OXFORD

Contrasting the form and strength of pre- and postcopulatory sexual selection in a flatworm

Lucas Marie-Orleach[1,2,3], Matthew D. Hall[4], Lukas Schärer[1]

[1]Evolutionary Biology, Zoological Institute, University of Basel, Basel, Switzerland
[2]CNRS, Université de Rennes 1, ECOBIO (Écosystèmes, Biodiversité, Évolution)—UMR 6553, Rennes, France
[3]Institut de Recherche sur la Biologie de l'Insecte, UMR 7261, CNRS—Université de Tours, Tours, France
[4]School of Biological Sciences, Monash University, Melbourne, Australia

Corresponding author. Institut de Recherche sur la Biologie de l'Insecte, UMR 7261, CNRS—Université de Tours, 37200 Tours, France.
Email: lucas.marie-orleach@univ-tours.fr

Abstract

Sexual traits may be selected during multiple consecutive episodes of selection, occurring before, during, or after copulation. The overall strength and form of selection acting on traits may thus be determined by how selection (co-)varies along different episodes. However, it is challenging to measure pre- and postcopulatory phenotypic traits alongside variation in fitness components at each different episode. Here, we used a transgenic line of the transparent flatworm *Macrostomum lignano* expressing green fluorescent protein (GFP) in all cell types, including sperm cells, enabling in vivo sperm tracking. We assessed the mating success, sperm-transfer efficiency, and sperm fertilizing efficiency of GFP(+) focal worms in which we measured 13 morphological traits. We found linear selection on sperm production rate arising from pre- and postcopulatory components and on copulatory organ shape arising from sperm fertilizing efficiency. We further found nonlinear (mostly concave) selection on combinations of copulatory organ and sperm morphology traits arising mostly from sperm-transfer efficiency and sperm fertilizing efficiency. Our study provides a fine-scale quantification of sexual selection, showing that both the form and strength of selection can change across fitness components. Quantifying how sexual selection builds up along episodes of selection allows us to better understand the evolution of sexually selected traits.

Keywords: Cryptic female choice, Linear and nonlinear selection, Mate choice, Measuring selection, Response surface methodology, Sperm competition.

Introduction

Sexual selection may act before, during, and after copulation, and it represents a compelling framework to explain the evolution of sexual traits involved in mate acquisition, copulation, and fertilization (Andersson, 1994; Birkhead et al., 2009; Prokuda & Roff, 2014). Traits under precopulatory sexual selection affect the opportunity to copulate. For instance, in the fungus beetles, *Bolitotherus cornutus*, males with longer horns have better access to females and therefore sire more offspring than their competitors (Conner, 1988). Traits under postcopulatory sexual selection affect the number of offspring sired per mating opportunity. For instance, in the cricket *Gryllus bimaculatus*, males producing more and smaller sperm sire more offspring than their competitors (Gage & Morrow, 2003). However, pre- and postcopulatory sexual selection may not be independent from one another. For instance, males that have better access to females may consistently sire more (e.g., McCullough et al., 2018; McDonald et al., 2017) or fewer (e.g., De Nardo et al., 2021) offspring per mating opportunity than their competitors. In fact, we know very little about how selection operates across pre- and postcopulatory fitness components, despite precopulatory traits often co-varying with postcopulatory sexual traits, either positively or negatively (Evans & Garcia-Gonzalez, 2016; Mautz et al., 2013).

Sexual selection studies have traditionally focused on either the interaction between pre- and postcopulatory fitness components (e.g., McCullough et al., 2018; McDonald et al., 2017; Morimoto et al., 2019) or on measuring selection for combinations of pre- and postcopulatory traits on total fitness (e.g., Devigili et al., 2015; House et al., 2020). Bridging this knowledge gap instead requires the formal comparison of selection acting on the combinations of phenotypic traits that determine pre- and postcopulatory fitness. If the strength or direction of selection changes across each pre- and postcopulatory fitness component, then the net selection acting on any given trait may be reinforced or negated depending on whether each episode acts synergistically or antagonistically. In considering indirect selection on correlated traits, it is also key to study how trait variation predicts fitness in components where selection is expected to occur. For instance, sperm traits may determine male reproductive success, not because they confer higher success in sperm competition, but instead because they are correlated with a trait under precopulatory sexual selection.

Partitioning episodes of sexual selection has long been a feature of sexual selection studies, but it is typically applied to the contrasting effects of male competition and female choice (reviewed in Hunt et al., 2009). Here the relationship between trait variation and relative fitness is used to formally assess

the strength of sexual selection, which is typically assessed through the regression coefficient of a linear regression of fitness on one or multiple traits (Arnold & Wade, 1984; Henshaw et al., 2018; Jones, 2009; Lande & Arnold, 1983). This approach has documented a strong selection of traits that were assumed to be involved in pre- and postcopulatory processes (reviewed in Prokuda & Roff, 2014; Simmons & Moore, 2009). Sexual selection may, however, act in more complex and nonlinear ways, including quadratic selection that either favors or opposes intermediate trait values and correlational selection that may act on multiple traits simultaneously (Henshaw & Zemel, 2017; Lande, 1979; Lande & Arnold, 1983). This body of literature has shown that sexual selection often acts on combinations of pre- and/or postcopulatory traits (e.g., Bentsen et al., 2006; Devigili et al., 2015; Hall et al., 2008; House et al., 2016, 2020; Oh & Shaw, 2013; Simmons et al., 2009). For instance, in the red flour beetle, *Tribolium castaneum*, male and female genitalia are under concave selection, i.e., intermediate sizes and shapes are favored by selection (House et al., 2020). Moreover, in the live-bearing fish, *Poecilia reticulata*, selection measured on male reproductive success indicates multivariate fitness landscapes involving both male morphological traits and sperm velocity (Devigili et al., 2015).

To our knowledge, formal comparisons of the strength and form of selection arising from pre- and postcopulatory components of sexual selection have only been obtained once, namely, in the broad-horned flour beetle *Gnatocerus cornutus* (House et al., 2016) (albeit in a non-competitive setting). That study found that precopulatory mating success and postcopulatory fertilization success favored similar male genital phenotypes, but that concave selection on male body size arose exclusively from precopulatory mating success. More studies along these lines are clearly needed. Arguably, the scarcity of such studies may be due to the (a) high sample sizes required to perform multivariate selection analyses (Green, 1991; Simon et al., 2022), and (b) the challenges in measuring pre- and postcopulatory traits alongside the relevant—and possibly multiple—pre- and postcopulatory fitness components. For instance, although postcopulatory processes are often measured as a single episode of selection, in many species male sperm competitiveness is in fact considered to consist of "(i) the relative number of sperm of different males that enter the fertilizing pool; and (ii) the relative fertilization efficiency of an ejaculate, after controlling for sperm number" (Pizzari & Parker, 2009).

Here we aim to study how phenotypic selection builds up along pre- and postcopulatory fitness episodes by using powerful features of the free-living flatworm model *Macrostomum lignano* (see below). Specifically, we studied the strength and form of selection acting on multiple morphological traits with respect to four pre- and postcopulatory fitness components. For this, we measured 13 reproductive morphology traits (including body size, testis size, ovary size, seminal vesicle size, four male copulatory organ traits, and five sperm traits), and used multivariate selection analyses to statistically test if pre- and postcopulatory sexual selection act synergistically or antagonistically, target similar or dissimilar trait combinations, and differ in the form of selection (directional vs. nonlinear).

Specifically, we sequentially exposed focal worms to three independent mating groups (Figure 1), and assessed, in each group, the individual success of the focal worm in four fitness components (as defined in Marie-Orleach et al., 2016, 2021). In brief, we decomposed focal male reproductive success into the following four multiplicative fitness components: (a) partner fecundity (F; i.e., the number of offspring produced by all potential partners through their female sex function), (b) mating success (MS; i.e., the proportion of copulations in which the focal worm was involved), (c) sperm-transfer efficiency (STE; i.e., the proportion of focal sperm among the sperm received by all potential partners, given the focal mating success), and (d) sperm fertilizing efficiency (SFE; i.e., the proportion of offspring produced that are sired by the focal worm, given the proportion of focal sperm received) (see Marie-Orleach et al., 2016, 2021 for more details). We could estimate the last two fitness components because in vivo tracking of sperm is feasible in this transparent worm (Janicke et al., 2013; Marie-Orleach et al., 2014; Wudarski et al., 2017). By using transgenic focal worms that express green fluorescent protein (GFP) in all cell types, including the sperm cells, one can easily distinguish sperm cells donated by GFP(–) wild-type competitors from those that are donated by the GFP(+) focal worms, directly inside the female reproductive tract of living partners. This provides powerful opportunities to quantify fitness components that are usually difficult to observe, such as the number of sperm cells that a GFP(+) focal individual has successfully transferred to partners, and the resulting number of sired offspring (Janicke et al., 2013; Marie-Orleach et al., 2014, 2016, 2021).

Furthermore, we collected data to characterize selection acting on trait combinations that are potentially important during pre- and postcopulatory selection (Figure 1). These traits include testis size, which has previously been found to predict mating success (Janicke & Schärer, 2009a; b), sperm production rate (Schärer & Vizoso, 2007), sperm transfer success (Janicke & Schärer, 2009a; Marie-Orleach et al., 2016), and male reproductive success in this worm (Marie-Orleach et al., 2016; Vellnow et al., 2018). We also measured body size, ovary size, Δ seminal vesicle size (i.e., the increase in seminal vesicle size of focal worms after two days in isolation, a measure of the sperm production rate), the size and shape of the male copulatory organ (the so-called stylet) using geometric morphometrics, and five different sperm morphology traits to test how these predict male fitness (Janicke & Schärer, 2009a, 2010; Marie-Orleach et al., 2016). To consider all 13 traits jointly in a single analysis would require the estimation of 13 linear and 81 nonlinear selection gradients, and sample sizes (i.e., >500 individuals for medium effect sizes (Green, 1991)) that are not logistically possible when quantifying postcopulatory selection in this system. For example, over 1,800 partner worms and 26,000 copulations were involved in the quantification of the mating success of focal worms alone. Instead, we grouped traits into sets of biologically related subsets, covering general mating, stylet morphology, and sperm characteristics, and analyzed each group separately (see *Material and methods*).

Intense postcopulatory sexual selection is predicted in simultaneous hermaphrodites (Charnov, 1979; Marie-Orleach et al., 2021; Schärer & Pen, 2013; Schärer et al., 2014). We expected testis size and Δ seminal vesicle size to be under directional positive selection, since both may positively contribute to higher success in sperm-transfer efficiency. In contrast, we expected ovary size to be under negative linear selection because a trade-off between the male and female sex functions may lead to negative correlations between male reproductive success and traits involved in the female sex
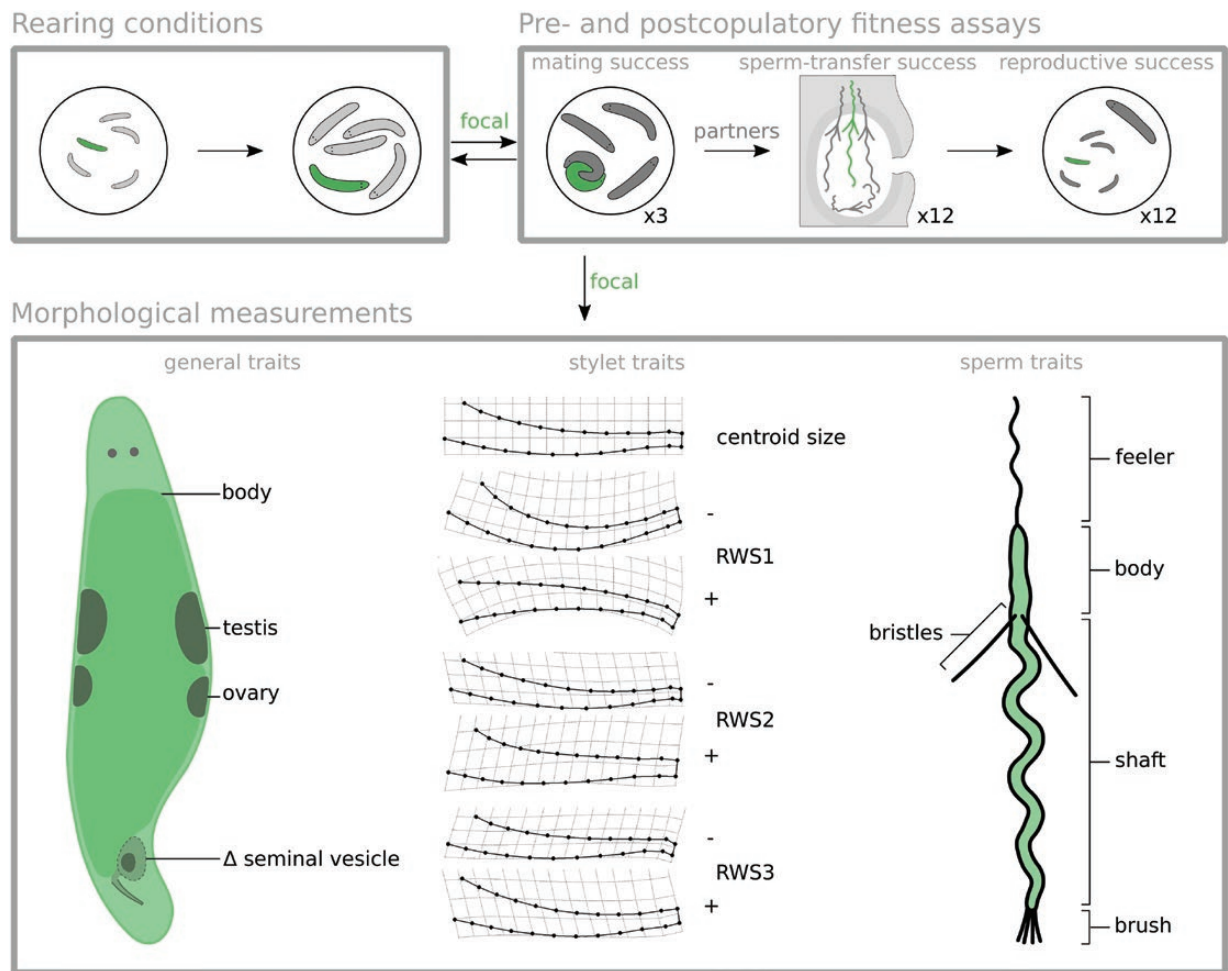
Figure 1. Overview of the experimental design used to measure pre- and postcopulatory morphological traits and fitness components. We allowed GFP(+) focal worms (green) to grow up under "Rearing conditions" with four GFP(−) partners, and then performed "Pre- and postcopulatory fitness assays" by subsequently exposing the focal worm to three independent groups of four worms to assess mating success, sperm-transfer success, and reproductive success, after which we performed "Morphological measurements" on different traits of the focal. We measured 13 traits that were divided into three trait sets. The 'general traits' set included the body, testis, ovary, and Δ seminal vesicle sizes (i.e., the change in seminal vesicle size after 2 days of isolation). The "stylet traits" set included the centroid size and the first three relative warp scores of a geometric morphometrics analysis. Diagrams show the configurations of the consensus stylet shape (top), the maximum and minimum values of the first relative warp score (RWS1), which mainly captured the overall stylet curvature, the second relative warp score (RWS2), which mainly captured the width of the stylet, and the third relative warp score (RWS3), which mainly captured the orientation of the tip. RWS1, RWS2, and RWS3 explained 55%, 18%, and 12% of the variation in stylet shape, respectively. The "sperm traits" set included the feeler, body, shaft, brush, and bristles sizes. See *Material and methods* for details.

function (Anthes et al., 2010). Several hypotheses have been proposed to explain the evolution of genitalia, predicting different forms of selection. Because interspecific studies usually find a lower allometric relationship for genitalia size than for other organs (i.e., the one-size-fits-all hypothesis), genitalia are often thought to be under concave selection (Eberhard et al., 1998; although other forms of selection may be expected on genitalia, see Arnqvist, 1997). Regardless of the form of selection, we expected selection on stylet traits to arise mostly from sperm-transfer efficiency. And finally, we expected directional selection on sperm traits arising from the sperm fertilizing efficiency fitness component. In particular, *M. lignano* displays a complex sperm morphology (Figure 1), including lateral stiff bristles that may be important for sperm competition (Schärer et al., 2011; Vizoso et al., 2010).

Material and methods

The data of the present study are associated with an experiment that is described in considerable detail elsewhere

(Marie-Orleach et al., 2021). In the previous paper, we used a variance-based approach measuring sexual selection exclusively through the variance in individual success (reported in Marie-Orleach et al., 2021). Here, we instead used a trait-based approach measuring sexual selection acting on morphological traits of the focal individuals. In the following, we briefly describe the parts of the material and methods shared with Marie-Orleach et al. (2021), and we also more fully explain the parts that are specific to the present study.

Model organism

The free-living flatworm, *Macrostomum lignano* (Macrostomorpha, Platyhelminthes) (Ladurner et al., 2005), is an outcrossing simultaneous hermaphrodite (Schärer & Ladurner, 2003) that is highly promiscuous (Janicke & Schärer, 2009a; Schärer et al., 2004). The copulation is reciprocal and consists of the intromission of the male copulatory organ (called stylet) into the female sperm-receiving and sperm-storage organ (called female antrum) of the partner (Schärer et al., 2004; Vizoso et al., 2010). Because the

worms are highly transparent, it is possible to observe and measure internal structures in vivo, such as testis and ovary size (Janicke et al., 2013; Schärer & Ladurner, 2003), stylet morphology (Janicke & Schärer, 2009a; Marie-Orleach et al., 2016), and the number of sperm cells stored inside the female antrum (Janicke et al., 2011; Marie-Orleach et al., 2016, 2021). Moreover, sperm traits can be measured after the amputation of a worm's tail plate (Janicke & Schärer, 2010). Little is known, however, about the biology of these flatworms under natural conditions. They are found in the upper intertidal zone of the Northern Adriatic and Aegean Sea (Ladurner et al., 2005; Schärer et al., 2020; Wudarski et al., 2020); but sexual selection studies were so far all carried out under laboratory conditions.

In vivo sperm, tracking is possible in *M. lignano* thanks to transgenic lines expressing GFP (Janicke et al., 2013; Marie-Orleach et al., 2014, 2021; Wudarski et al., 2017). Moreover, one can use the GFP marker to assign parentage in offspring (Marie-Orleach et al., 2014, 2021). Importantly, although worms from GFP(+) and GFP(−) lines differ significantly neither in body size, testis size, ovary size, or seminal vesicle size nor in their mating rate, siring ability, and female reproductive success (Marie-Orleach et al., 2014), we here always used GFP(+) worms as focal individuals and GFP(−) worms as partners to avoid potential artifacts in our analyses due to the expression of the GFP marker. Moreover, we used two outbred cultures, the GFP(+) BAS1 (Marie-Orleach et al., 2016; Vellnow et al., 2018) and the GFP(−) LS1 (Marie-Orleach et al., 2013). These cultures are expected to be genetically similar because BAS1 was established by introgression of the GFP marker into the LS1 culture, namely, by backcrossing a GFP(+) inbred line onto the LS1 outbred culture over nine generations (Marie-Orleach et al., 2016; Vellnow et al., 2018).

Experimental set-up

Our experimental set-up allowed us to measure the mating success, sperm-transfer success, and reproductive success of focal worms, for which we also measured 13 morphological traits (see below and Figure 1). For logistic reasons, the biological replicates used in the experiment were split into eight batches, treated sequentially (3–6 days apart), but in the following reporting we set day 1 as the first day for each batch.

Rearing conditions

To obtain same-aged individuals we, on day 1, allowed adult worms of the GFP(+) and GFP(−) cultures to lay eggs in Petri dishes for 24 h and, on day 6, we placed the resulting offspring in 24-well tissue culture plates to create 20 biological replicates per batch. Each biological replicate included (a) one group made up of one GFP(+) focal individual and four GFP(−) partners (called the A groups), and (b) three groups made up of four GFP(−) individuals (called the B, C, and D groups) that we used as partners of the focal worm (Figure 1). We then transferred all groups to new wells with fresh algae every 6–10 days. Importantly, in order to measure selection in individuals who had reached a steady state of sperm production, mating activity, sperm receipt, and egg production, we kept the worms in their A groups for several days after they had reached sexual maturity.

Mating success

We then estimated the focal individual's mating success on days 25–30 as explained in Marie-Orleach et al. (2021). In brief, we placed each focal individual in a mating chamber together with the four worms of its B group. To visually distinguish the focal worm from the partners, we placed all members of the A group into a well containing a blue vital dye for the 24 h before the mating trial (which does not affect sexual performance; Marie-Orleach et al., 2013). We then placed the five worms in an 8 μl drop of artificial sea water between two microscope slides and video-recorded all their interactions for 3 h. In total, we gathered 1,440 h of copulation interactions (analyzed anonymized with respect to treatment), which contained 26,723 copulations. For each mating group, we counted the total number of copulations (total matings) and the number of copulations in which the focal worm was involved (focal matings). After the mating trials, we transferred the focal worm back into its A group and isolated all four members of the B group to assess the sperm-transfer success of the focal worm during the mating trial.

Sperm-transfer success

We estimated the proportion of sperm cells received from the focal worm by its four partners following the protocol explained in Marie-Orleach et al. (2021). In brief, we recorded movies of the female antrum of each partner, first under bright-field illumination and then under epifluorescence illumination to, respectively, assess (a) the total number of sperm cells in all four potential partners (total sperm) and (b) the number of these sperm that were GFP(+) (focal sperm). The 3,840 resulting antrum movies were analyzed and anonymized with respect to treatment. The procedure used provides highly repeatable sperm counts (Marie-Orleach et al., 2014). Note that, when we could not assess total sperm in worms because they had eggs in their female antrum (783/1,920), we used the average total sperm counts computed from worms in which this could be estimated (i.e., 21 sperm cells).

This approximation could introduce some statistical noise in our data, potentially biasing our estimates of sperm-transfer efficiency and sperm-fertilizing efficiency. However, if our sperm-transfer success estimates were imprecise, one would expect a negative covariance between sperm-transfer efficiency and sperm-fertilizing efficiency. This is because an overestimation of sperm-transfer success leads to both an overestimation of sperm-transfer efficiency and an underestimation of sperm-fertilizing efficiency (and vice versa for an underestimation of sperm-transfer success). And because we found a non-significant covariance between sperm-transfer efficiency and sperm-fertilizing efficiency (Marie-Orleach et al., 2021), our estimates of sperm-transfer success should be fairly precise.

Reproductive success

We estimated the reproductive success of the focal individual by letting the partners lay eggs in isolation for 12 days, as explained in Marie-Orleach et al. (2021)—yielding 11,176 offspring in total. We counted the total number of offspring produced across all four potential partners (total offspring) and determined the number of these that were GFP(+) (focal offspring). We computed the proportion of offspring sired by the focal individual by dividing the sum of focal offspring by the sum of number of total offspring.

On the two subsequent days, we repeated all of the above-mentioned steps (*mating success*, *sperm-transfer success*, and *reproductive success*) by placing each focal

individual together with the four individuals of their C and D groups, respectively. We thus measured the same pre- and postcopulatory components of male reproductive success for each focal individual in three independent mating groups of four worms, sampled from the same pool of same-aged outbred worms.

Morphological measurements
On the next day, we assessed several morphological traits of the focal worms by following protocols described elsewhere (Janicke & Schärer, 2009a; Schärer & Ladurner, 2003). Briefly, these protocols consist of anesthetizing and squeezing the worm in between a microscope slide and a hemocytometer coverslip in a standardized way and taking digital pictures of the entire body, the two testes, the two ovaries, the seminal vesicle, and the stylet. For this, we used a Leica DM2500 microscope, an Imaging Source DFK 41AF02 camera, and BTV Pro 6.0b7. Focal worms were then isolated for 2 days, after which we measured them again to assess the morphological traits a second time, in order to (a) reduce measurement error and (b) estimate the increase in seminal vesicle size over 2 days (i.e., Δ seminal vesicle size), which can be used as a reliable proxy for the sperm production rate (Schärer & Vizoso, 2007). We analyzed the pictures anonymized with respect to treatment, using ImageJ, and we averaged the two measurements to assess the body size, testis size, and ovary size. Moreover, we assessed stylet morphology by using a geometric morphometric approach as in Janicke & Schärer (2009a), where we first analyzed all 320 stylet pictures in a single analysis, and then, for each focal worm, averaged the two measurements of stylet centroid size (CS) and of the first three relative warp scores (RWS). These RWS values together explained 85% of the total variance observed in stylet shape and mainly captured the overall curvature of the stylet (RWS 1), the width of the stylet (RWS2), and the orientation of the tip of the stylet (RWS 3). See Figure 1 for visualizations.

Immediately after the second measurement, we assessed five sperm traits following an established protocol (Janicke & Schärer, 2010), by amputating, squeezing, and thus rupturing the tail plate of the worm in a tiny drop, so that the sperm cells are released into the medium and become accessible for imaging. The pictures were then analyzed using ImageJ, anonymized with respect to treatment, to assess the length of the feeler, body, shaft, bristles, and brush (Janicke & Schärer, 2010). Note that we measured both bristles and used the averaged length. We assessed on average 7.7 sperm cells per individual (range from 1 to 11), and used the averaged values for the statistical analysis.

Penetrance of the GFP marker
Finally, as explained in Marie-Orleach et al. (2021), we estimated the penetrance of the GFP marker for each focal worm by pairing them with a virgin GFP(–) individual. We then assessed the GFP status of the resulting offspring (47.7 offspring screened per focal individual on average), which led to the exclusion of two focal worms. One produced 47% GFP(+) offspring, and the other one produced no offspring with the virgin GFP(–) individual. The other focal worms produced either 100% (n = 135) or between 90% and 100% GFP(+) offspring (n = 23). Because in our final dataset, the penetrance of the GFP marker of the worms did not significantly correlate with their mating success (Spearman's correlation; r_S = 0.01, N = 139, p = .93), sperm-transfer success

(r_S = −0.04, N = 139, p = .60), and paternity share (r_S = 0.08, N = 139, p = .36), we did not include it in the data analyses.

Data analysis
To compare the strength and form of selection acting on the measured morphological traits at pre- and postcopulatory fitness components, we measured (a) linear and (b) nonlinear selection on the original morphological traits, and (c) the nonlinear selection on composite traits that are more suited to test for nonlinear selection.

Following Marie-Orleach et al. (2021), we used four multiplicative fitness components, mating success (*MS*), sperm-transfer efficiency (*STE*), sperm fertilizing efficiency (*SFE*), and partner fecundity (*F*), which were computed over all 12 partner worms. Including the partner worm identity in the analysis would require another experimental design, in which we could visually distinguish each partner worm in the mating trials, which is currently not possible.

Before the analysis, all fitness data were relativized to a mean of 1 (Jones, 2009). Also, we transformed testis size (square root) and ovary size ($\log_{10}$) to account for data skewness, and then we standardized all morphological traits to a mean of 0 and a standard deviation of 1 (Jones, 2009). Importantly, we split our 13 morphological traits into three sets of traits. Such a procedure was necessary to avoid overfitting issues, which, in our case, would lead to poor power to detect nonlinear selection on the original morphological traits and spuriously significant nonlinear effects on the composite traits. Given our sample size, we restricted the number of traits to a maximum of five (which leads to 20 total predictors, including the quadratic and cross-product two-way interactions) (Green, 1991). This approach prevented us from accounting for potential correlational selection between traits belonging to different trait sets. The traits were split as follows: the *general traits set* (body size, testis size, ovary size, and Δ seminal vesicle size), the *stylet traits set* (CS, RWS1, RWS2, and RWS3), and the *sperm traits set* (feeler size, body size, shaft size, bristle size, and brush size). We split the 13 traits in this way due to a priori expectations. Namely, we expected traits of the *general traits set* to be involved mainly in *mating success* and *sperm-transfer efficiency*, those of the *stylet trait set* in *sperm-transfer efficiency*, and those of the *sperm traits set* in *sperm fertilizing efficiency*.

Our initial sample size was 160 replicates, but because we lost replicates due to developmental errors (n = 8), the penetrance of the GFP marker (n = 2), and missing measurements of one or more traits (n = 11), our final sample size was reduced to 139 replicates.

Selection on the original trait sets
We first studied whether there was linear and nonlinear selection acting overall traits contained in a given trait set. For this, we used a sequential model-building approach that statistically compares the fits of models with and without the terms of interest (see Appendix A of Chenoweth & Blows, 2005). Here, we specifically tested if the fits of linear models predicting fitness, and including only the intercept, are improved by adding, first, the linear terms of all traits contained in a trait set (i.e., testing for linear selection), and then, all quadratic and cross-product terms (i.e., testing for nonlinear selection). We did this analysis separately for all three trait sets (i.e., general, stylet, and sperm traits sets), and on relative male reproductive success and the four fitness components (i.e., *mRS**, *F**, *MS**, *STE**, and *SFE**).

Second, to test if linear and nonlinear selection measured in all three trait sets were consistent across the four fitness components, we used again the sequential model-building approach (Chenoweth & Blows, 2005). We built linear models predicting fitness in all four fitness components together, including the intercept and the linear terms, and tested if their fits were statistically improved by the addition of the interaction term *linear term × fitness component (Chenoweth & Blows, 2005)*. Significant interaction terms indicate that linear selection is different across fitness components (Chenoweth & Blows, 2005; House et al., 2016). We further tested if patterns of nonlinear selection were consistent over the fitness components by testing if the addition of the interaction terms *nonlinear terms × fitness component* statistically improved the fit of the models including the nonlinear terms (Chenoweth & Blows, 2005). When selection was inconsistent across fitness components, we tested if selection was different between each pair of fitness components. Note that because these analyses required using each replicate several times in a single analysis (i.e., once for each fitness component), we performed additional analyses including replicate IDs as a random effect, which did not qualitatively change the outcomes (data not shown).

We then estimated the linear selection gradients (β) for each trait, computed through the partial regression coefficients of multiple linear regression of *mRS**, *F**, *MS**, *STE**, and *SFE** separately on all morphological traits (Lande & Arnold, 1983), which indicated which specific morphological traits were experiencing linear selection. We also computed the matrix containing the quadratic and cross-product selection gradients, called the γ matrix, which we estimated through a multiple linear regression including all quadratic (γ_{ii}), and cross-product terms (γ_{ij}). We doubled the quadratic regression coefficients so that the estimates for concave and convex forms of selection correspond to the Lande and Arnold (1983) formulation (Stinchcombe et al., 2008). The γ matrix allowed us to determine which specific interaction terms were responsible for the nonlinear effect.

Nonlinear selection on composite traits
We studied nonlinear selection by using composite traits describing morphological trait variation, which were generated by canonical rotations of the γ matrix (Blows, 2007; Blows & Brooks, 2003; Phillips & Arnold, 1989). The canonical rotation of the γ matrix generates the so-called **M** matrix, which contains (a) the eigenvectors (m_i) describing the major axes of nonlinear selection (i.e., the loading of each original morphological trait), and (b) their eigenvalues (λ_i) describing the strength and form of nonlinear selection. A negative eigenvalue indicates a concave relationship between the fitness and the eigenvector (i.e., a "peak" in which high fitness is reached for intermediate values along the eigenvector) whereas positive eigenvalues indicate convex relationship (i.e., a "bowl" in which high fitness is reached for low and high values along the eigenvector). An advantage of the canonical analysis is that the composite traits of variation are orthogonal to one another, which makes the cross-product terms (γ_{ij}) null by definition. All nonlinear selection is thus captured by the quadratic terms (γ_{ii}), which provides better power to detect nonlinear selection (Blows & Brooks, 2003).

We then tested if there was nonlinear selection in each trait set, and if it was consistent over the four fitness components, this time using the composite traits in a sequential

model-building approach (Chenoweth & Blows, 2005; Garant et al., 2007; Parker et al., 2011). As above, we first tested if the additions of the quadratic terms significantly improved the fits of models predicting fitness and including only the linear terms. We tested if nonlinear selection was consistent across the four fitness components by testing if the addition of the interaction terms *nonlinear terms × fitness component* significantly improved the fit of models predicting fitness on all fitness components (Garant et al., 2007; Parker et al., 2011). As above, we ran additional analyses including replicate ID as a random effect, which again did not qualitatively change the outcomes (data not shown).

We then explored the eigenvalues and eigenvectors of the **M** matrix to determine whether the nonlinear selection was concave or convex, and which traits generated nonlinear selection. The statistical significance of the eigenvalues was tested by permutation of the fitness values (10,000 iterations) (Keagy et al., 2016).

Results

We present the results of the three trait sets sequentially, namely general traits, stylet traits, and sperm traits. And for each trait set we include three paragraphs presenting the results on (a) linear selection on total fitness, mRS*, and across fitness components; (b) nonlinear selection on total fitness, mRS*, using the original and composite traits; and (c) nonlinear selection across fitness components, using the original and composite traits.

General traits set
We found significant linear selection in the *general traits set* on *mRS** (Table 1), which was due to a significant positive selection gradient of Δ seminal vesicle size on *mRS** (Table 3; Figure 2A). Focal worms that replenished their seminal vesicle more quickly sired more offspring. Linear selection was not significantly different across the four fitness components (Supplementary Table S1A), which suggest that the effect of Δ seminal vesicle size on *mRS** arises from weak positive selection on multiple fitness components (*MS**, *STE**, and *SFE**) (although significantly so only in *MS**; Table 3).

We found no evidence of nonlinear selection in the general trait set on *mRS** using either the original morphological traits (Table 1) or the composite traits (Table 2).

Moreover, nonlinear selection did not seem to be different across fitness components (Supplementary Table S1). These results suggest that there is no nonlinear selection acting on body size, testis size, ovary size, and Δ seminal vesicle size, nor are there any interactions between these traits on *mRS**.

Stylet traits set
We found linear selection in the *stylet traits set* on *mRS** (Table 1), which arose due to a negative effect of stylet RWS3 on *mRS** (Table 3; Figure 2C). This suggested that worms with more bent stylet tips sired more offspring. Patterns of linear selection were different across fitness components (Supplementary Table S1A). Specifically, we found different patterns of linear selection on *STE** and *SFE**, and no significant linear selection on *F** and *MS** (Table 1 and Supplementary Table S2A). Interestingly, *STE** induced positive linear selection on RWS1 and negative linear selection on RWS2, whereas *SFE** induced negative linear selection on RWS3 (Table 3; Figure 2C). These results suggest that different

Table 1. Statistical outcomes testing for linear and non-linear selection on male reproductive success (*mRS**) and four male fitness components, partner fecundity (*F**), mating success (*MS**), sperm transfer efficiency (*STE**), and sperm fertilizing efficiency (*SFE**), on the original morphological traits of the general traits, stylet traits, and sperm traits. In the original morphological traits, non-linear terms includes both the quadratic and cross-product terms. Significant *P* values are indicated in bold (see *Material and Methods* for details).

Original morphological traits

		General traits set						Stylet traits set						Sperm traits set					
		res df	res SS	df	SS	*F*	*P*	res df	res SS	df	SS	*F*	*P*	res df	res SS	df	SS	*F*	*P*
*mRS**	intercept	138	92.1					138	92.1					138	92.1				
	+ linear terms	134	83.7	4	8.4	3.28	**0.013**	134	83.9	4	8.2	3.47	**0.010**	133	84.5	5	7.6	2.58	**0.030**
	+ non-linear terms	124	79.3	10	4.4	0.68	0.737	124	72.9	10	11.1	1.88	0.054	118	69.1	15	15.5	1.75	**0.048**
*F**	intercept	138	10.1					138	10.1					138	10.1				
	+ linear terms	134	10.0	4	0.0	0.08	0.987	134	9.5	4	0.5	1.83	0.128	133	9.5	5	0.6	1.71	0.138
	+ non-linear terms	124	9.1	10	1.0	1.36	0.208	124	8.9	10	0.7	0.96	0.478	118	8.2	15	1.2	1.18	0.290
*MS**	intercept	138	13.1					138	13.1					138	13.1				
	+ linear terms	134	11.3	4	1.8	5.22	**0.001**	134	12.4	4	0.7	1.92	0.111	133	12.6	5	0.5	1.16	0.333
	+ non-linear terms	124	10.7	10	0.6	0.71	0.710	124	11.5	10	0.9	0.96	0.480	118	11.1	15	1.4	1.00	0.460
*STE**	intercept	138	36.5					138	36.5					138	36.5				
	+ linear terms	134	35.1	4	1.4	1.3	0.272	134	33.0	4	3.5	3.90	**0.005**	133	33.5	5	2.9	2.20	0.058
	+ non-linear terms	124	33.5	10	1.6	0.6	0.830	124	27.6	10	5.4	2.44	**0.011**	118	31.3	15	2.3	0.57	0.890
*SFE**	intercept	135	74.6					135	74.6					135	74.6				
	+ linear terms	131	73.6	4	1.0	0.43	0.786	131	65.8	4	8.8	4.32	**0.003**	130	73.9	5	0.5	0.27	0.931
	+ non-linear terms	121	69.8	10	3.8	0.66	0.760	121	61.9	10	3.9	0.77	0.659	115	64.8	15	9.0	1.07	0.393

aspects of stylet morphology are selected on the two post-copulatory fitness components, *STE** and *SFE**. Specifically, worms with stylets curved away from the seminal vesicle (i.e., high RWS1 values) and narrow stylets (i.e., low RWS2 values) had a selective advantage with respect to *STE**, whereas worms with stylets with more bent tips (i.e., low RWS3 values) had a selective advantage with respect to *SFE**.

We found evidence for nonlinear selection acting on *mRS**, which was almost significant using the original morphological traits (Table 1), and highly significant using the composite traits (Table 2). This effect was due to an interaction between centroid size and RWS3 (Table 3), both of which had negative weights on m4, a composite trait capturing concave selection in *mRS** (Table 4). This interaction means that high *mRS** is achieved with intermediate m4 values, which may result from either intermediate centroid size and RWS3 values, or centroid size and RWS3 values that counteract each other (e.g., small stylets with straight tips, or large stylet with bent tips).

This nonlinear selection was not consistent across fitness components (Supplementary Table S1B), which was due to differences in *F**, *MS**, and *STE** (Supplementary Table S2B). We detected nonlinear selection in all these three fitness components (Table 2), though nonlinear effects were larger in *STE** compared to *F** and *MS** (Table 4; Figure 2D). Concave and convex nonlinear selection was found in *STE** (Table 4; Figure 3). The composite trait with the largest leigenvaluel was m4. It displayed concave selection and was mostly affected by centroid size (and RWS2 and RWS3 to a lesser extent), all with negative loadings (orange labels on Figure 3), and thus probably underlay the nonlinear effect we detected on *mRS**. The second significant composite trait was m1, which displayed convex selection and was influenced by centroid size and RWS3 but, importantly, with a positive (blue labels) and a negative (orange labels) loading, respectively (Figure 3). This result means that worms with

stylets that are small and have straight tips, or are large and have highly bent tips, had higher *STE** values. Although this may appear counterintuitive at first sight, the convex effect found on *STE** may concord with the concave effect found on *mRS**. This is because opposite values in centroid size and RWS3 (i.e., small stylets with straight tips, or large stylet with bent tips) lead to extreme values on m1 on *STE**, and to average values on m4 on *mRS**, which are both associated with high fitness values.

Sperm traits set

We found linear selection in the sperm traits set on *mRS** (Table 1), which arose from positive selection of brush size on *mRS** (Table 3; Figure 2E). Worms producing sperm cells with a longer brush sired on average more offspring. Patterns of linear selection were not significantly different across fitness components (Supplementary Table S1A), suggesting that the effect of brush size arises from weak selection on multiple fitness components (Table 3).

We found nonlinear selection in the original morphological traits (Table 1) and on the composite traits on *mRS** (Table 2). The quadratic terms in the γ matrix were significant for the sperm body size and shaft size, and there were significant interactions between feeler size and shaft size, and between brush size and bristle size (Table 3). Three composite traits significantly predicted *mRS** (Table 4; Figure 2F). The two composite traits with the largest leigenvaluesl, m5 and m4, were concave and mostly loaded by sperm body size and shaft size, respectively (Figure 4). The third significant composite trait, m1, was convex and seems to oppose brush size on the one side (i.e., positive loading) and bristle size, feeler size, and shaft size on the other side (i.e., negative loadings). This result may suggest that alternative sperm phenotypes may be correlated with high *mRS** values (i.e., sperm with either a long brush or with long bristles, feeler, and shaft).
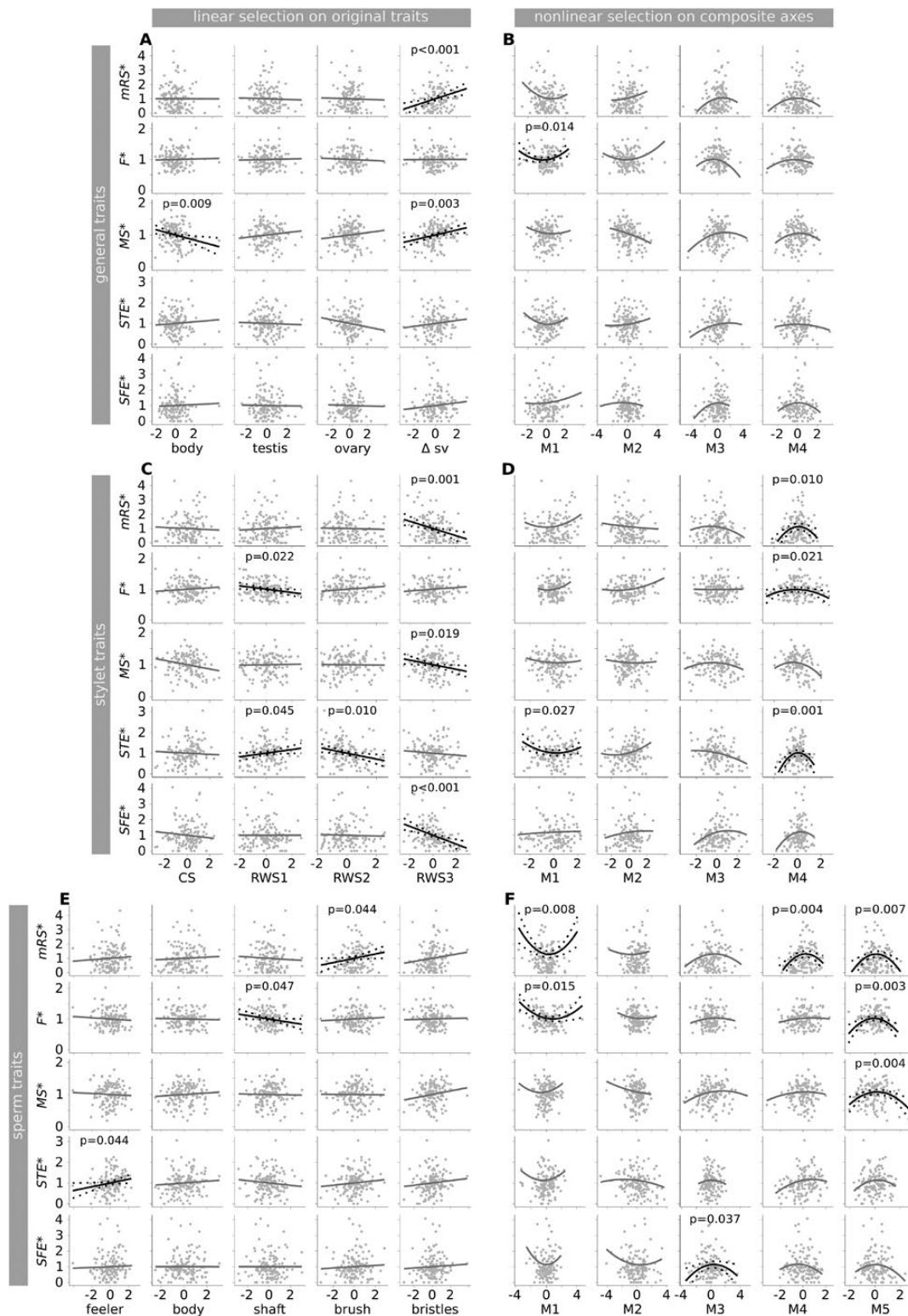
Figure 2. Linear selection (panels A, C, and E) and multivariate selection (panels B, D, and F) on the 13 morphological traits and the derived composite traits on male reproductive success, and the four fitness components. Dots represent observed values. Solid lines represent predicted marginal effects of the x-variable from models including either the linear effects of all traits of a trait set (morphological traits, see Table 3 for statistics) or the linear and quadratic effects of all composite traits of a trait set (composite traits, see Table 4 for statistics). Dotted lines represent the 95% confidence intervals. Fit lines are drawn in black when we found *p* values below .05, and in gray otherwise. Morphological and composite traits are standardized, and male relative success is relative (see *Material and methods*). See the supplementary information for a high-resolution version (Supplementary Figure S1).

Moreover, nonlinear selection was almost significantly different across fitness components (Supplementary Table S1). We found different patterns of nonlinear selection

in the sperm composite traits on *F** compared to *SFE** (Supplementary Table S2B), though both fitness components induced nonlinear selection (Table 2; Figure 2F). We found

Table 2. Statistical outcomes testing for linear and non-linear selection on male reproductive success (*mRS**) and four male fitness components, partner fecundity (*F**), mating success (*MS**), sperm transfer efficiency (*STE**), and sperm fertilizing efficiency (*SFE**), on the composite traits of the general traits, stylet traits, and sperm traits. Significant *P* values are indicated in bold.

Composite traits

		General traits set						Stylet traits set						Sperm traits set					
		res df	res SS	df	SS	*F*	*P*	res df	res SS	df	SS	*F*	*P*	res df	res SS	df	SS	*F*	*P*
mRS*	intercept	138	92.1					138	92.1					138	92.1				
	+ linear terms	134	83.7	4	8.4	3.44	**0.010**	134	83.9	4	8.2	3.64	**0.008**	133	84.5	5	7.6	2.80	**0.019**
	+ quadratic terms	130	79.3	4	4.4	1.79	0.134	130	72.9	4	11.1	4.94	**0.001**	128	69.1	5	15.5	5.73	**<0.001**
F*	intercept	138	10.1					138	10.1					138	10.1				
	+ linear terms	134	10.0	4	0.0	0.09	0.986	134	9.5	4	0.5	1.92	0.111	133	9.5	5	0.6	1.85	0.107
	+ quadratic terms	130	9.1	4	1.0	3.56	**0.009**	130	8.9	4	0.7	2.53	**0.044**	128	8.2	5	1.2	3.87	**0.003**
MS*	intercept	138	13.1					138	13.1					138	13.1				
	+ linear terms	134	11.3	4	1.8	5.47	**<0.001**	134	12.4	4	0.7	2.02	0.096	133	12.6	5	0.5	1.28	0.286
	+ quadratic terms	130	10.7	4	0.6	1.87	0.119	130	11.5	4	0.9	2.52	**0.045**	128	11.1	5	1.4	3.26	**0.008**
STE*	intercept	138	36.5					138	36.5					138	36.5				
	+ linear terms	134	35.1	4	1.4	1.37	0.249	134	33.0	4	3.5	4.09	**0.004**	133	33.5	5	2.9	2.39	**0.041**
	+ quadratic terms	130	33.5	4	1.6	1.51	0.203	130	27.6	4	5.4	6.41	**<0.001**	128	31.3	5	2.3	1.87	0.105
SFE*	intercept	135	74.6					135	74.6					135	74.6				
	+ linear terms	131	73.6	4	1.0	0.45	0.771	131	65.8	4	8.8	4.53	**0.002**	131	73.9	5	0.7	0.29	0.918
	+ quadratic terms	127	69.8	4	3.8	1.73	0.147	127	61.9	4	3.9	2.02	0.096	125	64.8	5	9.0	3.49	**0.006**

concave and convex selection on *F** (Table 4). The composite trait capturing concave selection, m4, was mostly influenced by sperm body size, while the composite trait capturing convex selection, m1, was mostly influenced by feeler size and shaft size. The only significant composite trait on *SFE**, m3, captured concave selection and was mostly influenced by brush size (Table 4; Figure 2F). Remarkably, other composite traits in *SFE** had larger eigenvalues than m3, but were not statistically significant. Altogether, our results on sperm traits showed nonlinear selection, but were spread over multiple composite traits and across fitness components, possibly rendering composite traits non-significant when considered individually.

Discussion

Although sexual selection is acknowledged to act along multiple pre- and postcopulatory episodes of selection (Arnold & Wade, 1984; Evans & Garcia-Gonzalez, 2016), studies measuring and comparing phenotypic selection on pre- and postcopulatory fitness components are scarce—and even more so with respect to studies accounting for nonlinear selection through multivariate selection analyses (but see House et al., 2016). By measuring phenotypic traits alongside pre- and postcopulatory fitness components, such as mating success, sperm-transfer efficiency, and sperm fertilizing efficiency, we could study how the form and the strength of selection materialize along these components. Our results show that selection may arise both from specific and multiple fitness components, and that morphological traits may be under different selective pressure with respect to different fitness components. In the following, we discuss the insights gained by measuring linear and nonlinear selection on pre- and postcopulatory fitness components.

Postcopulatory selection is usually measured through a single fitness component (e.g., De Nardo et al., 2021; Devigili

et al., 2015; House et al., 2016; McCullough et al., 2018; McDonald et al., 2017; Morimoto et al., 2019). Here, we could contrast phenotypic selection in two postcopulatory fitness components, sperm-transfer efficiency and sperm fertilizing efficiency, finding that these two components favored different shapes of the male copulatory organ. Namely, individuals having stylets that are curved away from the seminal vesicle (positive RWS1) and narrow (negative RWS2) had a better sperm-transfer efficiency, while sperm fertilizing efficiency was higher with more bent stylet tips (negative RWS3) (Figure 2C; Table 3; and Supplementary Table S2). Although only the latter effect was significant on male reproductive success, this outcome suggests that distinct episodes of selection can act on morphological traits after copulation. Our results also show nonlinear selection on stylet traits. The main source of nonlinear stylet traits was due to an interaction between centroid size and RWS3. Selection favored stylets that were either small and had straight tips or large and with bent tips, which arose from sperm-transfer efficiency. In *M. lignano*, stylet shape may be selected by allowing worms to interfere with previously received sperm (e.g., sperm removal), and/or to donate sperm at strategic places in the female antrum that increase sperm fertilization success. These two putative mechanisms may possibly induce contrasting selection on stylet shape. In our experimental design, however, because several copulations had occurred during our mating trials, it is possible that sperm's abilities to resist being displaced by subsequent mating partners can translate into high sperm-transfer efficiencies.

We found positive linear selection on Δ seminal vesicle size (Figure 2A; Table 3). In *M. lignano*, the seminal vesicle is a reliable proxy for the number of sperm it contains (Schärer & Vizoso, 2007), so the increase in seminal vesicle size of worms kept in isolation for two days can serve as a proxy for the sperm production rate. Thus, this result suggests that worms with higher sperm production rates sired more

Table 3. Linear and non-linear selection on each trait set, measured on male reproductive success (*mRS**) and on the four male fitness components, partner fecundity (*F**), mating success (*MS**), sperm transfer efficiency (*STE**), and sperm fertilizing efficiency (*SFE**). The table shows the standardized linear selection gradients (β), and the γ matrices showing the standardized quadratic and correlational selection gradients. Traits are body size (bod), testis size (tes), ovary size (ova), Δ s (Δ seminal vesicle size); stylet's centroid size (CS), first, second, and third relative warp scores (RWS1, RWS2, and RWS3); sperm feeler size (fee), body size (bod), shaft size (sha), brush size (bru), and bristle size (bri). Bold font stands for significant gradients and eigenvalues, and for loadings higher than |0.50| on canonical axes with significant eigenvalues. * *p* < .05; ** *p* < 0.01; *** *p* < .001.

Original morphological traits

| | | **General traits set** | | | | | | **Stylet traits set** | | | | | | **Sperm traits set** | | | | | |
		β	bod	tes	ova	Δ s		β	CS	RWS1	RWS2	RWS3		β	fee	bod	sha	bru	bri
mRS*	bod	0.00	0.02				CS	−0.04	−0.15				fee	0.05	−0.02				
	tes	−0.02	0.14	−0.12			RWS1	0.05	−0.19	−0.05			bod	0.04	0.07	**−0.58****			
	ova	−0.02	−0.07	−0.09	−0.02		RWS2	−0.02	−0.05	−0.12	−0.14		sha	−0.05	**0.23****	0.14	**−0.32***		
	Δ s	**0.25*****	0.11	0.05	0.03	0.06	RWS3	**−0.25****	**−0.35****	−0.08	−0.08	−0.22	bru	**0.15***	−0.09	0.04	−0.02	0.02	
													bri	0.11	−0.09	0.15	0.17	**−0.20***	−0.13
F*	bod	0.01	0.03				CS	0.02	0.08				fee	−0.02	0.01				
	tes	0.01	0.03	−0.03			RWS1	**−0.05***	−0.02	0.02			bod	−0.01	0.06	**−0.21****			
	ova	−0.01	0.01	−0.01	−0.06		RWS2	0.04	0.07	**−0.06***	0.00		sha	**−0.06***	0.01	0.07	−0.01		
	Δ s	0.00	−0.03	0.06	−0.03	0.05	RWS3	0.03	0.01	−0.01	0.04	0.01	bru	0.02	0.01	0.04	−0.01	−0.02	
													bri	0.01	−0.01	0.03	0.02	−0.03	0.01
MS*	bod	**−0.08****	0.02				CS	−0.07	−0.07				fee	−0.02	−0.03				
	tes	0.04	−0.02	−0.05			RWS1	0.01	0.00	0.03			bod	0.02	−0.00	−0.03			
	ova	0.04	−0.00	0.03	−0.06		RWS2	−0.00	0.03	−0.01	0.00		sha	−0.01	0.06	−0.01	−0.07		
	Δ s	**0.08****	0.02	−0.04	0.01	−0.01	RWS3	**−0.07***	−0.06	−0.00	−0.02	**−0.10***	bru	−0.00	**−0.07***	−0.00	0.05	−0.01	
													bri	0.05	−0.01	0.02	−0.01	−0.04	−0.01
STE*	bod	0.04	0.05				CS	−0.03	**−0.29***				fee	**0.09***	−0.03				
	tes	−0.02	0.04	−0.03			RWS1	**0.09***	−0.09	−0.04			bod	0.04	−0.03	−0.09			
	ova	−0.09	−0.07	−0.01	0.02		RWS2	**−0.14***	**−0.24***	−0.02	−0.09		sha	−0.06	0.08	0.00	−0.11		
	Δ s	0.07	−0.04	0.00	0.06	0.05	RWS3	−0.05	**−0.28*****	0.00	**−0.18****	−0.08	bru	0.05	−0.08	0.03	−0.08	−0.06	
													bri	0.05	0.02	0.02	0.13	0.05	−0.16
SFE*	bod	0.03	−0.04				CS	−0.08	−0.22				fee	0.02	−0.00				
	tes	−0.01	0.10	−0.10			RWS1	0.00	−0.06	−0.07			bod	0.00	−0.02	−0.23			
	ova	−0.02	0.02	−0.08	−0.12		RWS2	−0.02	−0.17	−0.01	−0.23		sha	0.00	**0.19***	0.04	−0.07		
	Δ s	0.09	0.07	−0.02	−0.01	−0.12	RWS3	**−0.28*****	−0.16	−0.06	−0.07	−0.21	bru	0.04	0.08	−0.11	0.04	0.05	
													bri	0.04	−0.13	0.18	0.00	−0.17	−0.12

Table 4. Linear and non-linear selection on each trait set, measured on male reproductive success (*mRS**) and on the four male fitness components, partner fecundity (*F**), mating success (*MS**), sperm transfer efficiency (*STE**), and sperm fertilizing efficiency (*SFE**). The table shows the eigenvalues (λ), and the M matrices of the eigenvectors from the canonical rotation of the γ matrices. Traits are body size (bod), testis size (tes), ovary size (ova), Δ s (Δ seminal vesicle size); stylet's centroid size (CS), first, second, and third relative warp scores (RWS1, RWS2, and RWS3); sperm feeler size (fee), body size (bod), shaft size (sha), brush size (bru), and bristle size (bri). Bold font stands for significant gradients and eigenvalues, and for loadings higher than |0.50| on canonical axes with significant eigenvalues. $*p < .05$; $**p < 0.01$; $***p < .001$.

Composite traits

General traits set

		β	λ	bod	tes	ova	Δ s
*mRS**	m1	0.23		-0.68	-0.42	0.28	-0.53
	m2	0.05		-0.10	-0.24	0.69	0.67
	m3	-0.12		-0.59	-0.11	-0.62	0.51
	m4	-0.22		0.43	-0.87	-0.24	0.01
*F**	m1	**0.09***		-0.21	0.39	-0.22	0.87
	m2	0.05		0.91	0.42	0.03	0.04
	m3	-0.07		-0.07	0.05	0.97	0.21
	m4	-0.08		0.36	-0.82	-0.02	0.45
*MS**	m1	0.05		0.72	-0.41	-0.07	0.56
	m2	-0.01		0.67	0.24	-0.09	-0.70
	m3	-0.03		0.18	0.52	0.80	0.25
	m4	-0.10		0.06	0.71	-0.60	0.37
*STE**	m1	0.16		-0.62	-0.18	0.57	0.52
	m2	0.03		0.53	0.39	0.01	0.75
	m3	-0.04		-0.17	-0.53	-0.73	0.41
	m4	-0.05		0.55	-0.73	0.39	-0.02
*SFE**	m1	0.05		0.78	0.55	-0.15	0.26
	m2	-0.05		-0.34	0.51	-0.67	-0.42
	m3	-0.13		-0.22	-0.20	-0.54	0.79
	m4	-0.25		0.47	-0.64	-0.49	-0.36

Stylet traits set

		β	λ	CS	RWS1	RWS2	RWS3
*mRS**	m1	0.21		0.71	-0.44	0.17	-0.53
	m2	0.01		-0.23	-0.74	0.45	0.45
	m3	-0.14		0.22	-0.37	-0.84	0.33
	m4	**-0.64****		-0.63	-0.35	0.26	-0.64
*F**	m1	0.15		0.70	-0.38	0.56	0.23
	m2	0.03		-0.65	-0.64	0.25	0.33
	m3	0.00		-0.04	0.50	0.04	0.86
	m4	**-0.07***		-0.29	0.44	0.79	-0.31
*MS**	m1	0.04		-0.26	0.71	-0.62	0.20
	m2	0.02		0.37	0.70	0.53	-0.30
	m3	-0.05		0.63	-0.12	-0.57	-0.51
	m4	-0.15		0.63	0.00	-0.01	0.78
*STE**	m1	**0.13***		0.53	-0.31	0.09	-0.78
	m2	0.07		0.41	-0.17	-0.86	0.24
	m3	-0.05		-0.15	-0.93	0.19	0.28
	m4	**-0.65*****		-0.72	-0.13	-0.46	-0.49
*SFE**	m1	-0.01		0.77	-0.19	-0.47	-0.40
	m2	-0.04		0.05	0.90	0.06	-0.43
	m3	-0.16		0.00	-0.35	0.69	-0.63
	m4	-0.51		-0.64	-0.17	-0.55	-0.51

Sperm traits set

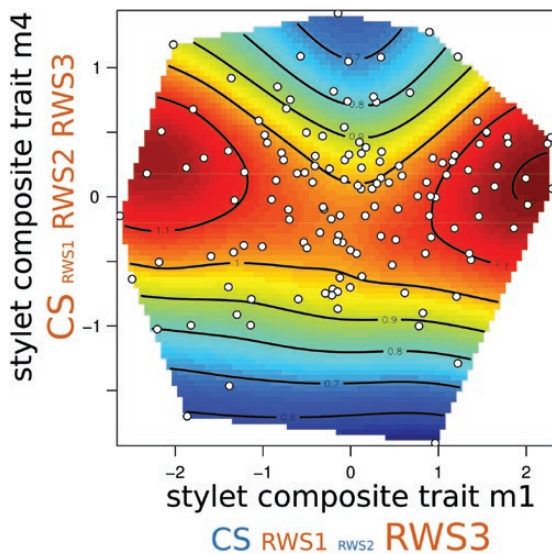|		β	λ	fee	bod	sha	bru	bri
*mRS**	m1	**0.24****		-0.46	-0.16	-0.40	0.62	-0.48
	m2	0.07		0.76	0.06	0.26	0.39	-0.45
	m3	-0.11		-0.17	0.33	0.41	0.64	0.54
	m4	**-0.56*****		0.43	0.08	-0.76	0.19	0.44
	m5	**-0.66*****		-0.10	0.92	-0.17	-0.16	-0.29
*F**	m1	**0.06***		0.65	0.37	0.58	0.08	0.33
	m2	0.04		-0.48	-0.04	0.26	-0.51	0.66
	m3	-0.03		0.54	-0.13	-0.39	-0.73	-0.03
	m4	-0.03		0.15	0.03	-0.61	0.40	0.66
	m5	**-0.25*****		0.20	-0.92	0.26	0.17	0.13
*MS**	m1	0.07		0.44	0.11	-0.14	-0.77	0.43
	m2	0.02		0.67	-0.29	0.48	-0.02	-0.49
	m3	-0.03		0.22	0.77	0.46	0.30	0.27
	m4	-0.05		0.01	-0.56	0.37	0.25	0.70
	m5	**-0.17****		-0.56	0.02	0.64	-0.51	-0.13
*STE**	m1	0.11		0.65	-0.13	0.53	-0.50	0.20
	m2	-0.03		0.13	-0.44	-0.44	-0.44	-0.63
	m3	-0.11		0.02	0.87	-0.09	-0.40	-0.26
	m4	-0.13		0.75	0.16	-0.34	0.53	-0.09
	m5	-0.30		0.04	-0.01	-0.63	-0.34	0.70
*SFE**	m1	0.33		0.51	-0.28	0.27	0.59	-0.48
	m2	0.09		-0.52	-0.34	-0.64	0.37	-0.26
	m3	**-0.14***		0.32	-0.33	-0.29	-0.69	-0.47
	m4	-0.27		-0.56	-0.41	0.66	-0.19	-0.23
	m5	-0.38		-0.21	0.73	0.04	-0.02	-0.65

Figure 3. The fitness surface of the composite stylet traits on sperm transfer efficiency (*STE**). The axes represent the two significant eigenvectors that had positive (m1) and negative (m4) eigenvalues (Table 4). The color of the fitness surface corresponds to the fitness value. The font sizes of the morphological traits below the eigenvector are proportional to the square root of the trait's |eigenvalue| to represent the respective loadings of each morphological trait on the composite traits. The color of the morphological traits indicates positive (blue labels) and negative (orange labels) eigenvalues.
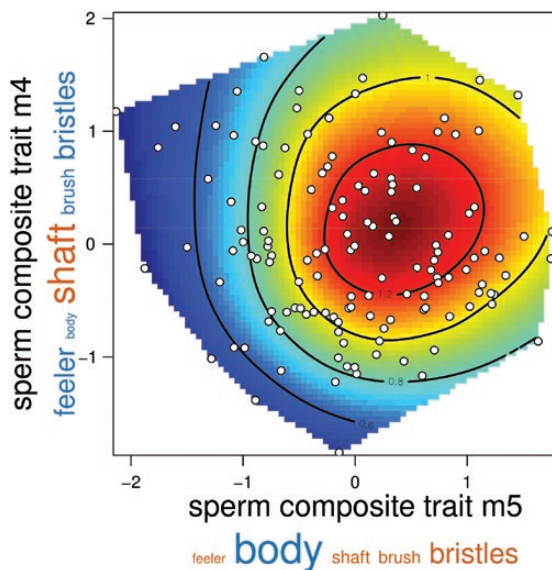
Figure 4. The fitness surface of the composite sperm traits on male reproductive success (*mRS**). The axes represent the two significant eigenvectors with the highest |eigenvalues|, m4 and m5, which were both negative (Table 4). The color of the fitness surface corresponds to the fitness value. The font sizes of the morphological traits below the eigenvector are proportional to the square root of the trait's |eigenvalue| to represent the respective loadings of each morphological trait on the composite traits. The color of the morphological traits indicates positive (blue labels) and negative (orange labels) eigenvalues.

offspring, which is predicted by sperm competition theory (Parker, 1970; Pizzari & Parker, 2009). However, we did not find evidence for a difference in selection across fitness components (Supplementary Table S1A), the effect of Δ seminal vesicle size was significantly positive only on mating success

(Figure 2A; Table 3). This result confirms a previous study, in which experimentally manipulated sperm production rate was shown to influence both mating success and male reproductive success (Sekii et al., 2013). Worms seem to adjust their mating rates according to the amount of sperm available to donate to partners, which illustrates that pre- and post-copulatory selection can be intertwined. Therefore, studies of finely decomposing sexual selection along fitness components are likely to provide critical insights into this interplay, while studies measuring phenotypic selection on total fitness are more likely to provide misleading interpretations about the biological mechanisms underlying the measured selection.

Surprisingly, we did not find selection on testis size (Figure 2A, Tables 3 and 4), which clearly contrasts with several previous findings in *M. lignano*. Bigger testes have been found to positively correlate with sperm production rate (Schärer & Vizoso, 2007), with the number of mating partners and the number of sperm transferred to partners (Janicke & Schärer, 2009a), with sperm-transfer efficiency and male reproductive success (Marie-Orleach et al., 2016), and with paternity share (Vellnow et al., 2018). Remarkably, testis size showed one of the lowest repeatability values between the two morphological measurements taken two days apart: body size, $r_I = 0.66$; testis size, $r_I = 0.36$; ovary size, $r_I = 0.35$; stylet centroid size, $r_I = 0.87$; stylet RWS1, $r_I = 0.68$; stylet RWS2, $r_I = 0.78$; stylet RWS3, and $r_I = 0.48$ (Lessells & Boag, 1987; Stoffel et al., 2017). In comparison, Schärer and Ladurner (2003) found a repeatability value of $r_I = 0.76$ for testis size but, in contrast to our repeated measures, theirs were performed in quick sequence. This could suggest that, in our experiment, the worms adjusted their sex allocation within the two days of isolation (through such adjustments generally seem to take longer to materialize; Brauer et al., 2007). However, using only the first measurement of testis size did not predict male reproductive success either (selection differential ± *SE*: −0.07 ± 0.07, *t* = −0.97, *p* = .336). It is thus unclear why, in our experiment, testis size showed low repeatability and did not correlate with male fitness.

We found that worms producing sperm cells with a longer brush sired more offspring (Figure 2E; Tables 3 and 4) (and we could not determine from which fitness component this effect specifically arose). It is currently unclear what are the biological functions of the sperm brush in *M. lignano*, and to our knowledge, no specific hypotheses have been proposed to date. Several species within the genus *Macrostomum* lack a sperm brush, and these species reproduce through traumatic insemination (i.e., sperm donors inject ejaculate through the epidermis of the sperm recipient, and sperm cells subsequently move through the recipient body to fertilize the eggs) (Brand et al., 2022; Schärer et al., 2011). The biological functions of the sperm brush may thus possibly only be relevant when, like in *M. lignano*, mating is reciprocal, and sperm deposition and fertilization occur inside the female antrum.

Moreover, we found evidence for nonlinear selection on sperm traits, which arose from multiple traits and multiple fitness components. We found concave selection on sperm body size and sperm shaft size, as well as convex selection on a composite trait opposing brush size to feeler size, bristle size, and shaft size (Table 4) (suggesting that alternative sperm phenotypes may be selected; Pizzari & Parker, 2009). Our analysis suggests that this nonlinear selection on sperm traits arose from partner fecundity and sperm fertilizing efficiency. In this experiment, focal effects on partner fecundity

are, however, unexpected given that a previous analysis showed that focal individuals did not influence the partner fecundity consistently over the three groups (Marie-Orleach et al., 2021). And, the significant composite trait on sperm fertilizing efficiency did not concord much with the composite traits affecting male reproductive success. Remarkably, however, other composite traits had stronger, yet non-significant, signs of nonlinear selection on sperm fertilizing efficiency. Therefore, we think that the unclear match between selection on male reproductive success and selection on its fitness components may be because sperm traits had relatively subtle effects (direct and/or indirect) on multiple fitness components, and these effects were statistically significant only when summed up all together in male reproductive success.

Finally, sexual selection is usually measured through two broadly defined approaches: (a) a trait-based approach, focusing on the relationship between variation in phenotypic traits and individual fitness, and (b) a variance-based approach, using exclusively variance in individual success (Arnold & Wade, 1984). We have now used both approaches to study pre- and postcopulatory sexual selection (the variance-based approach being reported in Marie-Orleach et al., 2021, and the trait-based approach in the current study), thus gaining complementary insights. By using the trait-based approach, we could estimate linear and nonlinear selection acting on pre- and postcopulatory fitness components. However, this approach is limited to the traits that can actually be measured in a study system. Unmeasured traits may also be under pre- and postcopulatory sexual selection, and could possibly change the interpretations of phenotypic selection (i.e., the missing trait problem in multivariate analysis; Morrissey et al., 2012). Measuring nonlinear selection on multiple traits also requires high sample sizes (Green, 1991), and the necessity of estimating selection on subsets of traits in this study, did not allow us to test for correlational selection among traits of different trait sets, which is an important limitation in our study.

In contrast, because the variance-based approach does not rely on phenotypic traits, these findings were not restricted by the traits measured, and so are possibly more informative about the total strength of selection arising from different fitness components (Marie-Orleach et al., 2021). Indeed we found that postcopulatory selection may be more intense than precopulatory selection in *M. lignano*. However, by focusing on variance in individual success only, one cannot know the form and strength of selection acting on any given trait. Going forward, we advocate using both the variance-based and the trait-based approaches. For instance, in the Trinidad guppy *Poecilia reticulata,* the variance in male reproductive success increased when individuals perceived a predation risk. However, this additional variance did not lead to higher selection on the measured male traits (Glavaschi et al., 2022), meaning that the additional opportunity for selection was either due to selection on unmeasured traits and/or due to stochastic effects (i.e., variance in individual success that is not due to the individuals' phenotypes).

Conclusions

We performed mating observations, in vivo sperm tracking, paternity analyses, and measured several morphological traits to study the strength and form of selection arising from pre- and postcopulatory sexual selection in the free-living flatworm *Macrostomum lignano*. We found evidence for linear selection on sperm production rate arising from multiple fitness components, and on (combinations of) stylet and sperm traits arising mostly from sperm-transfer efficiency and sperm fertilizing efficiency. Our results suggest that intense selection arises from different fitness components in *M. lignano*, overall inducing contrasting patterns of selection on combinations of traits. Thus, by contrasting the strength and form of selection on pre- and postcopulatory fitness components, our study shows that phenotypic selection can be highly complex. Phenotypic traits can be selected in combination with other traits before, during, and/or after copulation, and knowledge about how selection operates ultimately helps us to better understand the evolutionary dynamics of sexually selected traits. And we believe that promising research avenues to understand these dynamics will include a better understanding of (a) how the environment modulates the strength of pre- and postcopulatory sexual selection (Evans & Garcia-Gonzalez, 2016; García-Roa et al., 2020); and (b) the genetic variance and co-variances in the phenotypic traits, and as well as in the individual success in pre- and postcopulatory fitness components.

Supplementary material

Supplementary material is available online at *Evolution*.

Data availability

The data and R-scripts are available in the supplementary material and on the Dryad data repository (https://doi.org/10.5061/dryad.ffbg79d20).

Author contributions

L.M.O. and L.S. conceived the study and designed the experiment. L.M.O. performed the experiment and collected the data. L.M.O. performed the statistical analyses with help from M.D.H. L.M.O. wrote the manuscript with help from M.D.H. and L.S.

Acknowledgments

Funding

Conflict of interest: The authors declare no conflict of interest.

References

Andersson, M. (1994). *Sexual selection*. Princeton University Press.

Anthes, N., David, P., Auld, J. R., Hoffer, J. N. A., Jarne, P., Koene, J. M., Kokko, H., Lorenzi, M. C., Pélissié, B., Sprenger, D., Staikou, A., & Schärer, L. (2010). Bateman gradients in hermaphrodites: An extended approach to quantify sexual selection. *American Naturalist*, 176(3), 249–263. https://doi.org/10.1086/655218

Arnold, S. J., & Wade, M. J. (1984). On the measurement of natural and sexual selection: Theory. *Evolution*, *38*(4), 709–719. https://doi.org/10.1111/j.1558-5646.1984.tb00344.x

Arnqvist, G. (1997). The evolution of animal genitalia: Distinguishing between hypotheses by single species studies. *Biological Journal of the Linnean Society*, *60*(3), 365–379. https://doi.org/10.1111/j.1095-8312.1997.tb01501.x

Bentsen, C. L., Hunt, J., Jennions, M. D., & Brooks, R. (2006). Complex multivariate sexual selection on male acoustic signaling in a wild population of *Teleogryllus commodus*. *American Naturalist*, *167*(4), E102–E116. https://doi.org/10.1086/501376

Birkhead, T. R., Hosken, D. J., & Pitnick, S. (2009). *Sperm biology: An evolutionary perspective*. Elsevier Academic Press Inc.

Blows, M. W. (2007). A tale of two matrices: Multivariate approaches in evolutionary biology. *Journal of Evolutionary Biology*, *20*(1), 1–8. https://doi.org/10.1111/j.1420-9101.2006.01164.x

Blows, M. W., & Brooks, R. (2003). Measuring nonlinear selection. *American Naturalist*, *162*(6), 815–820. https://doi.org/10.1086/378905

Brand, J. N., Harmon, L. J., & Schärer, L. (2022). Frequent origins of traumatic insemination involve convergent shifts in sperm and genital morphology. *Evolution Letters*, *6*(1), 63–82. https://doi.org/10.1002/evl3.268

Brauer, V. S., Schärer, L., & Michiels, N. K. (2007). Phenotypically flexible sex allocation in a simultaneous hermaphrodite. *Evolution*, *61*(1), 216–222. https://doi.org/10.1111/j.1558-5646.2007.00018.x

Charnov, E. L. (1979). Simultaneous hermaphroditism and sexual selection. *Proceedings of the National Academy of Sciences*, *76*(5), 2480–2484. https://doi.org/10.1073/pnas.76.5.2480

Chenoweth, S. F., & Blows, M. W. (2005). Contrasting mutual sexual selection on homologous signal traits in *Drosophila serrata*. *American Naturalist*, *165*(2), 281–289. https://doi.org/10.1086/427271

Conner, J. (1988). Field measurements of natural and sexual selection in the fungus beetle, *Bolitotherus cornutus*. *Evolution*, *42*(4), 736–749. https://doi.org/10.1111/j.1558-5646.1988.tb02492.x

De Nardo, A. N., Roy, J., Sbilordo, S. H., & Lüpold, S. (2021). Condition-dependent interaction between mating success and competitive fertilization success in *Drosophila melanogaster*. *Evolution*, *75*(8), 2014–2026. https://doi.org/10.1111/evo.14228

Devigili, A., Evans, J. P., Di Nisio, A., & Pilastro, A. (2015). Multivariate selection drives concordant patterns of pre- and postcopulatory sexual selection in a livebearing fish. *Nature Communications*, *6*, 8291. https://doi.org/10.1038/ncomms9291

Eberhard, W. G., Huber, B. A., Rodriguez, R. L., Briceño, R. D., Salas, I., & Rodriquez, V. (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.

Evans, J. P., & Garcia-Gonzalez, F. (2016). The total opportunity for sexual selection and the integration of pre- and post-mating episodes of sexual selection in a complex world. *Journal of Evolutionary Biology*, *29*(12), 2338–2361. https://doi.org/10.1111/jeb.12960

Gage, M. J. G., & Morrow, E. H. (2003). Experimental evidence for the evolution of numerous, tiny sperm via sperm competition. *Current Biology*, *13*, 754–757.

Garant, D., Kruuk, L. E. B., McCleery, R. H., & Sheldon, B. C. (2007). The effects of environmental heterogeneity on multivariate selection on reproductive traits in female great tits. *Evolution*, *61*(7), 1546–1559. https://doi.org/10.1111/j.1558-5646.2007.00128.x

García-Roa, R., Garcia-Gonzalez, F., Noble, D. W. A., & Carazo, P. (2020). Temperature as a modulator of sexual selection. *Biological Review*, *95*, 1607–1629.

Glavaschi, A., Cattelan, S., Devigili, A., & Pilastro, A. (2022). Immediate predation risk alters the relationship between potential and realised selection on male traits in the Trinidad guppy *Poecilia reticulata*. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1982), 20220641. https://doi.org/10.1098/rspb.2022.0641

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research*, *26*(3), 499–510. https://doi.org/10.1207/s15327906mbr2603_7

Hall, M. D., Bussière, L. F., Hunt, J., & Brooks, R. (2008). Experimental evidence that sexual conflict influences the opportunity, form and intensity of sexual selection. *Evolution*, *62*(9), 2305–2315. https://doi.org/10.1111/j.1558-5646.2008.00436.x

Henshaw, J. M., Jennions, M. D., & Kruuk, L. E. B. (2018). How to quantify (the response to) sexual selection on traits. *Evolution*, *72*(9), 1904–1917. https://doi.org/10.1111/evo.13554

Henshaw, J. M., & Zemel, Y. (2017). A unified measure of linear and nonlinear selection on quantitative traits. *Methods in Ecology and Evolution*, *8*(5), 604–614. https://doi.org/10.1111/2041-210x.12685

House, C., Tunstall, P., Rapkin, J., Bale, M. J., Gage, M., Castillo, E., & Hunt, J. (2020). Multivariate stabilizing sexual selection and the evolution of male and female genital morphology in the red flour beetle. *Evolution*, *74*, 883–896.

House, C. M., Sharma, M. D., Okada, K., & Hosken, D. J. (2016). Pre and post-copulatory selection favor similar genital phenotypes in the male broad horned beetle. *Integrative and Comparative Biology*, *56*(4), 682–693. https://doi.org/10.1093/icb/icw079

Hunt, J., Breuker, C. J., Sadowski, J. A., & Moore, A. J. (2009). Male-male competition, female mate choice and their interaction: Determining total sexual selection. *Journal of Evolutionary Biology*, *22*(1), 13–26. https://doi.org/10.1111/j.1420-9101.2008.01633.x

Janicke, T., Marie-Orleach, L., De Mulder, K., Berezikov, E., Ladurner, P., Vizoso, D. B. B., & Schärer, L. (2013). Sex allocation adjustment to mating group size in a simultaneous hermaphrodite. *Evolution*, *67*, 3233–3242.

Janicke, T., Sandner, P., & Schärer, L. (2011). Determinants of female fecundity in a simultaneous hermaphrodite: The role of polyandry and food availability. *Evolutionary Ecology*, *25*(1), 203–218. https://doi.org/10.1007/s10682-010-9402-5

Janicke, T., & Schärer, L. (2009a). Determinants of mating and sperm-transfer success in a simultaneous hermaphrodite. *Journal of Evolutionary Biology*, *22*(2), 405–415. https://doi.org/10.1111/j.1420-9101.2008.01660.x

Janicke, T., & Schärer, L. (2009b). Sex allocation predicts mating rate in a simultaneous hermaphrodite. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1676), 4247–4253. https://doi.org/10.1098/rspb.2009.1336

Janicke, T., & Schärer, L. (2010). Sperm competition affects sex allocation but not sperm morphology in a flatworm. *Behavior, Ecology and Sociobiology*, *64*(9), 1367–1375. https://doi.org/10.1007/s00265-010-0951-y

Jones, A. G. (2009). On the opportunity for sexual selection, the Bateman gradient and the maximum intensity of sexual selection. *Evolution*, *63*(7), 1673–1684. https://doi.org/10.1111/j.1558-5646.2009.00664.x

Keagy, J., Lettieri, L., & Boughman, J. W. (2016). Male competition fitness landscapes predict both forward and reverse speciation. *Ecology Letters*, *19*(1), 71–80. https://doi.org/10.1111/ele.12544

Ladurner, P., Schärer, L., Salvenmoser, W., & Rieger, R. M. (2005). A new model organism among the lower Bilateria and the use of digital microscopy in taxonomy of meiobenthic Platyhelminthes: *Macrostomum lignano*, n sp (Rhabditophora, Macrostomorpha). *Journal of Zoological Systematics and Evolutionary Research*, *43*(2), 114–126. https://doi.org/10.1111/j.1439-0469.2005.00299.x

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. *Evolution*, *33*(1Part2), 402–416. https://doi.org/10.1111/j.1558-5646.1979.tb04694.x

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, *37*(6), 1210–1226. https://doi.org/10.1111/j.1558-5646.1983.tb00236.x

Lessells, C. M., & Boag, P. T. (1987). Unrepeatable repeatabilities—A common mistake. *Auk*, *104*, 116–121.

Marie-Orleach, L., Janicke, T., & Schärer, L. (2013). Effects of mating status on copulatory and postcopulatory behaviour in a simultaneous hermaphrodite. *Animal Behaviour*, *85*(2), 453–461. https://doi.org/10.1016/j.anbehav.2012.12.007

Marie-Orleach, L., Janicke, T., Vizoso, D. B., David, P., & Schärer, L. (2016). Quantifying episodes of sexual selection: Insights from a

transparent worm with fluorescent sperm. *Evolution*, *70*(2), 314–328. https://doi.org/10.1111/evo.12861

Marie-Orleach, L., Janicke, T., Vizoso, D. B., Eichmann, M., & Schärer, L. (2014). Fluorescent sperm in a transparent worm: Validation of a GFP marker to study sexual selection. *BMC Evolutionary Biology*, *14*, 148. https://doi.org/10.1186/1471-2148-14-148

Marie-Orleach, L., Vellnow, N., & Schärer, L. (2021). The repeatable opportunity for selection differs between pre- and postcopulatory fitness components. *Evolution Letters*, *5*(1), 101–114. https://doi.org/10.1002/evl3.210

Mautz, B. S., Møller, A. P., & Jennions, M. D. (2013). Do male secondary sexual characters signal ejaculate quality? A meta-analysis. *Biological Reviews of the Cambridge Philosophical Society*, *88*(3), 669–682. https://doi.org/10.1111/brv.12022

McCullough, E. L., Buzatto, B. A., & Simmons, L. W. (2018). Population density mediates the interaction between pre- and post-mating sexual selection. *Evolution*, *72*(4), 893–905. https://doi.org/10.1111/evo.13455

McDonald, G. C., Spurgin, L. G., Fairfield, E. A., Richardson, D. S., & Pizzari, T. (2017). Pre- and postcopulatory sexual selection favor aggressive, young males in polyandrous groups of red junglefowl. *Evolution*, *71*(6), 1653–1669. https://doi.org/10.1111/evo.13242

Morimoto, J., McDonald, G. C., Smith, E., Smith, D. T., Perry, J. C., Chapman, T., Pizzari, T., & Wigby, S. (2019). Sex peptide receptor-regulated polyandry modulates the balance of pre- and post-copulatory sexual selection in *Drosophila*. *Nature Communications*, *10*(1), 2083.

Morrissey, M. B., Parker, D. J., Korsten, P., Pemberton, J. M., Kruuk, L. E. B., & Wilson, A. J. (2012). The prediction of adaptive evolution: Empirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution*, *66*, 2399–2410.

Oh, K. P., & Shaw, K. L. (2013). Multivariate sexual selection in a rapidly evolving speciation phenotype. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1761), 20130482. https://doi.org/10.1098/rspb.2013.0482

Parker, G. A. (1970). Sperm competition and its evolutionary consequences in insects. *Biological Review*, *45*(4), 525–567. https://doi.org/10.1111/j.1469-185x.1970.tb01176.x

Parker, T. H., Wilkin, T. A., Barr, I. R., Sheldon, B. C., Rowe, L., & Griffith, S. C. (2011). Fecundity selection on ornamental plumage colour differs between ages and sexes and varies over small spatial scales. *Journal of Evolutionary Biology*, *24*(7), 1584–1597. https://doi.org/10.1111/j.1420-9101.2011.02289.x

Phillips, P. C., & Arnold, S. J. (1989). Visualizing multivariate selection. *Evolution*, *43*(6), 1209–1222. https://doi.org/10.1111/j.1558-5646.1989.tb02569.x

Pizzari, T., & Parker, G. A. (2009). Sperm competition and sperm phenotype. In T. R. Birkhead, D. J. Hosken, & S. Pitnick (Eds.), *Sperm biology: An evolutionary perspective* (pp. 207–245). Academic Press.

Prokuda, A. Y., & Roff, D. A. (2014). The quantitative genetics of sexually selected traits, preferred traits and preference: A review and analysis of the data. *Journal of Evolutionary Biology*, *27*(11), 2283–2296. https://doi.org/10.1111/jeb.12483

Schärer, L., Brand, J. N., Singh, P., Zadesenets, K. S., Stelzer, K. -P., & Viktorin, G. (2020). A phylogenetically-informed search for an alternative *Macrostomum* model species, with notes on taxonomy, mating behaviour, karyology and genome size. *Journal of Zoological Systematics and Evolutionary Research*, *58*, 41–65.

Schärer, L., Janicke, T., & Ramm, S. A. (2014). Sexual conflict in hermaphrodites. In W. R. Rice, & S. Gravilets (Eds.), *The genetics and biology of sexual conflict* (pp. 265–290). Cold Spring Harbor Press.

Schärer, L., Joss, G., & Sandner, P. (2004). Mating behaviour of the marine turbellarian *Macrostomum* sp.: These worms *suck*. *Marine Biology*, *145*, 373–380.

Schärer, L., & Ladurner, P. (2003). Phenotypically plastic adjustment of sex allocation in a simultaneous hermaphrodite. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1518), 935–941. https://doi.org/10.1098/rspb.2002.2323

Schärer, L., Littlewood, D. T. J., Waeschenbach, A., Yoshida, W., & Vizoso, D. B. (2011). Mating behavior and the evolution of sperm design. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1490–1495. https://doi.org/10.1073/pnas.1013892108

Schärer, L., & Pen, I. (2013). Sex allocation and investment into pre- and post-copulatory traits in simultaneous hermaphrodites: the role of polyandry and local sperm competition. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *368*(1613), 20120052. https://doi.org/10.1098/rstb.2012.0052

Schärer, L., & Vizoso, D. B. (2007). Phenotypic plasticity in sperm production rate: There's more to it than testis size. *Evolutionary Ecology*, *21*(3), 295–306. https://doi.org/10.1007/s10682-006-9101-4

Sekii, K., Vizoso, D. B., Kuales, G., De Mulder, K., Ladurner, P., & Schärer, L. (2013). Phenotypic engineering of sperm-production rate confirms evolutionary predictions of sperm competition theory. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1757), 20122711. https://doi.org/10.1098/rspb.2012.2711

Simmons, L. W., House, C. M., Hunt, J., & García-González, F. (2009). Evolutionary response to sexual selection in male genital morphology. *Current Biology*, *19*(17), 1442–1446. https://doi.org/10.1016/j.cub.2009.06.056

Simmons, L. W., & Moore, A. J. (2009). Evolutionary quantitative genetics of sperm. In T. R. Birkhead, D. J. Hosken, & S. Pitnick (Eds.), *Sperm biology: An evolutionary perspective* (pp. 405–434). Elsevier Academic Press Inc.

Simon, M. N., Marroig, G., & Arnold, S. J. (2022). Detecting patterns of correlational selection with sampling error: A simulation study. *Evolution*, *76*(2), 207–224. https://doi.org/10.1111/evo.14412

Stinchcombe, J. R., Agrawal, A. F., Hohenlohe, P. A., Arnold, S. J., & Blows, M. W. (2008). Estimating nonlinear selection gradients using quadratic regression coefficients: Double or nothing? *Evolution*, *62*(9), 2435–2440. https://doi.org/10.1111/j.1558-5646.2008.00449.x

Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *8*(11), 1639–1644. https://doi.org/10.1111/2041-210x.12797

Vellnow, N., Marie-Orleach, L., Zadesenets, K. S., & Schärer, L. (2018). Bigger testes increase paternity in a simultaneous hermaphrodite, independently of the sperm competition level. *Journal of Evolutionary Biology*, *31*(2), 180–196. https://doi.org/10.1111/jeb.13212

Vizoso, D. B., Rieger, G., & Schärer, L. (2010). Goings-on inside a worm: Functional hypotheses derived from sexual conflict thinking. *Biological Journal of the Linnean Society*, *99*(2), 370–383. https://doi.org/10.1111/j.1095-8312.2009.01363.x

Wudarski, J., Egger, B., Ramm, S. A., Schärer, L., Ladurner, P., Zadesenets, K. S., Rubtsov, N. B., Mouton, S., & Berezikov, E. (2020). The free-living flatworm *Macrostomum lignano*. *EvoDevo*, *11*(1), 1–8.

Wudarski, J., Simanov, D., Ustyantsev, K., De Mulder, K., Grelling, M., Grudniewska, M., Beltman, F., Glazenburg, L., Demircan, T., Wunderer, J., Qi, W., Vizoso, D. B., Weissert, P. M., Olivieri, D., Mouton, S., Guryev, V., Aboobaker, A., Schärer, L., Ladurner, P., & Berezikov, E. (2017). Efficient transgenesis and annotated genome sequence of the regenerative flatworm model *Macrostomum lignano*. *Nature Communications*, *8*(1), 2120. https://doi.org/10.1038/s41467-017-02214-8