

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

Running title: Stabilizing selection on genital morphology

Clarissa House^{1,*}, Philip Tunstall², James Rapkin², Mathilda J. Bale², Matthew Gage³, Enrique del Castillo⁴ 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.*

Corresponding author: E-mail: C.House@westernsydney.edu.au

Author contributions. CMH, MG, EC & JH conceived and designed the study. PT reared the beetles, conducted the mating assays, performed the dissections, morphometrics and collected the data. MJB reared the diet beetles, performed the dissections, morphometrics and collected the diet data. JR, EC & JH analysed the data. CMH & JH wrote the manuscript. All authors commented on earlier drafts.

Acknowledgements

CMH was funded by a Leverhulme Trust Fellowship, JR by a NERC studentship, EDC was partially funded by a NSF grant (CMII 1634878) and JH by a Royal Society University

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/evo.13912](https://doi.org/10.1111/evo.13912).

This article is protected by copyright. All rights reserved.

Fellowship (UF0762844) and research grants from NERC (NE/G00949X/1) and the ARC (DP180101708).

Data accessibility. Data will be deposited in Dryad once the manuscript is accepted for publication.

Competing interests. We have no competing interests.

Abstract

Male genitals are highly 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 selection targeting female genitals or how male and female genitals interact during mating. Here we use formal selection analysis to show that genital size and shape is subject to strong multivariate stabilizing sexual selection in both sexes of the red flour beetle, *Tribolium castaneum*. Moreover, we show significant sexual selection on the covariance between the sexes for specific aspects of genital shape suggesting that male and female genitalia also interact to determine the successful transfer of a spermatophore during mating. Our work therefore highlights the important role that both male and female genital morphology plays in determining mating success and that these effects can occur independently, as well as through their interaction. Moreover, it cautions against the overly-simplistic view that the sexual selection targeting genital morphology will always be directional in form and restricted primarily to males.

Key words. genitals, selection analysis, fitness peak, lock-and-key hypothesis

Introduction

Male genitals in animals with internal fertilization are widely regarded as being among the most 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: Stockley 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 (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 genitals have largely focussed on three main hypotheses: the lock-and-key, the pleiotropy and the sexual selection hypotheses (Hosken and Stockley 2004). Originally, the lock-and-key hypothesis proposes that genital divergence has evolved to prevent hybridization by ensuring that only males of the correct species were able to provide the right “key” for the female “lock” (Dufour 1844). More recently, however, Arnqvist (1997) proposed that a similar process could also operate within a species, with the poor alignment of male and female genitals reducing male mating success. 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 different

processes, most notably cryptic female choice for males with genitals that are better able to stimulate them during mating or through sperm competition and/or sexual conflict, all have the potential to drive genital divergence (Eberhard 1985; Hosken and Stockley 2004). While there is a degree of empirical support for both the lock-and-key (e.g. Wojcieszek and Simmons 2012, 2013; Anderson and Langerhans 2015) and pleiotropy (e.g. Arnqvist and Thornhill 1998; House and Simmons 2005) hypotheses, most support is currently for the sexual selection hypothesis (e.g. Hosken and Stockley 2004; Ah-King et al. 2014; Simmons 2014). This is not altogether surprising given the much larger number of studies testing this hypothesis (Ah-King et al. 2014) and highlights the need for more empirical studies focussed on the lock-and-key and pleiotropy hypotheses.

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 original lock-and-key hypothesis, it is predicted that a pattern of stabilizing natural selection prevents hybridization by ensuring that male genitals are species-specific and provide the correct “key” to fit the female “lock” (Dufour 1844). Moreover, within a species, the reduction in mating success that occurs when male genitals are poorly aligned with the average genital structure of females in the population is predicted to result in a pattern of stabilizing sexual selection targeting male genitals (Arnqvist 1997). 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). 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. 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 been shown to target some aspects of male genital morphology as well (seed bug: Tadler 1999; Dougherty and Shuker 2016; dung beetle: Simmons et al. 2009; millipede: Wojcieszek and Simmons 2012; broad horned beetle: House et al. 2016; water strider: Bertin and Fairbairn 2005). While this finding appears to support the sexual selection hypothesis, it is important to note that it is easier to statistically detect linear than stabilizing selection (Lande and Arnold 1983), and that there are also many empirical examples showing that sexual selection is not always linear in form (e.g. Blows et al. 2003; Chenoweth & Blows 2005; Brooks et al. 2005; Bentsen et al. 2006; Gerhardt and Brooks 2009; Steiger et al. 2013). Unfortunately, we currently lack similar formal estimates of selection for female genital morphology. In general, more comprehensive studies are needed (ideally exceeding several hundred individuals, Hersh and Phillips 2004) to characterize the pattern of linear and nonlinear selection acting on genital morphology, especially for females.

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). While peri-copulatory processes are likely to explain this outcome (Tyler and Tregenza 2012), the role that male and female genital morphology (or their interaction) play in the successful transfer of a spermatophore remains unknown. Here we use multivariate selection analysis (Lande and Arnold 1983) to characterize the strength and form of direct linear and nonlinear sexual selection acting independently on male and female genital size and shape in the red flour beetle *T. castaneum*. Having shown that the fitness surfaces for both sexes contain a peak, we then formally test whether this represents stabilizing selection. 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). To test this, we estimated the location of the global maxima and the 95% confidence region (CR) for each sex and mapped these alongside the fitness surfaces displaying the individual measures of genital size and shape. Finally, we estimated 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 for the successful transfer of a spermatophore during mating.

Materials and Methods

STOCK POPULATIONS AND REARING PROTOCOL

A total of six stock populations (~200 beetles per population) 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°C, 60% humidity and on a 16:8 hour light:dark cycle for 3 generations prior to the start of our experiment. Each generation, adult beetles were mixed at random between populations to ensure gene flow and the maintenance of genetic variation.

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 (10cm²; 25 cells per box, 2cm² 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

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 seconds, 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 seven 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. As we are interested in estimating the selection independently targeting male and female genitals, as well as the covariance between these structures during mating, we required intact genitals from both individuals in the interacting mating pair. When damage to the male or female genitals occurred, we therefore removed the mating pair from the data set leaving a final 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 were used as they each accounted for over 75% of the shape variation (Gutierrez et al. 2011). Our measurements of genital morphology were highly repeatable in both sexes (Table S1).

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 and female in our experiment, with one being assigned to the male and female in a pair that successfully obtained a mating and zero being assigned to the male and female in unsuccessful pairs. 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 (β), 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). In addition to these conventional models for each sex, we also fit a second multiple regression model to account for the interaction between male and female genitals during mating. In this model, we first estimated β using a first order linear multiple regression model that included CS and the four RW scores for each sex as the predictor variables, and relative fitness as the response

variable. We then estimated γ using a second order quadratic multiple regression model that included all linear, quadratic, and cross-product terms for each sex, as well as cross-product terms between each sex, as the predictor variables and relative fitness as the response variable. It was not our intention to interpret the selection gradients from this analysis but simply to demonstrate that including the genital morphology of both sexes and their interaction in the same model did not alter the sign or strength of the resulting selection gradients compared to the conventional models.

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) following the procedure outlined above.

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 Bisgaard 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).

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

According to Lande & Arnold (1983), a phenotypic trait subject to stabilizing selection is characterised by a negative standardized quadratic selection gradient (i.e. $\gamma < 0$). The problem with this definition is that a negative value of γ does not guarantee that a peak in fitness will exist within the range of existing phenotypes sampled (Mitchell-Olds and Shaw 1987). This scenario could arise, for example, if there is a monotonic increase in fitness across the range of phenotypes examined, meaning the location of the fitness peak must be based purely on extrapolation (Mitchell-Olds and Shaw 1987). Mitchell-Olds and Shaw (1987) therefore proposed that stabilizing selection should be limited to cases where extreme phenotypes have lower fitness and maximum fitness occurs at some intermediate point in the phenotypic distribution. While Mitchell-Olds and Shaw (1987) provide a number of statistical tests to support this definition of stabilizing selection, a simpler approach is to visually compare the location of the fitness peak to the distribution of the phenotypic data. To this end, we estimated the location of the global maxima (fitness peak) and the 95% confidence region (CR) for the fitness surface of each sex and mapped these alongside the same fitness surfaces displaying the individual measures of genital size and shape.

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 can find confidence regions on the location of the maxima of either quadratic polynomial models or of more flexible thin plate spline models. This approach is provided by the “OptRegionTps” function in the OPTIMAREGION package of R (del Castillo et al. 2016).

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 $\alpha = 0.95$). 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 $(x_1, x_2 \dots x_n)$ 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 of the fitness surface.

Characterizing sexual selection on the interaction between male and female genital morphology

Since 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 between-sex correlational gradients by fitting a multiple regression model that included the standardized linear, quadratic and cross-product terms for each sex and the standardized cross-product terms between the sexes 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 between-sex covariance matrix from this analysis as the interaction matrix and consider the terms in this matrix as being analogous to conventional Lande and Arnold (1983) selection gradients for subsequent interpretation. 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.

RESULTS

Geometric morphometric analyses yielded a measure of centroid size (CS) and four RW scores for each sex that together explained 78.40% and 85.38% of the total variation in male and female 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 aedeagus 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 negative quadratic 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 negative quadratic selection on CS, RW1 (length and width of the vaginal aperture) and RW4 and negative correlational selection on CS and RW1. Importantly, a full model that included the genital morphology of both sexes and their interaction produced quantitatively similar gradients (Table S2): that is, there were strong positive correlations between the linear ($r = 0.99$, $n = 10$, $P = 0.0001$), quadratic ($r = 0.99$, $n = 10$, $P = 0.0001$) and correlational (correlational: $r = 0.92$, $n = 20$, $P = 0.0001$) selection gradients from our conventional models and this full model.

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). The global maxima and their

associated 95% CRs 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 Figure 2A and 2B, respectively. We can therefore be confident in formally defining the observed pattern of nonlinear selection in the sexes as multivariate stabilizing sexual selection (Mitchell-Olds and Shaw 1987).

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).

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 Stockley 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, being characterised by a well-defined peak in fitness at intermediate values of genital size and shape in both sexes. 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. 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 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 for use in single species studies, with the form of selection targeting male genital morphology being an important criterion in this checklist (Arnqvist 1997). Our finding that the effects of male genital size and shape on mating success in *T. castaneum* generates a pattern of multivariate stabilizing sexual selection on this trait is 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. In insects, linear selection on male genital morphology (e.g. 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 (e.g. 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 more 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. 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, this produces 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 (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). This bias is likely to be further compounded by the fact that variation in male genital morphology is typically large across populations and species (Wojcieszek and Simmons 2012; Soto et al. 2013; Hosken et al. 2019) making them far easier to study. Our work, however, directly challenges these views 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 describing 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).

Although the strength and statistical significance of our estimated selection gradients indicates that stabilizing selection is the predominant form of selection targeting male and female genital morphology in *T. castaneum*, it is important to note that significant linear selection was also detected in both sexes for several aspects of genital shape (RW1, RW3 and RW4 in males and RW3 and RW4 in females) and also along the major eigenvectors of selection (\mathbf{m}_2 in males and \mathbf{m}_2 and \mathbf{m}_4 in females). Exactly how these aspects of genital shape provide an advantage to males and females during the transfer of a spermatophore is currently unclear, although it is possible that it helps better position the male genitals or the ability of the female genitals to accommodate these structures during the transfer of a spermatophore (e.g. Werner & Simmons 2008). It is important to highlight, however, that because we only examined a single episode of selection (i.e. mating success), this is likely to provide an incomplete measure of all the linear selection targeting male and female genital morphology, and consequently the total sexual selection acting on these traits. This is particularly true given that a number of post-copulatory processes in insects, including sperm competition (e.g. Danielsson & Askenmo 1999; Wenninger & Averill 2006) and cryptic female choice (e.g. Cordoba-Aguilar 1999), are known to impose significant linear selection on male genital morphology, although the contribution of female genitals to these episodes of selection remain largely unknown (Ah-King et al. 2014). Consequently, the selection gradients we provide should only be interpreted within the context of mating

success and not considered as estimates of total selection acting on male and female genital morphology.

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. castaneum*, 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, as isolating a mechanism requires functional studies (Brennan & Prum 2015; Ah-King et al. 2014; Simmons 2014). Unfortunately, only a handful of such studies exist for insects (Ronn et al. 2007; Werner & Simmons 2008; Polak & Rashed 2010; Kahn et al. 2010; Hotzy et al. 2012). For example, Werner and Simmons (2008) used histology to show that three of 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. castaneum* 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. However, detailed functional studies are clearly needed to confirm this.

The pattern of sexual selection we document for male and female genital morphology in *T. castaneum* is likely to provide a number of important insights for genital evolution. First, strong linear and stabilizing selection is predicted to deplete the additive

genetic variance of phenotypic traits (Lande, 1979). This is especially true for male and female genital morphology in *T. castaneum*, where our estimated quadratic selection gradients were almost twice the median strength ($\gamma = |0.16|$) reported across studies based on mating success (Kingsolver et al. 2001). Indeed, where stabilizing selection has been demonstrated it appears equally as strong, at least for insects ($\gamma = -0.28$, Tadler 1999; $\gamma = -0.19$ and -0.34 , House et al. 2016; $\gamma = -0.38$, Dougherty and Shuker 2016). Yet, genital morphology is known to evolve rapidly (e.g. Simmons et al. 2009; House et al. 2013; Hopwood et al. 2016) and contain as much additive genetic variance as non-genital traits (e.g. Arnqvist and Thornhill 1998; House and Simmons 2005; Higgins et al. 2009; Kamimura and Iwase 2010). This suggests that mechanism(s) must exist to preserve the additive genetic variance of genital morphology, as has been argued more generally for sexual traits subject to strong selection (Kirkpatrick and Ryan 1991). Second, in a population subject to persistent stabilizing selection in the absence of frequency-dependent selection, theory predicts that the population mean phenotype should evolve to match the peak in fitness (Simpson 1953; Lande 1979). Indeed, Estes and Arnold (2007) showed that a quantitative genetic model where the fitness optimum was able to evolve to the optimum within an adaptive zone with stable boundaries performed significantly better than five other competing models to explain the evolution of phenotypic means in an extensive database (Gingerich 2001). If the presence of stabilizing selection and the evolutionary response of the population mean is more common than generally appreciated for genitals, as might be expected under the lock-and-key hypothesis, it is relatively easy to envisage intra- and inter-specific 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 2012; Heinen-Kay and Langerhans 2013; Oneal and Knowles 2013) and experimental (e.g. Simmons et al. 2009; House et al. 2013; Hopwood et al. 2016) populations, as well as the extreme diversification in genitals observed across closely related species (e.g. Arnqvist 1998; Brindle and Opie 2016; Kuntner et al. 2016). Finally, theory predicts that correlational selection will generate a genetic correlation between the two traits by creating linkage disequilibrium (Lande 1984) or by favouring pleiotropic mutations (Lande 1980), although these mechanisms are not strictly required if correlational selection is strong and persistent

(Sinervo and Svensson 2002). Although we do not know of any formal models examining whether correlational selection can also generate intersexual genetic correlations, there is empirical evidence showing that artificial selection on the covariance between male and female traits can dramatically alter the strength of the intersexual genetic correlation between these traits (Delph et al. 2011; Stewart and Rice 2018). It is therefore possible that the correlational selection we document on the covariance between aspects of genital shape in male and female *T. castaneum* will also promote these trait being genetically correlated across the sexes. Ultimately, this will facilitate the co-evolution of male and female genital shape in *T. castaneum* (Lande 1980), and may explain why the sexual co-evolution of genitals has been so widely documented in both experimental evolution and comparative studies (reviewed in Brennan and Prum 2015).

In conclusion, our study shows that both male and female genital morphology is subject to strong multivariate stabilizing sexual 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 within species “lock-and-key” hypothesis of genital evolution (Arnqvist 1997), although we cannot determine at this stage whether this process is driven by a mechanical or sensory-based interaction (or both) during mating. The pattern of sexual selection we document for male and female genital morphology in *T. castaneum* is likely to have a number of important implications for genital evolution, including explaining the adaptive radiation of genitals across populations and the diversification of closely related species, as well as the co-evolution of male and female genital morphology.

References

- Ah-King, M., A. B. Barron and 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.
- 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.
- Arnqvist, G. and R. Thornhill. 1998. Evolution of animal genitalia: patterns of phenotypic and genotypic variation and condition dependence of genital and non-genital morphology in water strider (Heteroptera: Gerridae: Insecta). *Genet. Res.* 71: 193-212.
- Arnqvist, G. and I. Danielsson. 1999. Copulatory behaviour, genital morphology, and male fertilization success in water striders. *Evolution*: 53:147-156.
- 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.
- Bentsen, C.L., J. Hunt, M.D. Jennions and R. Brooks. 2006. Complex multivariate sexual selection on male acoustic signalling in a wild population of *Teleogryllus commodus*. *Am. Nat.* 167:E102-E116.
- Bertin, A. and D.J. Fairbairn. 2005. One tool, many uses: precopulatory sexual selection on genital morphology in *Aquarius remigis*. *J. Evol. Biol.* 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. 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.
- Blows, M.W. and A.A. Hoffmann. 2005. A reassessment of genetic limits to evolutionary change. *Ecology* 86:1371-1384.
- Bookmythe, I., M.L., Head, S. Keogh and M.D. Jennions. 2016. Fitness consequences of artificial selection on relative male genital size. *Nat. Comm.* 7:11597.
- Brennan, P.L.R., C. Clark and R.O. Prum. 2010. Explosive eversion and functional morphology of the duck penis supports sexual conflict in waterfowl genitalia. *Proc. R. Soc. B* 277:1309-1314.
- Brennan, P.L.R. and R.O. Prum. 2015. Mechanisms and evidence of genital coevolution: The roles of natural selection, mate choice, and sexual conflict. *Cold Spring Harb. Perspect. Biol.* 7:a017749.
- Brennan, P.L.R., I. Gereg, M. Goodman, D. Feng and R.O. Prum. 2017. Evidence of phenotypic plasticity of penis morphology and delayed reproductive maturation in response to male competition in waterfowl. *The Auk* 134:882-893.

- Brooks, 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.
- Chenoweth, S.F. and M.W. Blows. 2005. Contrasting mutual selection on homologous signal traits in *Drosophila serrata*. *Am. Nat.* 165:281-289
- Cordoba-Aguilar, A. 1999. Male copulatory sensory stimulation induces female ejection of rival sperm in a damselfly. *Proc. R. Soc. Lond. B* 266:779-784.
- Cordoba-Aguilar, A. 2002. Sensory trap as the mechanism of sexual selection in a damselfly genitalic trait (Insecta:Calopterygidae). *Am. Nat.* 160:594-601.
- Cordoba-Aguilar, A. 2009. Seasonal variation in genital and body size, sperm displacement ability, female mating rate, and male harassment in two calopterygid damselfies (Odonata: Calopterygidae). *Biol. J. Linn. Soc.* 96:815-829.
- Danielsson, I. and C. Askenmo. 1999. Male genital traits and mating interval affect male fertilization success in the water strider. *Behav. Ecol. Socio.* 46:149-156.
- del Castillo, E., J. Hunt and J. Rapkin. 2016. OptimaRegion: confidence regions for Optima. R package version 0.2.<https://CRAN.R-project.org/package=OptimaRegion>.
- Delph, L.F., Steven, J.C., Anderson, I.A., Herlihy, C.R. and E.D. Brodie III. 2011. Elimination of a genetic correlation between the sexes via artificial correlational selection. *Evolution* 65: 2872-2880.

- Dougherty, L.R. and D.M. Shuker. 2016. Variation in pre- and post-copulatory sexual selection on male genital size in two species of lygaeid bug. *Behav. Ecol. Socio.* 70:625-637.
- Dufour, L. 1844. Anatomie générale des diptères. *Ann. Sci. Nat.* 1:244-264.
- Eberhard, W.G., B.A. Huber, S. Rafael Lucas Rodriguez, R.D. Briceno, I. Salas and 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. 2001. Genitalic behavior in *Hybosciara gigantea* (Diptera: Sciaridae) and the evolution of species-specific genitalia. *J. Kans. Entomol. Soc.* 74:1-9.
- Endler, J.A. and A.L. Basolo. 1998. Sensory ecology, receiver biases and sexual selection. *TREE* 13:415-420.
- Estes, S. and S.J. Arnold. 2007. Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* 169:227-244.
- Fedina, T.Y. and S.M. Lewis. 2008. An integrative view of sexual selection in *Tribolium* flour beetles. *Biol. Rev.* 83:151-171.
- Foellmer, M.W. 2008. Broken genitals function as mating plugs and affect sex ratios in the orb-web spider *Argiope aurantia*. *Evolutionary Ecology Research* 10:449-462.
- 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.

- Gingerich, P.D. 2001. Rates of evolution on the time scale of the evolutionary process. In *Contemporary Microevolution: Rate, Pattern, and Process*, ed. AP Hendry, MT Kinnison, Dordrecht: Kluwer Acad., Genetica 112/113:127-144.
- 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 Scutigermorpha (Chilopoda). ZooKeys 66:49–66.
- Hansen, T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341-1351.
- Heinen-Kay, J.L. and R.B. Langerhans. 2013. Predation-associated divergence of male genital morphology in a livebearing fish. J. Evol. Biol. 26:2135-2146.
- Hersch, E.I. and P.C. Phillips. 2004. Power and potential bias in field studies of natural selection. Evolution 58: 479-485.
- Higgins, S.L., Hosken, D.J. and N. Wedell. 2009. Phenotypic and genetic variation in male genitalia in the seedbug, *Lygaeus equestris* (Heteroptera). Biol. J. Linn. Soc. 98: 400-405.
- Holwell, G.I., Winnick, C., Tregenza, T. and M.E. Herberstein. 2010. Genital shape correlates with sperm transfer success in the praying mantis *Ciulfina klassi* (Insecta: Mantodea). Behav. Ecol. Socio. 64:617-625.
- Hopwood, P.E., Head, M.L., Jordan, E.J., Carter, M. Davey, E., Moore, A.J. and N.J. Royle. 2016. Selection on an antagonistic behavioral trait can drive rapid genital coevolution in the burying beetle, *Nicrophorus vespilloides*. Evolution 70: 1180-1188.

- Hosken D.J., P. Stockley. 2004. Sexual selection and genital evolution. *Trends Ecol. Evol.* 19:87-93.
- Hotzy, C., M. Polak, J.L. Ronn and G. Arnqvist. 2012. Phenotypic engineering unveils the function of genital morphology. *Current. Biol.* 22:2258-2261.
- House, C.M. and Simmons, L.W. 2003. Genital morphology and fertilization success in the dung beetle *Onthophagus taurus*: an example of sexually selected male genitalia. *Proc. R. Soc. Lond. B* 270: 447-455.
- House, C.M. and Simmons, L.W. 2005. The evolution of male genitalia: patterns of genetic variation and covariation in the genital sclerites of the dung beetle *Onthophagus taurus*. *J. Evol. Biol.* 18: 1281-1292.
- House, C.M., Z. Lewis, D.J. Hodgson, N. Wedell, M.D. Sharma, J. Hunt and D.J. Hosken. 2013. Sexual and natural selection both influence male genital evolution. *PLoS ONE* 8: e63807.
- House, C.M., M.D. Sharma, K. Okada and D.J. Hosken. 2016. Pre and post-copulatory selection favour similar genital phenotypes in the male broad horned beetle. *Integr. Comp. Biol.* 56: 682-693.
- Hosken, D.J., Archer, C.R., House, C.M. and D.J. Hosken. 2019. Penis evolution across species: divergence and diversity. *Nat. Rev. Urol.* 16: 98-106.
- Hunt, J., C.J. Breuker, J.A. Sadowski and A.J. Moore. 2010. Male-male competition, female mate choice and their interaction: determining total sexual selection. *J. Evol. Biol.* 22:13-26.
- Johnson, S.G. 2014. The NLOpt nonlinear optimization package. <http://ab-initio.mit.edu/nlopt>.

- Kahn, A.T., B. Mautz and M.D. Jennions. 2010. Females prefer to associate with males with longer intromittent organs in mosquitofish. *Biol. Lett.* 6:55-58.
- Kamimura, Y. and R. Iwase. 2010. Evolutionary genetics of genital size and lateral asymmetry in the earwig *Euborellia plebeja* (Dermaptera: Anisolabididae). *Biol. J. Linn. Soc.* 101: 103-112.
- King, R.B., R.C. Jadin, M. Grue and H.D. Walley. 2009. Behavioural correlates with hemipenis morphology in new world natricine snakes. *Biol. J. Linn. Soc.* 98:110-120.
- Kingsolver, J.G., H.E. Hoekstra, J.M. Hoekstra, D. Berrigan, S.N. Vignieri, C.E. Hill, A. Hoang, P. Gibert and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* 157:245-261.
- Kirkpatrick, M. and M.J. Ryan. 1991. The evolution of mating preferences and the paradox of the lek. *Nature* 350: 33-38.
- Klaczko, J., T. Ingram and J. Losos. 2015. Genitals evolve faster than other traits in *Anolis* lizards. *J. Zool.* 295:44-48.
- Kuntner, M., R-C. Cheng, S. Kralj-Fišer, C. Liao, J.M. Schneider and M.A. Elgar. 2016. The evolution of genital complexity and mating rates in sexually size dimorphic spiders. *BMC Evolutionary Biology* 16:242.
- Lande, R. 1979. Quantitative genetical analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33:402-416.
- Lande, R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* 94:203-215.
- Lande, R. 1984. The genetic correlation between characters maintained by selection, linkage and inbreeding. *Genet. Res. Camb.* 44: 309-320.

- 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.
- 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.
- Manly, B.F.J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. Chapman and Hall, London, UK.
- Mayr, E. 1963. Animal species and evolution. Harvard Univ. Press
- Mitchell-Olds, T. and R.G. Shaw. 1987. Analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41:1149-1161.
- Oneal, E. and L.L. Knowles. 2013. Ecological selection as the cause and sexual differentiation as the consequence of species divergence. *Proc. R. Soc. B* 280: 20122236.
- Pai, A., B. Lauren and Y. Guiyun. 2005. Female multiple mating for fertility assurance in red flour beetles (*Tribolium castaneum*). *Can. J. Zool.* 83:913-919.
- 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, M. and A. Rashed. 2010. Microscale laser surgery reveals adaptive function of male intromittent genitalia. *Proc. R. Soc. B* 277:1371-1376.

- Polakow, 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
- Ramm, S.A. 2007. Sexual selection and genital evolution in mammals: A phylogenetic analysis of baculum length. Am. Nat. 169:360-369.
- 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.
- Ronn, J., M. Katvala and G. Arnqvist. 2007. Coevolution between harmful male genitalia and female resistance in seed beetles. PNAS 104:10921-10925
- 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. 2014. Sexual selection and genital evolution. Austral Entomology 53:1-17.
- Simpson, G.G. 1953. The major features of evolution. Columbia Univ. Press, Garden City, NY.
- Sinervo, B. and E. Svensson. 2002. Correlational selection and the evolution of genomic architecture. Heredity 89: 329-338.
- Soto, I.M., Carreira, V.P., Soto, E. M., Márquex, F., Lipko, P. and Hasson, E. 2013. Rapid divergent evolution of male genitalia among populations of *Drosophila buzzatii*. Evol. Biol. 40: 395-407.
- Steiger, S., G.D. Ower, J. Stökl, C. Mitchell, J. Hunt and S.K. Sakaluk. 2013. Sexual selection on cuticular hydrocarbons of male sagebrush crickets in the wild. Proc. R. Soc. B 280:20132353.

- Stewart, A.D. and W.R. Rice. 2018. Arrest of sex-specific adaptation during the evolution of sexual dimorphism in *Drosophila*. *Nature Ecology & Evolution* 2: 1507-1513.
- 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.
- Stockley, P. 2002. Sperm competition risk and male genital anatomy: comparative evidence for reduced duration of female sexual receptivity in primates with penile spines. *Evolutionary Ecology* 16: 123-137.
- Tadler, A. 1999. Selection of a conspicuous male genitalic trait in the seedbug *Lygaeus simulans*. *Proc. R. Soc. Lond. B* 266:1773-1777.
- Tyler, F. and T. Tregenza. 2012. Why do so many flour beetle copulations fail? *Entomologia Experimentalis et Applicata* 146:199-206.
- Van Lieshout, E. 2011. Male genital length and mating status differentially affect mating behaviour in an earwig. *Behav. Ecol. Socio.* 65:149-156.
- Van Lieshout, E. and M.A. Elgar. 2011. Longer exaggerated male genitalia confer defensive sperm-competition benefits in an earwig. *Evol. Ecol.* 25:351-362.
- Wenninger, E.J. and A.L. Averill. 2006. Influence of body and genital morphology on relative male fertilization success in oriental beetle. *Behav. Ecol.* 17:656-663.
- Werner, M. and L.W. Simmons. 2008. The evolution of male genitalia: functional integration of genital sclerites in the dung beetle *Onthophagus taurus*. *Biol. J. Linn. Soc.* 98:257-266.
- Wojcieszek, J.M. and L.W. Simmons. 2012. 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.
- 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/>.

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*. We provide r^2 estimates for the linear model used to estimate β and also for the second order quadratic model (that includes all linear, quadratic, and cross-product terms) to estimate γ in each sex. Randomization test: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

		γ				
	β	CS	RW1	RW2	RW3	RW4
A. Male	$r^2 = 0.030$			$r^2 = 0.106$		
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	$r^2 = 0.021$			$r^2 = 0.098$		
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 (θ_i) and quadratic (λ_i) gradient of sexual selection acting along each eigenvector (\mathbf{m}_i) are provided in the last two columns. The sign of λ_i describes the form of quadratic selection acting along each eigenvector, with a positive λ_i indicating disruptive selection and a negative λ_i indicating stabilizing selection. The strength of selection acting along each eigenvector is given by $|\lambda_i|$. Randomization test: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

	M					Selection	
	CS	RW1	RW2	RW3	RW4	θ_i	λ_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	0.062				
	RW1	0.092	0.016			
	RW2	-0.044	-0.029	-0.013		
	RW3	0.036	-0.046	-0.031	0.032	
	RW4	-0.025	-0.076*	0.007	-0.085*	0.031

Figure 1

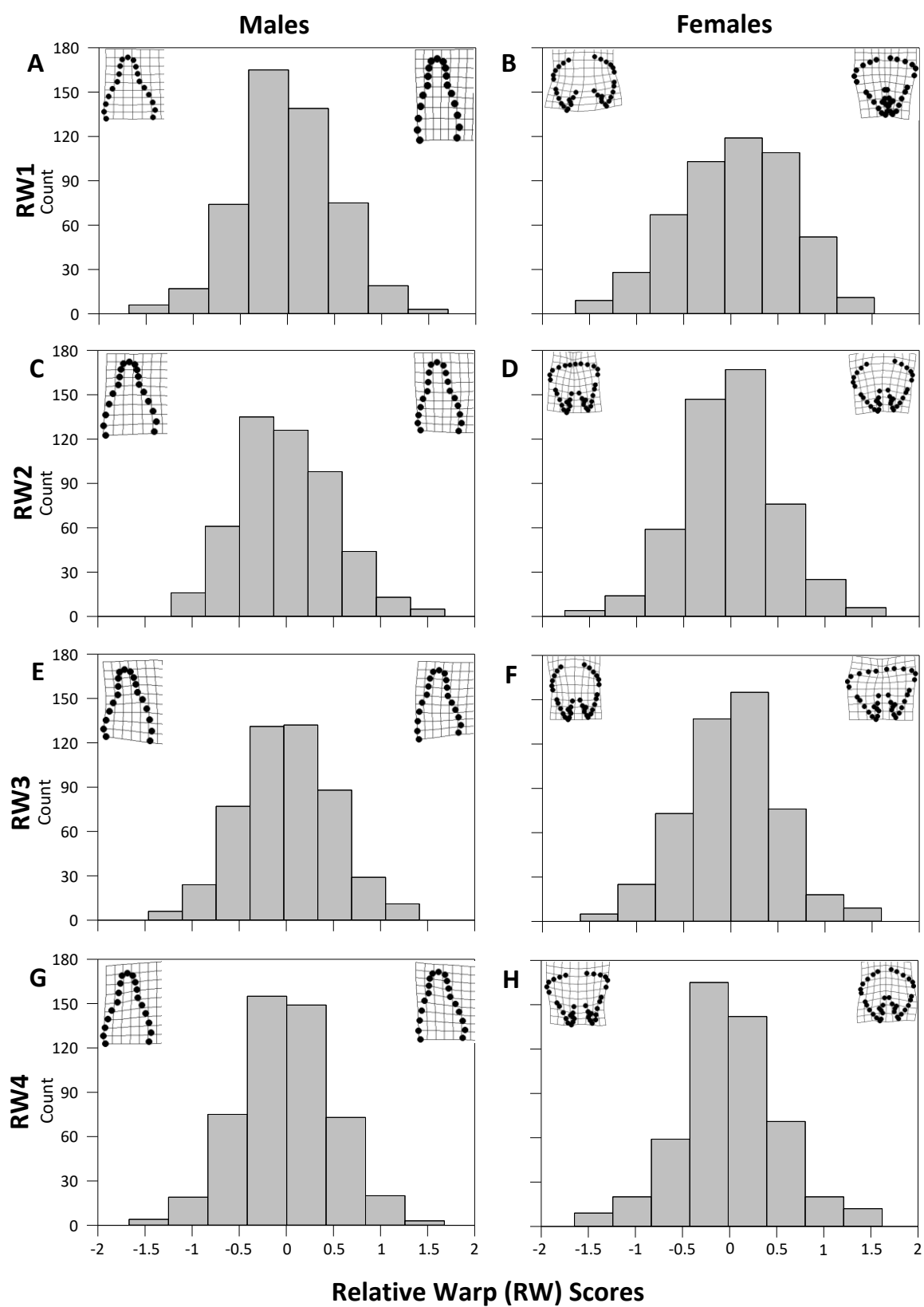


Figure 1. Frequency distribution of the four relative warp (RW) scores characterizing the variation in male (A, C, E and G) and female (B, D, F and H) genital shape. For each RW, we provide thin-plate spline visualizations (inset) that characterize a positive and negative score.

Figure 2

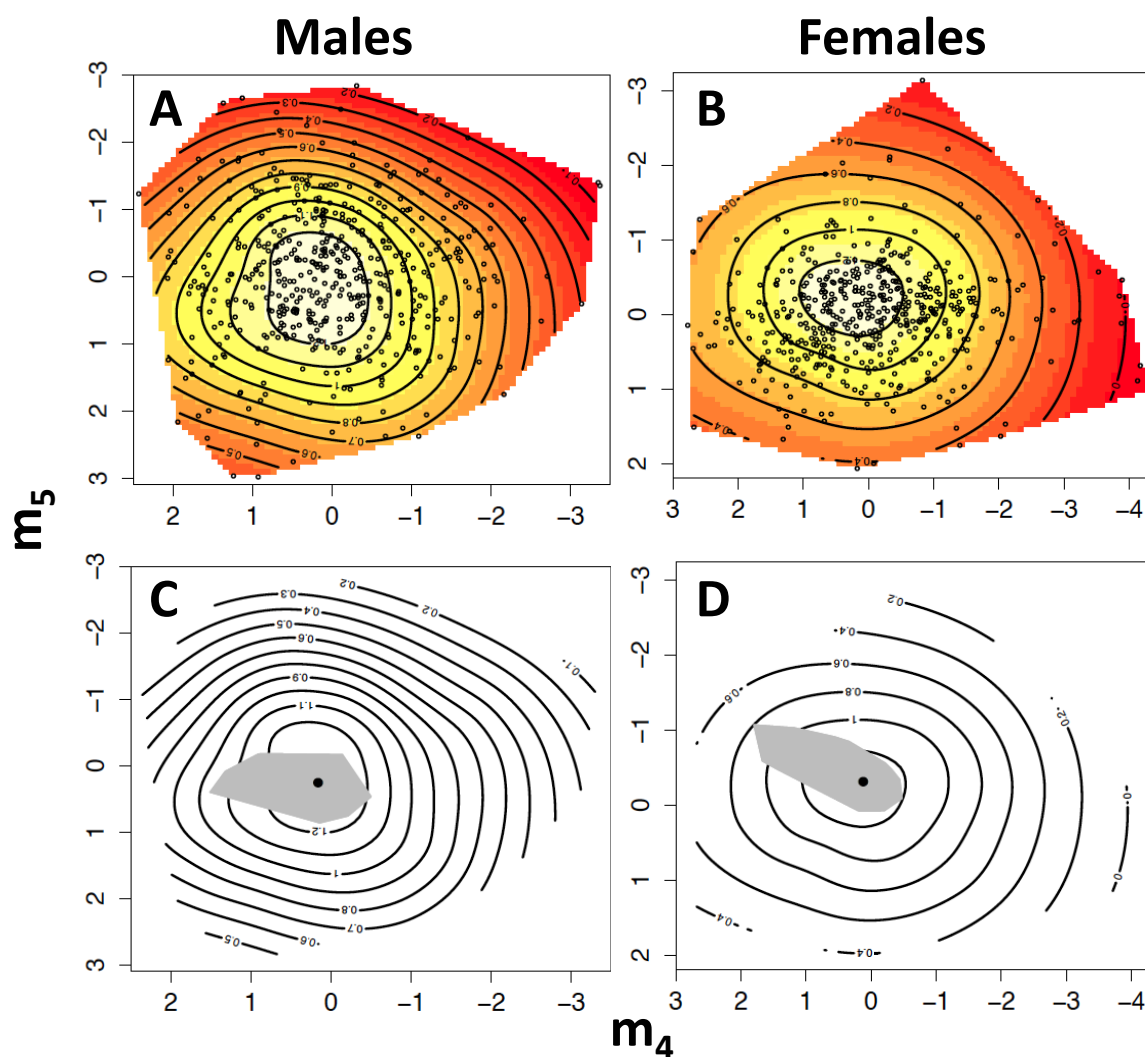


Figure 2. (A, B) Thin-plate spline visualizations (contour view) of the two major axes of nonlinear selection (m_4 and m_5) on the fitness surface for males and females, respectively. In each surface, white colouration represents regions of highest fitness, whereas red colouration represents regions of lowest fitness. Individual data points are provided as black circles on the surface. (C, D) 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 estimated location of the global maxima.

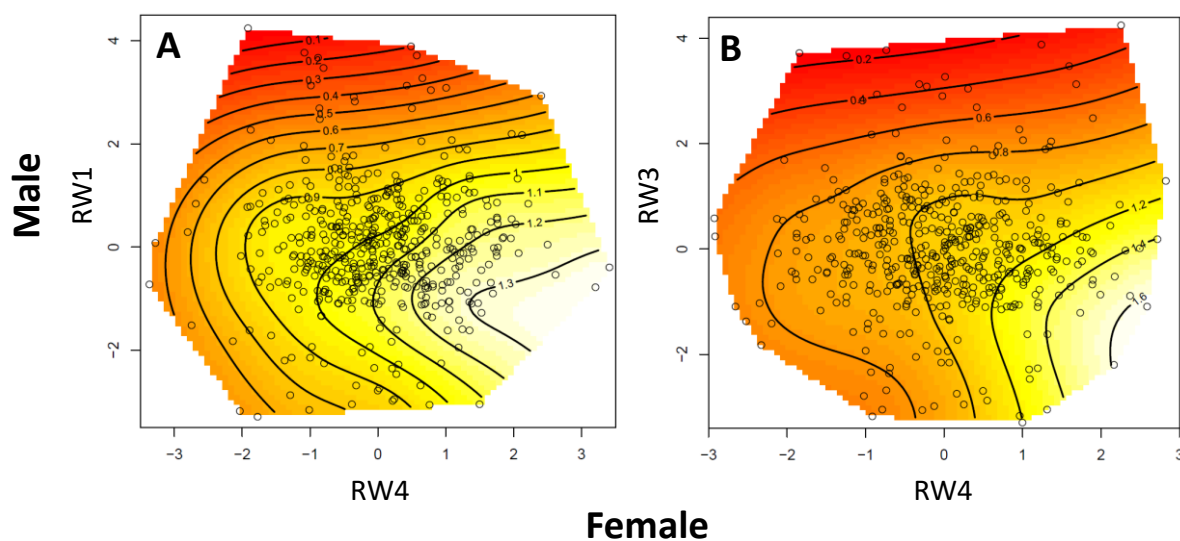
Figure 3

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.