

## **Evolution of animal genitalia: morphological correlates of fitness components in a water strider**

G. Arnqvist,<sup>1</sup> R. Thornhill<sup>2</sup> and L. Rowe<sup>3</sup>

<sup>1</sup>*Department of Animal Ecology, University of Umeå, S-901 87 Umeå, Sweden,  
e-mail: Goran.Arnqvist@animecol.umu.se*

<sup>2</sup>*Department of Biology, University of New Mexico, Albuquerque,  
NM 87131-1091, USA*

<sup>3</sup>*Department of Zoology, University of Toronto, Toronto, Ontario, M5S 1A1,  
Canada*

*Key words:* Genitalia; genitalic evolution; fluctuating asymmetry; sexual conflict; cryptic female choice.

### **Abstract**

Rapid divergence of male genitalia is one of the most general evolutionary trends in animals with internal fertilization, but the mechanisms of genital evolution are poorly understood. The current study represents the first comprehensive attempt to test the main hypotheses that have been suggested to account for genital evolution (the lock-and-key, sexual selection and pleiotropy hypotheses) with intraspecific data. We measure multivariate phenotypic selection in a water strider species, by relating five different components of fitness (mating frequency, fecundity, egg hatching rate, offspring survival rate and offspring growth rate) to a suite of genital and non-genital morphological traits (in total 48). Body size had a series of direct effects in both sexes. Large size in females was positively related to both fecundity and egg hatching rate. There was positive sexual selection for large size in males (mating frequency), which to some extent was offset by a reduced number of eggs laid by females mated to large males. Male genitalic morphology influenced male mating frequency, but the detected directional selection on genitalia was due to indirect selection on phenotypically correlated non-intromittent traits. Further, we found no assortative mating between male intromittent genitalia and female morphology. Neither did we find any indications of male genitalia conveying information of male genetic quality. Several new insights can be gained from our study. Most importantly, our results are in stark disagreement with the long standing lock-and-key hypothesis of genital evolution, as well as with certain models of

sexual selection. Our results are, however, in agreement with other models of sexual selection as well as with the pleiotropy hypothesis of genital evolution. Fluctuating asymmetry of bilaterally symmetrical traits, genital as well as non-genital, had few effects on fitness. Females with low fluctuating asymmetry in leg length produced offspring with a higher survival rate, a pattern most probably caused by direct phenotypic maternal effects. We also discuss the relevance of our results to sexual conflict over mating, and the evolution of sexual traits by coevolutionary arms races between the sexes.

## Introduction

Spectacular morphological diversification of genitalia is very wide spread among animals with internal fertilization (Eberhard, 1985). Even in closely related species, genitalic morphology typically differs greatly among species. Though one of the most general evolutionary trends, it is also one of the most poorly understood (Scudder, 1971; Eberhard, 1985, 1990, 1996; Shapiro and Porter, 1989). A number of different hypotheses have been proposed to account for the evolution of animal genitalia, but the empirical data necessary to distinguish among them is scarce. In particular, in-depth studies of phenotypic variability and selection on genitalic traits is lacking, rendering discussions of the causes of diversification speculative (Arnqvist, 1997b).

Previous attempts to distinguish between the various hypotheses for the evolution of intromittent genitalia have been based on interspecific comparative data (Eberhard, 1985; Shapiro and Porter, 1989). However, two types of intraspecific studies can provide unique information, that is required to illuminate the patterns and processes of genitalic evolution (Arnqvist, 1997b). First, thorough studies of patterns of inheritance and phenotypic plasticity of expression of genitalia are needed. Second, we need to directly estimate selection on genitalic traits in order to understand the processes responsible for genitalic evolution. This study is concerned with the latter line of investigation.

There are three main general hypotheses for the evolution of animal genitalia (Eberhard, 1985, 1990, 1993; Arnqvist, 1997b); a) the lock-and-key, b) the sexual selection and c) the pleiotropy hypotheses. Each hypothesis makes unique, yet untested, predictions about the form and mechanism of selection operating on genitalic traits (Arnqvist, 1997b). The lock-and-key hypothesis states that genitalia evolve via selection for pre-insemination reproductive isolation, so that male intromittent genitalia evolve to be specific and unique (the key) in order to fit appropriately in female genitalia (the lock). Under this hypothesis, stabilizing sexual selection on genitalic traits is predicted, manifested as a relationship between the morphology of male intromittent genitalia and mating/pairing success among males (Arnqvist, 1997b). Given that relative genitalic fit determines the outcome of mating attempts, another key prediction of the lock-and-key hypothesis is positive

true assortative mating (i.e., a homoscedastic relationship between male and female traits; see Arnqvist et al., 1996) by intromittent genitalic traits (Arnqvist, 1997b).

Alternatively, intromittent male genitalia may evolve via sexual selection (Eberhard, 1985, 1996; Andersson, 1994; Arnqvist, 1997b). This hypothesis assumes non-random fertilization success among males, based on genital morphology, which may be brought about by at least three different mechanisms. Males may differ in their ability to stimulate/titillate females, leading to cryptic female choice (Thornhill, 1983; Eberhard, 1985, 1996). Similarly, provided that there is sexual conflict over fertilization events (Rice, 1996), genital morphology may be related to male ability to coerce/control fertilizations (Lloyd, 1979; Arnqvist and Rowe, 1995; Alexander et al., 1997). Differences in genitalia may also be reflected in differences in ability to displace/dislocate sperm of competing males (Smith, 1984; Waage, 1984; Birkhead and Hunter, 1990). The sexual selection hypothesis predicts a relationship between genitalic morphology and relative post-copulatory fertilization success, as opposed to mating/pairing success (cf. the lock-and-key hypothesis above). Further, various versions of the hypothesis predicts that paternal genital morphology should affect offspring fitness in various ways. Under "good-genes" models of cryptic female choice (Eberhard, 1993, 1996), for example, genital morphology should relate to offspring viability (growth and/or survival). Under Fisherian or sensory-exploitation scenarios on the other hand, genital morphology should relate to offspring fitness only indirectly via a "sexy-sons" effect, where there is no natural selection advantage (Andersson, 1994; Eberhard, 1996).

The pleiotropy hypothesis, finally, holds that genitalic evolution is an indirect result of evolution of genetically correlated characters, via pleiotropic effects of genes that code for both genitalic traits and evolving general morphology (Mayr, 1963; Eberhard, 1985, 1990). Since phenotypic variation in genitalic traits *per se* is assumed to be neutral, pleiotropic effects on intromittent genitalia are not selected against and can thus accumulate. Hence, this hypothesis predicts indirect phenotypic selection on genitalic traits, mediated via phenotypic correlations across traits. We expect genitalic morphology to correlate with fitness in univariate estimates of selection (selection differentials). Yet, if the true targets of selection are included in multivariate analyses, no direct selection on genitalic traits (selection gradients) should be observed (see Arnqvist, 1997b). Selection on correlated traits could theoretically be of any form of selection.

In this study, we measure direct and indirect selection on a suite of genitalic and non-genitalic morphological traits in a water strider, by relating measures of relative fitness to multiple morphological traits using multivariate statistical analyses (Lande and Arnold, 1983; Endler, 1986). Selection on multiple components of fitness are analyzed both in the field and in the laboratory; including mating frequency, fecundity, fertility and offspring viability (growth and survival). The results are used primarily to distinguish between the proposed hypotheses of genitalic evolution, and to illustrate how data on the patterns of selection can increase our understanding of the evolutionary processes responsible for the evolution of animal genitalia.

## Materials and methods

The current study was performed on the water strider *Gerris incognitus* (Heteroptera; Gerridae), and it forms the second part of a comprehensive study of genital evolution in this species (Arnqvist and Thornhill, 1997). Water striders inhabit water surfaces of various aquatic habitats, and are predators/scavengers feeding mainly on arthropods trapped at the water surface (Andersen, 1982; Arnqvist, 1977a; Rowe et al., 1994). In terms of morphological divergence within the group, water striders of the genus *Gerris* are typical insects in the sense that relatively rapid evolutionary radiation of male genitalia is a particularly striking and consistent pattern (Andersen, 1982, 1993). The male genitalia 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 intromittent phallus consists of a proximal sclerotized phallosome, and an apical capsule (the vesica) which carries an armature of sclerites that has evolved particularly rapidly and divergently within the genus (see Fig. 2 and Andersen, 1982, 1993, for illustrations).

In the interest of reducing the length of the current contribution, we refer to the first part (Arnqvist and Thornhill, 1997) for a more detailed account than is given here of the study organism and its genital morphology, as well as the materials and methods of morphometrics, including the principals and methods of the multivariate shape analysis used (see below).

### *Mating frequency in the field*

To study the relationship between morphology and mating frequency, a natural monomorphic apterous population of *Gerris incognitus* from Bosque del Apache Wildlife Refuge, New Mexico, USA, was transferred on May 2–3, 1994, to a semi-natural experimental pond (area 28.3 m<sup>2</sup>) located at the University of New Mexico campus. Repeated sampling over a 24 h period revealed that >95% of the individuals in the natural source population were captured (Krebs, 1989). The source population (pond) was chosen to correspond closely with the environmental conditions in the experimental pond, according to habitat characteristics (similar size, shape, microhabitat structure, water depth, predator fauna and degree of shading). All individuals (males  $N=91$ , females  $N=82$ ) were marked individually with enamel paint, and released into the experimental pond. Following a two-day acclimation period, during which food (frozen *Drosophila* fruitflies) was added to the pond, the population was scanned for mating pairs once every 20 minutes from 8 AM to 6 PM during five consecutive days (corresponding to approximately 20% of the life span). Mating pairs were captured, checked for individual identity, and carefully released at the site of capture. Thus, the experiment provided longitudinal data on individual mating frequencies. Previous work on water striders has shown

that relatively short term estimates of sexual performance, even cross-sectional data, represent reliable estimates of more long term-measures (Preziosi and Fairbairn, 1996). All individuals were captured at the end of the experiment and preserved for subsequent morphometric analyses (see Arnqvist and Thornhill, 1997).

Our study does not provide data on relative post-mating fertilization success among males (cf. Eberhard, 1996). While non-random fertilization success certainly could dilute the relationship between mating frequency and reproductive fitness, previous studies of sperm utilization patterns in water striders have demonstrated high and relatively invariable degrees of last male sperm precedence (Arnqvist, 1988; Rubenstein, 1989).

#### *Fecundity, fertility and offspring viability in the laboratory*

To assess the relationship between morphology and fecundity, fertility and offspring growth/survival, a second set of adult *G. incognitus* were captured from a natural population neighbouring the one mentioned above, and brought into the laboratory. The sexes were kept separately during a female sperm depletion period, after which males and females were randomly introduced in pairs ( $N=61$ ) into containers provided with a styrox floater to serve as resting and oviposition substrate. The fecundity (number of eggs laid) and fertility (hatching rate of eggs) of each pair was monitored over a period of ten days, after which all individuals were preserved (frozen) for subsequent morphometric analyses. From each pair, twelve randomly selected full-sibling offspring were reared individually in plastic containers at two different food levels; a limited but sufficient food supply versus definite food shortage ( $N=6$  offspring from each pair and treatment). The survival until adulthood and the duration of the larval stage of each of the twelve offspring were recorded. For further details on methods, see Arnqvist and Thornhill (1997).

For each male-female pair, this experiment yielded data on four different fitness components; i) number of eggs produced over ten days, ii) the hatching rate of these eggs and iii) the subsequent survival rate of offspring. Since offspring growth rate depended on sex as well as food treatment, data on iv) average offspring development time for each pair was acquired by taking these factors into account, by calculating average residual offspring development time for each pair; residuals being generated in an analysis of variance model including sex, food availability and their interaction (Arnqvist and Thornhill, 1997).

#### *Geometric morphometrics*

Besides using a traditional linear approach to measure linear lengths, we applied landmark based geometric morphometrics to quantify complex variation in shape of body and genitalia (Bookstein, 1991, 1996; Rohlf, 1993; Rohlf and Marcus, 1993; Marcus et al., 1996). Two-dimensional digital morphometric landmark maps of the body and genitalia of water striders were attained by placing a digitizing tablet

(Summasketch<sup>®</sup> III) under a side mounted camera lucida attached to a dissecting microscope (Wild<sup>®</sup> M5). Geometric morphometrics constitute a novel multivariate approach to the study of complex morphologies, and it defines geometric shape variation as variance in the relative position of multiple landmarks. The method used here, thin-plate spline relative warp analysis, 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 displacement vector loadings on landmarks relative to the average configuration of landmarks. Our geometric morphometric analysis generated a set of multivariate shape dimensions (relative warps and uniform shape components; see Appendix), as well as an integrative measure of body size (the centroid size). The analysis was performed separately on the bilateral halves of the body and genitalia, which allowed us to also estimate fluctuating asymmetry in shape of these structures as the difference in shape score between the left and right sides. See the references above for general information on geometric morphometrics, and Arnqvist and Thornhill (1997) for a more detailed account of its application in the current study.

### *Statistical analyses*

When relating phenotypic variance to individual performance, many different statistical approaches may be adopted (Manly, 1985; Endler, 1986). The method applied here was introduced by Lande and Arnold (1983), and allows measurement of phenotypic selection acting on several traits simultaneously. It involves estimating univariate selection differentials and multivariate selection gradients by multivariate regression analysis (see below).

All measures of morphology were tested for normality by Kolmogorov-Smirnov tests. Measures of fluctuating asymmetry were transformed as  $x' = \log(1 + 10^c x)$ , where  $c$  is a constant ( $> 1$ ) specific for each variable chosen to minimize the sum of the absolute values of skewness and kurtosis of the distribution of the variable (Sokal and Rohlf, 1981; Berry, 1987; Swaddle et al., 1994; Arnqvist and Thornhill, 1997), and all other independent variables failing to meet distributional assumptions were transformed to comply with normality (log or square root transformations of a few variables). Model correctness was assessed for all models by visual inspection of plots of studentized residuals (variance functions and extreme residual values). In a few cases, cells with extreme residuals were detected, in which case models were reanalyzed excluding these observations. In no case, however, were these reanalyses qualitatively different from the analyses based on the full data set.

The fitness components used as dependent variable in our analyses were similarly tested for compliance with the normality assumption. Measures of hatching rate of eggs and survival rate of offspring were arcsine transformed to meet the assumption,

whereas no transformation was necessary for data on fecundity and development time of offspring (see below for mating frequency data).

All morphological traits were standardized prior to analysis. For each fitness component, individual measures of absolute fitness were transformed to relative fitness with a mean value of one. All selection coefficients reported are thus standardized selection coefficients (Lande and Arnold, 1983; Endler, 1986).

The dataset analyzed in the current paper consists of 48 phenotypic traits (see Appendix) and five different components of fitness. If all potential selection coefficients (differentials and gradients) were to be analyzed separately from this data set, two different problems would occur. First, parameter estimation of the multivariate selection gradients would be suspect due to multicollinearity (Belsley et al., 1980; Mitchell-Olds and Shaw, 1987; Arnqvist, 1992). Second, statistical inference would be suspect due to a possible increase in type I error rates (Holm, 1979; Cohen, 1988; Rice, 1989). Two different strategies may be adopted to avoid both of these problems (Endler, 1986). One is to reduce the number of dependent variables by principal component analysis or other multivariate methods (Lande and Arnold, 1983; Anholt, 1991; Schluter and Nychka, 1994). These methods are efficient at reducing the dimensionality, and thus the number of statistical tests, of the analysis, but results can be difficult to interpret biologically since multivariate composite variables (e.g. principal components) represent linear combinations of phenotypic traits. Another strategy to dramatically reduce the number of tests is to analyze groups of traits, by assessing the additional contribution of groups of functionally related morphological traits over and above that made by other traits, in sets of hierarchical models predicting relative fitness (Manly, 1985; Kleinbaum et al., 1988). Since our goal is to test *a priori* hypotheses about correlations between fitness components and genitalic versus non-genitalic morphological traits, we chose the latter strategy. By doing so, we control within-study type I error rate by reducing the number of primary statistical inferences by over 90 percent, while we simultaneously avoid deflating the statistical power which would lead to an increase in within-study type II error rate (Cohen, 1988). Nevertheless, we also report significances adjusted within-fitness-components by the sequential Bonferroni technique (Holm, 1979), for comparative purposes (see Tabs. 1 and 2).

Morphological traits were arranged into groups of related traits ( $N = 2-8$  in each group), each describing a unique component of overall morphology (see Appendix). Our primary statistical inference was made by analyzing the effects on fitness of each of these components of morphology in hierarchical regression models. General body size is known to correlate to various degrees with several different fitness components in water striders, such as mating frequency, fecundity, fertility and growth rate (Fairbairn, 1988; Arnqvist, 1997a; Arnqvist et al., 1996; Klingenberg and Spence, 1997; Preziosi et al., 1996; Rowe and Arnqvist, 1996), and to correlate phenotypically with many other traits (Arnqvist, 1992; Klingenberg and Zimmermann, 1992; Arnqvist and Thornhill, 1997). Thus, the contribution of a component of morphology (a trait group) to relative fitness was assessed by adding the variable group to a reduced model including only body size (centroid size of body). Multiple

**Table 1.** The effects of morphology on relative fitness for a suite of fitness components in female water striders, estimated in the field (A) and in the laboratory (B–E). Results given represent hierarchical log-likelihood tests and multiple partial  $F$ -tests of improvement in statistical fit of the addition of groups of morphological variables relative to a reduced model including only body size.

Trait/trait group	A. Mating frequency	B. Fecundity (# of eggs)	C. Hatching rate of eggs	D. Offspring survival rate	E. Offspring development time (days)
Average ( $\pm$ SD)	1.37 ( $\pm$ 1.47)	53.51 ( $\pm$ 16.31)	0.74 ( $\pm$ 0.21)	0.47 ( $\pm$ 0.18)	33.93 ( $\pm$ 3.02)
Body size	LR <sub>1</sub> <sup>a</sup> = 2.53 $P$ = 0.112	<b><math>F_{1,58} = 10.83</math></b> <b><math>P = 0.002^b</math></b>	<b><math>F_{1,58} = 6.73</math></b> <b><math>P = 0.012</math></b>	$F_{1,58} = 0.09$ $P = 0.770$	$F_{1,56} = 0.22$ $P = 0.64$
Body shape (8 variables)	LR <sub>8</sub> = 13.47 $P = 0.097$	$F_{8,50} = 0.89$ $P = 0.532$	$F_{8,50} = 1.55$ $P = 0.164$	$F_{8,50} = 1.13$ $P = 0.360$	$F_{8,48} = 1.81$ $P = 0.098$
Length of appendages (7 variables)	<b>LR<sub>7</sub> = 14.32</b> <b><math>P = 0.046</math></b>	$F_{7,51} = 0.84$ $P = 0.560$	$F_{7,51} = 0.29$ $P = 0.955$	$F_{7,51} = 1.07$ $P = 0.396$	$F_{7,49} = 1.86$ $P = 0.097$
FA in body shape (5 variables)	LR <sub>5</sub> = 9.55 $P = 0.089$	$F_{5,53} = 0.33$ $P = 0.893$	$F_{5,53} = 1.49$ $P = 0.209$	$F_{5,53} = 0.73$ $P = 0.604$	$F_{5,51} = 0.77$ $P = 0.576$
FA in length of appendages (5 variables)	LR <sub>5</sub> = 4.21 $P = 0.520$	$F_{5,53} = 1.69$ $P = 0.153$	$F_{5,53} = 0.78$ $P = 0.569$	<b><math>F_{5,53} = 2.93</math></b> <b><math>P = 0.021</math></b>	$F_{5,51} = 0.50$ $P = 0.775$

<sup>a</sup> LR represents the likelihood ratio test statistic, i.e., the deviance difference between hierarchical Poisson regression models.

<sup>b</sup> Test significant at  $\alpha = 0.05$  after further within-fitness-component adjustment with the sequential Bonferroni method (Holm, 1979).



**Table 2.** The effects of morphology on relative fitness for a suite of fitness components in male water striders, estimated in the field (A) and in the laboratory (B–E). Results given represent hierarchical log-likelihood tests and multiple partial  $F$ -tests of improvement in statistical fit of the addition of groups of morphological variables relative to a reduced model including only body size.

Trait/trait group	A. Mating frequency	B. Fecundity (# of eggs) <sup>b</sup>	C. Hatching rate of eggs <sup>b</sup>	D. Offspring survival rate	E. Offspring development time (days)
Average ( $\pm$ SD)	1.38 ( $\pm$ 1.29)	53.51 ( $\pm$ 16.31)	0.74 ( $\pm$ 0.21)	0.47 ( $\pm$ 0.18)	33.93 ( $\pm$ 3.02)
Body size	<b>LR<sub>1</sub><sup>a</sup> = 9.31</b> <b>P = 0.002<sup>c</sup></b>	<b>F<sub>1,57</sub> = 6.10<sup>b</sup></b> <b>P = 0.017</b>	F <sub>1,57</sub> = 0.07 <sup>b</sup> P = 0.792	F <sub>1,58</sub> = 0.02 P = 0.888	F <sub>1,56</sub> = 0.51 P = 0.478
Body shape (8 variables)	LR <sub>8</sub> = 6.10 P = 0.636	F <sub>8,49</sub> = 0.71 P = 0.681	F <sub>8,49</sub> = 2.06 P = 0.058	F <sub>8,50</sub> = 0.92 P = 0.508	F <sub>8,48</sub> = 1.42 P = 0.213
Length of appendages (7 variables)	LR <sub>7</sub> = 11.15 P = 0.132	F <sub>7,50</sub> = 0.41 P = 0.892	<b>F<sub>7,50</sub> = 3.14</b> <b>P = 0.008<sup>c</sup></b>	F <sub>7,51</sub> = 1.05 P = 0.409	F <sub>7,49</sub> = 1.26 P = 0.290
FA in body shape (5 variables)	LR <sub>5</sub> = 3.21 P = 0.668	F <sub>5,52</sub> = 0.28 P = 0.922	F <sub>5,52</sub> = 0.31 P = 0.905	F <sub>5,53</sub> = 1.17 P = 0.336	F <sub>5,51</sub> = 0.36 P = 0.873
FA in length of appendages (5 variables)	LR <sub>5</sub> = 5.69 P = 0.338	F <sub>5,52</sub> = 2.24 P = 0.064	F <sub>5,52</sub> = 1.12 P = 0.361	F <sub>5,53</sub> = 0.78 P = 0.569	F <sub>5,51</sub> = 0.45 P = 0.811
Size of genitalia (8 variables)	<b>LR<sub>8</sub> = 24.09</b> <b>P = 0.002<sup>c</sup></b>	F <sub>8,49</sub> = 1.62 P = 0.143	F <sub>8,49</sub> = 0.52 P = 0.836	F <sub>8,50</sub> = 0.58 P = 0.789	F <sub>8,48</sub> = 0.61 P = 0.765
Shape of genitalia (8 variables)	LR <sub>8</sub> = 11.08 P = 0.197	F <sub>8,49</sub> = 1.46 P = 0.196	F <sub>8,49</sub> = 0.58 P = 0.789	F <sub>8,50</sub> = 0.67 P = 0.715	F <sub>8,48</sub> = 0.53 P = 0.828
FA in size of genitalia (2 variables)	LR <sub>2</sub> = 2.26 P = 0.323	F <sub>2,55</sub> = 0.48 P = 0.621	F <sub>2,55</sub> = 1.08 P = 0.347	F <sub>2,56</sub> = 0.27 P = 0.764	F <sub>2,54</sub> = 0.86 P = 0.429
FA in shape of genitalia (4 variables)	LR <sub>4</sub> = 6.35 P = 0.174	F <sub>4,53</sub> = 0.68 P = 0.609	F <sub>4,53</sub> = 1.31 P = 0.278	F <sub>4,54</sub> = 0.55 P = 0.700	F <sub>4,52</sub> = 0.86 P = 0.494

<sup>a</sup> LR represents the likelihood ratio test statistic, i.e., the deviance difference between hierarchical Poisson regression models.

<sup>b</sup> Reduced model including both male and female body size, to control for maternal effects (see Tab. 1).

<sup>c</sup> Test significant at  $\alpha = 0.05$  after further within-fitness-component adjustment with the sequential Bonferroni method (Holm, 1979).

partial  $F$ -tests and, in the case of mating frequency, log-likelihood ratio tests were used to test whether improvement in model fit by addition of a variable group (traits  $z_{ij}$ ) over and above that made by body size ( $z$ ) alone was statistically significant, based on the multiple regression model  $w = a + \beta z + \sum \beta_{ij} z_{ij}$  where  $w$  represent relative fitness. These form our primary tests. If a trait group was found to be under phenotypic directional selection as indicated by the hierarchical tests, a conventional traitwise analysis of standardized linear selection differentials ( $s'$ ) and gradients ( $\beta'$ ) was performed to further characterize the pattern of selection (Lande and Arnold, 1983). An exception was sexual selection on shape of male genitalia (see Results below) where a traitwise analysis was performed despite a non-significant hierarchical test, because (a) it made the interpretation of sexual selection on genital size more lucid and (b) details on sexual selection on genitalia was of cardinal *a priori* interest. Explicit hypotheses of stabilizing selection were tested i) by addition of quadratic terms to the model above,  $w = a + \beta z + \sum \beta_{ij} z_{ij} + \sum \gamma_{ij} z_{ij}^2$ , where  $\gamma_{ij}$  represent non-linear selection gradients (Phillips and Arnold, 1989), and ii) by inspection of fitness functions estimated by nonparametric regression techniques (Schluter, 1998; Schluter and Nychka, 1994).

For data on fecundity, fertility, growth rate and survival of offspring, conventional least-squares regression were used for statistical analyses. Number of matings, however, is Poisson distributed, and we thus used Poisson regression models (generalized linear models) in all models predicting mating frequency. This method is appropriate for estimating regression models where the dependent variable is Poisson distributed (Hinde, 1982; McCullagh and Nelder, 1983; Mitchell-Olds and Shaw, 1987; Kleinbaum et al., 1988; Aitkin et al., 1989).

To test hypotheses about assortative mating, data on individual matings were compiled from the field experiment. As a measure of assortative mating, male morphology was correlated with female morphology. In the cases where a given male mated with more than one female, the average phenotypic value of his mates was correlated with male value. Statistical analyses reported in this paper were performed with GLIM (Hinde, 1982; Aitkin, et al., 1989) (Poisson regression analyses), GLMS (Schluter, 1988) (cubic spline regression), PP (Schluter and Nychka, 1994) (projection pursuit regression) and SYSTAT (1992) (other procedures). For multivariate shape analyses, we used the software packages TPSRW (Rohlf, 1993) and GRF-ND (Slice, 1994) (see Arnqvist and Thornhill, 1997, for details on morphometric methods).

## Results

The results of the hierarchical tests of contributions of different components of morphology to different components of fitness are summarized in Table 1 (females) and Table 2 (males).

*Morphology and relative fitness in females*

Female body size was found to be significantly related to both the number of eggs produced and the hatching rate of these eggs (see Tab. 1). Both fecundity and fertility selection were positive ( $s'$  [SE] = 0.11 [.03] for fecundity and  $s'$  [SE] = 0.12 [.05] for hatching rate).

Length of appendages in females was significantly related to mating frequency (Tab. 1). A multivariate analysis of the impact of these variables (Tab. 3) revealed that the detected effect was primarily due to a strong negative direct effect of the length of female abdominal spines on mating frequency, so that females with long abdominal spines mated less frequently (Fig. 1).

Fluctuating asymmetry in female length of appendages collectively related to offspring survival (Tab. 1). This resulted wholly from direct and independent effects of fluctuating asymmetry in length of all three legs (Tab. 4), so that females with more asymmetric legs produced less viable offspring, i.e. offspring with a lower survival rate.

The number of eggs produced by females (fecundity) did not correlate with offspring survival rate or offspring development time ( $r_p = 0.14$  and  $0.16$  respectively,  $P > 0.25$  in both cases). In contrast, fecundity did correlate with the hatching rate of eggs ( $r_p = 0.31$ ,  $P = 0.016$ ). This relationship was, however, due to an indirect correlation with size and was rendered non-significant when controlling for female body size (partial correlation  $r_{yx|z} = 0.19$ ,  $P > 0.1$ ) (see above).

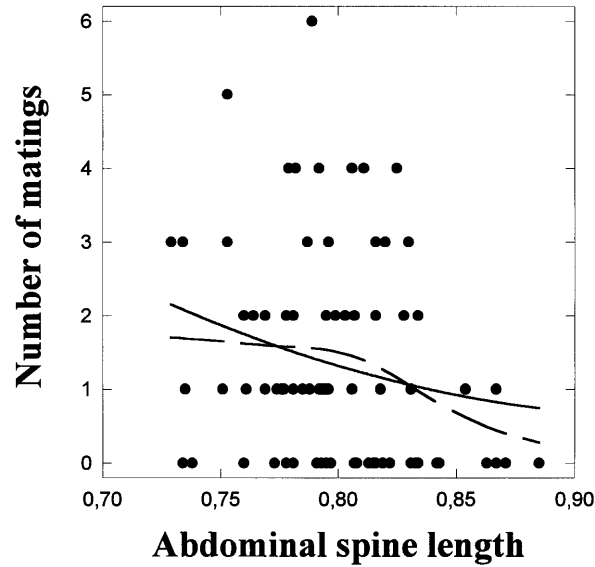
*Morphology and relative fitness in males*

There was strong positive sexual selection for male body size, as revealed by a significant relationship between body size and male mating frequency ( $s'$  [SE] = 0.29 [.10]) (Tab. 2 and Tab. 5). Further, after removing the effect of maternal body size

**Table 3.** Standardized directional selection differentials ( $s'$ ) and standardized directional selection gradients ( $\beta'$ ) for the impact of morphology of appendages on mating frequency in females.

	$s'$ [SE]	$P$	$\beta'$ [SE]	$P$
Body size (centroid size)	-0.16 [.10]	0.116	0.05 [.13]	0.713
Abdominal spine length	<b>-0.24 [.10]</b>	<b>0.021</b>	<b>-0.27 [.12]</b>	<b>0.027</b>
Distance between tips of abd. spines	-0.01 [.10]	0.910	-0.05 [.10]	0.648
Elevation angle of abdominal spines	-0.12 [.11]	0.269	-0.19 [.10]	0.073
Length of antennae	0.01 [.10]	0.928	0.34 [.18]	0.057
Length of forefemur	-0.11 [.10]	0.250	0.04 [.20]	0.823
Length of midleg	-0.19 [.10]	0.059	-0.19 [.25]	0.459
Length of hindleg	-0.21 [.11]	0.055	-0.30 [.26]	0.247

All parameter estimation (selection coefficients) and statistical inference based on uni- and multivariate Poisson regression analyses. Assessment of the overall fit of the multivariate model used for estimation of selection gradients made by analysis of deviance;  $LR_g = 16.85$ ,  $P < 0.05$ .



**Fig. 1.** The relationship between the length of the female abdominal spines and mating frequency. Fitness functions represent the fitted Poisson regression model (solid line) and a nonparametric cubic spline regression function (dashed line) (Schluter, 1988).

on fecundity (see above), male body size was negatively related to the number of eggs produced by females in pairs ( $s'$  [SE] =  $-0.10$  [.04]) (Tab. 2); thus, females laid more eggs when mated with small males.

Male genitalic morphology influenced relative male mating success (Tab. 2). Univariate analyses (selection differentials) showed significant sexual selection on several components of the configuration of the sclerotized armature of the vesical capsule (an intromittent structure). This was true for measures of size of the vesical sclerites (Tab. 5) as well as for measures of shape of the sclerites (Tab. 6). Selection

**Table 4.** Standardized directional selection differentials ( $s'$ ) and standardized directional selection gradients ( $\beta'$ ) for the effects on offspring survival of maternal fluctuating asymmetry (FA) in length of appendages.

	$s'$ [SE]	$P$	$\beta'$ [SE]	$P$
Body size (centroid size)	0.01 [.05]	0.770	0.03 [.05]	0.550
FA in abdominal spine length	0.00 [.05]	0.971	0.02 [.05]	0.617
FA in length of antennae	0.03 [.05]	0.513	0.03 [.05]	0.492
FA in length of forefemora	$-0.08$ [.05]	0.104	$-0.13$ [.05]	<b>0.011</b>
FA in length of midlegs	$-0.07$ [.05]	0.152	$-0.10$ [.05]	<b>0.047</b>
FA in length of hindlegs	0.02 [.05]	0.661	$-0.13$ [.05]	<b>0.009</b>

All parameter estimation (selection coefficients) and statistical inference based on uni- and multivariate conventional least-squares regression analyses. Assessment of the overall fit of the multivariate model used for estimation of selection gradients made by analysis of  $F_6 = 2.46$ ,  $P < 0.05$ .

**Table 5.** Standardized directional selection differentials ( $s'$ ) and standardized directional selection gradients ( $\beta'$ ) for the impact of size of male genitalia on relative mating success. Given is also each trait's contribution coefficient to the first single multivariate direction ( $a_1$ ) best explaining variation in male mating success, as revealed by a nonparametric projection pursuit regression analysis (see Fig. 3) (Schluter and Nychka, 1994).

	$s'$ [SE]	$P$	$\beta'$ [SE]	$P$	$a_1$ [SE] <sup>a</sup>
Body size (centroid size)	<b>0.29 [0.10]</b>	<b>0.004</b>	<b>0.52 [0.13]</b>	<b>&lt;0.001</b>	<b>0.72 [0.11]<sup>b</sup></b>
Length of 1:st genital segment	-0.07 [0.09]	0.425	-0.33 [0.11]	<b>0.004</b>	-0.30 [0.13] <sup>b</sup>
Length of proctiger	0.00 [0.09]	0.999	0.01 [0.12]	0.948	-0.11 [0.17]
Length of phallosome	0.01 [0.09]	0.939	-0.01 [0.12]	0.931	-0.04 [0.17]
Length of lateral vesical sclerites	0.00 [0.09]	0.981	0.25 [0.17]	0.135	0.13 [0.20]
Distance between lateral ves. sclerites	<b>0.22 [0.10]</b>	<b>0.031</b>	0.18 [0.11]	0.103	0.16 [0.18]
Length of ventral vesical sclerite	-0.23 [0.09]	<b>0.014</b>	0.01 [0.20]	0.972	-0.31 [0.19]
Length of dorsal vesical sclerite	0.00 [0.09]	0.962	0.36 [0.22]	0.103	0.28 [0.21]
Size of vesical capsule (centroid size)	-0.11 [0.09]	0.226	-0.71 [0.38]	0.065	-0.39 [0.22]

Parameter estimation and statistical inference of selection coefficients based on uni- and multivariate Poisson regression analyses. Assessment of the overall fit of the multivariate model used for estimation of selection gradients made by analysis of deviance; LR<sub>9</sub> = 33.39,  $P < 0.001$ .

<sup>a</sup> Standard errors based on 100 bootstrap replicates (smoothing parameter in  $[\lambda] = 3.0$ ).

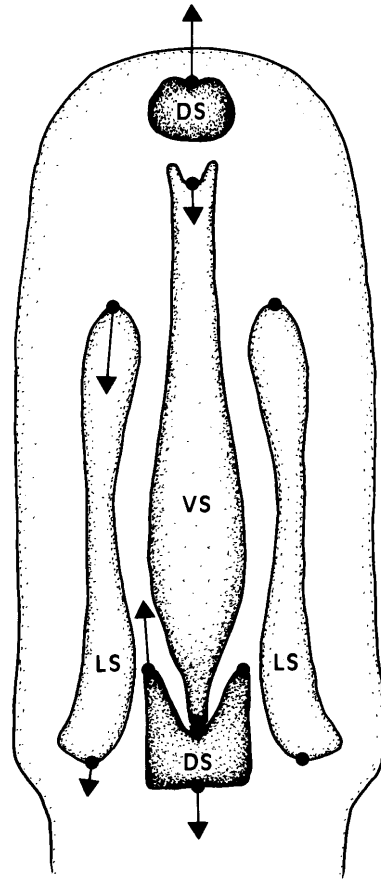
<sup>b</sup> Number of bootstrap replicates of different sign than the original estimate of coefficient  $\leq 5\%$ .

favoured a short ventral sclerite, and a wide distance between lateral sclerites (Tab. 5), suggesting univariate sexual selection for a relatively wide and short vesical capsule. This interpretation is confirmed by significant selection on the uniform shape component #2 of the vesical armature (Tab. 6). This shape component primarily describes a wide, and to some extent a short, collective configuration of vesical sclerites (phenotypic correlation between uniform shape component #2 and [a] the distance between lateral sclerites  $r_p = 0.903$ , [b] ventral sclerite length

**Table 6.** Standardized directional selection differentials ( $s'$ ) and standardized directional selection gradients ( $\beta'$ ) for the impact of multivariate shape of male genitalia (the vesical capsule) on relative mating success.

	$s'$ [SE]	$P$	$\beta'$ [SE]	$P$
Body size (centroid size)	<b>0.29 [0.10]</b>	<b>0.004</b>	<b>0.31 [0.10]</b>	<b>0.004</b>
Relative warp #1 score	0.03 [0.09]	0.769	-0.02 [0.12]	0.833
Relative warp #2 score	0.02 [0.09]	0.846	0.03 [0.26]	0.914
Relative warp #3 score	0.04 [0.09]	0.700	-0.07 [0.11]	0.489
Relative warp #4 score	<b>0.21 [0.09]</b>	<b>0.027</b>	0.23 [0.19]	0.232
Relative warp #5 score	-0.13 [0.09]	0.147	-0.01 [0.23]	0.955
Relative warp #6 score	-0.10 [0.09]	0.291	-0.13 [0.32]	0.694
Uniform shape component #1 score	-0.17 [0.09]	0.060	0.08 [0.19]	0.670
Uniform shape component #2 score	<b>0.25 [0.10]</b>	<b>0.015</b>	0.20 [0.11]	0.063

All parameter estimation (selection coefficients) and statistical inference based on uni- and multivariate Poisson regression analyses. Assessment of the overall fit of the multivariate model used for estimation of selection gradients made by analysis of deviance; LR<sub>9</sub> = 20.39,  $P < 0.025$ .



**Fig. 2.** The sclerites of the genital capsule, the vesica, in ventral view with the relative loadings of relative warp #4 (positive relative displacement of landmarks illustrated) superimposed on the average configuration of landmarks (the consensus configuration of sclerites). This shape component was found to be related to relative male mating success, and it parameterizes simultaneous variation in the length of the U-shaped dorsal sclerite (DS), the relative length of the proximal horns of the dorsal sclerite, the position of the distal tip of the ventral sclerite (VS), and the length and relative position of the lateral sclerites (LS). Filled circles represent the landmarks used for multivariate shape analysis. Loadings of relative warp #4 only indicated for one of the bilateral sides of the vesica.

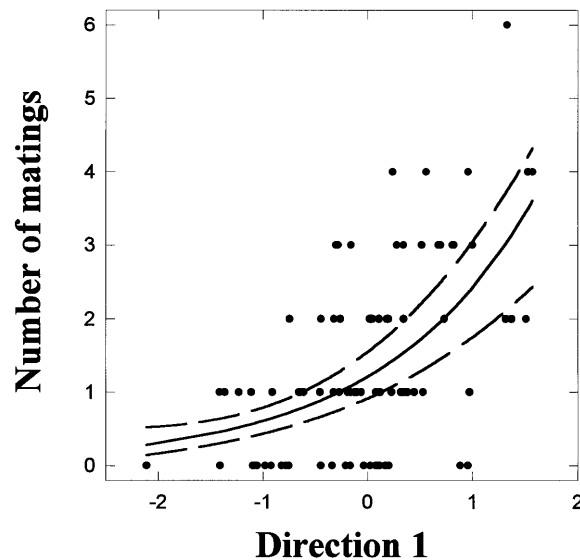
$r_p = -0.245$ ). There was also selection on more complex components of shape of the vesical sclerites. The significant positive selection differential for relative warp #4 (Tab. 6) again indicates selection for a short ventral sclerite, a relatively drawn out dorsal sclerite with long ‘‘horns’’ and to some extent for short lateral sclerites (see Fig. 2).

However, all of the detected sexual selection on the sclerotized armature of the vesical capsule was apparently due to indirect selection on phenotypically correlated traits: none of the selection gradients for morphology of vesical sclerites remained

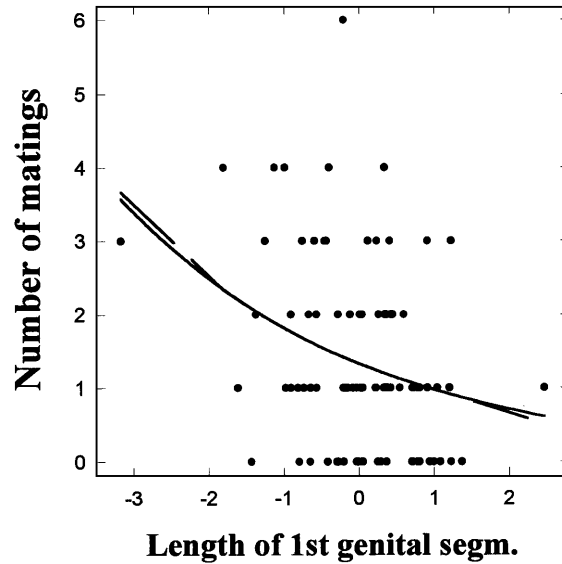
significant in the multivariate analyses (Tab. 5 and Tab. 6). These analyses all suggest that the targets of sexual selection was body size and the length of the 1st genital segment (a non-intromittent structure). There were strong direct selection on both of these phenotypically correlated ( $r_p = 0.569$ ) traits, acting in opposite directions (Tab. 5). A nonparametric projection pursuit regression analysis (Schluter and Nychka, 1994) confirmed the result of the Poisson regression analysis, and identified body size and the length of the 1st genital segment as most important traits in explaining variation in male mating frequency (Tab. 5 and Fig. 3). Thus, sexual selection favored large males with a relatively short 1st genital segment (Fig. 4), and the detected effects on the morphology of vesical sclerites were apparently due to indirect selection.

We found no evidence of stabilizing sexual selection on genitalic morphology. Expansion of Poisson regression models of male mating frequency to include also quadratic terms of genital traits (see materials and methods) did not improve model fit either for size of genitalia ( $LR_8 = 10.73$ ,  $P = 0.217$ ) or shape of genitalia ( $LR_8 = 4.98$ ,  $P = 0.759$ ). Further, the projection pursuit regression analysis did not indicate any intermediate fitness optima or convex trend in the fitness function (Fig. 3).

Finally, paternal leg length was significantly related to the hatching rate of eggs (Tab. 2). An analysis of selection differentials and gradients revealed that eggs fathered by males with relatively long legs exhibited a higher hatching success, and



**Fig. 3.** Male mating frequency as a function of genital morphology and body size. Direction 1 is a linear combination ( $a_1$ ) of 9 standardized traits (see Tab. 5 for relative coefficients of contribution for each trait), estimated to be most closely related to relative mating success. Fitness function represents the nonparametric cubic spline regression function ( $\pm 1$  SE) of predicted mating frequency (Schluter and Nychka, 1994).



**Fig. 4.** The relationship between the relative length of 1st genital segment, measured as the residual length after accounting for body size, and male mating frequency. Fitness functions represent the fitted Poisson regression model (solid line) and a nonparametric cubic spline regression function (dashed line) (Schluter, 1988).

that this effect seemed to be due primarily to direct selection on fore femur length (Tab. 7).

#### *Assortative mating*

We found no evidence of assortative mating by body size in *G. incognitus* (measured as the centroid size of body) ( $N = 58$  in all cases below,  $r = 0.014$ ,  $P > 0.5$ ) (cf. Arnqvist et al., 1996). To assess any potential positive assortative mating by genitalic traits, as predicted under the lock-and-key hypothesis, we correlated the eight different measures of size of male genitalia with i) female body size, ii) external width of female genitalia and iii) the length of female abdominal spines. The latter, located just in front of the genitalia, is known to represent a female counteradaptation to male harassment, that functions to increase the efficiency by which unwanted suitors are repelled (Arnqvist and Rowe, 1995). Neither female body size nor width of genitalia correlated with any measures of male genitalia ( $|r| < 0.126$  in all cases). In contrast, the length of female abdominal spines correlated with male genitalia in mating pairs. The correlation coefficients were negative in seven out of eight cases, and significantly so for the length of 1:st genital segment ( $r = -0.262$ ,  $P < 0.05$ ), the length of dorsal vesical sclerite ( $r = -0.290$ ,  $P < 0.05$ ) and the size of vesical capsule (centroid) ( $r = -0.296$ ,  $P < 0.05$ ). To assess



the independent effects of each of the eight different measures of size of male genitalia, we examined the partial correlation coefficients for all eight relationships. The only relationship that remained significant, when controlling for all seven non-focal measures of genitalic size, was the partial correlation between the length of female abdominal spines and the length of the non-intromittent 1st genital segment of males ( $r_{yx|z1-z7} = -0.337$ ,  $P = 0.015$ ). Hence, females with large pregenital structures (spines) tended to mate assortatively with males with a short 1st genital segment.

## Discussion

An understanding of how morphological variation in genitalia relates to components of male and female fitness is required to disentangle the various hypotheses for the evolution of animal genitalia (Eberhard, 1993, 1996; Arnqvist, 1997b). Most importantly, data on patterns of sexual selection on genital traits are needed to elucidate the mechanisms of genital evolution. The current study not only showed that size and shape of male water strider genitalia were in non-trivial way related to male reproductive success, but it also revealed a number of other interesting relationships between components of male and female morphology and fitness. In the following discussion, we compare the results of the current study with the predictions and assumptions of the different hypotheses of genitalic evolution. We also discuss the possible role of sexual conflict in evolution of sexual traits in this taxa, and the role of body size and fluctuating asymmetry.

**Table 7.** Standardized directional selection differentials ( $s'$ ) and standardized directional selection gradients ( $\beta'$ ) for the impact of morphology of appendages in males on hatching rate of eggs.

	$s'$ [SE]	$P$	$\beta'$ [SE] <sup>a</sup>	$P$
Body size (centroid size)	0.06 [0.05]	0.217	-0.17 [0.10]	0.086
Abdominal spine length	0.04 [0.05]	0.404	-0.02 [0.07]	0.755
Distance between tips of abdominal spines	-0.05 [0.05]	0.358	-0.03 [0.05]	0.570
Elevation angle of abdominal spines	0.05 [0.05]	0.301	0.10 [0.05]	0.051
Length of antennae	0.04 [0.05]	0.452	-0.12 [0.07]	0.086
Length of forefemur	<b>0.12 [0.05]</b>	<b>0.011</b>	<b>0.21 [0.10]</b>	<b>0.042</b>
Length of midleg	<b>0.11 [0.05]</b>	<b>0.021</b>	0.18 [0.12]	0.147
Length of hindleg	<b>0.14 [0.05]</b>	<b>0.004</b>	0.04 [0.13]	0.723

All parameter estimation (selection coefficients) and statistical inference based on uni- and multivariate conventional least-squares regression analyses. Assessment of the overall fit of the multivariate model used for estimation of selection gradients made by analysis of variance;  $F_9 = 3.35$ ,  $P < 0.01$ .

<sup>a</sup> Multiple regression model used for estimation of selection gradients also included female body size (centroid size), to control for maternal effects.

*Selection on genitalia – predictions and patterns**The lock-and-key-hypothesis*

The results of this study are clearly not in support of the long-standing lock-and-key hypothesis, which stipulates that an appropriate genitalic fit determines relative male mating success. Three lines of evidence are inconsistent with this hypothesis (Arnqvist, 1997b). First, and most importantly, we found no evidence that the morphology of male intromittent structures directly affected male mating/pairing success. We did find evidence for univariate selection on components of the intromittent organ, but these were all due to indirect effects of selection on correlated non-intromittent traits (see below). Thus, this study also illustrates the importance of measuring multiple traits, and of evaluating the results with multivariate methods: our results could otherwise easily have been misinterpreted as supporting the lock-and-key hypothesis. Second, we found no evidence of stabilizing selection on genitalic traits. As demonstrated both in the selection coefficient analysis and the projection pursuit analysis, sexual selection on genitalic morphology was directional. Third, we found no evidence of positive assortative mating by intromittent genitalic traits, as predicted by the lock-and-key hypothesis. However, since our measures of female genital morphology were external, rather than internal, this latter test assumes a phenotypic correlation between internal and external female genital morphology, and is thus not an ideal test.

*The sexual selection hypothesis*

The central assumption of the various versions of the sexual selection hypothesis, that selection operates via non-random fertilization success (rather than mating success) among males, is very difficult to assess (cf. Eberhard, 1996). Not only must one show a relationship between male genitalic morphology and male fertilization success, which would require determination of paternity for a large number of offspring from multiply mated females, but one must also determine the mechanisms by which variation is brought about (cryptic female choice, sexual conflict or sperm competition) (Arnqvist, 1997b). The current study does not provide the data necessary to test this specific assumption. However, our results do provide a basis for assessing other important assumptions of certain versions of the sexual selection hypothesis.

Models of genital evolution via good-genes cryptic female choice assume that genitalic traits are condition dependent and honest indicators of male genotypic quality (Johnstone, 1995). Hence, the model predicts a relationship between genitalic conformation and offspring growth and/or viability (Moore, 1994; Arnqvist, 1997b). However, in *G. incognitus* neither size or shape, nor fluctuating asymmetry, of genitalic traits in males were related to any of our three measurements of offspring quality (egg hatching rate, offspring survival rate and offspring development time). In summary, we found no indications of genitalia conveying any information about male quality, and hence we have no support for the good-genes models of female choice. Other versions of the cryptic female choice hypothesis, notably Fisherian and sensory exploitation scenarios, do not rely upon a relationship between

genitalic morphology and offspring viability (Eberhard, 1993, 1996; Arnqvist, 1997b). These models instead predict a relationship between paternal genitalic conformation and male offspring mating success, via the “sexy-sons” effect. Our study of *G. incognitus* provides some support for this possibility. We found indirect sexual selection on male genitalic morphology. Assuming that there is also direct cryptic female choice on male intromittent genitalic characters, fitness advantages experienced by males with the preferred genitalic conformation may result in higher reproductive success in male offspring fathered by these males, since genitalic morphology is heritable (Arnqvist and Thornhill, 1997).

#### *The pleiotropy hypothesis*

This hypothesis predicts indirect, but no direct, selection on genitalic traits (Arnqvist, 1997b). Our results are in good agreement with this prediction. We found selection on several components of both size and shape of the intromittent vesical sclerites, the part of male genitalia that have evolved most divergently and rapidly within this genus (Andersen, 1982, 1993). However, the detected selection was wholly due to indirect effects of selection on phenotypically correlated non-intromittent traits, as predicted by the pleiotropy hypothesis (Arnqvist, 1997b).

#### *Hypotheses of genitalic evolution – conclusions*

Our study of selection on genitalic traits in *G. incognitus* (this study), and patterns of variation and inheritance of these traits (Arnqvist and Thornhill, 1997), provides the first comprehensive single-species study in which the different hypotheses for the evolution of animal genitalia are explicitly assessed (Eberhard, 1993; Arnqvist, 1997b). Below, we discuss how our overall results agree with these hypotheses.

It is clear that our results are in stark disagreement with the assumptions and predictions of the lock-and-key hypothesis, which is still considered a valid explanation for genitalic evolution among some taxonomists despite its inconsistency with comparative data (Eberhard, 1985; Shapiro and Porter, 1989). Both the pattern of variation and inheritance and that of sexual selection failed to conform with the microevolutionary scenario envisioned by the lock-and-key hypothesis. In short, male genitalia in *G. incognitus* are not the invariant and highly canalized structures predicted under the lock-and-key hypothesis (Arnqvist and Thornhill, 1997), and we found no indications of a “key” function of the intromittent parts of male genitalia. Since our study is the first of its kind, it is at present not possible to draw any general conclusions from intraspecific data. However, if future studies will show results similar to this study (Liu et al., 1996), we will have a wholly new foundation on which to reject the lock-and-key hypothesis of genitalic evolution (cf. Shapiro and Porter, 1989).

Models of sexual selection are notoriously difficult to test empirically (Kirkpatrick and Ryan, 1991; Andersson, 1994; Johnstone, 1995; Andersson and Iwasa, 1996). However, it is clear that our results do not conform to the cryptic female choice good-genes scenario of genitalic evolution (Arnqvist, 1997b). In particular,

we failed to find any relation between paternal genital morphology and offspring viability. The findings of a lower degree of condition dependence in expression of genitalic traits compared to other traits, and that of different male genitalic genotypes not differing in their ability to cope with environmental stress during development (Arnqvist and Thornhill, 1997), are also in disagreement with predictions of good-genes models. Likewise, the cryptic female choice sensory exploitation scenario suggested by Eberhard (1993) is not well supported by our results. The findings of fairly high levels of genetic variation in genitalia, as well as that of condition dependence in expression of genitalic traits, are in disagreement with this particular hypothesis (Arnqvist and Thornhill, 1997). In contrast, our results are consistent with Fisherian, or self-reinforcing (Andersson, 1994), models of cryptic female choice (assuming that genital evolution is halted by a cost of producing complex genitalia). The patterns of phenotypic and genotypic variation of genitalia are in agreement with its predictions, and the patterns of selection can be accommodated under this hypothesis. The same is, however, true for sexual selection generated both by sexual conflict (Lloyd, 1979; Alexander et al., 1997) and by sperm competition among males.

Our results are also in agreement with the pleiotropy hypothesis (Mayr, 1963), which holds that genitalia evolve as a result of direct selection on non-intromittent traits which are genetically correlated with intromittent traits. The general level of phenotypic and genotypic variation in genitalia are in agreement with this hypothesis (Arnqvist and Thornhill, 1997). More importantly, the pleiotropy hypothesis makes two more restrictive assumptions, both of which are upheld in *G. incognitus*. First, the hypothesis predicts only indirect selection on intromittent structures of the genitalia, caused by phenotypic correlations with other traits which are under direct selection. Our findings of indirect sexual selection on the intromittent vesical sclerites, caused by phenotypic correlations with non-intromittent traits, are in agreement with this assumption. Second, a limited set of pleiotropic genes is assumed to determine both genital morphology as well as other components of morphology. The finding of fairly high genetic correlations between the shape of the intromittent vesical sclerites and general body length in this species (range = 0.19 – 0.32) suggests that pleiotropic genes are at least partially responsible for genital morphology, and hence provides some support for this assumption (Arnqvist and Thornhill, 1997).

Thus, based on the results of this study and those in our companion paper (Arnqvist and Thornhill, 1997), we conclude that our findings of microevolutionary patterns and processes in *G. incognitus* are clearly incompatible with the lock-and-key as well as the good-genes cryptic female choice hypotheses of genital evolution. Further, the female choice model suggested by Eberhard (1993), assuming no cost of genitalia, is not supported by our results (Arnqvist, 1997b). The hypotheses of evolutionary mechanisms for genitalic evolution that are in agreement with all of our observed results are i) sexual selection by either a classical Fisherian cryptic female choice process, sexual conflict or sperm competition, and ii) the pleiotropy hypothesis. Further discrimination between these hypotheses will require detailed studies of the very mechanism of selection under sexual selection, i.e. non-random

fertilization success among males based on their genitalic morphology and how such a pattern is brought about (Arnqvist, 1997b). To our knowledge, no study has yet convincingly shown non-random fertilization success among males based on their genitalia *per se* (cf. Watson, 1991a, b; LaMunyon and Eisner, 1993; Eberhard, 1996; Dickinson, 1997).

Our study of microevolutionary patterns and processes of genitalia in *G. incognitus* illustrate that thorough single species studies have a lot to contribute to the challenging task of discriminating between the different hypotheses of genitalic evolution, and that these studies gain invaluable insights into the mechanisms of genitalic evolution. We believe that a pluralistic approach, where studies such as this are conducted in concert with comparative studies and studies of functional morphology, will prove to be the most fruitful path towards a future consensus on the mechanisms of genitalic evolution in animals.

### *Sexual conflict and coevolutionary arms races between the sexes*

In certain taxa, a conflict between the sexes over control of reproductive decisions may drive the coevolution of male and female sexual structures, including at least external genitalia (Eberhard, 1985, 1996; Arnqvist, 1989, 1997a; Rowe et al., 1994; Sakaluk et al., 1995; Alexander et al., 1997). In a series of laboratory experiments, Arnqvist and Rowe (1995) showed that the highly modified (prolonged and elevated) pregenital abdominal spines of *G. incognitus* females represent female counteradaptations to male harassment, that reduce the cost to females of avoiding superfluous and costly matings. The results of the current field study, where natural variation in spine length was studied, are in agreement with these previous experiments, where spine length was experimentally manipulated. First, the length of abdominal spines in females was directly related to female mating frequency. As predicted, females with long spines mated relatively less frequently than those with short spines. Second, Arnqvist and Rowe (1995) found an apparent non-linearity in the effect of spine length on mating frequency, in that the effect of an increase in spine length from average length seemed to reduce mating frequency slightly more than the corresponding reduction in spine length from average increased mating frequency. Again, the results of the current field study mimicked those of previous laboratory experiments (see cubic spline in Fig. 1).

Under a coevolutionary arms-race scenario, an increased spine length in females would generate directional selection for a higher grasping ability in males. Though the relationship between grasping ability and genital morphology is unclear (Arnqvist, 1997a), it seems reasonable to expect selection for large structures in males, particularly for external non-intromittent components of the genitalia that might be related to grasping/clasping females. The cylindrical 1:st genital segment, for example, is protruded and used in part to pinch females during premating struggles (cf. Arnqvist, 1992). In accordance, total genital length has been found to be under positive direct sexual selection in the water strider *Aquarius remigis* (Preziosi and Fairbairn, 1996). Surprisingly enough, two facts suggest that short, rather than

long, non-intromittent genitalia may be related to high grasping ability in *G. incognitus*. First, we found direct negative sexual selection on the length of the 1st genital segment. Second, we found negative assortative mating between female spine length and the length of the 1st genital segment in males, so that females with long spines tended to mate with males with short non-intromittent genitalia. At present, however, an unambiguous interpretation of these results is not possible since necessary detailed information on the “mechanics” and functional morphology of male external genitalia are lacking (Arnqvist, 1997a).

#### *Environmental dependency of estimates of selection*

Estimates of phenotypic selection (or more general; any relation between performance and morphology) are always specific for the environmental conditions under which observations or experiments are made (Endler, 1986; Grant and Grant, 1989; Arnqvist, 1992). In the current study, phenotypic selection was estimated in two different environments; one semi-natural and one in the laboratory. First, it must, of course, be kept in mind that our estimates of selection are no different than any other estimates: they explicitly represent the adaptive landscape in these particular environments, and the extent to which our results quantitatively reflect the relationship between performance and morphology in other environments is unknown. Second, if the environmental conditions under which observations are made are more “favourable” than in most environments, relationships between performance and morphology might be weakened and selection thus generally underestimated. Since our laboratory environment closely reflected field conditions in terms of abiotic conditions (light, temperature, etc), and since offspring were reared under sub-optimal biotic conditions (food stress), we do not see any reason to believe that our estimates are generally negatively biased.

#### *The role of body size and fluctuating asymmetry*

Body size *per se* had a suite of effects on various components of fitness in both sexes. In females, large females had higher fecundity and a higher hatching rate of their eggs. This has also been found in other water strider species (Fairbairn, 1988; Klingenberg and Spence, 1997), and probably represent direct advantages of a large body size and/or an underlying variation in overall phenotypic condition (Preziosi et al., 1996).

Since the degree of fluctuating asymmetry of bilaterally symmetrical traits to some extent represents a measure of an individual's ability to cope with environmental stress, and thus of its overall quality, we would expect fluctuating asymmetry to correlate with various fitness components (see Møller and Pomiankowski, 1993; Watson and Thornhill, 1994). However, we failed to find any general and extensive effects of fluctuating asymmetry on fitness. The only result which were consistent with expectations based on theory, was that females with relatively low

degrees of fluctuating asymmetry in leg length gave rise to offspring with a higher survival rate. Two facts suggest that this effect is due to direct phenotypic maternal effects rather than an underlying variation in genotypic quality. First, the positive effect of low fluctuating asymmetry on offspring survival was present only for females. No corresponding effects of paternal fluctuating asymmetry was observed. Second, Arnqvist and Thornhill (1997) did not find any detectable levels of genetic variation in fluctuating asymmetry of leg length in this species. Thus, it seems that maternal fluctuating asymmetry in leg length parameterizes some purely environmental component of female phenotypic condition which is transferred to offspring, perhaps as nutritional quantity/quality of the eggs, and that affects subsequent offspring survival. These results call for some caution when interpreting relations between fluctuating asymmetry and components of fitness as revealing of genetic, as opposed to phenotypic, quality (cf. Møller and Pomiankowski, 1993; Swaddle et al., 1994; Thornhill and Watson, 1994).

Large body size in males was associated with a high mating frequency. Similar patterns of sexual selection on body size has been found in a range of related species, and a large male mating advantage can be considered a typical and general pattern within this group of insects (Arnqvist et al., 1996). In contrast to some other studies (e.g., Preziosi and Fairbairn, 1996), however, sexual selection on body size in *G. incognitus* did apparently not result merely from selection on correlated characters.

Male size also affected mate fecundity and hatching rate of eggs. Females paired with small males tended to lay more eggs. Interestingly enough this intriguing pattern has also been observed in *Drosophila* (Pitnick, 1991). In both these taxa, large males have a general advantage in terms of mate acquisition, but small males tend to copulate for longer times (Pitnick, 1991; Rowe and Arnqvist, 1996). However, it is unclear whether the effect on mate fecundity is an indirect effect of male behaviour (e.g., copulation duration, amount of sperm transferred) or if females choose small males to father their eggs by varying oviposition rates (Thornhill, 1983, Eberhard, 1996). In the current study, however, males with long legs tended to father eggs with a higher hatching rate. This provides at least an indication of that large males produce offspring of superior quality, in which case active female choice of small males would seem less likely (Eberhard, 1996). In either case, the mate acquisition fitness advantage of large size in males may be balanced to some extent by a reduced number of eggs produced by each mating (cf. Pitnick, 1991).

### Acknowledgements

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. We are most greatly grateful to Stephanie Weinstein for digitizing all morphometric data. We also thank P. Watson and four anonymous reviewers 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, The Swedish Institute and The Canadian NSERC. Logistic support was received from the Department of Biology, University of New Mexico, and the Department of Zoology, University of Toronto, which is gratefully acknowledged.

## References

- Aitkin, M., D. Andersson, B. Francis and J. Hinde. 1989. *Statistical Modelling in GLIM*. Clarendon Press, Oxford.
- Alexander, R. D., D. C. Marshall and J. R. Cooley. 1997. Evolutionary perspectives on insect mating, pp 4–31. *In* J. C. Choe and B. J. Crespi (Eds.), *The Evolution of Mating Systems in Insects and Arachnids*. Cambridge University Press, Cambridge.
- Andersen, N. M. 1982. *The Semiaquatic Bugs (Hemiptera, Gerromorpha): Phylogeny, Adaptations, Biogeography and Classification*. Entomograph 3. Scandinavian Science Press, Klampenborg, Denmark.
- Andersen, N. M. 1993. Classification, phylogeny, and zoogeography of the pond skater genus *Gerris* Fabricius (Hemiptera: Gerridae). *Can. J. Zool.* 71: 2471–2508.
- Andersson, M. 1994. *Sexual Selection*. Princeton Univ. Press, Princeton, NJ.
- Andersson, M. and Y. Iwasa. 1996. Sexual selection. *Trends in Ecol. Evol.* 11: 53–58.
- Anholt, B. R. 1991. Measuring selection on a population of damselflies with a manipulated phenotype. *Evolution* 45: 1091–1106.
- Arnqvist, G. 1988. Mate guarding and sperm displacement in the water strider *Gerris lateralis* Schumm. (Heteroptera: Gerridae). *Freshwater Biol.* 19: 269–274.
- Arnqvist, G. 1989. Sexual selection in water strider: the function, nature of selection and heritability of a male grasping apparatus. *Oikos* 56: 344–350.
- Arnqvist, G. 1992. Spatial variation in selective regimes: sexual selection in the water strider, *Gerris odontogaster*. *Evolution* 46: 914–929.
- Arnqvist, G. 1997a. The evolution of water strider mating systems: causes and consequences of sexual conflicts, pp 146–163. *In* J. C. Choe and B. J. Crespi (Eds.), *The Evolution of Mating Systems in Insects and Arachnids*. Cambridge University Press, Cambridge.
- Arnqvist, G. 1997b. The evolution of animal genitalia: distinguishing between hypotheses by single species studies. *Biol. J. Linn. Soc.* 60: 365–379.
- Arnqvist, G. and L. Rowe. 1995. Sexual conflict and arms races between the sexes: a morphological adaptation for control of mating in a female insect. *Proc. R. Soc. Lond. B.* 261: 123–127.
- Arnqvist, G., L. Rowe, J. Krupa and A. Sih. 1996. Assortative mating by size: a meta-analysis of mating patterns in water striders. *Evol. Ecol.* 10: 265–284.
- Arnqvist, G. and R. Thornhill. 1997. Evolution of animal genitalia: patterns of phenotypic and genotypic variation and condition dependence of genital and non-genital morphology in a water strider (Heteroptera: Gerridae). Submitted.
- Belsley, D. A., E. Kuh and R. E. Welsch. 1980. *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley, New York.
- Berry, D. A. 1987. Logarithmic transformations in ANOVA. *Biometrics* 43: 439–456.
- Bookstein, F. L. 1991. *Morphometric Tools for Landmark Data*. Cambridge Univ. Press, Cambridge.
- Bookstein, F. L. 1996. Biometrics, biomathematics and the morphometric synthesis. *Bull. Math. Biol.* 58: 313–365.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioural Sciences*. Lawrence Erlbaum Ass., Hillsdale.
- Dickinson, J. L. 1997. Multiple mating, sperm competition, and cryptic female choice in the leaf beetles (Coleoptera: Chrysomelidae). pp 164–183. *In* J. C. Choe and B. J. Crespi (Eds.), *The Evolution of Mating Systems in Insects and Arachnids*. Cambridge University Press, Cambridge.



- Eberhard, W. G. 1985. *Sexual Selection and the Evolution of Animal Genitalia*. Harvard Univ. Press, Cambridge, Mass.
- Eberhard, W. G. 1990. Animal genitalia and female choice. *Amer. Sci.* 78: 134–141.
- Eberhard, W. G. 1993. Evaluating models of sexual selection: genitalia as a test case. *Am. Nat.* 142: 564–571.
- Eberhard, W. G. 1996. *Female Control: Sexual Selection by Cryptic Female Choice*. Princeton Univ. Press, Princeton, NJ.
- Endler, J. A. 1986. *Natural Selection in the Wild*. Princeton Univ. Press, Princeton, NJ.
- Fairbairn, D. J. 1988. Sexual selection for homogamy in the *Gerridae*: an extension of Ridley's comparative approach. *Evolution* 42: 1212–1222.
- Grant, B. R. and P. R. Grant. 1989. Natural selection in a population of Darwin's finches. *Am. Nat.* 133: 377–393.
- Hinde, J. 1982. Compound Poisson regression models, pp 109–127. *In* R. Gilchrist (Ed.), *GLIM 82: Proceedings of the International Conference on Generalised Linear Models*. Springer-Verlag, New York.
- Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6: 65–70.
- Johnstone, R. A. 1995. Sexual selection, honest advertisement and the handicap principle: reviewing the evidence. *Biol. Rev.* 70: 1–65.
- Kirkpatrick, M. and M. J. Ryan. 1991. The evolution of mating preferences and the paradox of the lek. *Nature* 350: 33–38.
- Kleinbaum, D. G., L. L. Kupper and K. E. Muller. 1988. *Applied Regression Analysis and Other Multivariate Methods*. PWS-KENT, Boston, MA.
- Klingenberg, C. P. and J. R. Spence. 1997. On the role of body size for life history evolution. *Ecol. Entomol.* 22: 55–67.
- Klingenberg, C. P. and Zimmermann, M. 1992. Static, ontogenetic, and evolutionary allometry: a multivariate comparison in nine species of water striders. *Am. Nat.* 140: 601–620.
- Krebs, C. J. 1989. *Ecological Methodology*. Harper & Row, New York.
- LaMuyon, C. W. and T. Eisner. 1993. Postcopulatory sexual selection in an arctiid moth (*Utetheisa ornatrix*). *Proc. Natl. Acad. Sci.* 90: 4689–4692.
- Lande, R. and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37: 1210–1226.
- Liu, J., J. M. Mercer, L. F. Stam, G. G. Gibson, Z. -B. Zeng and C. C. Laurie. 1996. Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritana*. *Genetics* 142: 1129–1145.
- Manly, B. F. 1985. *The Statistics of Natural Selection on Animal Populations*. Chapman and Hall, New York.
- Marcus, L. F., M. Corti, A. Loy, G. J. P. Naylor and D. Slice (Eds.). 1996. *Advances in Morphometrics*. NATO ASI Series A: Life Sciences vol. 284. Plenum Press, New York.
- Mayr, E. 1963. *Animal Species and Evolution*. Harvard Univ. Press, Cambridge, Mass.
- McCullagh, P. and J. A. Nelder. 1983. *Generalized Linear Models*. Chapman and Hall, New York.
- Mitchell-Olds, T. and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41: 1149–1161.
- Moore, A. J. 1994. Genetic evidence for the "good genes" process of sexual selection. *Behav. Ecol. Sociobiol.* 35: 235–241.
- Møller, A. P. and A. Pomiankowski. 1993. Fluctuating asymmetry and sexual selection. *Genetica* 89: 267–279.
- Philips, P. C. and S. J. Arnold. 1989. Visualizing multivariate selection. *Evolution* 43: 1209–1222.
- Pitnick, S. 1991. Male size influences mate fecundity and remating interval in *Drosophila melanogaster*. *Anim. Behav.* 41: 735–745.
- Preziosi, R. F. and D. J. Fairbairn. 1996. Sexual size dimorphism and selection in the wild water strider *Aquarius remigis*: body size, components of body size and male mating success. *J. evol. biol.* 9: 317–336.

- Preziosi, R. F., D. J. Fairbairn, D. A. Roff and J. M. Brennan. 1996. Body size and fecundity in the water strider *Aquarius remigis*: a test of Darwin's fecundity advantage hypothesis. *Oecologia* 108: 424–431.
- 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.
- Rohlf, F. J. 1993. Relative warp analysis and an example of its application to mosquito wings, pp 131–159. *In* L. F. Marcus, E. Bello and A. Garcia-Valdecasas (Eds.), *Contributions to Morphometrics*. Monografias del Museo Nacional de Ciencias Naturales, Madrid, Spain.
- Rohlf, F. J., and L. F. Marcus. 1993. A revolution in morphometrics. *Trends Ecol. Evol.* 8: 129–132.
- Rowe, L. and G. Arnqvist. 1996. Analysis of the causal components of assortative mating in water striders. *Behav. Ecol. Sociobiol.* 38: 279–286.
- Rowe, L., G. Arnqvist, A. Sih and J. J. Krupa. 1994. Sexual conflict and the evolutionary ecology of mating patterns: water striders as a model system. *Trends Ecol. Evol.* 9: 289–293.
- Rubenstein, D. I. 1989. Sperm competition in the water strider *Gerris remigis*. *Anim. Behav.* 38: 631–636.
- Sakaluk, S. K., P. J. Bangert, A.-K. Eggert, C. Gack and L. V. Swanson. 1995. The gin trap as a device facilitating coercive mating in sagebrush crickets. *Proc. R. Soc. Lond. B.* 261: 65–71.
- Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. *Evolution.* 42: 849–861.
- Schluter, D. and D. Nychka. 1994. Exploring fitness surfaces. *Am. Nat.* 143: 597–616.
- Scudder, G. G. E. 1971. Comparative morphology of insect genitalia. *Ann. Rev. Entomol.* 16: 379–406.
- Shapiro, A. M. and A. H. Porter. 1989. The lock-and-key hypothesis: evolutionary and biosystematic interpretation of insect genitalia. *Ann. Rev. Entomol.* 34: 231–245.
- Slice, D. 1994. Generalized rotational fitting of n-dimensional landmark data (computer software). State Univ. of New York at Stony Brook.
- Sokal, R. R. and F. J. Rohlf. 1981. *Biometry*. Freeman & Co., San Francisco.
- Swaddle, J. P., M. S. Witter and I. C. Cuthill. 1994. The analysis of fluctuating asymmetry. *Anim. Behav.* 48: 986–989.
- SYSTAT, 1992. SYSTAT for Windows: Statistics, Version 5 Edition. SYSTAT Inc, Evanston, Ill.
- Thornhill, R. 1983. Cryptic female choice and its implications in the scorpionfly *Harpobittacus nigriceps*. *Am. Nat.* 122: 765–788.
- Watson, P. J. 1991a. Multiple paternity as genetic bet-hedging in female Sierra dome spiders, *Linyphia litigiosa* (Linyphiidae). *Anim. Behav.* 41: 343–360.
- Watson, P. J. 1991b. Multiple paternity and first mate sperm precedence in the sierra dome spider, *Linyphia litigiosa* Keyserling (Linyphiidae). *Anim. Behav.* 41: 135–148
- Watson, P. J. and R. Thornhill. 1994. Fluctuating asymmetry and sexual selection. *Trends Ecol. Evol.* 9: 21–25.

Received 9 September 1996;

revised 25 November 1996;

accepted 6 January 1997.

## Appendix. Functional trait groups included in the analysis.

### 1. *Body shape*:

Relative warp #1 score

Relative warp #2 score

Relative warp #3 score  
Relative warp #4 score  
Relative warp #5 score  
Relative warp #6 score  
Uniform shape component #1 score  
Uniform shape component #2 score

2. *Length of appendages:*

Abdominal spine length  
Distance between tips of abdominal spines  
Elevation angle of abdominal spines  
Length of antennae  
Length of forefemur  
Length of midleg  
Length of hingleg

3. *Fluctuating asymmetry in body shape:*

FA in relative warp #1 score  
FA in relative warp #2 score  
FA in relative warp #3 score  
FA in relative warp #4 score  
FA in relative warp #5 score  
FA in relative warp #6 score

4. *Fluctuating asymmetry in length of appendages:*

FA in abdominal spine length  
FA in length of antennae  
FA in length forefemora  
FA in length midlegs  
FA in length hindlegs

5. *Size of genitalia:*

Length of 1st genital segment  
Length of proctiger  
Length of phalloteca  
Length of lateral vesical sclerites  
Distance between lateral vesical sclerites  
Length of ventral vesical sclerite  
Length of dorsal vesical sclerite  
Size of vesical capsule (centroid size)

6. *Shape of genitalia (vesical capsule):*

Relative warp #1 score  
Relative warp #2 score  
Relative warp #3 score  
Relative warp #4 score  
Relative warp #5 score  
Relative warp #6 score  
Uniform shape component #1 score  
Uniform shape component #2 score

7. *Fluctuating asymmetry in size of genitalia:*

FA in length of lateral vesical sclerites  
FA in size of vesical capsule

8. *Fluctuating asymmetry in shape of genitalia (vesical capsule):*

FA in relative warp #1 score  
FA in relative warp #2 score  
FA in relative warp #3 score  
FA in relative warp #4 score  
FA in uniform shape component #1 score