Multivariate stabilizing sexual selection and the evolution of male and female genital morphology in the red flour beetle

Clarissa House[1,5], Philip Tunstall[2], James Rapkin[2], Mathilda J. Bale[2], Matthew Gage[3], Enrique del Castillo[4] and John Hunt[1,2]

[1]School of Science and Health and Hawkesbury Institute for the Environment, Western Sydney University, Hawkesbury Campus, Locked Bag 1797, Richmond, NSW 2753, Australia.
[2]Centre for Ecology and Conservation, College of Life & Environmental Sciences, University of Exeter, Penryn Campus, Cornwall, TR10 9EZ, UK.
[3]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.
[4]Department of Industrial Engineering and Department of Statistics, 357 Leonhard Building, Pennsylvania State University, University Park, Pennsylvania State, 16802, USA.
[5]E-mail: C.House@westernsydney.edu.au

Running title: Stabilizing selection on genital morphology

KEY WORDS: Genital morphology, selection analysis, selection gradient, individual fitness surface, multivariate stabilizing selection

Male genitals are strikingly divergent in animals with internal fertilization. Most studies attempting to explain this diversity have focussed on testing the major hypotheses of genital evolution (the lock-and-key, pleiotropy and sexual selection hypotheses) and quantifying the form of selection targeting male genitals has played an important role in this endeavour. However, we currently know far less about the selection that targets female genitals, how male and female genitals interact during mating or the longer-term evolutionary dynamics of genitals. Here we show that both male and female genital morphology are under strong sexual selection during mating in the red flour beetle, *Tribolium castaneum*. However, contrary to the sexual selection hypothesis that predicts strong directional sexual selection acting on genital morphology, we found that genital size and shape were subject to multivariate stabilizing sexual selection in both sexes. Moreover, there is significant sexual selection on the covariance between the sexes for specific aspects of genital shape suggesting that male and female genital shape interact to determine the successful transfer of a spermatophore during mating. Finally, we show that the mean genital morphology of an independent sample of males and females reared on four different experimental diets were located within the 95% confidence region of the global maxima on the fitness surface for each sex indicating that genital morphology has evolved to the peak of fitness in both sexes. Collectively, our work highlights the important role that both male and female genital morphology plays in determining mating success and shows that these effects can occur independently, as well as through their interaction. Moreover, our work shows how a complex pattern of multivariate stabilizing selection can shape the long-term evolutionary dynamics of genital morphology in both sexes and cautions against the overly simplistic view that the sexual selection targeting genital morphology will always be directional in form.

Male genitals in animals with internal fertilization are widely regarded as being the most highly divergent and variable of all morphological structures, to the extent that genital morphology is often used to distinguish between closely related species that are otherwise indistinguishable (reviewed in Hosken and Stockley 2004; Simmons 2014). Not surprisingly, the proximate mechanism(s) responsible for this variation has puzzled evolutionary biologists for several decades and has been the topic of investigation in almost all animal taxa (e.g. insects: House and Simmons 2003, House et al. 2013; spiders: Foellmer 2008, Kunter et al. 2016; reptiles: King et al. 2009, Klaczko et al. 2015; fish: Langerhans et al. 2005, Booksmythe et al. 2016; birds: Brennan et al. 2010, 2017; mammals: Stockly 2002, Ramm 2007). More recently, there have been a growing number of studies showing that female genitals may often be as variable as male genitals and may also evolve as rapidly (e.g. Simmons 2014; Ah-King et al. 2014). Despite this, there is a strong under-representation of studies on female genitalia that has worsening with time (Ah-King et al. 2014), as well as a general lack of studies examining how complex interactions between the sexes can shape genital coevolution (Ah-King et al. 2014; Brennan and Prum 2015). A more detailed understanding of genital evolution therefore requires a greater focus on female genitals and how they interact with male genitals during mating (Ah-King et al. 2014; Brennan and Prum 2015).

Historical explanations for the evolution of genitalia have largely focussed on three main hypotheses: the lock-and-key, the pleiotropy and the sexual selection hypotheses (Hosken and Stockly 2004). The lock-and-key hypothesis proposes that genital divergence is the result of natural selection acting on genital morphology to prevent hybridization by ensuring that only males of the correct species are able to provide the right “key” for the female “lock”(Dufour 1844). The pleiotropy hypothesis proposes that the divergence in genital morphology is due to the pleiotropic effects of selection on other non-genital traits (Mayr 1963). Thus, genital divergence is considered a neutral process with genitals only evolving because they are genetically correlated with other non-genital traits that are the target of selection (Mayr 1963). Finally, the sexual selection hypothesis proposes that a number of processes, most notably cryptic female choice for males with genitals that are most able to stimulate them during mating, sperm competition and/or sexual conflict, all have the potential drive genital divergence (Eberhard 1985; Hosken and Stockly 2004). While there is some empirical support for the lock-and-key (e.g. Arnqvist 1997; McPeek et al. 2008; Wojcieszek and Simmons 2012a, b; Anderson and Langerhans 2015) and pleiotropy (e.g. Arnqvist et al. 1997; Arnqvist and Thornhill, 1998; Grieshop and Polak, 2014; LaVasseur-Viens and Moehring 2014) hypotheses, their general applicability has been questioned (Hosken and Stockly 2004). For example, the levels of variation observed in male and female genitals within and across populations (Eberhard 1985; Eberhard 2001) and the way that many genitals are known to functionally interact during mating (e.g. Siva-Jothy et al. 2996; Werner and Simmons 2008; Mouginot et al. 2015) challenge the lock-and-key hypothesis, while uncertainty over why pleiotropy should disproportionately influence genitals more than other traits challenges the pleiotropy hypothesis. In contrast, there appears to be widespread support for the sexual selection hypothesis and the general view that this process is a key driver of genital evolution (e.g. Arnqvist 1997; Hosken and Stockley 2004; Mendez and Cordoba-Aguilar 2004; Simmons 2014).

One important criterion that has been used to discriminate between these alternate hypotheses is the form of selection targeting male genitals (Arnqvist 1997; Hosken and Stockley 2004). According to the lock-and-key hypothesis, males with genitals that are poorly aligned with the average genital structure of females in the population are predicted to have reduced mating success, resulting in a pattern of stabilizing selection targeting male genitals (Arnqvist 1997). According to the sexual selection hypothesis, variation in male genital morphology is related to fertilization success, with males having extreme genitals being the most successful due to their stimulatory, competitive and/or coercive ability (Arnqvist 1997). Sexual selection is therefore predicted to impose strong linear (or directional) selection on male genitals (Arnqvist 1997; Hosken and Stockley 2004). In contrast, according to the pleiotropy hypothesis male genital morphology does not correlate with fitness and therefore should not experience any direct selection (Arnqvist 1997). However, if phenotypically correlated with other traits under selection, male genitals can experience indirect selection and this can take any form (Arnqvist 1997). The ability to empirically quantify the strength and form of selection acting on male genitals has been greatly enhanced by the use of multivariate selection analysis (Lande and Arnold 1983) and insects have played a key role in this endeavour (but also see Mautz et al. 2013; Devigli et al. 2015; Head et al. 2015 for other taxa). The majority of studies on insects have documented linear selection on male genitals (e.g. damselflies: Cordoba-Aguilar 1999, 2002, 2009; water strider: Arnqvist and Danielsson 1999, Danielsson and Askenmo 1999, Bertin and Fairbairn

2005; praying mantis: Holwell et al. 2010; oriental beetle: Wenninger and Averill 2006; earwig: van Lieshout 2011; van Lieshout and Elgar 2011), although stabilizing selection has also been shown to target some aspects male genital morphology as well (seed bug: Tadler 1999; Dougherty and Shuker 2016; dung beetle: Simmons et al. 2009; millipede: Wojcieszek and Simmons 2011; broad horned beetle: House et al. 2016; water strider: Bertin and Fairbairn 2005). At first glance, this finding appears to add further support for the sexual selection hypothesis but it is important to note that it is statistically far easier to detect linear than stabilizing selection (Hunt et al. 2010) and that there are also many empirical examples showing that sexual selection is not always linear in form (e.g. Lebas et al. 2003; Blows et al. 2003; Chenoweth & Blows 2005; Brooks et al. 2005; Bentsen et al. 2006; Gerhardt and Brooks 2009; Wheeler et al. 2012; Steiger et al. 2013; Oh and Shaw 2013). Unfortunately, we currently do not have similar formal estimates of selection available for female genitals.

Differences in the form of selection targeting male genitals is expected to have important consequences for how these traits evolve (Lande 1979, Lande and Arnold 1983), yet the dynamics of genital evolution has received far less empirical attention. Assuming sufficient additive genetic variance, the strong linear selection proposed in the sexual selection hypothesis is predicted to result in either a rapid increase or decrease in mean genital morphology in the population (Lande 1979). In support of this prediction, studies applying direct artificial selection to male genitals (Booksmythe et al. 2016a) and manipulating the intensity of sexual selection using experimental evolution (Simmons et al. 2009; House et al. 2013) have shown rapid changes in mean genital morphology, and comparative studies have also shown a positive relationship between the intensity of sexual selection and mean genital morphology across species (Arnqvist 1998; Stockley 2002; Ramm 2007; Rowe and Arnqvist 2011; Brindle and Opie 2016). For example, bidirectional artificial selection on mean relative gonopodium length in the mosquito fish (*Gambusia holbrooki*) resulted in a consistent divergence in the trait across replicate lines in as little as 4 generations (Booksmythe et al. 2016). The genetic correlation between male genital morphology and other traits targeted by selection that is proposed by the pleiotropy hypothesis is predicted to drive the coevolution of the two traits (Lande 1979). In the simplest case, this coevolution can occur when genital morphology is genetically correlated with other male traits (e.g. Booksmthye et al. 2016a; Hopwood et al. 2016). For example, bidirectional artificial selection on repeated mating rate in male burying beetles (*Nicrophorus vespilloides*) resulted in significant divergence in genital shape (Hopwood et al. 2016). Coevolution can also occur when male genital morphology is genetically correlated with traits in females (i.e. intersexual pleiotropy). This includes the coevolution of male and female genital morphology that has been documented within species using experimental evolution (e.g. Simmons and Garcia-Gonzalez 2011) and a across species using comparative (e.g. Brennan et al. 2007; Kuntner et al. 2009; Lupše et al. 2016), as well as the coevolution of male genital morphology with other important phenotypic traits in females (e.g. brain size, Booksmythe et al. 2016b; mating rate, Kuntner et al. 2016). Finally, the strong stabilizing selection proposed by the lock-and-key hypothesis is predicted to decrease the variance in male genital morphology around the mean of the population (Lande 1979). Moreover on a fitness surface characterised by a single peak in fitness, it is predicted that in the absence of frequency-dependent selection, the population mean will evolve uphill until it converges on the fitness peak (Simpson 1953; Lande 1976, 1979). While this remains a central prediction in evolutionary biology, surprisingly few empirical studies exist (Armbruster 1990; Benkman 1993, 2003; Schluter 2000; Brooks et al. 2005) and there are currently no studies testing this prediction for male or female genital morphology.

The red flour beetle, *Tribolium castaneum* (Coleoptera: Tenebrionidae) is a model species in the study of sexual selection (reviewed in Fedina & Lewis 2008). This species is highly polygamous and will mate every few minutes, yet as high as 55% of mating attempts do not produce viable offspring (Lewis and Iannini 1995; Bloch Qazi et al. 1996; Pai et al. 2005; Fedina and Lewis 2008). This failure has been attributed to both peri (i.e. occurring during copulation) and post-copulatory processes and these two temporal aspects of mating behaviour have been shown to influence reproductive success in different ways (Fedina and Lewis 2006; Fedina 2007; Fedina and Lewis 2007; Tyler and Tregenza 2013). Peri-copulatory processes are known to result in the outright failure of the male to successfully transfer a spermatophore to the bursa copulatrix of the female and the production of viable offspring (Tyler and Tregenza 2013). In contrast, a number post-copulatory processes (i.e. sperm storage, sperm competition, female spermathecal morphology, cryptic female choice and female remating behaviour) have been shown to bias paternity after a spermatophore has been successfully transferred (Fedina and Lewis 2006; Fedina 2007; Fedina and Lewis 2007). While there is some (albeit conflicting) evidence to suggest that this bias in paternity is associated with specific male traits, such as courtship “leg-rubbing” behaviour (e.g. Edvardsson and Arnqvist 2000, 2005; Fedina and Lewis 2006) and olfactory cues (e.g. Lewis and Austad 1994), similar traits have not been linked to the successful transfer of a spermatophore (Fedina and Lewis 2008), although attractive males are more likely to transfer a spermatophore when mating after an unattractive male (Fedina and Lewis 2007). Furthermore, it is unknown whether the genital morphology of males or females, or their interaction, influence the successful transfer of a spermatophore in *T. castaneum*, although it seems likely given this is the primary function of the genitals.

In this study we use multivariate selection analysis (Lande and Arnold 1983) to characterize the strength and form of direct linear and nonlinear sexual selection acting on male and female genital size and shape in the red flour beetle *T. castaneum.* Having quantified the sexual selection that independently targets genital morphology in each sex, next we estimate the sign and strength of correlational selection that targets the covariance between male and female genital size and shape. The standardized selection gradients from this analysis therefore measure the importance of the interaction between male and female genital morphology to the successful transfer of a spermatophore during mating. As our analysis found that multivariate stabilizing selection was the dominant form of direct selection operating on genital morphology in both sexes, we conclude by testing the key prediction that the mean genital morphology in the population will evolve to the peak of the fitness surface. To do this, we used a nonparametric bootstrapping method to estimate the global maxima (i.e. fitness peak) and its 95% confidence region for the fitness surface of each sex. We then reared an independent sample of larvae to adulthood on four different experimental diets, measured male and female genital size and shape and mapped the mean genital morphology from these samples onto the fitness surfaces for each sex. If genital morphology in the sexes have evolved to the peak of the fitness surface, we predict that the mean genital morphology of males and females from our independent samples reside within the 95% confidence region for the global maxima on their respective fitness surfaces.

Materials and Methods

STOCK POPULATIONS AND REARING PROTOCOL

A total of 6 stock populations of the widely used Georgia 1 (GA1) “wild-type” strain of *T. castaneum* were originally derived from the Beeman Lab (US Grain Marketing Production Research Centre). These populations were cultured in *ad libitum* standard medium (95% white flour and 5% bakers’ yeast) and maintained at 30^{o}C, 60% humidity and on a 16:8 hour light:dark cycle. All populations consisted of over 200 beetles, maintained with overlapping generations and free mate choice.

EXPERIMENTAL PROCEDURE

Male and female beetles used in this experiment were taken at random from the stock populations. To ensure virginity, pupae were collected from the stock populations over a two week period. A set of nested sieves were used to separate the pupae from the adults and medium. The pupae were then removed from the sieve with soft grip tweezers and their sex determined under a microscope. Each pupa was then placed into an individual cell of a unisex, square plastic transparent box (10cm^2; 25 cells per box, 2cm^2 per cell) with each cell, half-filled with medium. The boxes were checked daily and the eclosion date for emerging adults was recorded to ensure that only virgin adults aged 7-21 days were used in the mating trials (Attia and Tregenza 2004).

MATING TRIALS

Prior to mating, sexually mature adult males were marked with a blue gel pen on one of their elytra, to allow for quick recognition. Mating trials were conducted at 22±1 °C in a mating arena that consisted of 2 x 2 cm cells in a 25 cell box that was lined with paper to provide traction (Tyler and Tregenza 2012). In every trial, a female was first placed into one of the mating arena cells followed by a male. The time of male introduction and the start and end of mating was recorded. Males typically make multiple mounting attempts however we define mating as a mounting that lasted longer than 30 s as shorter mating attempts are unlikely to result in the transfer of a spermatophore (Tyler and Tregenza 2012). Following a mating longer than 30 seconds, the male was removed and frozen (n = 535). To verify whether a mating attempt had been successful or unsuccessful each female was placed in a

60ml breeding pot (sized 67 x 34mm) that contained 30ml of standard media to oviposit under the standard incubation conditions. After 7 days, each female was removed and frozen (*n* = 535). Forty days later each pot was checked for the presence or absence of offspring (now newly eclosed beetles) to verify whether mating was successful or failed. Mating pairs were classified as successfully mated if mating resulted in offspring and received a fitness score of one (*n* = 216). Mating pairs were classified as unsuccessfully mated if no offspring were produced and received a fitness score of zero (*n* = 282).

DISSECTIONS AND GEOMETRIC MORPHOMETRICS

The male genitalia were removed from the abdomen and mounted on a microscope slide in a drop of Hoyer's solution. The female genitalia were squeezed out of the body by gently pressing the abdomen and mounted in a drop of Hoyer's solution whilst still attached to the body. The genitalia are delicate and prone to damage during dissection furthermore we required both individuals from the interacting mating pair as we are interested in the covariance between the male and female structure. When damage to the genitalia occurred we removed the mating pair from the data set leaving a sample size of 498. All genitalia were placed in a consistent, longitudinal orientation and digital images were taken using a Leica DFC295 digital microscope-camera that was mounted on a dissecting Leica M125 microscope (Figure S1 and S2). Due to the complexity of the male and female genitalia, geometric morphometric (GM) analysis was used to quantify the variation in the size and shape of the outline of the male aedeagus and female vagina and supporting structures. A description of the programs used to digitize the male and female genitalia and analyse the GM data is described in Figure S1 and S2. Although our shape analysis for males and females returned a total of 19 and 36 RW scores respectively, only the first four where used as they each accounted for over 75% of the shape variation (Gutierrez 2011).

For a subset of 25 experimental males and females sampled from the parental population, we measured the repeatability of digitization of two images of the same genitalia using the R code provided in Wolak et al. (2012). The male and female genitalia both consist of thin membranous tissue so it is not possible to re-mount specimens and test the repeatability of mounting. All of our measurements of male and female genital morphology were repeatable (Males: centroid size = 0.953, 95% CIs: 0.916, 0.989; RW1 = 0.936, 95% CIs: 0.887, 0.985; RW2 = 0.777, 95% CIs: 0.619, 0.933; RW3 = 0.745, 95% CIs:

0.568, 0.921; RW4 = 0.744, 95% CIs: 0.567, 0.921; Females: centroid size = 0.993, 95% CIs: 0.988, 0.998; RW1 = 0.992, 95% CIs: 0.994, 0.999; RW2 = 0.963, 95% CIs: 0.934, 0.991; RW3 = 0.988, 95% CIs: 0.978, 0.997; RW4 = 0.940, 95% CIs: 0.895, 0.986).

STATISTICAL ANALYSIS

Characterizing Linear and Nonlinear Sexual Selection on Male and Female Genital Size and Shape

We used standard multivariate selection analysis (Lande and Arnold 1983) to evaluate the strength and form of linear and nonlinear selection acting on male and female genital size and shape. An absolute fitness score was assigned to each male in our experiment, with one being assigned to males that successfully obtained a mating and zero being assigned to males that were unsuccessful. Following Lande and Arnold (1983), this absolute fitness score was transformed to relative fitness by dividing by the mean absolute fitness of the population.

To estimate the standardized linear selection gradients ($\boldsymbol{\beta}$), a first order linear multiple regression model was fitted using CS and the first four RW scores describing the variation in male and female genital shape as the predictor variables, and relative fitness as the response variable (Lande and Arnold 1983). We then used a second order quadratic multiple regression model that included all linear, quadratic, and cross-product terms to estimate the matrix of nonlinear selection gradients (γ) that describes the curvature of the fitness surface. Quadratic regression coefficients are known to be underestimated by a factor of 0.5 using standard multiple regression analysis, so we doubled the quadratic selection gradients derived from this model (Stinchcombe et al. 2008).

As relative fitness does not conform to a normal distribution, we used a resampling procedure to assess the significance of our standardized selection gradients (Mitchell-Olds and Shaw 1987). We randomly shuffled relative fitness scores across male and female pairs in our dataset to obtain a null distribution for each selection gradient where there is no relationship between our measures of genital size and shape and relative fitness. We used a Monte Carlo simulation to determine the proportion (p) of times (out of 10,000 iterations) that each gradient pseudo-estimate was equal to or less than the original estimated gradient, and this was used to calculate a two-tailed probability value (as $2p$ if $p < 0.5$ or as

2(1 - p) if $p > 0.5$) for each selection gradient in the model (Manly 1997). We conducted separate randomization tests for the linear multiple regression model and the full quadratic model (including linear, quadratic and correlational terms).

As the strength of nonlinear selection gradients can be underestimated by interpreting the size and significance of individual γ coefficients (Blows and Brooks 2003), we explored the extent of nonlinear selection acting on male and female genital size and shape by conducting a canonical analysis of the γ matrix to locate major eigenvectors of the fitness surface in each sex (Phillips and Arnold 1989). For each sex, we used the permutation procedure outlined in Reynolds et al. (2010) to determine the strength and significance of nonlinear selection operating along the eigenvectors of γ. This procedure, however, does not estimate the strength of linear selection operating along the eigenvectors of γ and we therefore used the “double regression” method of Biasgaard and Ankenman (1996) to estimate this form of selection acting along each eigenvector. The strength of linear selection along each eigenvector ($\mathbf{m}_i$) is given by theta (θ_i) whereas the strength of nonlinear selection is given by their eigenvalue (λ_i).

We used thin-plate splines (Green and Silverman 1994) to visualize the major eigenvectors of the fitness surface extracted from the canonical rotation of the γ for males and females. We used the “*Tps*” function in the FIELDS package of R (version 2.13.0, www.r-project.org) to fit the thin-plate splines, and visualized splines as a contour-map using the value of smoothing parameter (λ) that minimized the generalized cross-validation score (Green and Silverman 1994).

Sexual selection on the interaction between male and female genital morphology

As we measured the genital size and shape of both males and females in each interacting pair, as well as the outcome of this interaction, we were able to estimate the sign and strength of the correlational selection operating on the covariance between these traits across the sexes. We estimated these correlational gradients by fitting a linear multiple regression model using the standardized cross-product terms as the predictor variables and relative fitness as the response variable. As these gradients essentially measure how genital size and shape interacts between the sexes to determine fitness, we refer to the resulting covariance matrix from this analysis as the interaction matrix. We used the resampling and thin-plate spline procedures outlined above to test the statistical significance and to visualize the standardized correlational selection gradients, respectively.

Estimating the Location of the Global Maxima on the Fitness Surface and its 95% Confidence Region

Existing methods for finding the confidence region associated with the location of the maxima of a regression function rely on the assumption that the data is normally distributed (Peterson et al. 2002), which is clearly not the case for our measure of fitness. Here, we use nonparametric bootstrapping method that we have previously developed (del Castillo et al. 2016) that is not based on any distributional assumptions and uses a flexible regression spline approach. This approach is provided by the "OptRegionTps" function in the OPTIMAREGION package of R (del Castillo et al. 2016; see Rapkin et al. 2018 for an application of this package).

In brief, a quadratic polynomial model was fit to the data using ordinary least squares regression implemented in the "lm" function of R, yielding a fitted response surface $\hat{y}(x)$ and residuals $r_i = y(x) - \hat{y}(x)$. We then applied bootstrapping to the residuals to create bootstrapped realizations $y^*(x)$= $\hat{y}(x)$ + r^* for each data point in our data set(x). For each simulated set of $y^*(x)$, we fit a quadratic polynomial and found parameter estimates γ^*. Following Yeh and Sing (1997), we repeated this procedure 1,000 times and computed Tukey's data depth for each generated γ^* vector, keeping the $100(1 - \alpha)$ % deepest (where in our case a = 0.95). This provides an approximate nonparametric bootstrap 95% confidence region for the quadratic polynomial coefficients (γ). The responses $y^*(x)$ that corresponded to the parameter vectors γ^* lying inside of their confidence region were then maximized numerically using the NLOPTR package of R (Johnson 2014; Ypma 2014) with respect to the regressors $(x_1, x_2 \ldots x_n)$ yielding the bootstrapped response global maxima (x^*). The nonparametric bootstrapped confidence region for the location of the global maximum of the fitness function is computed as the convex hull of all the bootstrapped maxima (x^*) that were found. We use the centroid (average) of all the maxima found as our point estimate of the global maxima (or peak) of the fitness surface.

Testing Whether the Population Mean is at the Peak of the Fitness Surface

To determine if males and females, on average, reside at the peak of the fitness surface for genital size and shape, we obtained four independent samples of beetles from each sex to map onto the fitness surfaces. We individually reared 50 larvae taken at random from our stock populations on four different diets varying in the ratio of protein to carbohydrates and total nutrition (P:C ratio, % total nutrition): diet 1 (1:3, 60%), diet 2 (1:3, 84%), diet 3 (1:8, 60%) and diet 4 (1:8, 84%). At eclosion, 18 beetles of each sex were taken at random from each diet (*n* = 72 beetles per sex) and stored in individual eppendorfs at -20°C. These beetles were dissected and the genitalia removed, imaged and measured using the protocol outlined above to extract CS and RW scores. It is important to note that CS and RW scores were extracted from a single data set that also contained the beetles used in our multivariate selection to ensure that these measures of genital size and shape were comparable across data sets. In order to map the mean genital size and shape of males and females from these independent samples onto their respective fitness surfaces, it was necessary to project CS and RW scores for each sex into the same eigenspace as the fitness surface. This was achieved by multiplying these trait values for each beetle by the eigenvectors provided in Table 2. This produced a unique set of eigenscores for each male and female beetle in the data set.

For each sex, we tested for differences among these independent samples along each of the dimensions of the fitness surface using Multivariate Analysis of Variance (MANOVA). Post-hoc univariate ANOVAs were used to determine which (if any) dimensions were contributing to any overall multivariate effects. To assess whether the mean of the four independent samples of males and females reside at the peak of their respective fitness surfaces, we visually compared the mean eigenscore of each sample with the 95% confidence region of the peak. If the 95% confidence interval for the sample mean (as determined by 1,000 bootstraps) overlaps the 95% confidence region of the peak, this is taken as evidence that the sample mean resides on the peak of the fitness surface.

RESULTS

Geometric morphometric analyses of 498 male and female genitalia yielded centroid size (CS) and 4 RW scores that collectively explained 78.40% and 85.38% of the total variation in genital shape, respectively. In males, RW1 explained 39.25% of the total variance in genital shape with negative values corresponding to a short, wide aedeagus and positive values to a long, narrow aedeagus (Figure 1A). RW2 explained a further 15.04% of this total variance with negative values of RW2 corresponding to an anteriorly shortened tip of the aedaegus and positive values to an anteriorly lengthened tip (Figure 1C). RW3 explained 12.79% of the total variance in male genital shape with negative values corresponding to an anti-clockwise twist of the anterior tip of the aedeagus and positive values to a clockwise twist of the anterior tip of the aedeagus (Figure 1E). RW4 explained the remaining 11.32% of the total variance in male genital shape with negative values corresponding to a compression of the left-side, posterior of the aedeagus and positive values to a similar compression but on the right-side of the aedeagus (Figure 1G).

In females, RW1 explained 59.26% of the total variance in genital shape with negative values corresponding to a wide, short vaginal aperture and positive values to a narrow, elongated vaginal aperture (Figure 1B). RW2 explained a further 14.26% of the total variation in female genital shape with negative values corresponding to narrower, longer supportive structures of the vagina and positive values to broader, curved supportive structures (Figure 1D). RW3 explained 7.82% of the total variance in female genital shape with negative values corresponding to an extreme posterior elongation of the supportive structures of the vagina and positive values to an extreme posterior broadening of these supportive structures (Figure 1F). RW4 explained the remaining 4.04% of the total variance in female genital shape with negative values corresponding to a shorter, wider vaginal aperture and posterior broadening of the supportive structures and positive values to a narrower, tapering vaginal aperture and posterior elongation of the supportive structures (Figure 1H).

Standardized linear, quadratic and correlational selection gradients for genital size and shape in males and females are presented in Tables 1A and B, respectively. In males, there was significant linear selection favouring increased values of RW1 (long, narrow aedeagus), RW3 (clockwise twist of the anterior tip of the aedeagus) and RW4 (posterior compression on the right-side of the aedeagus) (Table 1A). There was also significant stabilizing selection on CS and RW2 (length of the tip of the aedeagus) (Table 1A). In females, there was significant linear selection favouring reduced values of RW3 (extreme posterior elongation of the supportive structures of the vagina) and RW4 (shorter, wider vaginal aperture and posterior broadening of the supportive structures) (Table 1B). There was also significant stabilizing selection on CS, RW1 (length and width of the vaginal aperture) and RW4 and negative correlational selection on CS and RW1.

We conducted a canonical rotation of the γ matrices presented in Table 1 to locate the major dimensions of nonlinear sexual selection for male and female genital size and shape. The resulting $\mathbf{M}$ matrices of eigenvectors and their associated eigenvalues are presented in Tables 2A and B, respectively. In males, three of the five eigenvectors ($\mathbf{m}_3$ to $\mathbf{m}_5$) had negative eigenvalues, whereas the remaining two eigenvectors ($\mathbf{m}_1$ and $\mathbf{m}_2$) had positive eigenvalues (Table 2A). However there is only significant nonlinear selection operating on $\mathbf{m}_4$ and $\mathbf{m}_5$ demonstrating that the fitness surface is best described as a multivariate peak in shape (Figure 2A). There was also negative linear selection operating on $\mathbf{m}_2$ which largely favours an increase in RW1 and RW4 (Table 2A). In females, four of the five eigenvectors ($\mathbf{m}_2$ to $\mathbf{m}_5$) had negative eigenvalues, whereas the remaining eigenvector ($\mathbf{m}_1$) had a positive eigenvalue (Table 2B). As shown for males, significant nonlinear selection was only detected on $\mathbf{m}_4$ and $\mathbf{m}_5$ indicating that the fitness surface for females is also best described as a multivariate peak in shape (Figure 2B). There was also significant positive linear selection on $\mathbf{m}_4$ which largely favours a reduction in RW4 and negative linear selection on $\mathbf{m}_2$ which largely favours a reduction in RW3 (Table 2B).

Table 3 provides the interaction matrix of standardized correlational selection gradients for genital size and shape across the sexes. There was significant negative correlational selection on RW1 in males and RW4 in females (Table 3) and inspection of the thin-plate spline (Figure 3A) showed that fitness was highest at negative values of RW1 in males and positive values of RW4 in females (Figure 3A). Consequently, the fitness of an interacting male and female beetle is highest when males have a short, wide aedeagus (Figure 1A) and females have a narrower, tapering vaginal aperture and posterior elongation of the associated supportive structures (Figure 1H). There was also significant negative correlational selection on RW3 in males and RW4 in females (Table 3) and inspection of the thin-plate splines (Figure 3B) showed that fitness was highest at negative values of RW3 in males and positive values of RW4 in females. As a result, the fitness of an interacting pair is highest when males have an anti-clockwise twist to the anterior tip of the aedeagus (Figure 1E) and females have a narrower, tapering vaginal aperture and posterior elongation of the supportive structures (Figure 1H).

Having shown that the fitness surfaces for males and females contained a multivariate peak, we next estimated the global maxima and its 95% confidence region for $\mathbf{m}_4$ and $\mathbf{m}_5$ on each fitness surface. These are presented in Figures 2C and D for males and females, respectively. It has been argued that a fitness surface containing a peak should only be referred to as stabilizing selection if the peak resides within the phenotypic space sampled (Mitchell-Olds and Shaw 1987). The global maxima for $\mathbf{m}_4$ and $\mathbf{m}_5$ in males ($\mathbf{m}_4$ = 0.222, $\mathbf{m}_5$ = 0.275; Figure 2C) and females ($\mathbf{m}_4$ = 0.202, $\mathbf{m}_5$ = -0.236; Figure 2D) both exist within the distribution of phenotypic data sampled in our experiment and we can therefore formally define the observed pattern of nonlinear selection in the sexes as multivariate stabilizing selection (Figure 2A and B). We mapped the mean and 95% confidence intervals for male and female beetles reared on four independent diets along these two major axes of the fitness surface and these can be visualized in Figures 2C and D. In both sexes, the 95% confidence interval for each dietary mean overlapped the confidence region for the global maxima.

MANOVA revealed an overall multivariate effect of diet on the size and shape of the genitals along the eigenvectors of the fitness surface for both sexes and post-hoc ANOVAs showed that this overall effect was driven by significant differences in $\mathbf{m}_3$ for males and $\mathbf{m}_2$ and $\mathbf{m}_3$ for females (Table 4). Importantly, diet did not influence genital size and shape along $\mathbf{m}_4$ or $\mathbf{m}_5$ in either sex (Table 4). Collectively, this provides compelling evidence that the mean genital size and shape of male and female beetles reared on different diets reside on the peak of the fitness surface. Interestingly, even though the means of these samples are aligned with the peak of the fitness surface, many individual males (Figure 2E) and females (Figure 2F) from each of the diets fall outside the 95% confidence region for the peak. This shows that there is variation in morphology despite the operation of stabilizing selection and suggests that males and females are not somehow mechanically or physiologically constrained to occupy this region of the fitness surface.

Discussion

Explaining why the genital morphology of males are so highly divergent in species with internal fertilization has intrigued evolutionary biologists for decades (Hosken and Stockely 2004; Simmons 2014). Here we show that male and female genital size and shape play an important role in the successful transfer of a spermatophore during mating in the red flour beetle (*Tribolium castaneum*). This imposes strong sexual selection on male and female genital morphology that is multivariate stabilizing in form and characterised by a well-defined peak in fitness at intermediate values of genital size and shape. We also found that sexual selection targeted the covariance between the sexes for specific aspects of genital shape indicating that the interaction between male and female genitals also plays an important role in the successful transfer of a spermatophore during mating. The mean genital morphology of males and females taken from an independent sample of beetles reared on different diets were found to reside on their respective fitness peaks suggesting that genital morphology has evolved from a region of low to high fitness in both sexes. Collectively, our work highlights the important yet often ignored role that female genital morphology plays in determining mating success (Ah-King et al. 2014; Simmons 2014) and shows that these effects can occur independently, as well as through their interaction with male genital morphology. Moreover, our work shows how a complex pattern of multivariate stabilizing selection can shape the evolution of genital morphology in both males and females and cautions against the overly simplistic view that the sexual selection targeting genital morphology will always be directional in form (Arnqvist 1997).

Over two decades ago, Arnqvist (1997) provided a set of criteria to distinguish between the three major hypotheses of genital evolution (lock-and-key, pleiotropy and sexual selection hypotheses) for use in single species studies. An important criterion in this checklist is the form of selection targeting male genital morphology (Arnqvist 1997). More specifically, it has been proposed that stabilizing selection is the dominant form of selection targeting male genital morphology under the lock-and-key hypothesis, directional selection is the dominant form of selection under the sexual selection, whereas according to the pleiotropy hypothesis there will be no direct selection targeting male genital morphology (Arnqvist 1997). It is important, however, to make a subtle distinction on how stabilizing selection is proposed to operate in the lock-and-key hypothesis. In the original proposal of the lock-and-key hypothesis (Dufour 1848), male genital morphology is subject to stabilizing natural selection to prevent hybridization by ensuring that only males of the correct species are able to provide the right “key” for the female “lock”. Arnqvist (1997), however, proposed that because the poor alignment of male and female genitals within a species will reduce male mating success, this will also generate a pattern of stabilizing sexual selection on male genital morphology. Thus, while both scenarios predict that male genital morphology will be under stabilizing selection, the mode of selection is fundamentally different (i.e. natural versus sexual selection). The results of our multivariate selection analysis on the effects of male genital size and shape on mating success in *T. castaneum* is therefore in general agreement with Arnqvist's (1997) within species view of the lock-and-key hypothesis and demonstrates a clear and important role for sexual selection in male genital evolution in this species. Furthermore, it illustrates the inherent difficulty in distinguishing between the lock-and-key and sexual selection hypotheses, especially when sexual selection is largely non-linear in form, and supports the view that these hypothesis for genital evolution are not exclusive (Simmons 2014).

Multivariate selection analysis (Lande and Arnold 1983) has proved a powerful approach in quantifying the strength and form of selection acting on male genital morphology and formal estimates of selection gradients now exist for numerous species. In insects where the majority of estimates exist, linear selection on male genital morphology (Cordoba-Aguilar 1999, 2002, 2009; Arnqvist and Danielsson 1999, Danielsson and Askenmo 1999, Bertin and Fairbairn 2005; Wenninger and Averill 2006; Holwell et al. 2010; van Lieshout 2011; van Lieshout and Elgar 2011) appears more common than stabilizing selection (Tadler 1999; Simmons et al. 2009; Wojcieszek and Simmons 2011; House et al. 2016; Dougherty and Shuker 2016). This pattern also appears true more generally for selection on male sexual traits, although it should be noted that most experimental designs have far greater power to detect linear than non-linear forms of selection (Hunt et al. 2009). Interestingly, stabilizing sexual selection does appear common in a number of signaller-receiver systems (e.g. Greenfield 2002). For example, female mate choice exerts multivariate stabilizing selection on temporal and spectral components of the male advertisement call in grasshoppers (Butlin et al. 1985), field crickets (Brooks et al. 2005) and anurans (Gerhardt 1991; Gerhardt and Brooks 2009; Polakow et al. 1995). In these systems, the transmission of an acoustic signal is constrained by the biophysics of signal generation and emission and how the sensory organs of the receiver perceive and process the signal (Endler and Basolo 1998). As the sensory system of the female can only detect call components within a narrow range, they are not attracted to calls that are produced outside of this range, resulting in a pattern of multivariate stabilizing selection acting on male call structure (Endler and Basolo 1998). While the most parsimonious explanation for the pattern of multivariate stabilizing sexual selection we observe on male genital morphology in *T. castaneum* is the “mechanical fit” of genitals, we cannot rule out the possibility that a similar “sensory based” lock-and-key process does not exist. Indeed, Eberhard et al. (1998) has argued that male genitals are perceived by the female using tactile channels, with an intermediate male genital morphology being favoured because they best stimulate the average female in the population (the “one-size-fits-all” hypothesis). Distinguishing between mechanical and sensory lock-and-key processes, however, has proven difficult (Eberhard et al. 1998) and clearly more work is needed before this can be achieved in *T. castaneum*.

While the evolution of male genital morphology has been the subject of intense research, female genitals have been relatively understudied and this pattern appears to be getting worse with time (Ah-King et al. 2014). This bias in genital research likely stems from the long held view that males play the dominant role in sex and that female genitals are largely invariant (Ah-King et al. 2014). Our work, however, directly challenges this view by showing that female genital morphology in *T. castaneum* is far from invariant and that the existing variation in this trait is also subject to strong multivariate stabilizing sexual selection. In fact, the nonlinear selection gradients that describe the pattern of multivariate stabilizing sexual selection acting on female genital morphology were as strong as those reported for male genital morphology, further highlighting the equally important role that females play in determining the outcome of mating in this species. While the major models of genital evolution do not provide any clear predictions regarding the strength and form of selection acting on female genitals, we believe that the pattern of sexual selection we document for female *T. castaneum* adds further support to an important role for a “lock-and-key” process in the evolution of genital morphology in this species. The operation of the “lock-and-key” process centres on the alignment of the male and female genitals during mating, with the optimal male genital morphology being the one that, on average, most closely aligns with the average female genital morphology in the population. As any deviation from this optimal morphology decreases the fit with the female genitals and reduces subsequent mating success, stabilizing selection is predicted to target male genital morphology (Arnqvist 1997). However, because the successful outcome of mating depends on both sexes, the poor alignment of genitals and the resulting reduction in mating success can also be caused by the female genitals deviating from the “average” in the population, which should also generate stabilizing selection on female genital morphology. Furthermore, as the “lock-and-key” process is based on the interaction between male and female genitals during mating, sexual selection should also target the covariance between these traits across the sexes: a finding that is also supported by our study. Collectively, our work highlights the value of formally estimating sexual selection on genital morphology in both sexes and supports recent claims that our understanding of genital evolution will continue to be hampered until the persisting male bias in genital research is addressed (Ah-King et al. 2014).

Our work shows that male and female genital morphology not only has important independent effects on the successful transfer of a spermatophore during mating in *T. castaneum* but also that the interaction between these traits influences mating success. That is, sexual selection targets male and female genital morphology directly, as well as indirectly via the covariance between these traits. To our knowledge, this is the first time such an approach has been used to study genital evolution, despite offering a novel means to assess how important the interaction of male and female genital morphology is to mating success. In *T. castenum*, we found significant negative correlational selection gradients for two aspects of genital shape: RW4 in females with RW1 and RW3 in males. Biologically, this means that a mating pair will have a higher success in transferring a spermatophore when a male with either a short, wide aedeagus (RW1) or an anti-clockwise twist to the anterior tip of the aedeagus (RW3) mates with a female having a narrower, tapering vaginal aperture and posterior elongation of the supportive structures (RW4). Explaining these relationships at this stage would be purely speculative and highlights the limitation of our approach: correlational selection gradients do not identify a mechanism. Isolating a mechanism would require functional studies (Brennan & Prum 2015; Ah-King et al. 2014; Simmons 2014) and unfortunately only a handful such studies exist for insects (Ronn et al. 2007; Werner & Simmons 2008; Polak & Rashed 2009; Kahn et al. 2010; Hotzy et al. 2012). For example, Werner and Simmons (2008) used histology to show that three the genital sclerites in male dung beetles (*Onthophagus taurus*) form a functionally integrated unit that generates the tubular-shaped spermatophore and delivers its opening to the female’s spermathecal duct, whereas a fourth serves as a holdfast device during mating. It is possible that a similar mechanical process enables a short, wide aedeagus in *T. casteneum* or one with an anti-clockwise twist to the anterior tip more efficiently deliver a spermatophore or better anchors the male when mating to a female with a narrower, tapering vaginal aperture with posterior elongation of the supportive structures, but functional studies are clearly be needed to confirm this. Irrespective of the exact mechanism, theory (e.g. Cheverud 1984, 1996; Blows and Hoffman 2005) predicts that correlational selection will generate a genetic correlation between the two traits and a meta-analysis across 22 animal and plant species provides empirical support for this prediction (Roff and Fairbairne 2012). It is therefore likely that the aspects of genital shape that are subject to correlational selection in *T. castaneum* will also be genetically correlated across the sexes. Ultimately, this will facilitate the co-evolution of male and female genital shape in *T. castaneum* (Lande 1980), as has been well documents in a range of other animal species using both experimental evolution and comparative approaches (reviewed in Brennan and Prum 2015). We are currently using a half-sibling breeding design to estimate the genetic architecture of genital size and shape within and across the sexes in *T. castaneum*, as well as experimental evolution to examine the co-evolutionary dynamics of these traits.

In a population that has experienced a history of strong and persistent stabilizing selection and the absence of frequency-dependent selection, theory predicts that the population mean should evolve to match the peak in fitness (Simpson 1953; Lande 1976, 1979). Despite being central too much of evolutionary biology (Fear and Price 1998; Arnold 2003; Dietrich and Skipper 2012; Skipper and Dietrich 2012) and receiving formal mathematical support (Lande 1976; Bürger and Lande 1994), surprisingly few empirical studies have actually tested this prediction (Travis 1989; but see Armbruster 1990; Benkman 1993, 2003; Schluter 2000; Brooks et al. 2005). One reason for this is the difficulty in formally estimating stabilizing selection, especially in natural populations where designs including several hundred individuals are required to provide accurate estimates (Hersch and Phillips 2004; Hunt et al. 2010). Consequently, while many estimates are likely to be biased (Hersch and Phillips 2004), the general consensus is that stabilizing selection is weak in natural populations (Kingsolver et al. 2001; Kingsolver and Diamond 2011). Our selection analysis of genital morphology in *T. casteneum*, however, challenges this view by documenting quadratic gradients in both sexes that are much stronger than the average reported across studies based on mating success (median $\gamma = |0.16|$, Kingsolver et al. 2001). Given this strong stabilizing sexual selection, which appears consistent with estimates from other genital studies in insects (e.g. Tadler 1999; Bertin and Fairbairn 2005; House et al.

2016; Dougherty and Shuker 2016), it is perhaps not surprising that we found the mean genital morphology of male and female *T. castaneum* reside within the 95% confidence region of the stationary point (peak) on their respective fitness surfaces. What is more surprising is that this convergence on the peak occurred even though we estimated mean genital morphology from an independent sample of males and females that had been reared on very different diets during development and suggests that stabilizing selection is not only strong but is also likely to be both persistent and have a stable peak location over evolutionary time in this species. Indeed, Estes and Arnold (2007) found that a quantitative genetic model where the fitness optimum was able to move within an adaptive zone with stable boundaries performed significantly better than five other competing models in explaining the evolution of phenotypic means using an extensive database (Gingerich 2001) of phenotypic divergence across species. Importantly, this model performed better across all timescales examined, ranging from a few to as many as 10 million generations, suggesting that the process of the population mean evolving to the fitness peak we observe may have important implications for both the micro and macroevolution of genital morphology (Estes and Arnold 2007). For example, if populations evolve in response to stabilizing selection imposed by stable adaptive optima it is relatively easy to envisage intra and interspecific differences in genital morphology evolving as a result of variation in the location of the adaptive optima (Hansen 1997). This would certainly help explain the adaptive radiation in genital morphology frequently observed across natural (e.g. Wojcieszek and Simmons 2012b; Heinen-Kay and Langerhans 2013; Oneal and Knowles 2013) and experimental (e.g. Simmons et al. 2009; House et al. 2013) populations, as well as the extreme diversification in genital observed across closely related species (e.g. Arnqvist 1998; Brindle and Opie 2016; Kuntner et al. 2016).

In conclusion, our study shows that male and female genital morphology is subject to strong multivariate stabilizing selection in *T. castaneum* but that sexual selection also targets the covariance between the sexes for aspects of genital shape, indicating that how the genitals interact during mating is also important to the successful transfer of a spermatophore in this species. Both findings provide empirical support for the “lock-and-key” hypothesis of genital evolution, although we cannot determine at this stage whether this process is driven by a mechanical or sensory-based interaction (or both) during mating. Moreover, our finding that the mean genital morphology of male and female beetles reared on different diets has converged on the peak on their respective fitness landscape also shows that this pattern of multivariate stabilizing selection can shape the long-term evolutionary dynamics of genital morphology in *T. castaneum*. The ability of the population mean to converge on the peak in fitness is likely to have important implications for the adaptive radiation of genital morphology across populations of *T. castaneum*, as well as the diversification of male and female genitals across *Tribolium* species.

Acknowledgements

CMH was funded by a Leverhulme Trust Fellowship, JR by a NERC studentship, MG by XXX, EDC was partially funded by a NSF grant (CMII 1634878) and JH by a Royal Society University Fellowship (UF0762844) and research grants from NERC (NE/G00949X/1) and the ARC (DP180101708).

References

Ah-King, M., A. B. Barron, M. E. Herberstein. 2014. Genital evolution: Why are females still understudied? PLOS Biol. 12:e1001851.

Anderson, C.M. and R.B. Langerhans. 2015. Origins of female genital diversity: predation risk and lock-and-key explain rapid divergence during an adaptive radiation. Evolution 69:2452-2467.

Andersson, M. 1994. Sexual selection. Princeton Univ. Press, Princeton, N.J.

Armbruster, W.S. 1990. Estimating and testing the shapes of adaptive surfaces: The morphology and pollination of *Dalechampia* blossoms. Am. Nat. 135:14-31.

Arnqvist, G. 1997. The evolution of animal genitalia: distinguishing between hypotheses by single species studies. Biol. J. Linn. Soc. 60:365-379.

Arnqvist, G., R. Thornhill and L. Rowe. 1997. Evolution of animal genitalia: morphological correlates of fitness components in a water strider. J. Evol. Biol. 10:613-640.

Attia, F.A. and T. Tregenza. 2004. Divergence revealed by population crosses in the red flour beetle *Tribolium castaneum*. Evol. Ecol. Res. 6: 927–935.

Benkman, C. W. 1993. Adaptation to single resources and the evolution of crossbill (Loxia) diversity. Ecol. Monogr. 63:305-325.

Benkman, C. W. 2003. Divergent selection drives the adaptive radiation of crossbills. Evolution 57:1176-1181.

Bertin, A., and D.J. Fairbairn. 2005. One tool, many uses: precopulatory sexual selection on genital morphology in *Aquarius remigis.* J. Evol. Biol. 2005; 18:949-961.

Bisgaard, S. and B. Ankenman. 1996. Standard errors for the eigenvalues in second-order response surface models. Technometrics 38:238-246.

Bloch Qazi, M.C., J.T. Herbeck and S.M. Lewis. 1996. Mechanisms of sperm transfer and storage in the red flour beetle (Coleoptera: Tenebrionidae). Ann. Entomol. Soc. Am. 89:892-897.

Blows, M.W. 2002. Interaction between natural and sexual selection during the evolution of mate recognition. Proc. R. Soc. Lond. B 269:1113-1118.

Blows, M.W. and R. Brooks. 2003. Measuring nonlinear selection. Am. Nat. 162:815-820.

Blows, M.W., R. Brooks and P.G. Kraft. 2003. Exploring complex fitness surfaces: Multiple ornamentation and polymorphism in male guppies. Evolution 57:1622-1630.

Boughman, J.W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. Nature 411:944-948.

Brook, R., J. Hunt, M. W. Blows, M. J. Smith, L. F. Bussière, and M. D. Jennions. 2005. Experimental evidence for multivariate stabilizing sexual selection. Evolution 59:871-880.

Butlin, R.K., G.M. Hewitt and S.F. Webb. 1985. Sexual selection for intermediate optimum in *Chorthippus brunneus* (Orthoptera: Acrididae). Anim. Behav. 33:1281-1292.

Cayetano, L., A.A. Maklakov, R.C. Brooks and R. Bonduriansky. 2011. Evolution of male and female genitalia following release from sexual selection. Evolution 65:2171-2183.

Chenoweth, S.F. and M.W. Blows. 2005. Contrasting mutual selection on homologous signal traits in *Drosophila serrata*. Am. Nat. 165:281-289

Conner, J., and S. Via. 1992. Natural selection on body size in Tribolium: possible genetic constraints on adaptive evolution. Heredity 69:73–83.

Dietrich, M. R. and R. A. Jr. Skipper. 2012. A shifting terrain: A brief history of the adaptive landscape. Pp 3-14 *in* E. I. Svensson and R. Calsbeek, eds. *The adaptive landscape in evolutionary biology.* Oxford University Press, Oxford, OX2 6DP.

Dougherty LR, Shuker DM. 2016. Variation in pre- and post-copulatory sexual selection on male genital size in two spcies of lygaeid bug. Behav. Ecol. Socio: *In press*

Dufour, L. 1844. Anatomie générale des diptères. Ann. Sci. Nat. 1:244-264.

Eberhard WG. 1985. *Sexual selection and animal genitalia*. Harvard University Press, Cambridge MA.

Eberhard, W.G., B.A. Huber, S. Rafael Lucas Rodriguez, R.D. Briceno, I. Salas, V. Rodriguez. 1998. One size fits all? Relationships between the size and degree of variation in genitalia and other body parts in twenty species of insects and spiders. Evolution 52:415-431.

Eberhard, W.G., R.L. Rodriguez and M. Polihronakis. 2009. Pitfalls in understanding the functional significance of genital allometry. J. Evol. Biol. 22:435-445.

Eberhard, W.G. 2010. Evolution of genitalia: theories, evidence, and new directions. Genetica 138:5-18.

Endler, J.A. and A.E. Houde. 1995. Geographic variation in female preferences for male traits in *Poecilia reticulate*. Evolution 49:456-468.

Fear, K.K. and T. Price. 1998. The adaptive surface in ecology. Oikos 82:440-448.

Fedina, T.Y. and S.M. Lewis. 2006. Proximal traits and mechanisms for biasing paternity in the red flour beetle *Tribolium castaneum.* Behav. Ecol. Sociobiol. 60:844-853.

Fedina, T.Y. 2007. Cryptic female choice during spermatophore transfer in *Tribolium castaneum* (Coleoptera: Tenebriodae). J. Insect Phys. 53:93-98.

Fedina, T.Y. and S.M. Lewis. 2007. Female mate choice across mating stages and between sequential mates in flour beetles. J. Evol. Biol. 20:2138-2143.

Fedina, T.Y. and S.M. Lewis. 2008. An integrative view of sexual selection in *Tribolium* flour beetles. Biol. Rev. 83:151-171.

Gerhardt, H.C. 1991. Female mate choice in treefrogs: static and dynamic acoustic criteria. Anim. Behav. 42:615-635.

Gerhardt, H.C. and R. Brooks. 2009. Experimental analysis of multivariate female choice in gray treefrogs (*Hyla versicolor*): Evidence for directional and stabilizing selection. Evolution 63:2504-2512.

Green, P.J. and B.W. Silverman. 1994. Nonparametric regression and generalised linear models. Chapman and Hall, London.

Gutierrez, B.L., Macleod, N. and G.D. Edgecombe. 2011. Detecting taxonomic signal in an under-utilised character system: geometric morphometrics of the forcipular coxae of Scutigeromorpha (Chilopoda). ZooKeys 66:49–66.

Higgie, M. and M.W. Blows. 2007. Are traits that experience reinforcement also under sexual selection? Am. Nat. 170:409-420.

House, C.M. and L. Simmons. 2005. Relative influence of male and female genital morphology on paternity in the dung beetle *Onthophagus taurus.* Behav. Ecol. 16:889-897.

Hosken DJ, P. Stockley. 2004. Sexual selection and genital evolution. Trends Ecol. Evol. 19:87-93.

Hotzy, C. and G. Arnqvist. 2009. Sperm competition favours harmful males in seed beetles. Current Biol. 19:404-407.

Hotzy, C., M. Polak, J.L. Ronn and G. Arnqvist. 2010. Phenotypic engineering unveils the function of genital morphology. Current. Biol. 22:2258-2261.

Hunt J, C.J. Breuker, J.A. Sadowski, A.J. Moore. 2009. Male-male competition, female mate choice and their interaction: determining total sexual selection. J. Evol. Biol. 22:13-26.

Jones, A.G., Ratterman, N.L., and K.A. Paczolt. 2012. The adaptive landscape in sexual selection research. Pp. 110-125 *in* E.I. Svensson and R. Calsbeek, eds. *The adaptive landscape in evolutionary biology.* Oxford Univ. Press, Oxford, U.K.

Johnson, S.G. 2014. The NLopt nonlinear optimization package. http://ab-initio.mit.edu/nlopt.

Kawano, K. 2002. Character displacement in giant rhinoceros beetles. Am. Nat. 159:255-271.

Kawano, K. 2003. Character displacement in stag beetles (Coleoptera: Lucanidae). Ann. Entomol. Soc. Am. 96:503-511.

Kingsolver, J.G. and S.E. Diamond. 2011. Phenotypic selection in natural populations: What limits directional selection? Am. Nat. 177:346-357.

Kuntner, M., J.A. Coddington, and J.M. Schneider. 2009. Intersexual arms race? Genital coevolution in nephilid spiders (Araneae, Nephilidae). Evolution 63:1451-1463.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314-334.

Lande, R. 1979. Quantitative genetical analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402-416.

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

Langerhans, R.B., C.A. Layman and T.J. DeWitt. 2005. Male genital size reflects a tradeoff between attracting mates and avoiding predators in two live-bearing fish species. PNAS 102:7618-7623

LeBas, N., L.R. Hockham and M.G. Ritchie. 2003 Nonlinear and correlational sexual selection on 'honest' female ornamentation. Proc. R. Soc. Lond. B. 270:2159-2165.

Lewis, S.M. and J. Iannini. 1995. Fitness consequences of differences in male mating behaviour in relation to female reproductive status in flour beetles. Anim. Behav. 50:1157-1160.

Maan, M.E. and O. Seehausen. 2011. Ecology, sexual selection and speciation. Ecol. Lett doi:10.1111/j.1461-0248.2011.01606.x

Macagno, A.L.M., A. Pizzo, H.F. Parzer, C. Palestrini, A. Rolando, A.P. Moczek. 2011. Shape – but not size – codivergence between male and female copulatory structures in *Onthophagus* beetles. PLoS ONE 6:12 e28893.

Manly, B.F.J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. Chapman and Hall, London, UK.

McGlothlin, J.W., P.G. Parker, V. Jr Nolan and E.D. Ketterson. 2005. Correlational selection leads to genetic integration of body size and an attractive plumage traits in dark-eyed juncos. Evolution 59:658-671.

McPeek, M.A., L. Shen, J.Z. Torrey, and H. Farid. 2008. The tempo and mode of three-dimensional morphological evolution in male reproductive structures. Am. Nat. 171:E158-E178.

McPeek, M.A., L. Shen, J.Z. Torrey, and H. Farid. 2009. The correlated evolution of three-dimensional reproductive structures between male and female damselflies. Evolution 63:73-83.

Mendez, V. and A. Cordoba-Aguilar. 2004. Sexual selection and animal genitalia. Trends Ecol. Evol. 19:224-225.

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

Pai, A., B. Lauren and Y. Guiyun. 2005. Female multiple mating for fertility assurance in red flour beetles (*Tribolium casteneum*). Can. J. Zool. 83:913-919.

Panhuis, T.M., R. Butlin, M. Zuk, T. Tregenza. 2001. Sexual selection and speciation. Trends. Ecol. Evol. 16:364-371.

Parker, G.A. 1970. Sperm competition and its evolutionary consequences in the insects. Biol. Rev. 45:525-567.

Peterson, J., S. Cahya and E. Del Castillo. 2002. A General Approach to Confidence Regions for Optimal Factor Levels of Response Surfaces. Biometrics 58: 422-431.

Phillips, P.C. and S.J. Arnold. 1989. Visualizing multivariate selection. Evolution 43:1209-1222.

Polak, D.A., P.Y. Backwell, N. Caithness and M.D. Jennions. 1995. Stabilizing or directional selection in signalling systems: investigations in a population of painted reed frogs, *Hyperolius marmoratus.* S. Afr. J. Sci. 91:270-273

Preziosi, R.F. and D.J. Fairbairn. 1996. Sexual size dimorphism and selection in the wild in the waterstrider *Aquarius remigis*: Body size, components of body size and male mating success. J. Evol. Biol. 9:317-336.

Preziosi, R.F. and D.J. Fairbairn. 2000. Lifetime selection on adult body size and components of body size in a waterstrider: Opposing selection and maintenance of sexual size dimorphism. Evolution 54:558-566.

Reynolds, R.J., D.K. Childers, N.M. Pajewski. 2010. The distribution and hypothesis testing of eigenvalues from the canonical analysis of the gamma matrix of quadratic and correlation selection gradients. Evolution 64:1076-1085.

Rohlf, F.J. 2006. *tpsUtil, file utility program*, Version 1.38. Stony Brook NY: Department of Ecology and Evolution. State University of New York.

Rohlf, F.J. 2008. *tpsRelw, relative warps analysis*, Version 1.46. Stony Brook NY: Department of Ecology and Evolution. State University of New York.

Rohlf, F.J. 2009. *tpsDig, digitize landmarks and outlines*, Version 2.14. Stony Brook NY: Department of Ecology and Evolution. State University of New York.

Ronn, J., M. Katvala and G. Arnqvist. 2007. Coevolution between harmful male genitalia and female resistance in seed beetles. PNAS 104:10921-10925

Rowe, L. and G. Arnqvist. 2011. Sexual selection and the evolution of genital shape and complexity in water striders. Evolution 66:40-54.

Rundle, H. D., S.F. Chenoweth, P. Doughty and M.W. Blows. 2005 Divergent selection and the evolution of signal traits and mating preferences. PLoS Biol. 3:1988-1995 (e368)

Schilthuizen, M. 2003. Shape matters: the evolution of insect genitalia. Proc. Exper. Appl. Entomol. Nev Amsterdam 14:9-15.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford Univ. Press, New York, N.Y.

Simmons, L.W., House, C.M., Hunt, J., Garcia-Gonzalez, F. 2009. Evolutionary response to sexual selection in male genital morphology. Current Biol. 19:1442-1446.

Simmons, L.W., and F. Garcia-Gonzalez. 2011. Experimental coevolution of male and female genital morphology. Nat. Comm. 2:374 DOI:10.1038/ncomms1379

Simmons, L.W. 2014. Sexual selection and genital evolution. Austral Entomology 53:1-17.

Simpson, G.G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.

Simpson, G.G. 1953. The major features of evolution. Columbia Univ. Press, Garden City, NY.

Skipper, R. A. Jr and M. R. Dietrich. 2012. Sewall wrights adaptive landscape: Philosophical reflections on heuristic value. Pp 16-23 *in* E. I. Svensson and R. Calsbeek, eds. *The adaptive landscape in evolutionary biology.* Oxford University Press, Oxford, OX2 6DP.

Song, H. and S.R. Bucheli. 2009. Comparison of phylogenetic signal between male genitalia and non-genital characters in insect systematics. Cladistics 25:1-13.

Sota, T. and T. Tanabe. 2009. Multiple speciation events in an arthropod with divergent evolution in sexual morphology. Proc. R. Soc. B. 277:689-696

Stinchcombe, J.R., A.F. Agrawal, P.A. Hohenlohe, S.J. Arnold and M.W. Blows. 2008. Estimating non-linear selection gradients using quadratic regression coefficients: double or nothing? Evolution 62:2435-2440.

Svensson, E.I., and R. Calsbeek. 2012. The adaptive landscape in evolutionary biology. Oxford Univ. Press, Oxford, U.K.

Tadler, A. 1999. Selection of a conspicuous male genitalic trait in the seedbug *Lygaeus simulans.* Proc. R. Soc. Lond. B 266:1773-1777.

Takami, Y. and T. Sota. 2007. Rapid diversification of male genitalia and mating strategies in *Ohomopterus* ground beetles. J. Evol. Biol. 20:1385-1395.

Tanabe, T. and T. Sota. 2013. Both male and female novel traits promote the correlated evolution of genitalia between the sexes in an arthropod. Evolution 68:441-452.

Telford, S.R. and M.D. Jennions. 1998. Establishing cryptic female choice in animals. Trends Ecol. Evol. 13:217-218.

Turelli, M., N.H. Barton and J.A. Coyne. 2001. Theory and speciation. Trends. Ecol. Evol. 16:330-343.

Tyler, F. and T. Tregenza. 2012. Why do so many flour beetle copulations fail? Entomologia Experimentalis et Applicata 146:199-206.

Wojcieszek, J.M. and L.W. Simmons. 2012a. Male genital morphology influences paternity success in the millipede *Antichiropus variabilis*. Behav. Ecol. Sociobiol. 65:1843–1856.

Wojcieszek, J.M. and L.W. Simmons. 2012b. Evidence for stabilizing selection and slow divergent evolution of male genitalia in a millipede (*Antichiropus variabilis*). Evolution 66:1138-1153

Wojcieszek, J.M. and L.W. Simmons. 2013. Divergence in genital morphology may contribute to mechanical reproductive isolation in a millipede. Ecol. Evol. 3:334-343.

Wolak, M.E., D.J. Fairbairne and Y.R. Paulsen. 2012. Guidelines for estimating repeatability. Methods in Ecology Evolution 3:129-137.

Woutersen T. and J.C. Ham. 2013. Confidence Sets for Continuous and Discontinuous Functions of Parameters. Center for Microdata Methods and Practice (Cemmap) working paper CWPP23/13, Dept. of Economics, University College London, October 31.

Yeh, A.B. and K. Singh. 1997. Balanced Confidence Regions Based on Tukey's Depth and the Bootstrap. J. R. Statist. Soc. B 59:639-652.

Ypma, J. 2014. Introduction to nloptr: an R interface to NLopt, https://cran.r-project.org/web/packages/nloptr/.

FIGURE LEGENDS

Figure 1. Frequency distribution of the four relative warp (RW) scores characterizing the variation in male (A, C, E and G) and female (B, D, F and H) genital shape. For each RW, we provide thin-plate spline visualizations (inset) that characterize a positive and negative score.

Figure 2. (A, B) Thin-plate spline visualizations (contour view) of the two major axes of nonlinear selection ($\mathbf{m_4}$ and $\mathbf{m_5}$) on the fitness surface for males and females, respectively. In each surface, white colouration represents regions of highest fitness, whereas red colouration represents regions of lowest fitness. Individual data points are provided as black circles on the surface. (B,C) Thin-plate spline visualizations mapping the 95% confidence region of the global maxima (grey region) on the fitness surface for males and females, respectively. In each surface, the solid black dot represents the global maxima and the coloured symbols (closed) represent the means of the four diet treatments and the coloured bars their 95% confidence interval, where the blue circle is diet 1, the red square is diet 2, the green diamond is diet 3 and the orange triangle is diet 4. (E,F) Thin-plate spline visualizations mapping the 95% confidence region of the global maxima (grey region) on the fitness surface for males and females, respectively. In each surface, the solid black dot represents the global maxima and the coloured symbols (closed) represent the means of the four diet treatments (as defined above). The open coloured symbols represent the actual data points for each treatment so their location on the fitness surface can be observed.

Figure 3. Thin-plate spline visualizations (contour view) showing the two significant cases of negative correlational selection operating on the covariance between male and female genital shape: (A) RW1 in males and RW4 in females and (B) RW3 in males and RW4 in females. In both instances, white colouration represents regions of highest fitness, whereas red colouration represents regions of lowest fitness. Individual data points are provided as black circles on the surface.

Tables

Table 1. The vector of standardized linear selection gradients (β) and the matrix of standardized quadratic and correlational selection gradients (γ) for successful matings in male and female *Tribolium castaneum*. Randomization test: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

		γ				
	β	CS	RW1	RW2	RW3	RW4
A. Male						
CS	0.030	-0.290***				
RW1	0.110*	-0.061	-0.018			
RW2	-0.043	-0.015	-0.015	-0.244**		
RW3	0.136**	-0.065	-0.052	0.093	0.100	
RW4	0.110*	-0.024	0.063	0.060	0.015	-0.010
B. Female						
CS	-0.019	-0.236**				
RW1	0.037	-0.242*	-0.364*			
RW2	0.033	-0.137	-0.137	0.082		
RW3	-0.110*	0.057	-0.003	-0.017	-0.044	
RW4	-0.092*	0.015	-0.052	-0.045	-0.028	-0.140**

Table 2. The **M** matrix of eigenvectors from the canonical analysis of γ for successful matings in male and female *Tribolium castaneum*. The linear ($\boldsymbol{\theta}_i$) and quadratic ($\boldsymbol{\lambda}_i$) gradient of sexual selection acting along each eigenvector are provided in the last two columns. The quadratic selection gradient ($\boldsymbol{\lambda}_i$) of each eigenvector ($\mathbf{m}_i$) is equivalent to the eigenvalue. Randomization test: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

	M					**Selection**	
	CS	RW1	RW2	RW3	RW4	$\boldsymbol{\theta}_i$	$\boldsymbol{\lambda}_i$
A. Male							
$\mathbf{m}_1$	0.120	0.229	-0.249	-0.927	-0.105	-0.098	0.149
$\mathbf{m}_2$	0.172	-0.654	-0.123	-0.024	-0.726	-0.144**	0.063
$\mathbf{m}_3$	-0.190	0.682	-0.136	0.253	-0.645	0.038	-0.077
$\mathbf{m}_4$	-0.024	-0.072	-0.951	0.210	0.214	0.084	-0.279***
$\mathbf{m}_5$	0.959	0.222	0.003	0.176	0.021	0.079	-0.317**
B. Female							
$\mathbf{m}_1$	0.299	0.093	-0.929	0.150	0.128	-0.061	0.149
$\mathbf{m}_2$	0.496	-0.430	0.232	0.718	-0.007	-0.096*	-0.009
$\mathbf{m}_3$	0.418	-0.450	0.071	-0.576	0.535	-0.008	-0.060
$\mathbf{m}_4$	0.375	-0.191	-0.071	-0.359	-0.830	0.099*	-0.175***
$\mathbf{m}_5$	-0.591	-0.753	-0.271	0.043	-0.089	-0.022	-0.610*

Table 3. Interaction matrix containing the standardized correlational selection gradients operating on the covariance between male and female genital size and shape. Randomization test: * $P < 0.05$.

		Males				
		CS	RW1	RW2	RW3	RW4
Females	CS	0.029				
	RW1	0.073	0.017			
	RW2	-0.034	0.015	-0.024		
	RW3	0.024	-0.048	-0.041	0.065	
	RW4	-0.062	-0.071*	0.026	-0.099*	0.029

Table 4. Multivariate Analysis of Variance (MANOVA) comparing the means genital size and shape of the four independent dietary samples of (A) males and (B) females along each of the five eigenvectors derived by canonical rotation of γ for each sex. In each sex, univariate ANOVA and means (± 95% confidence intervals) are also provided for each eigenvector.

A. Males	MANOVA (Pillai's trace = 0.535, $F_{15,198}$ = 2.864, P = 0.0001)						
	Means (95% CIs)					ANOVA	
	Diet 1	Diet 2	Diet 3	Diet 4		$F_{3,71}$	P
m_1	0.195 (-0.181,0.592)	-0.387 (-0.878,0.094)	0.172 (-0.302,0.587)	0.052 (-0.324,0.524)		1.371	0.259
m_2	0.074 (-0.348,0.462)	0.310 (-0.276,0.809)	-0.440 (-0.779,-0.116)	-0.013 (-0.466,0.378)		1.856	0.145
m_3	0.463 (0.027,0.935)	0.272 (-0.272,0.819)	0.256 (-0.130,0.621)	-0.949 (-1.287,-0.570)		7.559	0.0001
m_4	-0.116 (-0.523,0.267)	0.359 (-0.119,0.827)	-0.148 (-0.353,0.065)	-0.124 (-0.545,0.315)		1.440	0.239
m_5	-0.229 (-0.454,0.001)	0.324 (-0.165,0.841)	-0.219 (-0.527,0.094)	0.090 (-0.307,0.475)		1.803	0.155
B. Females	MANOVA (Pillai's trace = 0.610, $F_{15,198}$ = 3.368, P = 0.0001)						
	Means (95% CIs)					ANOVA	
	Diet 1	Diet 2	Diet 3	Diet 4		$F_{3,71}$	P
m_1	-0.088 (-0.496,0.294)	0.446 (-0.012,0.888)	-0.281 (-0.549,-0.019)	-0.133 (-0.521,0.276)		2.551	0.063
m_2	-0.340 (-0.660,-0.029)	0.492 (-0.073,1.048)	-0.490 (-0.868,-0.119)	0.250 (-0.321,0.743)		3.662	0.017
m_3	-0.380 (-0.921,0.083)	-0.348 (-0.935,0.236)	-0.207 (-0.593,0.211)	0.903 (0.492,1.292)		5.826	0.001
m_4	0.241 (-0.098,0.612)	0.181 (-0.367,0.711)	-0.324 (-0.667,0.015)	-0.162 (-0.574,0.195)		1.587	0.201
m_5	-0.108 (-0.481,0.288)	-0.070 (-0.457,0.312)	0.120 (-0.152,0.389)	0.076 (-0.238,0.356)		0.402	0.752

Figure 1

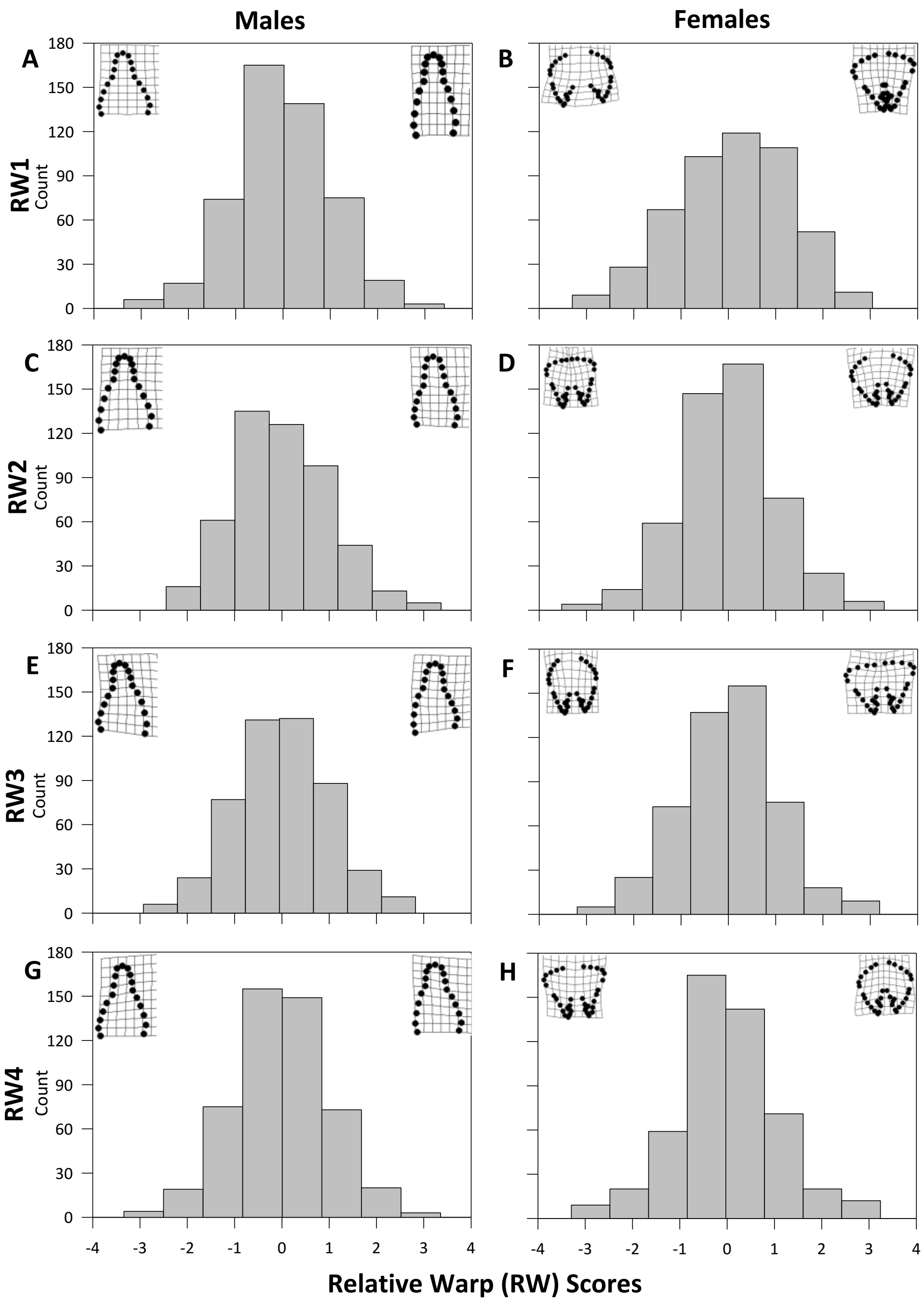

Males
Females
A
B
C
D
E
F
G
H
RW1
RW2
RW3
RW4
Count
0
30
60
90
120
150
180
-4
-3
-2
-1
0
1
2
3
4
Relative Warp (RW) Scores

Figure 2

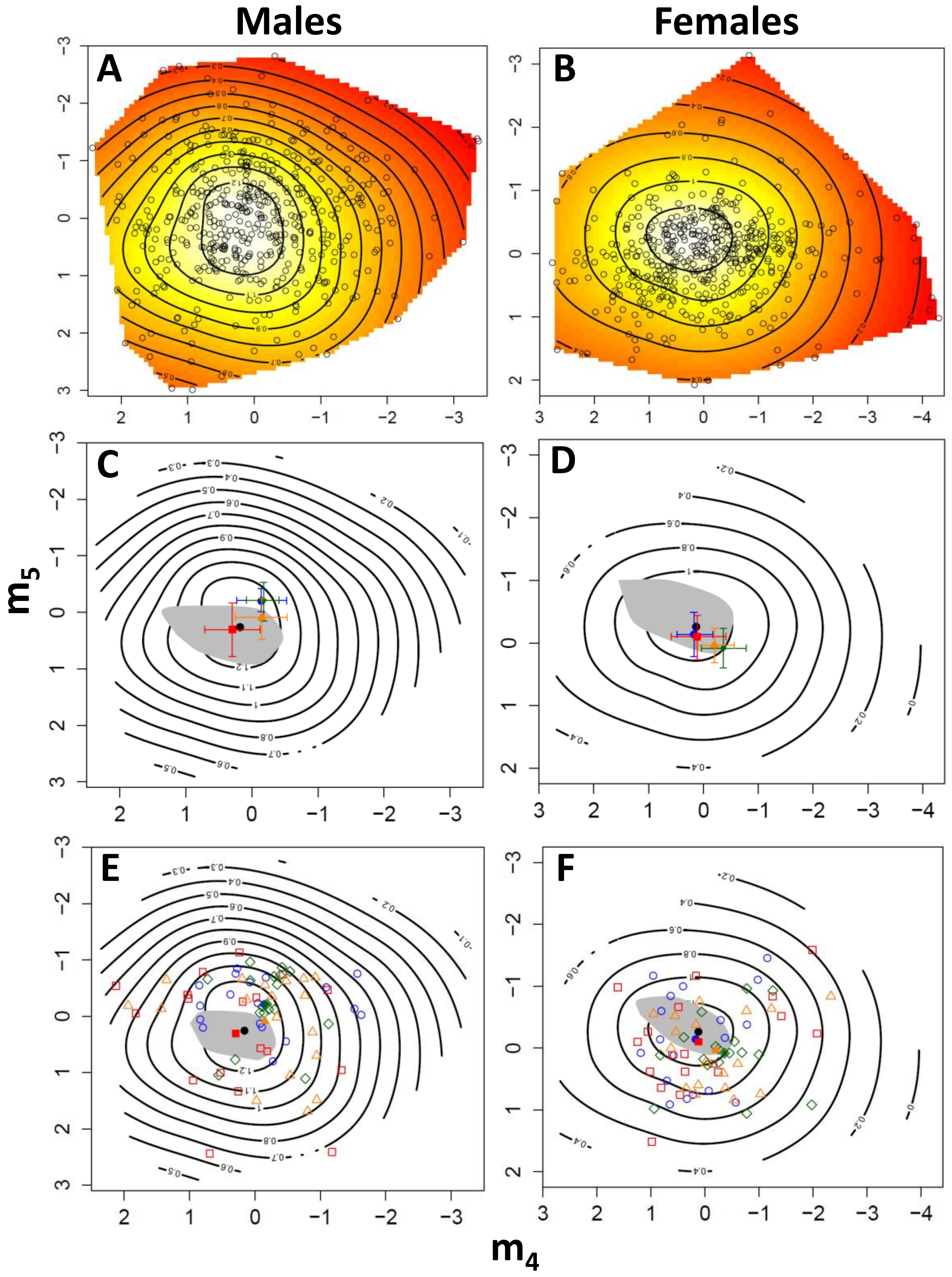
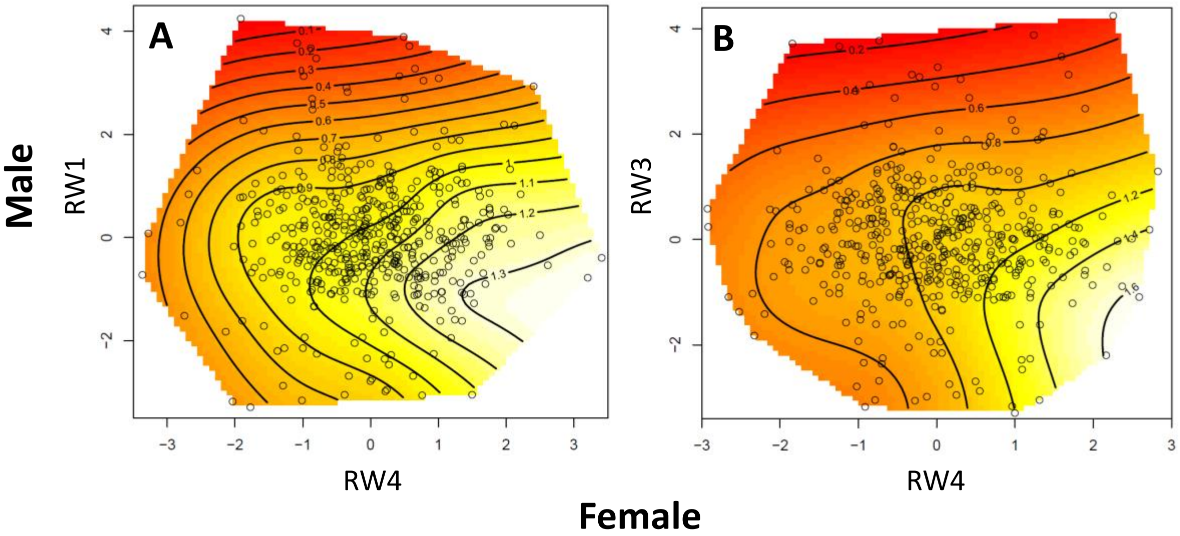

Males
Females
A
B
C
D
E
F
m5
m4
-3
-2
-1
0
1
2
3
2
1
0
-1
-2
-3
-4
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
0.1
0

Figure 3


A
B
Male
RW1
RW3
RW4
RW4
Female
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
0.2
0.4
0.6
0.8
1.2
1.4
1.6
4
2
0
−2
−3
−2
−1
0
1
2
3