Thus, we directly compared the phenotypic variances in both generations across traits. Since the repeatabilities offer upper limits on narrow-sense heritabilities, we were also able to assess potential systematic biases by comparing these parameters. In addition, the genotype × environment interactions in the full-sib analysis to a certain extent reflects how robust genetic variances are to environmental variations, which could potentially cause differences between parental (field) and offspring (laboratory) generations in additive genetic variances (Lande, 1987).

Narrow-sense heritabilities were estimated in regular linear regressions of mean offspring on mid-parental (sire for genitalic traits) values, where narrow-sense heritability was derived directly from the regression coefficient (β), as $h^2 = \beta$ ($h^2 = 2\beta$ for genitalic traits) (Falconer & Mackay, 1996). Again, these analyses were restricted to include only families with four or more surviving offspring. Homogeneity of variance was assessed by visual inspection of residual plots.

We also estimated phenotypic and genetic correlations between general and genital morphology, based on sire–offspring data. Due to the very large number of potential correlations between traits, we restricted our analyses by choosing a subset of traits *a priori*. General traits were chosen on the basis of their functional significance: five size-related traits (body and legs) known to be under selection in congeneric species (Spence & Andersen, 1994; Arnqvist, 1997*a*) were included. Genital traits were chosen on the basis of their repeatabilities; only the seven traits exhibiting a repeatability $> 0{\cdot}8$ were included. Genetic correlations were estimated as:

$$r_a = \frac{COV_{x1z2} + COV_{x2z1}}{\sqrt[2]{(COV_{x1z1}\, COV_{x2z2})}},$$

where COV_{x1z2} denotes the covariance of trait 1 in sires and trait 2 in offspring, COV_{x2z1} the covariance of trait 2 in sires and trait 1 in offspring, COV_{x1z1} the covariance of trait 1 in sires and trait 1 in offspring and COV_{x2z2} the covariance of trait 2 in sires and trait 2 in offspring (Pirchner, 1983; Becker, 1984; Falconer & Mackay, 1996). Standard errors of the genetic correlations were approximated as:

$$SE(r_a) = \frac{1 - r_a^2}{\sqrt{2}} \qquad \frac{SE(h_1^2)\, SE(h_2^2)}{h_1^2\, h_2^2},$$

where h_1^2 denotes the heritability of trait 1 and h_2^2 the heritability of trait 2 (Falconer & Mackay, 1996). Genetic correlations were tested individually with *t*-tests, and a conservative overall assessment of H_0 (all genetic correlations are equal to zero) was made with a *t*-test of the mean of the genetic correlations.

(viii) *Statistical error rates*

Since this study is concerned with multiple and complex aspects of morphology, we report a large number of statistical estimates, primarily for quantitative genetic parameters. Each of these estimates is accompanied by a measure of dispersion and a test of significance. When multiple statistical significance tests are reported, a monotonic increase in group-wide statistical type I error rate will automatically occur (Rice, 1989). Compensation for such an increase in type I error rate is often made in articles focusing on single estimates, e.g. by the sequential Bonferroni method (Holm, 1979). However, as such compensation is made, a loss of statistical power and an increased type II error rate is inevitable. This is one of the most serious general problems of conventional inferential statistics: the inability simultaneously to control type I and type II statistical errors (Cohen, 1988; Arnqvist & Wooster, 1995). In the current paper, we face this dilemma by group-wide comparisons. Our main goal was to compare patterns of inheritance and condition dependence of genitalic versus general morphological traits. We base the interpretation of our results, and found our conclusions, on group-wide comparisons of overall patterns of the magnitudes of parameter estimates. Thus, we report non-adjusted significance levels in our tables, but in no case are our conclusions based on significances of these single estimates.

The statistical analyses reported in this paper were performed with SAS (1992) (REML estimation of

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

variance components in mixed model analyses of variance) and SYSTAT (1992) (other procedures).

3. Results

Differences in the degree of phenotypic variation exhibited by general and genital morphological traits were assessed by comparing coefficients of variation for all linear scalar traits in parental individuals (Lande, 1977). Average coefficients of variation in males were 3·94% (range 2·7–7·3) for general morphological traits and 5·91% for genital traits (range 3·2–16·5) (Appendix C). Thus, the degree of phenotypic variation was of similar magnitude for both types of traits ($t = 1{\cdot}10$, d.f. = 13, $P > 0{\cdot}25$).

Data on offspring growth and survival confirmed that the experimental food treatment had the desired effect. Food-stressed offspring had a lower survivorship (mean = 25·9%, SD = 20·3) than offspring provided with a more adequate food supply (mean = 42·9%, SD = 22·7) (paired t-test on arcsine transformed survival rates; $t = 5{\cdot}48$, d.f. = 57, $P < 0{\cdot}001$). Offspring experiencing food stress during ontogeny also had a longer development time from hatching to adulthood (mean development time: 36·05 days, SD = 2·90, versus 32·37 days, SD = 4·12) (only offspring where exact times of both hatching and moulting into the adult stage were recorded were included in analysis; $t = 7{\cdot}34$, d.f. = 234, $P < 0{\cdot}001$).

Food stress during development had a cascade of effects on adult morphology (Tables 2, 3). The largest effect occurred on measures of adult body size (range of F-values: 206·3–270·3) and on length of appendages (range of F-values: 45·4–62·8), whereas more variable effects were recorded for measures of body shape (range of F-values: 1·0–85·6). Measures of genitalic size showed somewhat milder effect sizes (range of F-values: 0·7–33·0). Shape of genitalia did not appear to be affected by food stress (range of F-values: 0·0–5·2). Contrary to our expectations based on theory of

Table 2. *The degree of condition dependence, measured as the magnitude of effect of food stress, of a number of general (non-genitalic) morphological traits*

Trait/trait group	F-value (d.f. = 1,46)	P value
Linear measures		
Body length	246·39	< 0·001
Thorax width	206·32	< 0·001
Length of abdominal spines	62·02	< 0·001
Distance between tips of abdominal spines	9·34	0·004
Elevation angle of abdominal spines	17·54	< 0·001
Length of 1st antennal segment	62·77	< 0·001
Forefemur length	62·52	< 0·001
Midleg length	45·43	< 0·001
Hindleg length	58·56	< 0·001
FA in abdominal spine length	0·08	> 0·5
FA in length of 1st antennal segment	0·80	0·375
FA in forefemur length	0·03	> 0·5
FA in midleg length	2·91	0·095
FA in hindleg length	1·16	0·286
Shape analysis of body		
Centroid size	270·30	< 0·001
Score relative warp number 1	85·61	< 0·001
Score relative warp number 2	47·91	< 0·001
Score relative warp number 3	4·42	0·041
Score relative warp number 4	38·94	< 0·001
Score relative warp number 5	2·15	0·149
Score relative warp number 6	6·46	0·014
Score uniform shape component number 1	44·26	< 0·001
Score uniform shape component number 2	0·96	0·333
FA in score relative warp number 1	3·01	0·089
FA in score relative warp number 2	1·46	0·233
FA in score relative warp number 3	0·03	0·857
FA in score relative warp number 4	1·94	0·170
FA in score relative warp number 6	1·31	0·259

F-values represent the effect of food treatment in multivariate mixed model analyses of variance (see text).

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

Table 3. *The degree of condition dependence, measured as the magnitude of effect of food stress, of a number of male genitalic morphological traits*

Trait/trait group	*F*-value (d.f. = 1,46)	*P* value
Linear measures		
Length of 1st genitalic segment	17·45	< 0·001
Proctiger length	6·60	0·013
Length of phallotheca	11·29	0·001
Length of lateral sclerites	7·74	0·008
Distance between lateral sclerites	0·67	0·416
Length of ventral sclerite	1·13	0·298
Length of dorsal sclerite	32·97	< 0·001
FA in length of lateral sclerites	0·16	0·687
Shape analysis of genital capsule		
Centroid size	13·17	< 0·001
Score relative warp number 1	1·83	0·182
Score relative warp number 2	5·24	0·027
Score relative warp number 3	0·00	> 0·5
Score relative warp number 4	0·23	> 0·5
Score relative warp number 5	2·23	0·142
Score relative warp number 6	0·40	> 0·5
Score uniform shape component number 1	3·54	0·066
Score uniform shape component number 2	0·01	> 0·5
FA in centroid size	1·93	0·171
FA in score relative warp number 1	1·88	0·177
FA in score relative warp number 3	5·02	0·030
FA in score relative warp number 4	1·45	0·235
FA in score uniform shape component number 1	0·05	0·827

F-values represent the effect of food treatment in multivariate mixed model analyses of variance (see text).

condition dependence and FA, food stress did not cause any measurable effects on the degree of FA in any traits, either in size-related traits or for measures of shape. Effect sizes for effects of food stress on FA were generally very low and insignificant (range of *F*-values: 0·0–5·0). To assess the validity of this interpretation of the pattern, we performed a three-way analysis of variance of log-transformed *F*-values, using three different dichotomous characteristics of each trait as factors (genital/non-genital, FA/non-FA and shape/size). In accordance with our interpretation, this analysis indicated that genitalic traits showed somewhat lower effects of food stress compared with non-genitalic traits ($F_{1,42} = 8{\cdot}09$, $P = 0{\cdot}007$) and that measures of FA showed much lower degrees of condition dependence compared with size and shape traits *per se* ($F_{1,42} = 25{\cdot}75$, $P < 0{\cdot}001$), whereas measures of size and shape were affected by food stress to a similar degree ($F_{1,42} = 1{\cdot}09$, $P > 0{\cdot}25$).

In theory, the lack of effects of food stress on FA could in part result from non-random mortality, with respect to FA, in the food-stressed group. This possibility was assessed by comparing the variances in FA in the two offspring treatment groups, across all 16 morphological measures of FA. Variance in measures of FA did not differ in general between the two groups of offspring (Wilcoxon signed rank $Z = 0{\cdot}310$, $P = 0{\cdot}756$; paired *t*-test of log-transformed variances, $t = 1{\cdot}211$, d.f. = 15, $P = 0{\cdot}245$). Offspring provided with an adequate food supply exhibited higher variance in nine cases, and the opposite was true for the remaining seven cases.

We found no evidence of a correlation between average trait size/magnitude and the absolute value of FA in any of the traits exhibiting FA, either in parental or offspring males (Table 7). The average correlation coefficient across all traits was −0·012 (SE = 0·028, range −0·175 to 0·155) in parental males and 0·002 (SE = 0·023, range −0·138 to 0·161) in offspring males. Correlation coefficients did not differ in parental and offspring males (paired *t*-test; $t = 0{\cdot}411$, d.f. = 15, $P > 0{\cdot}5$), and were not correlated with each other ($r = 0{\cdot}102$, $n = 16$, $P > 0{\cdot}5$).

Broad-sense and narrow-sense heritability estimates were positively correlated across traits ($r = 0{\cdot}443$, $n = 50$, $P < 0{\cdot}001$). Contrary to our expectations, however, narrow-sense heritability estimates were overall slightly higher than broad-sense heritability estimates (paired *t*-test, $n = 50$, $t = 4{\cdot}36$, $P < 0{\cdot}01$), but both were considerably lower than the estimated

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

Table 4. *General (non-genitalic) morphology*

Trait/trait group	σ^2_g (SE)	σ^2_{ge} (SE)	σ^2_e (SE)[a]	V_G/V_P	h^2 (SE)
Linear measures					
Body length	0·072 (0·048)	0[b]	0·928 (0·081)	0·14	0·34 (0·14)*
Thorax width	0·090 (0·066)	0·029 (0·076)	0·881 (0·085)	0·18	0·37 (0·14)**
Length of abdominal spines	0·082 (0·048)†	0	0·918 (0·079)	0·16†	0·21 (0·15)
Distance between tips of abdominal spines	0·123 (0·053)*	0	0·877 (0·075)	0·25*	0·14 (0·15)
Elevation angle of abdominal spines	0·034 (0·053)	0·054 (0·075)	0·911 (0·087)	0·07	0
Length of 1st antennal segment	0·355 (0·102)***	0·015 (0·053)	0·630 (0·062)	0·71***	0·63 (0·12)***
Forefemur length	0·239 (0·080)**	0·018 (0·067)	0·743 (0·074)	0·48**	0·63 (0·12)***
Midleg length	0·205 (0·070)**	0	0·795 (0·069)	0·41**	0·62 (0·12)***
Hindleg length	0·281 (0·086)***	0·001 (0·055)	0·718 (0·070)	0·56***	0·49 (0·13)***
FA in abdominal spine length	0·002 (0·033)	0	0·998 (0·086)	0·00	0·15 (0·15)
FA in length of 1st antennal segment	0	0	1·000 (0·082)	0	0
FA in forefemur length	0·002 (0·047)	0·013 (0·069)	0·985 (0·092)	0	0·09 (0·15)
FA in midleg length	0·019 (0·039)	0	0·981 (0·086)	0·04	0·15 (0·14)
FA in hindleg length	0·086 (0·048)†	0	0·914 (0·079)	0·17†	0
Shape analysis of body					
Centroid size	0·056 (0·046)	0	0·944 (0·082)	0·11	0·33 (0·14)*
Score relative warp number 1	0·188 (0·079)*	0·040 (0·075)	0·772 (0·075)	0·38*	0·51 (0·13)***
Score relative warp number 2	0·234 (0·078)**	0	0·765 (0·066)	0·47**	0·34 (0·14)*
Score relative warp number 3	0·125 (0·072)†	0·090 (0·071)	0·785 (0·073)	0·25†	0·48 (0·13)***
Score relative warp number 4	0·006 (0·046)	0·011 (0·068)	0·983 (0·091)	0·01	0·16 (0·15)
Score relative warp number 5	0·105 (0·053)*	0	0·895 (0·078)	0·21*	0·50 (0·13)***
Score relative warp number 6	0·126 (0·069)†	0·051 (0·076)	0·823 (0·079)	0·25†	0·24 (0·14)
Score uniform shape component number 1	0·118 (0·052)*	0	0·882 (0·076)	0·24*	0·30 (0·14)*
Score uniform shape component number 2	0·136 (0·059)*	0	0·864 (0·074)	0·27*	0·42 (0·14)**
FA in score relative warp number 1	0·021 (0·032)	0	0·979 (0·083)	0·04	0·01 (0·15)
FA in score relative warp number 2	0·016 (0·053)	0·074 (0·073)	0·910 (0·083)	0·03	0
FA in score relative warp number 3	0·019 (0·044)	0·043 (0·064)	0·938 (0·085)	0·04	0·03 (0·15)
FA in score relative warp number 4	0	0	1·000 (0·080)	0	0
FA in score relative warp number 6	0	0	1·000 (0·079)	0	0

Estimates of variance components and their standard errors derived from full-sib analyses, corresponding to genotypic, genotype × environment and environmental sources of phenotypic variance. All variance components below have been scaled to a sum of unity for each trait. Given are also estimates of broad-sense heritabilities (V_G/V_P) based on full-sib resemblance, and estimates of narrow-sense heritabilities h^2 (V_A/V_P) based on parent–offspring resemblance.
†$P < 0{\cdot}1$, *$P < 0{\cdot}05$, **$P < 0{\cdot}01$, ***$P < 0{\cdot}001$.
[a] No significance levels are given for the residual variance component.
[b] Negative estimates of variance components and narrow-sense heritabilities are denoted with '0'.

repeatabilities, across all traits. The average difference between the estimated heritability and repeatability was 0·55 (SD = 0·19) for broad-sense estimates (paired *t*-test, $n = 50$, $t = 20{\cdot}37$, $P < 0{\cdot}001$) and 0·40 (SD = 0·23) for narrow-sense estimates (paired *t*-test, $n = 50$, $t = 12{\cdot}27$, $P < 0{\cdot}001$). We assessed the possibility that our narrow-sense heritability estimates were systematically inflated as a result of consistently higher phenotypic variances in the parental generation (see Section 2) in two ways. First, we performed F_{max}-tests of equality of variances among the three groups involved (parents, high food offspring and low food offspring) separately for the two sexes across all traits (excluding measures of FA) (Sokal & Rohlf, 1981). The mean maximum variance ratio (F_{max}) was 1·51 ($n = 52$). The number of cases where the variance ratio exceeded the upper 5% point of the F_{max} distribution ($F_{max[crit.]} = 1{\cdot}85$) did not differ from the expected number under the null hypothesis of equal variances across all traits (Fisher's exact test, $P = 0{\cdot}319$). Secondly, to test for directionality, we performed a contingency table test of the frequencies of occurrence of highest variance in the three groups. Across all traits, the three groups did not differ in frequency of occurrence of highest variance among groups ($\chi^2 = 1{\cdot}653$, d.f. = 2, $P > 0{\cdot}25$). Thus, the phenotypic variances in parental and offspring generations did not differ in general across traits (cf. Coyne & Beecham, 1987), and none of the groups had consistently higher phenotypic variance than any other.

Overall, positive single estimates of heritability were frequently observed for both size-related traits and multivariate shape scores, in both genital and non-genital traits (Tables 4, 5). However, this was not the case for measures of FA. This interpretation of the pattern was again confirmed in three-way analyses of variance of heritability of traits. For both broad- and

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

Table 5. *Genital morphology*

Trait/trait group	σ^2_g (SE)	σ^2_{ge} (SE)	σ^2_e (SE)[a]	V_G/V_P	h^2 (SE)
Linear measures					
Length of 1st genital segment	0·055 (0·129)	0·143 (0·169)	0·802 (0·121)	0·11	0·51 (0·29)
Proctiger length	0·101 (0·110)	0·169 (0·123)	0·730 (0·098)	0·20	0·76 (0·28)**
Length of phallotheca	0[b]	0·116 (0·098)	0·884 (0·124)	0	0·56 (0·29)
Length of lateral sclerites	0·234 (0·128)†	0·017 (0·113)	0·749 (0·105)	0·47†	0·73 (0·28)*
Distance between lateral sclerites	0	0·036 (0·072)	0·964 (0·124)	0	0·26 (0·30)
Length of ventral sclerite	0·184 (0·087)*	0	0·816 (0·100)	0·37*	0·28 (0·30)
Length of dorsal sclerite	0·164 (0·089)†	0	0·836 (0·104)	0·33†	0·90 (0·27)**
FA in length of lateral sclerites	0·075 (0·071)	0	0·0925 (0·114)	0·15	0
Shape analysis of genital capsule					
Centroid size	0·179 (0·127)	0·065 (0·147)	0·756 (0·109)	0·36	0·41 (0·29)
Score relative warp number 1	0	0·035 (0·077)	0·965 (0·127)	0	0·40 (0·29)
Score relative warp number 2	0	0	1·000 (0·109)	0	0·25 (0·29)
Score relative warp number 3	0·028 (0·051)	0	0·972 (0·115)	0·06	0·03 (0·30)
Score relative warp number 4	0·037 (0·082)	0·031 (0·111)	0·932 (0·126)	0·07	0
Score relative warp number 5	0·074 (0·070)	0	0·926 (0·114)	0·15	0·84 (0·27)**
Score relative warp number 6	0	0	1·000 (0·109)	0	0·68 (0·28)*
Score uniform shape component number 1	0	0·064 (0·081)	0·936 (0·124)	0	0·53 (0·29)
Score uniform shape component number 2	0	0·006 (0·073)	0·994 (0·129)	0	0·44 (0·29)
FA in centroid size	0·070 (0·103)	0·019 (0·142)	0·911 (0·139)	0·14	0
FA in score relative warp number 1	0·093 (0·073)	0	0·907 (0·112)	0·19	0
FA in score relative warp number 3	0·044 (0·066)	0	0·956 (0·118)	0·09	0·09 (0·30)
FA in score relative warp number 4	0·010 (0·059)	0	0·990 (0·122)	0·02	0·72 (0·28)*
FA in score uniform shape component number 1	0·070 (0·069)	0	0·930 (0·114)	0·14	0·27 (0·30)

Estimates of variance components and their standard errors derived from full-sib analyses, corresponding to genotypic, genotype × envrionment and environmental sources of phenotypic variance. All variance components below have been scaled to a sum of unity for each trait. Given are also estimates of broad-sense heritabilities (V_G/V_P) based on full-sib resemblance, and estimates of narrow-sense heritabilities h^2 (V_A/V_P) based on parent–offspring resemblance.
†$P < 0{\cdot}1$, *$P < 0{\cdot}05$, **$P < 0{\cdot}01$, ***$P < 0{\cdot}001$.
[a] No significance levels are given for the residual variance component.
[b] Negative estimates of variance components and narrow-sense heritabilities are denoted with '0'.

narrow-sense heritabilities (log-transformed) the single significant factor was whether the trait measured FA or not ($F_{1,42} = 6{\cdot}10$, $P = 0{\cdot}018$ for broad-sense heritability, and $F_{1,42} = 24{\cdot}83$, $P < 0{\cdot}001$ for narrow-sense heritability). Thus, excluding measures of FA, average broad-sense heritability for non-genital traits ($n = 18$) was 0·29 (SE = 0·04) and average narrow-sense heritability was 0·37 (SE = 0·05). For genital traits ($n = 16$), average broad-sense heritability was 0·13 (SE = 0·03) and average narrow-sense heritability was 0·47 (SE = 0·05). The average broad-sense heritability of measures of size ($n = 16$) was 0·28 (SE = 0·04) and average narrow-sense heritability was 0·46 (SE = 0·05). For measures of shape ($n = 18$), average broad-sense heritability was 0·16 (SE = 0·04) and average narrow-sense heritability was 0·38 (SE = 0·05). Thus, in conclusion, measures of genital and general morphology both exhibited overall weak to intermediate heritability, irrespective of whether they measured components of size or shape. In contrast, the magnitude of heritability of measures of FA differed from that of other traits, and did not show non-zero heritabilities.

We found no evidence of genotype × environment interactions for any traits. Most estimates of V_{GE} were very low (scaled $\sigma^2_{ge} < 0{\cdot}1$ for 47 of 50 traits; Tables 4, 5). Thus, while food stress had a large impact on adult morphology (see above), different genotypes did not seem to differ in their morphological response to this stress, i.e. they did not differ appreciably in their ability to cope with a stressful environment.

There was a positive relationship between the magnitude of the estimates of phenotypic and genetic correlations between general morphological and genital traits ($r = 0{\cdot}59$, $n = 35$, $P < 0{\cdot}001$) (Table 6). Overall, phenotypic correlations were slightly larger (mean = 0·25, SE = 0·03) than genetic correlations (mean = 0·19, SE = 0·05), but not significantly so (paired t-test, $t = 1{\cdot}52$, $n = 35$, $P = 0{\cdot}138$). There was evidence of non-zero overall genetic correlation between genital and general morphology, even if the fact that non-positive genetic correlations are potentially informative is disregarded (t-test of H_0: all genetic correlations are equal to zero; $t = 3{\cdot}84$, $P < 0{\cdot}001$). Of particular interest is the apparent lack of phenotypic correlation, but a consistently negative

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

Table 6. *Estimates of phenotypic correlations* (*top part*) *and genetic correlations* (*bottom part*) *between non-genital morphological traits* (*horizontally*) *and genital traits* (*vertically*)

	Thorax width	Body length	Forefemur length	Midleg length	Hindleg length
Length of 1st genital segment	0·55 (0·12)***	0·60 (0·12)***	0·55 (0·12)***	0·60 (0·12)***	0·52 (0·13)***
Proctiger length	0·53 (0·13)***	0·49 (0·13)***	0·39 (0·14)**	0·40 (0·14)**	0·37 (0·14)*
Length of phallotheca	0·12 (0·15)	0·15 (0·15)	0·15 (0·15)	0·03 (0·15)	0·01 (0·15)
Length of lateral sclerites	0·26 (0·14)	0·20 (0·15)	0·30 (0·14)*	−0·25 (0·14)	0·34 (0·14)*
Length of ventral sclerite	0·31 (0·14)*	0·25 (0·14)	0·27 (0·14)	0·21 (0·15)	0·17 (0·15)
Length of dorsal sclerite	0·33 (0·14)*	0·30 (0·14)*	0·23 (0·14)	0·25 (0·14)	0·26 (0·14)
Score relative warp number 5 (shape)	0·02 (0·15)	0·06 (0·15)	0·02 (0·15)	0·06 (0·15)	−0·16 (0·15)
Length of 1st genital segment	0·17 (0·32)	0·67 (0·19)**	0·61 (0·15)***	0·71 (0·12)***	0·51 (0·20)*
Proctiger length	0·33 (0·24)	0·53 (0·20)**	0·39 (0·16)*	0·44 (0·15)**	0·60 (0·14)***
Length of phallotheca	0·40 (0·26)	0·35 (0·29)	0·50 (0·17)**	0·30 (0·21)	0·24 (0·24)
Length of lateral sclerites	0·05 (0·27)	0·26 (0·26)	0·26 (0·18)	0·19 (0·19)	0·25 (0·21)
Length of ventral sclerite	0·00 (0·45)	0·34 (0·41)	−0·19 (0·31)	−0·28 (0·30)	−0·14 (0·36)
Length of dorsal sclerite	−0·07 (0·24)	0·08 (0·25)	0·08 (0·17)	0·17 (0·17)	0·16 (0·19)
Score relative warp number 5 (shape)	−0·12 (0·24)	−0·19 (0·25)	−0·32 (0·16)*	−0·26 (0·17)	−0·29 (0·19)

Numbers within brackets represent estimated standard errors.
*$P < 0·05$, **$P < 0·01$, ***$P < 0·001$.

Table 7. *Phenotypic correlations between trait size and the absolute value of fluctuating asymmetry*, *for all traits exhibiting fluctuating asymmetry*. *Separate estimates are given for parental* ($n = 61$) *and offspring* ($n = 198$) *males*

	Parental males	Offspring males
General morphology		
Abdominal spine length	0·021	−0·083
Length of 1st antennal segment	−0·175	−0·060
Forefemur length	−0·070	0·073
Midleg length	−0·029	0·161
Hindleg length	0·128	0·074
Score relative warp number 1	−0·017	0·060
Score relative warp number 2	−0·161	−0·138
Score relative warp number 3	0·151	−0·061
Score relative warp number 4	−0·017	0·097
Score relative warp number 6	0·121	−0·074
Genital morphology		
Length of lateral sclerites	−0·137	−0·116
Centroid size	0·039	0·050
Score relative warp number 1	0·155	−0·085
Score relative warp number 3	0·040	0·098
Score relative warp number 4	−0·095	0·062
Score uniform shape component number 1	−0·141	−0·019

genetic correlation, between genitalic shape (score of relative warp number 5) and general morphology (Table 6).

The repeatabilities of linear measures of size were consistently very high, for both general and genital traits (mean = 0·91, SD = 0·07, range 0·75–1·00) (see Appendixes A and B). Repeatabilities of multivariate shape scores were somewhat lower and more variable, for both types of traits (mean = 0·70, SD = 0·19, range 0·13–0·92). Measures of asymmetry showed somethat lower yet and also variable degrees of repeatabilities (mean = 0·54, SD = 0·18, range 0·10–0·86). Again, the results of a three-way ANOVA of the repeatabilities (square-root transformed) were in agreement with this interpretation. Traits measuring multivariate shape were less repeatable compared with linear size measures ($F_{1,42} = 6·10$, $P = 0·018$), irrespective of whether they measured genital or non-genital traits ($F_{1,42} = 2·55$, $P > 0·1$), and measures of FA were less repeatable than other traits ($F_{1,42} = 12·63$, $P < 0·001$). For general morphological traits, the repeatabilities estimated for males and females were highly correlated ($r = 0·89$, $n = 32$, $P < 0·01$). Further, as expected, the degree of repeatability was positively related to the estimated heritability across all traits. The correlation coefficient with repeatability was 0·52 for broad-sense heritability and 0·55 for narrow-sense heritability ($n = 50$, $P < 0·001$ in both cases).

4. Discussion

This is the first in-depth study of the intraspecific patterns of inheritance and phenotypic plasticity of genitalic morphology in any species. Our data are

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

novel in many ways, and several new insights can be gained. We put our results to five tasks. First, we summarize our main findings. Secondly, we assess how our overall results agree with the various hypotheses for genitalic evolution. Third, we compare the relative merits of linear and multivariate morphometrics. Fourth, we compare different measures of condition dependence, and discuss their implications. Fifth, we discuss some of the methodological implications of our results.

(i) *General conclusions*

Genitalic morphology in our species exhibited moderate levels of phenotypic variation but, more importantly, genitalia were as phenotypically variable as general morphological traits. Further, both size and shape of genitalia exhibited sizable additive genetic components of phenotypic trait variation which, again, were of similar magnitudes to those of general morphological traits. Thus, we may conclude that male genitalia in *G. incognitus* were as variable, both phenotypically and genotypically, as the other types of traits measured. This is, to our knowledge, the first quantification of intraspecific genetic variation in genitalic morphology. Interestingly enough, our conclusions are largely in agreement with previous comparative work on *Drosophila*, where the genetic basis for interspecific differences in genitalic conformation has been shown to be polygenic and largely additive (Coyne, 1983, 1985; Coyne & Kreitman, 1986; Liu *et al*., 1996).

In general, genital traits showed somewhat lower levels of condition dependence compared with general morphological traits. However, two lines of evidence show that phenotypic expression of genitalia was indeed condition dependent. First, the effects sizes of our experimental food stress on genitalic conformation were in several cases high, especially for measures of size of genitalia, showing that food stress indeed affected genitalic morphology. Secondly, the fairly high phenotypic correlations between genitalic size and general size of body and appendages also indicates condition dependence. We conclude that development of genitalia was neither highly canalized nor invariant to environmental conditions, but rather exhibited moderate levels of condition dependence.

Measures of genitalic morphology, both size and shape, were genetically correlated with measures of body size and leg length. Actually, our estimates of genetic correlations were of similar magnitude to those of phenotypic correlations (cf. Cheverud, 1988; Roff, 1996). Assuming that transient linkage plays a minor role, we can thus conclude that genitalic and general morphology are to a certain extent influenced by the same set of pleiotropic genes in *G. incognitus* (Falconer & Mackay, 1996).

(ii) *Assessment of hypotheses for genitalic evolution*

On the basis of comparative studies, several previous authors have concluded that the lock-and-key hypothesis, the most widespread and long-standing hypothesis, is in poor agreement with the patterns of genitalic function and divergence across animal taxa (Mayr, 1963; Scudder, 1971; Eberhard, 1985, 1990, 1993; Shapiro & Porter, 1989). We found that genitalia are as phenotypically variable as are other traits, and that this variation is as additive genetic in nature. Further, genitalic morphology is phenotypically plastic and responds to differences in environmental conditions. This general picture is in contrast to the image of the invariant 'key' envisioned by the lock-and-key hypothesis (Eberhard, 1985; Arnqvist, 1997*b*), and our results in terms of the patterns of inheritance and condition dependence of genitalia are hence not compatible with the lock-and-key hypothesis.

Although it is notoriously difficult to distinguish empirically between various models of sexual selection (Kirkpatrick & Ryan, 1991; Andersson, 1994; Johnstone, 1995; Andersson & Iwasa, 1996), traits under sexual selection have been shown generally to exhibit relatively high levels of phenotypic and genetic variance (Pomiankowski & Møller, 1995) and several mechanisms may generate condition dependence in trait expression (Andersson, 1994; Rowe & Houle, 1996). Thus, in a very general sense, our results are in agreement with the sexual selection hypothesis of general evolution (cf. Arnqvist, 1997*b*).

Eberhard (1993) suggested that genitalic elaboration conveys no costs, and that genitalia should thus be particularly prone to evolve rapidly by a run-away process generated by a sensory exploitation mechanism (cryptic female choice). Under this particular 'Fisherian' sexual selection scenario, phenotypic evolution of genitalic traits would be halted only by a lack of genetic variance, rather than by any antagonistic natural selection. Additive genetic variance would thus rapidly be exhausted. Our results does not support this somewhat unrealistic version of the sexual selection hypothesis, since we found fairly high levels of genetic variation in genitalic morphology.

Our results are also in general agreement with the pleiotropy hypothesis of Mayr (1963) (Arnqvist, 1997*b*). The pleiotropy hypothesis is based on the assumption that the set of genes determining genital morphology also affects other traits by pleiotropic effects. Our finding of genetic correlations between genital shape/size and general morphology indicates that such a pattern exists in *G. incognitus*. Thus, if components of general morphology (such as body size or leg length) evolve, genitalia will also tend to evolve as a correlated response. Moreover, since genetic correlations are primarily caused by the basic func-

tional genetic 'architecture' of an organism (Houle, 1991; Falconer & Mackay, 1996), which is likely to be shared between closely related species, this pattern of inheritance could potentially contribute to evolutionary diversification of genitalia within the genus *Gerris*.

In conclusion, the patterns of variation, inheritance and condition dependence of genitalia in *G. incognitus* presented here are in disagreement with the long-standing lock-and-key hypothesis. They are, however, in general agreement with the sexual selection hypothesis and the pleiotropy hypothesis. In a related and complementary study, Arnqvist *et al.* (1997) studied the patterns of phenotypic selection on genitalic traits in *G. incognitus*. Their results were similarly in disagreement with the lock-and-key hypothesis, and hence yielded conclusions very similar to those drawn in the current study. Further discrimination between the various hypotheses based on these data alone is difficult, since many of the predictions are relatively weak (for a discussion see Arnqvist, 1997*b*).

(iii) *Morphometric methods*

Our study is one of the very first to address the quantitative genetics of multivariate shape (but see Liu *et al.*, 1996). Even though our linear measures of size of various traits and our multivariate shape scores measure in part wholly different components of morphology, they exhibited a similar pattern of phenotypic and genotypic variation. Morphological shape measures were generally somewhat less affected by food stress compared with measures of size, but the amount of additive genetic components in phenotypic trait variation was similar for both types of morphological measures.

The repeatabilities of our multivariate shape scores were reasonably high (average 0·70), but were, nevertheless, lower than the repeatabilities of linear measures of size – a pattern that is expected and may be general (see Arnqvist & Mårtensson, 1998). Multivariate measures of shape potentially capture more complex and comprehensive aspects of morphology than do simple linear measures of size (Rohlf & Marcus, 1993). In the light of this, it is encouraging that shape measures and size measures gave essentially the same result when patterns of genital and general morphological variation were compared, even though the greatest benefit in intraspecific studies from this characteristic of multivariate shape analysis may be reached in studies where morphology is related to performance in various ways (Arnqvist *et al.*, 1997). In either case, we have shown that it is possible to measure and capture complex morphological variation accurately even in very small traits by means of landmark-based multivariate shape analysis.

(iv) *Condition dependence*

In the domain of sexual selection, much attention is currently being given to condition-dependent expression of sexual characters (Andersson, 1994; Rowe & Houle, 1996), especially with regard to biological handicap models (Johnstone, 1995). However, much of this research is limited to correlative studies, which are weakened by potential confounding effects, and very few have assessed the degree of condition dependence in a suite of sexual traits by experimentally altering the environmental conditions (Johnstone, 1995). Several insights can be gained from our experiment. First, to manipulate environmental conditions experimentally allows for unambiguous assessments of the degree of condition dependence in a suite of traits for a given magnitude of a known source of stress. This is essential when any causal relationships are sought. Secondly, virtually all traits are condition dependent to some extent. Thus, we believe it is very important to study multiple traits and to focus on the magnitude of condition dependence rather than the mere existence of condition dependence. The fact that a potential morphological signal/ornament (such as genital size or shape) is 'significantly' revealing of phenotypic condition is not necessarily particularly informative. Other traits (for example body size) may be much more revealing of phenotypic quality, in which case we would expect female perception systems to evolve to home in on the latter rather than the former.

Thirdly, provided females receive no direct benefits from males (Price *et al.*, 1993), an indicator trait (a handicap) should be revealing of genotypic, as opposed to phenotypic, quality for a good-genes/handicap sexual selection process to operate (Johnstone, 1995). In other words, condition dependence is assumed to be heritable (Grafen, 1990; Rowe & Houle, 1996). A major benefit of controlled laboratory experiments, such as that reported in this paper, is that they allow direct estimations of the extent to which different male genotypes differ in their condition dependence by investigating the genotype × environment interaction terms. Body size in *G. incognitus* may serve as an example. Size was strongly revealing of environmental condition experienced, and also found to be heritable. Thus, choosing large mates would on average select for mates of high phenotypic condition. However, different male genotypes did not differ in their response to food stress. In other words, there was no evidence of any variance in genotypic quality, in the sense that different genotypes did not differ in their ability to cope with harsh environmental conditions (food stress).

Fourthly, there is currently much discussion about measures of FA in bilaterally symmetrical traits, and its utility as an indicator of individual condition and

quality (Møller & Pomiankowski, 1993; Swaddle *et al*., 1994; Watson & Thornhill, 1994). First, we wish to stress the importance of quantifying measurement error, by estimating repeatabilities of measures of FA, which are often not presented (Swaddle *et al*., 1994; Merilä & Björklund, 1995). In our case, measures of FA were in some cases to a large extent composed of measurement error. Overall, almost 50% of the between-individual variation in FA was due to measurement error, though the repeatability varied greatly between traits (range 14–90%). Obviously, this illustrates the difficulty of measuring low levels of FA, and that information of repeatability is critical in interpreting any relationships between FA of any given trait and individual performance in empirical studies.

Despite our relatively dramatic food treatment, which had effects on survivorship and growth rate as well as many other components of morphology, we failed to find any effects of food stress on FA. Considering that other studies have shown effects of stress on FA (Palmer & Strobeck, 1986; Møller & Pomiankowski, 1993; Swaddle *et al*., 1994; Watson & Thornhill, 1994), this result is surprising. It could certainly in part be due to the influence of measurement error in our measures of FA, but this can not account for the measures of FA that showed high repeatabilities but no detectable effects of food stress. We conclude that FA in the traits measured was apparently not affected by food stress in *G. incognitus*. While this certainly does not eliminate the possibility that other forms of stress (e.g. parasites) may affect FA, it illustrates the fact that measures of FA can not be assumed to be universal and integrative measures of phenotypic quality. Experimental manipulations, such as those presented here, are necessary to truly reveal the causal factors (if any) of individual variation in FA (Møller, 1992).

(v) *Methodological implications*

Our estimates of narrow-sense heritability were systematically somewhat higher than the estimates of broad-sense heritability. This result is, of course, puzzling, since true narrow-sense heritabilities can not exceed the true value of broad-sense heritabilities. In theory, this could be due either to inflated narrow-sense estimates or to depressed broad-sense estimates. Four facts strongly suggest that the former is not the case. First, phenotypic variances did not differ across generations (see Section 3). Secondly, our narrow-sense estimates were generally much lower than the upper limit set by the repeatabilities (less than half on average). Thirdly, we found no genotype × environment interactions across laboratory environments (cf. Lande, 1987). Fourthly, narrow-sense estimates based on parent–offspring resemblance are generally most reliable (Falconer & Mackay, 1996). Thus, we suggest that our broad-sense heritability estimates are underestimates of the true values of V_G/V_P. This could in part result from founding our broad-sense estimates on full-sib data, which are less reliable and in some cases can render depressed estimates of genetic parameters (Arnold, 1994). However, an even more important factor is undoubtedly constraints imposed by the computational methods involved. REML estimators of heritabilities are namely known to be potentially negatively biased, as a result of the inability to handle negative estimates of variance components in maximum likelihood estimation (the non-negativity constraint) (Shaw, 1987). A comparison across traits supports this interpretation: average difference between estimates of narrow- and broad-sense heritabilities was 0·18 ($n = 32$, SE = 0·04) for traits in which the REML models produced negative estimates of variance components, but only 0·08 ($n = 18$, SE = 0·04) for traits that were not affected by the non-negativity constraint, suggesting a negative bias in the order of at least 10%. Our results hence strongly indicate that this bias is systematic and can be considerable. Thus, while REML estimation allows for a very flexible and relatively accurate analysis of variance components even in cases where the data are unbalanced, we believe that interpretation of REML estimations of variance components when REML estimation involves negative estimates should be done with caution (see also Shaw, 1987).

We also wish to stress the utility and importance of estimating repeatabilities of morphological characters. These set upper limits to the heritabilities (see above) and hold quantitative information about the influence of measurement error in trait variation. This is key for traits prone to be associated with high degrees of measurement error (e.g. FA or small-scale traits such as genitalia), but is also imortant for more complex measures of morphology such as multivariate measures of shape. For such measures, certain shape components obviously reflect measurement error more than others. For example, while most of our multivariate measures of shape captured primarily true between-individual variation in shape, this was not always the case. For at least one measure of shape (genitalia; score of relative warp number 2), phenotypic variance consisted almost wholly of measurement error.

The current contribution represents the first extensive intraspecific study of the genetic control of genital morphology (Eberhard, 1985). The data presented here allowed us tentatively to assess various hypotheses for the evolution of animal genitalia, but also yielded a number of additional insights. In particular, our results do not support the long-standing lock-and-key hypothesis and our study thus

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

illustrates the utility and importance of in-depth studies of the pattern of variability and inheritance of genital versus general morphology. We hope that more empirical studies will follow, allowing for a future consensus on the evolutionary processes responsible for genitalic evolution.

Appendix A

Estimates of repeatability for general (non-genitalic) morphological traits, estimated separately for females and males. The critical limit of a significant ($P < 0{\cdot}05$) between-individuals variation (unadjusted) equals a repeatability $\geqslant 0{\cdot}23$. Given also are P-values of t-tests of $H_0: \mu = 0$ and of Kolmogorov–Smirnov tests of distributional normality for all measures of asymmetry.

Trait/trait group	Repeatability		$H_0: \mu = 0$ (t-test)	Normality (K–S test)
	Females	Males		
Linear measures				
Body length	0·99	0·99	—	—
Thorax width	0·90	0·93	—	—
Length of abdominal spines	0·91	0·83	—	—
Distance between tips of abdominal spines	0·93	0·88	—	—
Elevation angle of abdominal spines	0·57	0·72	—	—
Length of 1st antennal segment	0·95	0·93	—	—
Forefemur length	0·96	0·96	—	—
Midleg length	0·95	0·96	—	—
Hindleg length	0·96	0·97	—	—
Asymmetry in abdominal spine length	0·68	0·56	$P > 0{\cdot}1$	$P > 0{\cdot}6$
Asymmetry in length of 1st antennal segment	0·47	0·61	$P > 0{\cdot}2$	$P > 0{\cdot}2$
Asymmetry in forefemur length	0·86	0·64	$P > 0{\cdot}6$	$P > 0{\cdot}1$
Asymmetry in midleg length	0·69	0·60	$P > 0{\cdot}2$	$P > 0{\cdot}2$
Asymmetry in hindleg length	0·65	0·70	$P > 0{\cdot}4$	$P > 0{\cdot}1$
Shape analysis of body				
Centroid size	1·00	0·99	—	—
Score relative warp number 1	0·90	0·89	—	—
Score relative warp number 2	0·92	0·90	—	—
Score relative warp number 3	0·82	0·66	—	—
Score relative warp number 4	0·89	0·68	—	—
Score relative warp number 5	0·83	0·75	—	—
Score relative warp number 6	0·57	0·63	—	—
Score uniform shape component number 1	0·62	0·61	—	—
Score uniform shape component number 2	0·78	0·82	—	—
Asymmetry in centroid size	0·58	0·69	$P < 0{\cdot}001$	$P > 0{\cdot}05$
Asymmetry in score relative warp number 1	0·63	0·51	$P > 0{\cdot}05$	$P > 0{\cdot}4$
Asymmetry in score relative warp number 2	0·62	0·67	$P > 0{\cdot}05$	$P > 0{\cdot}1$
Asymmetry in score relative warp number 3	0·77	0·57	$P > 0{\cdot}1$	$P > 0{\cdot}1$
Asymmetry in score relative warp number 4	0·56	0·50	$P > 0{\cdot}2$	$P > 0{\cdot}05$
Asymmetry in score relative warp number 5	0·12	0·31	$P < 0{\cdot}001$	$P < 0{\cdot}001$
Asymmetry in score relative warp number 6	0·23	0·33	$P > 0{\cdot}05$	$P > 0{\cdot}05$
Asymmetry in score uniform shape comp. number 1	0·41	0·46	$P < 0{\cdot}001$	$P < 0{\cdot}05$
Asymmetry in score uniform shape comp. number 2	0·56	0·58	$P < 0{\cdot}001$	$P > 0{\cdot}2$

https://doi.org/10.1017/S0016672398003279 Published online by Cambridge University Press

Appendix B

Estimates of repeatability for genital morphological traits in males. The critical limit for a significant ($P < 0{\cdot}05$) between-individuals variation (unadjusted) equals a repeatability $\geqslant 0{\cdot}23$. Given also are P-values of t-tests of $H_0: \mu = 0$ and of Kolmogorov–Smirnov tests of distributional normality for all measures of asymmetry.

Trait/trait group	Repeatability	$H_0: \mu = 0$ (t-test)	Normality (K–S test)
Linear measures			
Length of 1st genital segment	0·94	—	—
Proctiger length	0·94	—	—
Length of phallotheca	0·91	—	—
Length of lateral sclerites	0·82	—	—
Distance between lateral sclerites	0·79	—	—
Length of ventral sclerite	0·86	—	—
Length of dorsal sclerite	0·84	—	—
Asymmetry in length of 1st genital segment	0·83	$P < 0{\cdot}001$	$P > 0{\cdot}4$
Asymmetry in length of lateral sclerites	0·54	$P > 0{\cdot}8$	$P > 0{\cdot}8$
Shape analysis of genital capsule			
Centroid size	0·75	—	—
Score relative warp number 1	0·74	—	—
Score relative warp number 2	0·13	—	—
Score relative warp number 3	0·35	—	—
Score relative warp number 4	0·46	—	—
Score relative warp number 5	0·82	—	—
Score relative warp number 6	0·62	—	—
Score uniform shape component number 1	0·67	—	—
Score uniform shape component number 2	0·76	—	—
Asymmetry in centroid size	0·44	$P > 0{\cdot}5$	$P > 0{\cdot}2$
Asymmetry in score relative warp number 1	0·55	$P > 0{\cdot}05$	$P > 0{\cdot}3$
Asymmetry in score relative warp number 2	0·10	$P < 0{\cdot}001$	$P > 0{\cdot}1$
Asymmetry in score relative warp number 3	0·20	$P > 0{\cdot}4$	$P > 0{\cdot}3$
Asymmetry in score relative warp number 4	0·52	$P > 0{\cdot}1$	$P > 0{\cdot}9$
Asymmetry in score relative warp number 5	0·61	$P < 0{\cdot}001$	$P > 0{\cdot}4$
Asymmetry in score relative warp number 6	0·33	$P < 0{\cdot}001$	$P > 0{\cdot}4$
Asymmetry in score uniform shape component number 1	0·80	$P > 0{\cdot}6$	$P > 0{\cdot}1$
Asymmetry in score uniform shape component number 2	0·44	$P < 0{\cdot}001$	$P > 0{\cdot}8$

Appendix C

Summary of phenotypic variation in linear scalar general and genital morphometric traits. Data shown are based on parental individuals ($n = 61$ in all cases). Mean represents the arithmetic mean and CV represents the coefficient of variation.

	Males				Females			
Trait/trait group	Mean	SE_{mean}	CV (%)	SE_{CV}	Mean	SE_{mean}	CV (%)	SE_{CV}
General morphology								
Body length (mm)	7·06	0·029	3·2	0·29	8·00	0·034	3·3	0·30
Thorax width (mm)	2·25	0·008	2·7	0·21	2·65	0·010	3·0	0·28
Length of abdominal spines (mm)	1·00	0·007	5·3	0·48	1·35	0·005	4·8	0·44
Distance between tips of abdominal spines (mm)	0·78	0·007	7·3	0·66	0·62	0·016	20·6	1·86
Elevation angle of abdominal spines (degrees)	5·62	0·599	—	—	29·82	0·634	—	—
Length of 1st antennal segment (mm)	1·22	0·005	3·2	0·29	1·29	0·007	4·2	0·38
Forefemur length (mm)	2·07	0·008	3·0	0·26	2·14	0·008	3·0	0·27
Midleg length (mm)	8·51	0·036	3·3	0·30	9·11	0·039	3·4	0·30
Hindleg length (mm)	6·81	0·032	3·6	0·32	7·25	0·038	4·0	0·36
Genital morphology								
Length of 1st genital segment (mm^{-1})	3·86	0·017	3·5	0·32				
Proctiger length (mm^{-1})	2·23	0·009	3·5	0·31				
Length of phallotheca (mm^{-1})	1·50	0·006	3·2	0·29				
Length of lateral sclerites (mm^{-1})	0·78	0·005	4·9	0·44				
Distance between lateral sclerites (mm^{-1})	0·28	0·006	16·5	1·49				
Length of ventral sclerite (mm^{-1})	0·89	0·007	6·0	0·54				
Length of dorsal sclerite (mm^{-1})	1·19	0·006	3·8	0·35				

We wish to thank J. F. Rohlf for modifying his thin-plate spline relative warp analysis computer program TPSRW to meet our requirements, and for allowing us to use it. Similarly, D. Slice kindly allowed us access to his morphometric program GRF-ND. R. G. Shaw gave invaluable statistical advice. We are most grateful to Stephanie Weinstein for digitizing all morphometric data. We thank N. Barton, L. Rowe, P. Watson and an anonymous reviewer for constructive comments on earlier versions of this paper. J. Elmberg provided laboratory assistance. This study was made possible by financial support from The Swedish Natural Science Research Council, The Crafoord Foundation (The Royal Swedish Academy of Sciences), The Fulbright Commission and The Swedish Institute. Logistic support was received from the Department of Biology, University of New Mexico, which is gratefully acknowledged.

References

Alexander, R. D., Marshall, D. C. & Cooley, J. R. (1997). Evolutionary perspectives on insect mating. In *Social Competition and Cooperation in Insects and Arachnids*, vol. I, *Evolution of Mating Systems* (ed. J. C. Choe & B. J. Crespi), pp. 4–31. Cambridge: Cambridge University Press.

Andersen, N. M. (1982). *The Semiaquatic Bugs* (*Hemiptera, Gerromorpha*)*: Phylogeny, Adaptations, Biogeography and Classification*. Entomonograph 3. Klampenborg, Denmark: Scandinavian Science Press.

Andersen, N. M. (1993). Classification, phylogeny, and zoogeography of the pond skater genus *Gerris* Fabricius (Hemiptera: Gerridae). *Canadian Journal of Zoology* **71**, 2473–2508.

Andersson, M. (1994). *Sexual Selection*. Princeton, NJ: Princeton University Press.

Andersson, M. & Iwasa, Y. (1996). Sexual selection. *Trends in Ecology and Evolution* **11**, 53–58.

Arnold, S. J. (1994). Multivariate inheritance and evolution: a review of concepts. In *Quantitative Genetic Studies of Behavioral Evolution* (ed. C. R. B. Boake), pp. 17–48. Chicago: University of Chicago Press.

Arnqvist, G. (1988). Mate guarding and sperm displacement in the water strider *Gerris lateralis* Schumm. (Heteroptera: Gerridae). *Freshwater Biology* **19**, 269–274.

Arnqvist, G. (1989). Multiple mating in a water strider: mutual benefits or intersexual conflict? *Animal Behaviour* **38**, 749–756.

Arnqvist, G. (1992). Spatial variation in selective regimes: sexual selection in the water strider, *Gerris odontogaster*. *Evolution* **46**, 914–929.

Arnqvist, G. (1997*a*). The evolution of water strider mating systems: causes and consequences of sexual conflicts. In *Social Competition and Cooperation in Insects and Arachnids*, vol. I, *Evolution of Mating Systems* (ed. J. C. Choe & B. J. Crespi), pp. 146–163. Cambridge: Cambridge University Press.

Arnqvist, G. (1997*b*). The evolution of animal genitalia: distinguishing between hypotheses by single species studies. *Biological Journal of the Linnean Society* **60**, 365–379.

Arnqvist, G. & Mårtensson, T. (1998). Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measure of shape. *Acta Zoologica* (in press).

Arnqvist, G. & Rowe, L. (1995). Sexual conflict and arms races between the sexes: a morphological adaptation for control of mating in a female insect. *Proceedings of the Royal Society of London, Series B* **261**, 123–127.

Arnqvist, G. & Wooster, D. (1995). Meta-analysis: synthesizing research findings in ecology and evolution. *Trends in Ecology and Evolution* **10**, 236–240.

Arnqvist, G., Thornhill, R. & Rowe, L. (1997). Evolution of animal genitalia: morphological correlates of fitness components in a water strider. *Journal of Evolutionary Biology* **10**, 613–640.

Auffray, J.-C., Alibert, P., Renaud, S., Orth, A. & Bonhomme, F. (1996). Fluctuating asymmetry in *Mus musculus* subspecific hybridization: traditional and Procrustes comparative approach. In *Advances in Morphometrics* (ed. L. F. Marcus, M. Corti, A. Loy, G. Naylor & D. Slice), pp. 275–284. New York: Plenum Press.

Becker, W. A. (1984). *Manual of Quantitative Genetics*, 4th edn. Pullman, Wash.: Academic Enterprises.

Berry, D. A. (1987). Logarithmic transformations in ANOVA. *Biometrics* **43**, 439–456.

Birkhead, T. R. & Hunter, F. M. (1990). Mechanisms of sperm competition. *Trends in Ecology and Evolution* **5**, 48–52.

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data*. Cambridge: Cambridge University Press.

Bookstein, F. L. (1996). A standard formula for the uniform shape component in landmark data. In *Advances in Morphometrics* (ed. L. F. Marcus, M. Corti, A. Loy, G. Naylor & D. Slice), pp. 153–168. New York: Plenum Press.

Cheverud, J. M. (1988). A comparison of genetic and phenotypic correlations. *Evolution* **42**, 958–968.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Coyne, J. A. (1983). Genetic basis of differences in genital morphology among three sibling species of *Drosophila*. *Evolution* **37**, 1101–1118.

Coyne, J. A. (1985). Genetic studies of three sibling species of *Drosophila* with relationship to theories of speciation. *Genetical Research* **46**, 169–192.

Coyne, J. A. & Beecham, E. (1987). Heritability of two morphological characters within and among natural populations of *Drosophila melanogaster*. *Genetics* **117**, 727–737.

Coyne, J. A. & Kreitmann, M. (1986). Evolutionary genetics of two sibling species, *Drosophila simulans* and *D. sechellia*. *Evolution* **40**, 673–691.

Eberhard, W. G. (1985). *Sexual Selection and the Evolution of Animal Genitalia*. Cambridge, Mass.: Harvard University Press.

Eberhard, W. G. (1990). Animal genitalia and female choice. *American Scientist* **78**, 134–141.

Eberhard, W. G. (1993). Evaluating models of sexual selection: genitalia as a test case. *American Naturalist* **142**, 564–571.

Eberhard, W. G. (1996). *Female Control: Sexual Selection by Cryptic Female Choice*. Princeton, NJ: Princeton University Press.

Falconer, D. S. (1973). Replicated selection for body weight in mice. *Genetical Research* **22**, 291–321.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Essex: Longman.

Grafen, A. (1990). Sexual selection unhandicapped by the Fisher process. *Journal of Theoretical Biology* **144**, 473–516.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320–340.

Heming-van Battum, K. E. & Heming, B. S. (1989). Structure, function, and evolutionary significance of the reproductive system in males of *Hebrus ruficeps* and *H. pusillus* (Heteroptera, Gerromorpha, Hebridae). *Journal of Morphology* **202**, 281–323.

Henderson, C. R. (1985). MIVQUE and REML estimation of additive and nonadditive genetic variances. *Journal of Animal Science* **61**, 113–121.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65–70.

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

Johnstone, R. A. (1995). Sexual selection, honest advertisement and the handicap principle: reviewing the evidence. *Biological Reviews* **70**, 1–65.

Kaitala, A. (1987). Dynamic life-history strategy of the water strider *Gerris thoracicus* as an adaptation to food and habitat variation. *Oikos* **48**, 125–131.

Kearsey, M. J. (1965). Biometrical analysis of a random mating population: a comparison of five experimental designs. *Heredity* **20**, 205–235.

Kirkpatrick, M. & Ryan, M. J. (1991). The evolution of mating preferences and the paradox of the lek. *Nature* **350**, 33–38.

Lande, R. (1977). On comparing coefficients of variation. *Systematic Zoology* **26**, 214–216.

Lande, R. (1987). Appendix to Coyne and Beecham. *Genetics* **117**, 737.

Lessels, C. M. & Boag, P. T. (1987). Unrepeatable repeatabilities: a common mistake. *Auk* **104**, 116–121.

Liu, J., Mercer, J. M., Stam, L. F., Gibson, G. C., Zeng, Z.-B. & Laurie, C. C. (1996). Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritana*. *Genetics* **142**, 1129–1145.

Lloyd, J. E. (1979). Mating behavior and natural selection. *Florida Entomologist* **62**, 17–23.

Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. & Slice, D. E. (eds.) (1996). *Advances in Morphometrics*. New York: Plenum Press.

Mather, K. (1953). Genetical control of stability in development. *Heredity* **7**, 297–336.

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, Mass.: Harvard University Press.

McCullagh, P. & Nelder, J. A. (1983). *Generalized Linear Models*. New York: Chapman and Hall.

Merilä, J. & Björklund, M. (1995). Fluctuating asymmetry and measurement error. *Systematic Biology* **44**, 97–101.

Møller, A. P. (1992). Parasites differentially increase the degree of fluctuating asymmetry in secondary sexual characters. *Journal of Evolutionary Biology* **5**, 691–699.

Møller, A. P. & Pomiankowski, A. (1993). Fluctuating asymmetry and sexual selection. *Genetica* **89**, 267–279.

Palmer, A. R. & Strobeck, C. (1986). Fluctuating asymmetry: measurement, analysis and pattern. *Annual Review of Ecology and Systematics* **17**, 391–421.

Pirchner, F. (1983). *Population Genetics in Animal Breeding*. New York: Plenum Press.

Pomiankowski, A. & Møller, A. P. (1995). A resolution of the lek paradox. *Proceedings of the Royal Society of London, Series B* **260**, 21–29.

Price, T., Schluter, D. & Heckman, N. E. (1993). Sexual selection when the female directly benefits. *Biological Journal of the Linnean Society* **48**, 187–211.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

Rice, W. R. (1996). Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381**, 232–234.

Roff, D. A. (1996). The evolution of genetic correlations: an analysis of patterns. *Evolution* **50**, 1392–1403.

Rohlf, F. J. (1993). Relative warp analysis and an example

of its application to mosquito wings. In *Contributions to Morphometrics* (ed. L. F. Marcus, E. Bello & A. Garcia-Valdecasas), pp. 131–159. Madrid, Spain: Monografias del Museo Nacional de Ciencias Naturales.

Rohlf, F. J. & Marcus, L. F. (1993). A revolution in morphometrics. *Trends in Ecology and Evolution* **8**, 129–132.

Rowe, L. & Houle, D. (1996). The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society of London*, *Series B* **263**, 1415–1421.

Rowe, L., Arnqvist, G., Sih, A. & Krupa, J. J. (1994). Sexual conflict and the evolutionary ecology of mating patterns: water striders as a model system. *Trends in Ecology and Evolution* **9**, 289–293.

Rubenstein, D. I. (1989). Sperm competition in the water strider *Gerris remigis*. *Animal Behaviour* **38**, 631–636.

SAS (1992). *Software: Changes and Enhancements*, *Release 6.07 Technical Report P-229*. Cary, NC: SAS Institute.

Scudder, G. G. E. (1971). Comparative morphology of insect genitalia. *Annual Review of Entomology* **16**, 379–406.

Searle, S. R., Casella, G. & McCulloch, C. E. (1992). *Variance Components*. New York: Wiley.

Shapiro, A. M. & Porter, A. H. (1989). The lock-and-key hypothesis: evolutionary and biosystematic interpretation of insect genitalia. *Annual Review of Entomology* **34**, 231–245.

Shaw, R. G. (1987). Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution* **41**, 812–826.

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

Slice, D. (1994). *Generalized Rotational Fitting of n-Dimensional Landmark Data*. Stony Brook, New York: State University of New York.

Smith, C. A. B. (1980). Estimating genetic correlations. *Annals of Human Genetics* **43**, 265–284.

Smith, D. R., Crespi, B. J. & Bookstein, F. L. (1997). Fluctuating asymmetry in the honey bee, *Apis mellifera*: effects of ploidy and hybridization. *Journal of Evolutionary Biology* **10**, 551–574.

Smith, R. L. (ed.) (1984). *Sperm Competition and the Evolution of Animal Mating Systems*. Orlando, Fla.: Academic Press.

Sokal, R. R. & Rohlf, F. J. (1981). *Biometry*. San Francisco: W. H. Freeman.

Spence, J. R. & Andersen, N. M. (1994). Biology of water striders: interactions between systematics and ecology. *Annual Review of Entomology* **39**, 101–128.

Swaddle, J. P., Witter, M. S. & Cuthill, I. C. (1994). The analysis of fluctuating asymmetry. *Animal Behaviour* **48**, 986–989.

SYSTAT (1992). *SYSTAT for Windows: Statistics*, *Version 5 Edition*. Evanston, Ill.: SYSTAT.

Thornhill, R. (1983). Cryptic female choice and its implications in the scorpionfly *Harpobittacus nigriceps*. *American Naturalist* **122**, 765–788.

Van Valen, L. (1962). A study of fluctuating asymmetry. *Evolution* **16**, 125–142.

Waage, J. K. (1984). Sperm competition and the evolution of Odonate mating systems. In *Sperm Competition and the Evolution of Animal Mating Systems* (ed. R. L. Smith), pp. 251–290. Orlando, Fla.: Academic Press.

Watson, P. J. & Thornhill, R. (1994). Fluctuating asymmetry and sexual selection. *Trends in Ecology and Evolution* **9**, 21–25.