doi:10.1111/evo.12861

Quantifying episodes of sexual selection: Insights from a transparent worm with fluorescent sperm

Lucas Marie-Orleach,[1,2,3] Tim Janicke,[1,4] Dita B. Vizoso,[1] Patrice David,[4] and Lukas Schärer[1]

[1]*Zoological Institute, University of Basel, Basel, Switzerland*

[2]*Centre for Biological Diversity, University of St Andrews, St Andrews, United Kingdom*

[3]*E-mail: lmo2@st-andrews.ac.uk*

[4]*Centre d'Écologie Fonctionnelle et Évolutive, Montpellier, France*

Received August 5, 2015
Accepted December 22, 2015

Sexual selection operates through consecutive episodes of selection that ultimately contribute to the observed variance in reproductive success between individuals. Understanding the relative importance of these episodes is challenging, particularly because the relevant postcopulatory fitness components are often difficult to assess. Here, we investigate different episodes of sexual selection on the male sex function, by assessing how (precopulatory) mating success, and (postcopulatory) sperm-transfer efficiency and sperm-fertilizing efficiency contribute to male reproductive success. Specifically, we used a transgenic line of the transparent flatworm, *Macrostomum lignano*, which expresses green fluorescent protein (GFP) in all cell types, including sperm cells, enabling in vivo sperm tracking and paternity analysis. We found that a large proportion of variance in male reproductive success arose from the postcopulatory episodes. Moreover, we also quantified selection differentials on 10 morphological traits. Testis size and seminal vesicle size showed significant positive selection differentials, which were mainly due to selection on sperm-transfer efficiency. Overall, our results demonstrate that male reproductive success in *M. lignano* is not primarily limited by the number of matings achieved, but rather by the ability to convert matings into successful fertilizations, which is facilitated by producing many sperm.

KEY WORDS: **Opportunity for selection, quantification of sexual selection, selection gradients, sperm competition, variance decomposition.**

Bateman (1948) famously introduced a framework for quantifying the strength of sexual selection that is mainly based on the linear relationship between mating success and reproductive success (later called the "Bateman gradient"; Lande and Arnold 1983; Arnold and Wade 1984; Arnold and Duvall 1994; Jones 2009; Anthes et al. 2010). This framework assumes that sexual selection arises primarily from competition for mating partners and mate choice, as originally defined by Darwin (1871). However, it is now widely acknowledged that, in promiscuous species, sexual selection can continue after mating, via sperm competition (Parker 1970, 1998) and/or cryptic female choice (Charnov 1979; Thornhill 1983; Eberhard 1996). Thus, postcopulatory sexual selection needs to be integrated when quantifying sexual selection (Eberhard 2009; Birkhead 2010; Jennions and Kokko 2010; Collet et al. 2014).

Beyond the historical distinction between pre- and postcopulatory sexual selection, it is becoming increasingly clear that sexual selection is more complex still. Specifically, sexual selection may act along multiple episodes of an individual's reproduction (e.g., fighting against sexual competitors, courting and choosing mates, transferring ejaculates, and competing against competitors' ejaculate) during which different kinds of traits may be under selection (e.g., sexual behaviors, weapons, ornaments, ejaculate production, sperm behavior, and morphology; reviewed in Andersson 1994; Birkhead et al. 2009). Consequently, a more fine-scaled understanding of sexual selection requires its

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

decomposition into multiple episodes. This can be achieved by decomposing an individual's reproductive success into different fitness components (here defined as the individual's success during a selection episode; Arnold and Wade 1984; Webster et al. 1995; Collet et al. 2012, 2014; Rose et al. 2013; Pélissié et al. 2014; Devigili et al. 2015; Janicke et al. 2015). Using this analytical approach allows us to investigate the relative importance of, and the potential interactions between, different episodes of selection (Pizzari et al. 2002; Jones 2009; Anthes et al. 2010). In addition, a complete understanding of sexual selection also requires that one (1) identifies the selective forces operating on specific phenotypic traits and (2) assesses the relative importance of different episodes acting on those traits. The fact that it is difficult to quantify all relevant pre- and postcopulatory processes in the same study system has until now made this a challenging task.

Several recent studies have decomposed male reproductive success along two components (Collet et al. 2012; Pischedda and Rice 2012; Pélissié et al. 2014; Devigili et al. 2015; Janicke et al. 2015). In these studies, mating success—inferred from either behavioral observation ("copulatory mating success," Collet et al. 2012; Pélissié et al. 2014; Janicke et al. 2015) or paternity analysis ("genetic mating success," Pischedda and Rice 2012; Devigili et al. 2015)—was assessed together with the resulting male reproductive success. This permitted male reproductive success to be decomposed into mating success and paternity share, corresponding to a pre- and a postcopulatory episode of selection, respectively. But this kind of decomposition is arguably incomplete because postcopulatory sexual selection is usually thought to be determined by two ejaculate features (Pizzari and Parker 2009). First, the relative number of sperm that enter the female reproductive tract often plays a critical role in sperm competition so that, in many species, selection acts on the efficiency to transfer and store sperm in the sperm recipient (hereafter called sperm-transfer efficiency; reviewed in Parker 1998). Second, the morphology and the behavior of the sperm (Snook 2005) and the seminal fluid transferred along with the sperm (Chapman 2001; Arnqvist and Rowe 2005) have the potential to affect the efficiency with which each successfully transferred sperm is converted into successful fertilization, thereby biasing the fertilization success toward particular donors (hereafter called sperm-fertilizing efficiency; Pizzari and Parker 2009). Because (1) sperm transfer and storage, and (2) sperm recruitment and fertilization occur in sequence and may involve different traits, one can consider them as being two distinct episodes of sexual selection. However, disentangling them requires to study the sperm inside the female reproductive tract in vivo, which is challenging. Owing to this difficulty we, to our knowledge, currently lack quantitative studies that simultaneously consider how mating success, sperm-transfer efficiency, and sperm-fertilizing efficiency affect male reproductive success.

Here, we report a study that aims at quantifying sexual selection along consecutive episodes of selection by using two distinct model formulations that allow us to decompose male reproductive success (1) into mating success and postmating success, and (2) into mating success and two postcopulatory fitness components, sperm-transfer efficiency, and sperm-fertilizing efficiency (Fig. 1). For this, we used the simultaneously hermaphroditic flatworm *Macrostomum lignano*, a species in which a recently established transgenic line was shown to express green fluorescent protein (hereafter called GFP) in all cell types, including the sperm cells (Demircan 2013; Marie-Orleach et al. 2014). In combination with the worms' transparency, these fluorescent sperm allow one to visualize sperm received from a GFP-expressing individual (hereafter GFP[+]) inside the female sperm-storage organ of a living wild-type sperm recipient (hereafter GFP[–]) (Janicke et al. 2013). In the present study, we recorded male reproductive performance of focal GFP(+) worms (hereafter focals) in a competitive context by measuring (1) their copulatory activity, (2) the resulting number of sperm cells successfully transferred to the female sperm-storage organ of the GFP(–) partners, and (3) the resulting number of offspring sired. Moreover, we measured a suite of morphological traits in the focals (including gonad size, male copulatory stylet morphology, and sperm morphology) to estimate the selection differentials on male reproductive success as well as on each of the studied fitness components. Given that sexual selection in simultaneous hermaphrodites has been argued to be shifted toward postcopulatory episodes of selection compared to gonochorists (Michiels 1998; Schärer and Pen 2013), we expected that postcopulatory fitness components are the prevailing determinants of male reproductive success. Also, we expected testis size, male copulatory stylet morphology, and sperm morphology to predict sperm-transfer success, and sperm morphology to predict sperm-fertilizing success.

Materials and Methods

MODEL ORGANISM

The free-living flatworm *M. lignano* (Macrostomorpha, Platyhelminthes) inhabits the intertidal zone of the Northern Adriatic Sea (Ladurner et al. 2005). Laboratory cultures are maintained at 20°C in glass Petri dishes in f/2 medium (Andersen et al. 2005) and are fed with the diatom *Nitzschia curvilineata*, and individuals from two outbred cultures, one GFP(+) and one GFP(–), were used in this study (see Supporting Information A and next section for details). *Macrostomum lignano* is an outcrossing and promiscuous simultaneous hermaphrodite (Schärer and Ladurner 2003; Janicke and Schärer 2009a). Copulations are frequent (about six copulations per hour) and consist of the reciprocal insertion of the male copulatory organ (hereafter called stylet) into the female sperm-receiving and sperm-storage organ (hereafter called female antrum) of the partner (Schärer et al. 2004). Individuals

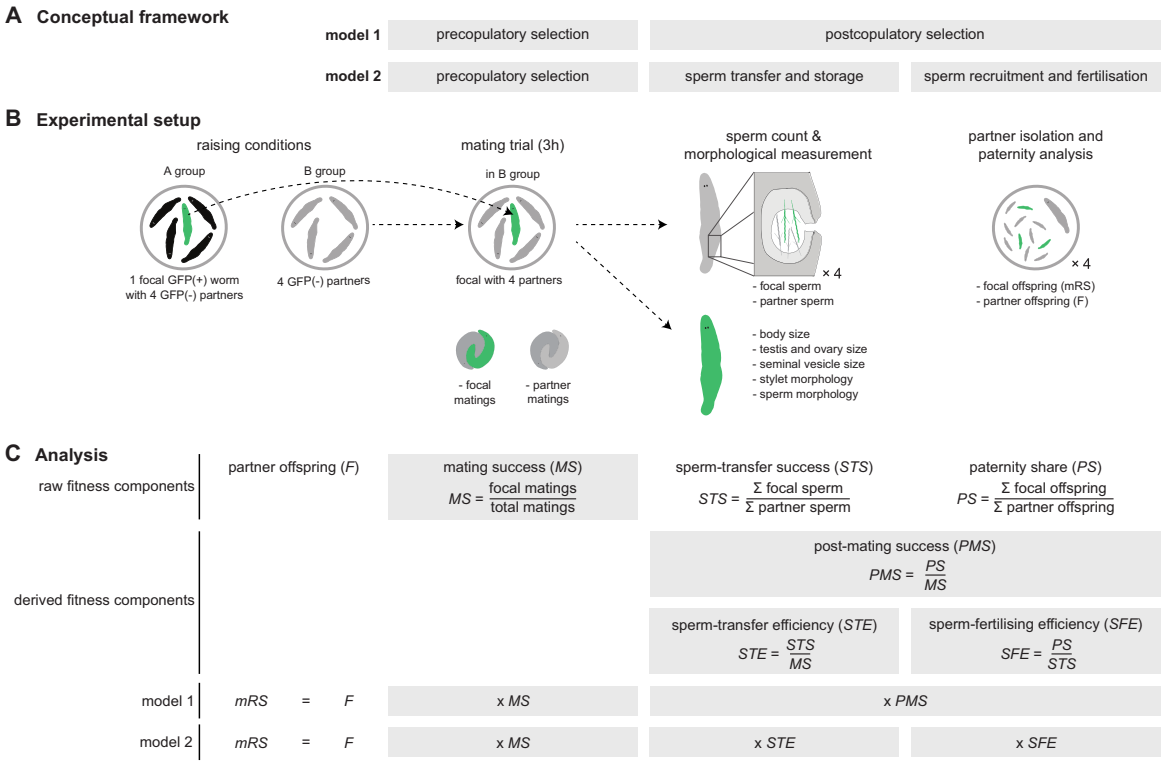

Figure 1. Overview over (A) the experimental rationale, (B) the observations and measurements, and (C) the obtained raw and derived measures, namely mating success (*MS*), sperm-transfer success (*STS*), focal offspring (*mRS*), paternity share (*PS*), partner fecundity (*F*), postmating success (*PMS*), sperm-transfer efficiency (*STE*), and sperm-fertilizing efficiency (*SFE*). Although the focal worms were dyed with a blue dye in the mating experiment, we here depict them green to represent their GFP(+) status.

trade-off their resource allocation between the male and female sex functions (Schärer et al. 2005; Janicke and Schärer 2009b), with more male-biased individuals having a higher sperm production rate that is reflected by larger testes and a faster replenishment of the seminal vesicle (Schärer and Vizoso 2007). Some of the received sperm is stored in the female antrum and can then be displaced by subsequent mating partners (Marie-Orleach et al. 2014). The sperm have a complex morphology—notably including stiff lateral bristles—that may be an evolutionary response to sexual conflict over the fate of received ejaculate (Vizoso et al. 2010; Schärer et al. 2011). The transparency of the worm allows us to measure a range of traits in vivo, such as gonad size (Schärer and Ladurner 2003), stylet morphology (Janicke and Schärer 2009a), and the number of received sperm in the female antrum (Janicke et al. 2011). Additionally, sperm morphology can be measured by amputating the tail plate of the worm (Janicke and Schärer 2010), which can later regenerate it within a few days (Egger et al. 2006).

GFP TECHNIQUES

This study requires us to discriminate between competing sperm of different donors inside recipients in vivo and to assess the resulting paternity share. Both can be achieved in *M. lignano* by taking advantage of recently established transgenics that express GFP ubiquitously, including in the sperm cells (see Supporting Information A). Sperm of a GFP(+) individual show a GFP(+) signal when observed under epifluorescence illumination, which allows us to quantify in vivo the number and proportion of GFP(+) sperm stored together with GFP(–) sperm in a multiply mated GFP(–) recipient (Janicke et al. 2013; Marie-Orleach et al. 2014). Moreover, the GFP marker is inherited by offspring, allowing us to measure the paternity success of the GFP(+) individual. An earlier study indicated that the inheritance pattern of the GFP marker deviated in a few cases from the Mendelian expectations for a single dominant and homozygous diploid locus, presumably due to an underlying karyotype polymorphism (K. Zadesenets et al., unpubl. ms.; see Marie-Orleach et al. 2014 for more details). To obtain accurate estimates of the paternity success of given focals, we needed to account for the likelihood at which each GFP(+) focal transmits the GFP marker to its offspring, which we determined as the proportion of GFP(+) offspring sired by a GFP(+) focal when mated with a single virgin GFP(–) individual (hereafter called penetrance; see Supporting Information B). Offspring production, mating rate, and morphology were previously found not to differ significantly between GFP(+) and GFP(–) individuals (Marie-Orleach et al. 2014).

EXPERIMENTAL SET-UP

In this study, we sought to (1) quantify the relative contribution of subsequent episodes of sexual selection to the variance in male reproductive success, and (2) identify the episodes during which selection is likely to operate on specific traits. Mating rates are very high in this species (Schärer et al. 2004; Janicke and Schärer 2009b; Marie-Orleach et al. 2013) so that mating as a virgin poorly reflects an individuals' lifetime reproductive behavior. We therefore preferred to use sexually experienced worms that had been allowed to interact and copulate with other worms for some time prior to the experiment, so that focal individuals and their potential partners had likely reached a biologically realistic steady state of production of sperm and eggs, and donation and receipt of ejaculate. We then tested individuals in a similar competitive context to assess selection operating in these specific conditions. To this end, each focal was assigned to two groups of four individuals during the entire experiment, namely a group in which the focal was raised and maintained in a steady state (hereafter called A group), and a group in which we tested the reproductive performance of the focal (hereafter called B group) (Fig. 1). An earlier study found that the number of mating partners (inferred from sperm-transfer success) in groups of five worms is on average 3.2 mates (Janicke et al. 2013), which suggests that the group size we used here induces an intermediate level of polyandry.

Rearing conditions

On day 1, we sampled 500 GFP(+) and 1000 GFP(–) adults from the mass cultures and distributed them for egg laying onto, respectively, five and 10 glass Petri dishes filled with f/2 medium and ad libitum algae. On day 3, we removed all adults, thus limiting the age differences of the resulting juveniles to 48 h. On day 9, we sampled the by now hatched juveniles to create 72 A groups comprising one GFP(+) and four GFP(–) individuals, and 72 B groups comprising only four GFP(–) individuals. Groups were placed in wells of 24-well tissue culture plates (TPP AG, Switzerland) filled with 1.5 mL of 32‰ artificial sea water (ASW), and maintained under specific food conditions (adjusted per capita and day by counting diatoms using a Neubauer-improved counting chamber, Marienfeld GmbH, Germany). For logistic reasons, the experiment was split into four batches, each including one-quarter of the replicates. For the sake of clarity, we only report the days on which the first batch was processed (the three other batches were always processed on the three subsequent days).

From day 9 to 70, the experiment included two experimental phases that did not yield informative data (reported in Supporting Information C for completeness). During these two phases, all replicates were treated in the same way, so no biases can result from these earlier holding conditions.

Estimating mating success

On day 70, we performed mating trials of each GFP(+) focal with its B group partners (Fig. 1). For this, we transferred all A group worms into fresh wells containing the food color Patent Blue V (also called E-131; Werner Schweizer AG, Switzerland; 0.25 mg/mL). A 24-h exposure allows us to visually distinguish colored from noncolored worms without significantly affecting the mating rate or the offspring production (Marie-Orleach et al. 2013). On day 71, we transferred the now blue focal and its B group partners in 8 μl drops into observation chambers (see Schärer et al. 2004) with five groups per chamber. We then filmed the mating interactions for 3 h, at 1 frame/s, using digital video cameras (DFK 41AF02, The Imaging Source Europe GmbH, Bremen, Germany) and the software BTV Pro 6.0b7 (http://www.bensoftware.com/) (see Movie Clip S1). We used KMPlayer version 1.5.1 (http://kmplayer.com) to analyze each movie and estimate the number of copulations involving the focal (hereafter focal matings) and the total number of copulations (hereafter total matings, see Fig. 1 for an overview over all fitness components). Each movie was assessed twice to evaluate our consistency in assessing mating interactions. This procedure indicated a very high repeatability in both focal matings (intraclass correlation coefficient, $r_i = 0.97$, $F_{51,52} = 66.79$, $P < 0.001$) and total matings (intraclass correlation coefficient, $r_i = 0.94$, $F_{51,52} = 31.35$, $P < 0.001$). The means of both observations were used in the following analyses.

Importantly, as we could not distinguish the four noncolored partners, we estimated the mating success based on the number of matings (as in Anthes et al. 2010; Fritzsche and Arnqvist 2013), as opposed to the commonly used approach relying on the estimation of the number of mating partners (e.g., Collet et al. 2012; Pélissié et al. 2012).

Estimating sperm-transfer success

Next, we assessed the number of sperm stored in the female antrum of all individuals (Fig. 1), as previously reported (Janicke et al. 2011). We recorded a first movie of the female antrum of all individuals by focusing through the preparation under differential interference contrast illumination, visualizing the total number of sperm in storage (see Movie Clips S2, S4). For the GFP(–) individuals, we then recorded a second female antrum movie under epifluorescence illumination to visualize the number of GFP(+) sperm, that is, the number of sperm successfully transferred by the focal (see Movie Clips S3, S5). We used a Leica DM2500 microscope (Leica Microsystems, Heerbrugg, Switzerland) equipped with an epifluorescence light source, a GFP filter cube (11513890, Leica Microsystems), and a digital microscope camera (Leica DFC360 FX, Leica Microsystems). Movies were recorded using the Leica Application Suite 4.1.0 (Leica Microsystems). We analyzed the movies using KMPlayer

and counted, for each female antrum, the total number of sperm and GFP(+) sperm. This yielded, for each replicate, the total number of sperm counted in the female antra of the four potential partners (partner sperm) and the total number of GFP(+) sperm in those partners (focal sperm, Fig. 1). Counts of both the total number of sperm and the number of GFP(+) sperm show high repeatabilities (Janicke et al. 2011; Marie-Orleach et al. 2014). However, we encountered some worms for which we could only assess the number of GFP(+) sperm cells due to the presence of a ripe egg in the female antrum ($N = 57$ out of 208 in total), which prevents reliable counts of the total number of sperm cells. For these individuals, we used the average number of total sperm cells computed from the counts of all GFP(–) individuals without eggs in the female antrum (excluding these individuals yielded qualitatively similar results). Note that, given the presence of substantial sperm displacement within 24 h (Marie-Orleach et al. 2014), we can safely assume that the sperm recipients to no longer carried GFP(+) sperm from the earlier phase that ended 11 days earlier. Thus, our sperm-transfer success estimate relies on the sperm that were transferred by the focal individual during the 3-h mating trial, and that remained successfully stored in the sperm recipients at the time of observation.

Measuring morphological traits

Next, we took micrographs of each focal to assess body, testis, ovary, and seminal vesicle size (see Schärer and Ladurner 2003), and stylet morphology (see Janicke and Schärer 2009a), by using a Leica DM2500 microscope, an Imaging Source DFK 41AF02 camera, and BTV Pro 6.0b7. We analyzed micrographs with ImageJ 1.45s (http://rsb.info.nih.gov/ij/), and we used geometric morphometrics (Zelditch et al. 2004) to assess the stylet centroid size, and stylet shape based on the first three relative warp scores (hereafter RWS; see Janicke and Schärer 2009a for details). RWS1 mainly captured the general stylet curvature, RWS2 the width of the stylet and the orientation of the stylet tip, and RWS3 the orientation of the stylet tip (see Supporting Information D for visualization). To account for data skewness, we used square root transformation for seminal vesicle size, and cube root transformation for testis and ovary size. All these measurements show good repeatabilities, ranging from 0.57 to 0.97 (Schärer and Ladurner 2003; Janicke and Schärer 2009a; T. Janicke and L. Schärer, unpubl. data).

Estimating reproductive success

Next, all individuals were maintained in isolation until day 82, and we counted and assessed the GFP status of the produced offspring. Thereby, we could assess the total number of offspring produced by the four partners through their female sex function (hereafter partner fecundity, *F*), of which the GFP(+) offspring were sired by the GFP(+) focal individual. We could then assess the number of offspring produced by each focal, both through its own female sex function, and through its male sex function (hereafter male reproductive success, *mRS*, Fig. 1). See Supporting Information B for information on the penetrance of the marker.

Estimating sperm morphology

On day 82, we characterized the morphology of the sperm of the focals, following Janicke and Schärer (2010). We took micrographs of about 10 sperm per individual from which we determined two sperm traits (averaged over all measured sperm per individual), namely sperm length and bristle length, which have, respectively, a high (0.96) and moderate but significant (0.46) repeatability (Janicke and Schärer 2010). We cube root transformed the sperm length data for analysis.

DATA ANALYSIS

We started with 72 replicates, but lost seven, six, and seven replicates, respectively, due to handling errors, developmental problems, and production of not enough offspring to reliably assess the penetrance of the GFP marker (< 10 offspring), yielding a final sample size of 52 replicates.

Our experiment allowed us to measure copulatory activity, sperm transfer, and offspring production by the focal and its four partners (see "raw fitness components" in Fig. 1). First, we estimated a focal's performance relative to that of the partners within its group—the focal's direct reproductive competitors—in obtaining matings (mating success, *MS*), successfully transferring sperm to partners (sperm-transfer success, *STS*), and siring offspring (paternity share, *PS*; see Fig. 1). Second, and as advocated by Jones (2009), we expressed these data relative to all focal worms by dividing the focal values by the mean values of all focals so that the means of all fitness measurements equal 1. This enables decomposing the variance observed in relative male reproductive success along subsequent fitness components (Pélissié et al. 2014), and also facilitates the comparison of the slopes of the linear regressions performed between our different fitness measurements. We use asterisks to denote these "relative to groups" data, that is, *MS**, *STS**, *PS** (see also Pélissié et al. 2014). Finally, we standardized all morphological traits (i.e., mean = 0 and SD = 1; Jones 2009) to facilitate the comparison between our morphological traits. Expressing the data in this way also facilitates comparisons with other studies (Jones 2009).

Decomposition of the variance in male reproductive success

We decomposed the variance observed in *mRS** along subsequent fitness components (Arnold and Wade 1984; Webster et al. 1995; Collet et al. 2012, 2014; Pélissié et al. 2014), by using two different deterministic models (see Fig. 1).

Model (1) is a decomposition of the variance (V) in *mRS** along three fitness components: partner fecundity (*F**), mating success (*MS**), postmating success (*PMS**), and their covariances (COV).

$$V(mRS^*) \approx V(F^*) + V(MS^*) + V(PMS^*) + \text{covariances} \quad (1)$$

in which

$$\text{covariances} = 2\text{COV}(F^*, MS^*) + 2\text{COV}(F^*, PMS^*) + 2\text{COV}(MS^*, PMS^*).$$

Model (2) includes the number of sperm cells successfully transferred to the partners (i.e., sperm-transfer success, *STS*), which allows us to further decompose postmating success (*PMS**) into sperm-transfer efficiency (*STE**) and sperm-fertilizing efficiency (*SFE**). Thus:

$$V(mRS^*) \approx V(F^*) + V(MS^*) + V(STE^*) + V(SFE^*) + \text{covariances} \quad (2)$$

in which

$$\text{covariances} = 2\text{COV}(F^*, MS^*) + 2\text{COV}(F^*, STE^*) + 2\text{COV}(F^*, SFE^*) + 2\text{COV}(MS^*, STE^*) + 2\text{COV}(MS^*, SFE^*) + 2\text{COV}(STE^*, SFE^*).$$

For all variances and covariances, we computed the 95% percentile confidence intervals by bootstrapping (10,000 iterations). Importantly, because *PMS**, *STE**, and *SFE** were not directly observed but rather derived from direct observations (Fig. 1), a certain amount of the variance observed in these components is expected to arise simply due to sampling error (i.e., sampling a finite number of sperm and offspring to count the proportion of GFP(+) types among them). For example, even if each copulation were assumed to have the same probability of fertilizing a given egg, we would still have observed variance in *PMS** because of the fertilization lottery. The expected amount of error variance depends on how good the fitness component estimates are (e.g., the more offspring sampled, the better our estimate of *PMS**). As a consequence, such sampling errors could have artificially inflated V(*PMS**), V(*STE**), and V(*SFE**). Therefore, we accounted for this error variance by computing the variance expected from a binomial sampling error (see Supporting Information E for details), which we then subtracted from the observed variance of their respective fitness components (see also Pélissié et al. 2012). The remaining variance can be considered as a conservative estimate of biologically meaningful variation in the individual performances of the focals.

We tested for differences between variances by using a pairwise signed difference test. Specifically, we bootstrapped the variances observed in each fitness component (10,000 iterations), and calculated the differences between the bootstrapped variances of two fitness components for each iteration. We then counted the occurrences of positive and negative differences, and used the less frequent occurrence to derive a *P*-value of the pairwise comparison as: $P = (2 \times \text{occurrence})/10,000$.

Finally, to test whether the covariances between fitness components of models (1) and (2) significantly differed from zero, we tested the pairwise correlations by using Spearman's rank correlation coefficient.

Selection gradients and morphological traits

We investigated the linear relationship between mating success (*MS**) and paternity share (*PS**), which we further decomposed into the linear relationships between mating success (*MS**) and sperm-transfer success (*STS**), and sperm-transfer success (*STS**) and paternity share (*PS**). For this, we performed ordinary least squares linear regressions. Additionally, for each morphological trait, we computed the total male selection differential (i.e., its effect on *mRS**), as well as the selection differentials on each of the different fitness components of models (1) and (2) (i.e., *F**, *MS**, *PMS**, *STE**, and *SFE**). For this, we performed ordinary least squares linear regressions.

Comparing different estimators of male mating success

Finally, we investigated the relationships between male reproductive success (*mRS*) and different estimators of male mating success. For this, we used the following four measures of mating success: the number of copulations the focal performed (focal matings), the number of sperm cells successfully transferred to the partners (focal sperm), the number of mating partners that had at least one sperm from the focal in storage (sperm mating success, *sMS*), and the number of mating partners producing at least one offspring that was sired by the focal (genetic mating success, *gMS*). We performed ordinary least squares linear regressions.

The linear regressions were carried out in JMP 10.0.1 (SAS Institute Inc., Cary, NC), and the statistical analyses for the variance decomposition were carried out in Mathematica 9.0 (Wolfram Research, Inc., Champaign, Illinois).

Results

DECOMPOSITION OF THE VARIANCE IN MALE REPRODUCTIVE SUCCESS

The decomposition of the variance in male reproductive success (*mRS**) according to model (1) indicated that partner fecundity (*F**), mating success (*MS**), and postmating success (*PMS**) accounted for 16%, 9%, and 46% of the variance observed, respectively (Fig. 2A). Importantly, the variance arising from *PMS** was significantly larger than the variance arising from *MS**, and

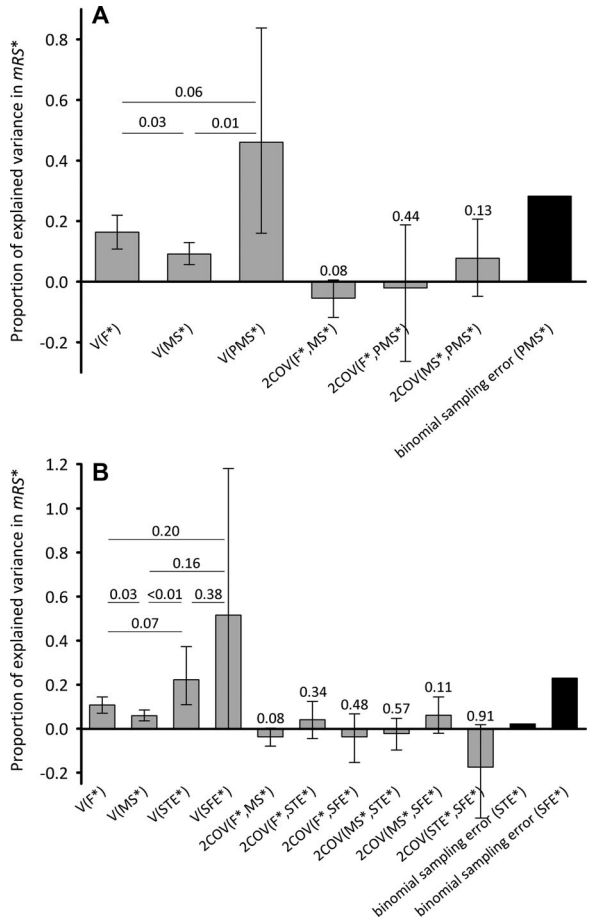

Figure 2. **Decomposition of the variance in male reproductive success (*mRS**) along different fitness components and their covariances (see Fig. 1 for details on parameters). (A) The decomposition following model (1) along three multiplicative fitness components, partner fecundity (*F**), mating success (*MS**), and postmating success (*PMS**), after subtracting the binomial sampling error that arose from *PMS** (black bar). (B) The decomposition following model (2) along four multiplicative fitness components, partner fecundity (*F**), mating success (*MS**), sperm-transfer efficiency (*STE**), and sperm-fertilizing efficiency (*SFE**), after subtracting the binomial sampling errors that arose from *STE** and *SFE** (black bars). Error bars represent the bootstrapped 95% percentile confidence intervals. The numbers indicate either *P* values of the pairwise comparisons between fitness components (connected by lines), or *P* values testing if covariances are different from zero. See Methods for details.

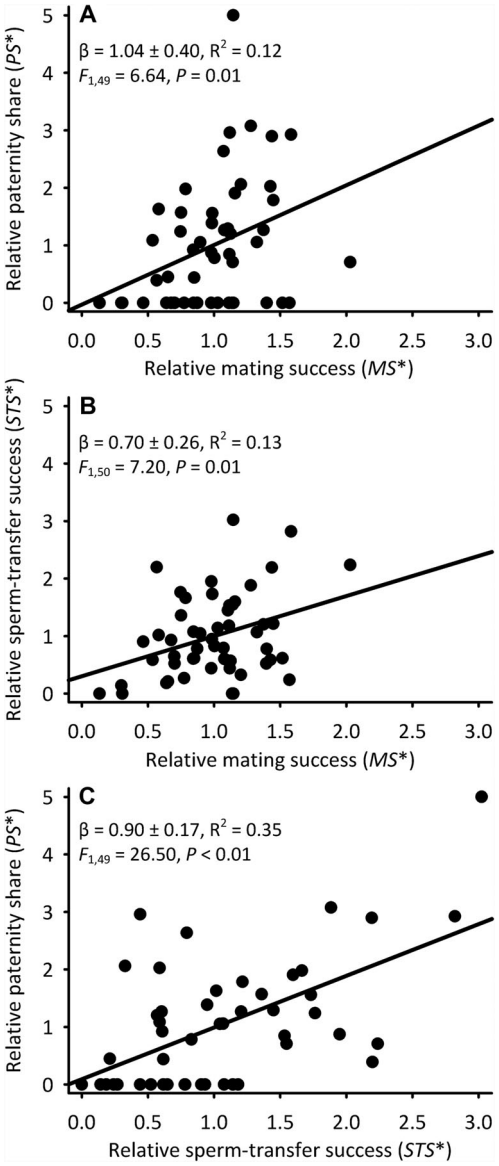

Figure 3. **Linear regressions of (A) paternity share (*PS**) on mating success (*MS**), (B) sperm-transfer success (*STS**) on mating success (*MS**), and (C) paternity share (*PS**) on sperm-transfer success (*STS**) (see Fig. 1 for details on parameters).

it tended to be larger than the variance arising from *F**, whereas the variance arising from *MS** was smaller than that due to *F**. The remaining 29% of variance was due to the binomial error variance in *PMS** and the covariances between fitness components. None of the covariances were significantly different from zero (Fig. 2A).

The decomposition according to model (2) shows a similar picture, with *F**, *MS**, sperm-transfer efficiency (*STE**), and sperm-fertilization efficiency (*SFE**) accounting for 11%, 6%, 22%, and 52% of the variance observed in *mRS**, respectively (Fig. 2B). The variance arising from *STE** tended to be larger than the variance arising from *F**, and was significantly larger than the variance arising from *MS**. Although *SFE** accounted for the largest portion of variance, it was not significantly different from those arising from the other components. This was probably due to the generally small numbers of offspring produced, which made our estimates of *SFE** relatively error-prone. Two percent and 23% of the variance was due to the binomial sampling error arising from *STE** and *SFE**, respectively. The covariances between fitness components were overall negative, accounting for

–16% of the variance, but none were significantly different from zero (Fig. 2B).

The variance observed in *mRS** was 1.15, whereas the variances predicted by models (1) and (2) were 1.48 and 2.23, respectively. The discrepancy between V(*mRS**) and the variances predicted by the models arises from the skewed distributions of our data, especially in *SFE**, which suggests that model (2) needs to be interpreted with some caution.

SELECTION GRADIENTS AND MORPHOLOGICAL TRAITS

We found a positive relationship between mating success (*MS**) and paternity share (*PS**) (Fig. 3A), suggesting that individuals that copulated relatively more also sired relatively more offspring. The decomposition of this relationship showed that individuals that copulated relatively more managed to successfully transfer relatively more sperm in their partners (*MS** vs. *STS**; Fig. 3B), and that individuals that managed to successfully transfer relatively more sperm sired relatively more offspring (*STS** vs. *PS**; Fig. 3C). In contrast, *MS** was not related to the proportion of offspring produced via the focal's female function ($\beta = 0.29 \pm 0.42$, $R^2 = 0.01$, $F_{1,50} = 0.47$, $P = 0.50$). We do not discuss this result in much detail because, unlike for the male function, female reproductive success resulted not only from copulations performed during the staged mating trial (i.e., with the B group worms), but also from copulations performed before the mating trial (i.e., with the A group worms). We therefore did not expect mating success during the mating trial to be a good predictor of the focal's female reproductive success, given the steady-state experimental paradigm we used here.

The analyses of the morphological traits showed positive male selection differentials for testis size (Fig. 4A) and seminal vesicle size (Fig. 4B), which both mainly arose from selection on *STE** (Fig. 4C, D; Table 1). In other words, individuals with bigger testes and seminal vesicles sired more offspring, presumably because they successfully transferred relatively more sperm cells per copulation. Moreover, stylet centroid size showed a significant male selection differential. The selection differentials on all the measured morphological traits are summarized in Table 1.

COMPARING DIFFERENT ESTIMATORS OF MALE MATING SUCCESS

The relationships between male mating success and male reproductive success depended greatly on how the former was measured (Fig. 5). When it was based on the raw number of focal matings, male mating success was a rather poor and nonsignificant predictor of male reproductive success (Fig. 5A). Note, however, that when the differences among groups in total matings and partner offspring are accounted for (i.e., using *MS** and *PS** instead of focal matings and *mRS*), the relationship is significant (see Fig. 3A). In contrast, when mating success was inferred from the sperm data, mating success became a significant predictor of male reproductive success, both when the measure was based on focal sperm (Fig. 5B) and when it was based on the number of mates to whom the focals successfully transferred sperm (*sMS*, Fig. 5E). Finally, when mating success was inferred from the number of individuals in which the focals successfully sired offspring (*gMS*), mating success was a strong predictor of male reproductive success (Fig. 5F), but this is at least in part due to the fact that this involves an autocorrelation (see legend of Fig. 5).

Discussion

Our decomposition analysis showed that mating success accounted for a relatively small part of the variance observed in male reproductive success, which instead mainly arose from postcopulatory success. We further found that mating success had a positive effect on paternity share, which was mediated via a positive relationship between mating success and sperm-transfer success. Moreover, individuals with bigger testes and bigger seminal vesicles sired more offspring, probably because they managed to successfully transfer relatively more sperm per copulation. Finally, we showed that using different estimators of male mating success can greatly influence the relationship between mating success and male reproductive success, which illustrates a potential risk of misinterpreting Bateman gradients. In the following, we discuss these different findings in turn.

DECOMPOSITION OF THE VARIANCE IN MALE REPRODUCTIVE SUCCESS

Our results show that most of the variance observed in male reproductive success arose from postmating success, accounting for 46%, whereas only 9% arose from mating success (model 1). Previous studies that have decomposed the variance in male reproductive success have found contrasting results about the relative importance of pre- and postcopulatory episodes of selection. Specifically, several studies have found that postmating success explained a relatively large portion of the variance in male reproductive success, namely 46% in the red junglefowl *Gallus gallus* (Collet et al. 2012), 37.5% in the livebearing fish *Poecilia reticulata* (Devigili et al. 2015), 36% in the simultaneously hermaphroditic freshwater snail *Physa acuta* (after accounting for the mating order inferred from behavioral observations, but see below, Pélissié et al. 2014; see also Janicke et al. 2015 for another study in the same species). Also, in the particular case of a species with male-pregnancy, Rose et al. (2013) found that 28.4% and 17.5% of variance in male reproductive success arose from the number of eggs transferred per mate and embryo survivorship, respectively. In contrast, Pischedda and Rice (2012) found that in the fruit fly *Drosophila melanogaster*, after accounting

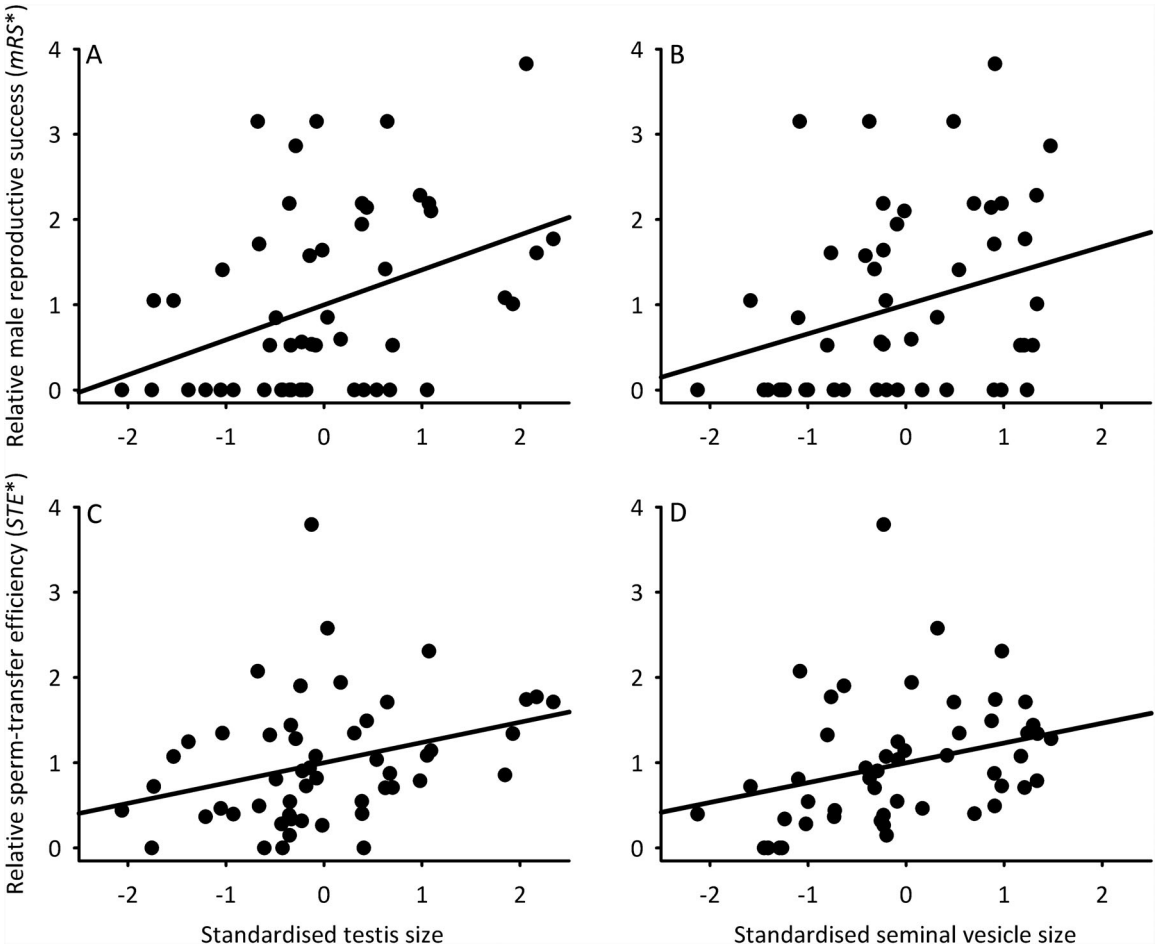

Figure 4. Effects of standardized testis size and seminal vesicle size on (A, B) male reproductive success (*mRS), and (C, D) sperm-transfer efficiency (*STE**). See Table 1 for statistics.**

for the mating order inferred from paternity analysis, only 2% of the variance in male reproductive success arose from postmating success.

In simultaneous hermaphrodites, sexual selection has been argued to be shifted toward the postcopulatory level (Charnov 1979; Michiels 1998; Schärer and Pen 2013) for the following reason. If we assume that mating success is a stronger predictor of male fitness than female fitness, then all individuals in the population may tend to search the proximity of potential partners and attempt to preferentially assume the male role when a mating opportunity arises. This may lead to a conflict over mating roles, as it presumably exposes simultaneously hermaphroditic sperm recipients to more copulations than gonochoristic females (Charnov 1979; Michiels and Newman 1998; Lange et al. 2013). This conflict can arguably be resolved if both mating partners agree to also assume the less-preferred female mating role to have the opportunity to assume the preferred male mating role, leading to reciprocal copulation (or unilateral copulation with conditional reciprocity). Such reciprocity has two consequences: first, individuals likely have to deal with a surplus of received sperm, which may be undesirable for the sperm recipient and thus may be dealt with by postcopulatory sperm selection or removal. Second, male fitness may then not be primarily limited by the number of matings achieved but rather by the ability to successfully transfer sperm to their partners and to have them used for fertilization by the partner (Charnov 1979; Michiels 1998; Schärer and Pen 2013; Schärer et al. 2014). Therefore, postcopulatory sexual selection is expected to be prevalent in simultaneous hermaphrodites (see, e.g., Koene and Schulenburg 2005; Chase and Blanchard 2006; Schärer et al. 2011), which is fully supported by our data.

A major new contribution of our study is the decomposition of the postmating success into sperm-transfer efficiency and sperm-fertilizing efficiency (model 2). The variance arising from sperm-transfer success was relatively large and was significantly higher than that arising from mating success. Sperm-transfer efficiency is expected to depend on several components, including the ability of individuals to transfer sperm to partners (see the section on "Selection gradients on morphological traits"), the

Table 1. Selection differentials of the 10 measured morphological traits on male reproductive success (*mRS**), and on the different fitness components, partner fecundity (*F**), mating success (*MS**), postmating success (*PMS**), and its components, sperm-transfer efficiency (*STE**), and sperm-fertilizing efficiency (*SFE**).

Traits	*mRS**	*F**	*MS**	*PMS**	*STE**	*SFE**
Body size	0.12 ± 0.15	0.12 ± 0.07	0.00 ± 0.05	–0.09 ± 0.15	0.02 ± 0.10	–0.24 ± 0.19
	$P = 0.41$	$P = 0.09$	$P > 0.99$	$P = 0.54$	$P = 0.84$	$P = 0.21$
Testis size	**0.41 ± 0.14**	0.13 ± 0.08	0.03 ± 0.05	0.17 ± 0.15	**0.24 ± 0.10**	–0.06 ± 0.19
	***P* = 0.01**	$P = 0.05$	$P = 0.53$	$P = 0.26$	***P* = 0.02**	$P = 0.76$
Ovary size	0.01 ± 0.15	0.03 ± 0.07	0.03 ± 0.05	–0.16 ± 0.15	–0.09 ± 0.09	–0.09 ± 0.19
	$P = 0.97$	$P = 0.66$	$P = 0.54$	$P = 0.29$	$P = 0.38$	$P = 0.62$
Seminal vesicle size	**0.34 ± 0.14**	0.06 ± 0.07	0.09 ± 0.05	0.20 ± 0.15	**0.23 ± 0.01**	–0.04 ± 0.20
	***P* = 0.02**	$P = 0.36$	$P = 0.08$	$P = 0.18$	***P* = 0.02**	$P = 0.86$
Stylet centroid size	**0.34 ± 0.14**	0.08 ± 0.07	0.08 ± 0.05	0.00 ± 0.15	0.02 ± 0.10	0.04 ± 0.19
	***P* = 0.02**	$P = 0.25$	$P = 0.11$	$P = 0.98$	$P = 0.88$	$P = 0.84$
Stylet RWS 1	–0.08 ± 0.15	0.04 ± 0.07	–0.03 ± 0.05	–0.08 ± 0.15	0.11 ± 0.10	–0.27± 0.19
	$P = 0.61$	$P = 0.58$	$P = 0.63$	$P = 0.58$	$P = 0.28$	$P = 0.16$
Stylet RWS 2	0.05 ± 0.15	0.01 ± 0.07	0.05 ± 0.05	–0.14 ± 0.15	–0.10 ± 0.10	0.19 ± 0.18
	$P = 0.74$	$P = 0.89$	$P = 0.31$	$P = 0.36$	$P = 0.33$	$P = 0.30$
Stylet RWS 3	–0.19 ± 0.15	0.01 ± 0.07	–0.03± 0.05	–0.09 ± 0.15	–0.05 ± 0.10	–0.13 ± 0.19
	$P = 0.20$	$P = 0.88$	$P = 0.52$	$P = 0.55$	$P = 0.62$	$P = 0.49$
Sperm length	0.21 ± 0.15	0.00 ± 0.17	0.05 ± 0.05	0.07 ± 0.15	0.09 ± 0.11	0.08 ± 0.20
	$P = 0.17$	$P > 0.99$	$P = 0.31$	$P = 0.65$	$P = 0.40$	$P = 0.69$
Sperm bristle length	0.21 ± 0.15	0.02 ± 0.07	0.08 ± 0.05	0.14 ± 0.15	0.15 ± 0.11	0.06 ± 0.20
	$P = 0.18$	$P = 0.83$	$P = 0.14$	$P = 0.37$	$P = 0.17$	$P = 0.75$

Significant *P* values are indicated in bold. All of the significant relationships remain significant when tested with the nonparametric Spearman's correlation test, except for the relationship between stylet centroid size and *mRS ($r_S = 0.22$; $N = 52$; $P = 0.12$). *P* values do not account for multiple testing because the study is considered exploratory and aims at guiding future research. Asterisks stand for relative data. See the Methods for details.**

interactions between ejaculates of different donors (e.g., resisting displacement by consecutive partners), and the interactions between ejaculates and the reproductive tract of the sperm recipients (e.g., seminal fluid effects). The opportunity for selection on the other postmating component, sperm-fertilizing efficiency, seemed to be even larger, but given the large confidence limits of that estimate in our data, the relative importance of this fitness component needs to be interpreted carefully (a larger sample size and longer progeny arrays might have helped to improve our accuracy in measuring this component). The outcome of this episode of selection is expected to mainly depend on interactions with the ejaculates of competing sperm donors and with the sperm recipient, as well as on zygotic, embryonic, and juvenile development. Regardless of the underlying traits, the large opportunity for selection in sperm-transfer efficiency (and possibly in sperm-fertilizing efficiency) observed in our data suggests that these fitness components may capture important episodes of selection in *M. lignano*.

Importantly, one may have expected either positive or negative covariances between different episodes (see, e.g., Pélissié et al. 2014). For instance, variance in overall individual quality could have led to positive covariance between mating success and postmating success, if high-quality individuals could achieve both higher copulation rates and transfer more sperm and/or produce ejaculate with a higher fertilizing efficiency. Alternatively, trade-offs between sperm quality and quantity could potentially have led to a negative covariance between sperm-transfer efficiency and sperm-fertilizing efficiency. We found no significant covariances between any of the estimated fitness components. This should, however, also be interpreted with some caution because we did not restrict food availability to individuals in the later stage of the experiment, which could affect the emergence of such trade-offs (van Noordwijk and De Jong 1986; Schärer et al. 2005).

The decomposition of variance using either of the two models could potentially lead to artifacts for the following reasons. First, because several components are defined as ratios, extreme values could be observed when the denominators approach zero, and thereby inflate the variance. Our method deals with such potential artifacts because a disproportionate influence of extreme values would increase both the binomial error (which is subtracted from the observed variance) and the confidence intervals, thus avoiding spurious conclusions. Second, because the same estimates are at times included in the computation of several components (e.g., STS is both in the nominator of STE and in the denominator of SFE), artifactual negative covariances could arise in case of measurement error. This is because measurement error

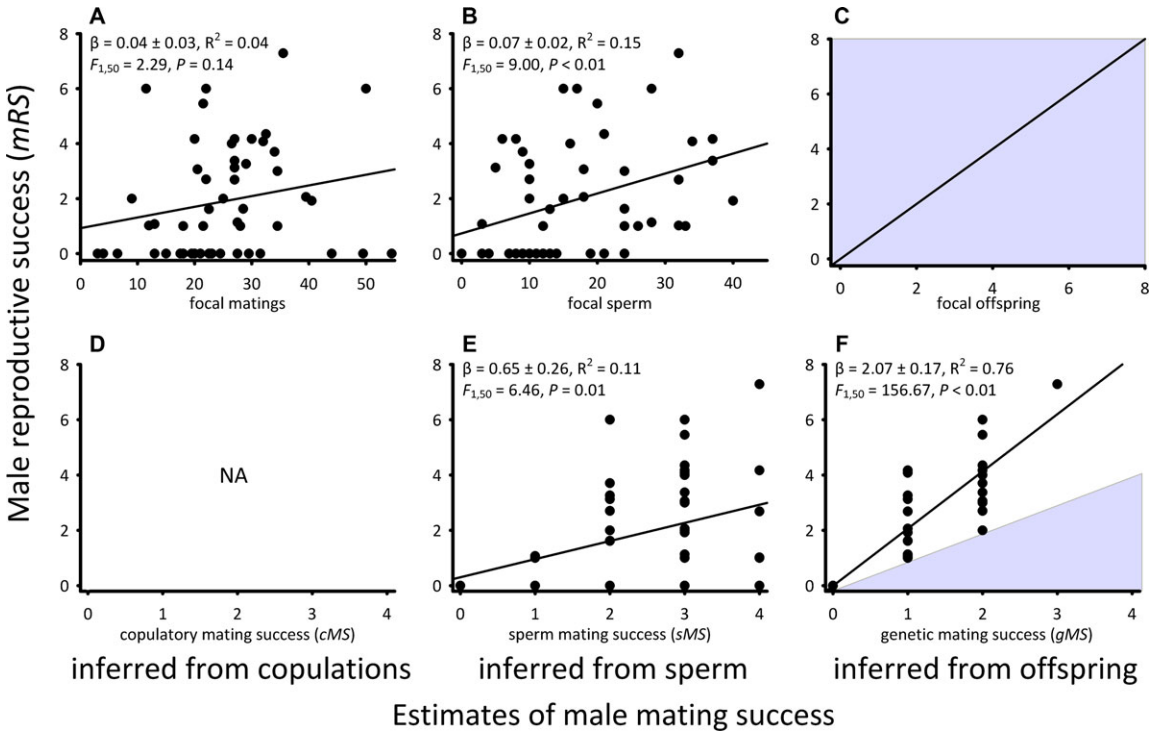

Figure 5. **Linear relationships between male reproductive success (*mRS*) and different estimators of male mating success (here the values are not relativized to more clearly illustrate the structure of the data). Male mating success of the focals could in theory be inferred from the observation of copulations (A, D) of successfully transferred sperm (B, E) and of offspring sired (C, F). Moreover, the upper row (A–C) uses counts of matings, sperm, or offspring, whereas the lower row (D–F) uses the numbers of mates with whom the focal mated (copulatory mating success, *cMS*), to whom it transferred sperm (sperm mating success, *sMS*), and with whom it sired offspring (genetic mating success, *gMS*). (C) and (D) are shown for completeness, but are not represented because (C) shows a perfect fit as the variables on the *x*- and *y*-axes are identical, and estimates of the *x*-axis in (D) are not available (NA) in this study. The gray regions represent areas where, by definition, it is impossible to have any datapoints, thereby creating a complete autocorrelation in (C) and a partial autocorrelation in (F).**

creates nonbiological variance shared by the denominator of one component and the denominator of the other. However, because none of our covariances were significant, we believe that measurement errors did not strongly affect our results in this way.

In general, interpreting variance decomposition of reproductive success calls for considerable care. First, variance arising from a fitness component only represents the upper limit (opportunity) for selection—and not necessarily the actual strength of selection (Jennions et al. 2012). For example, the specific mating interactions observed in a mating assay may to some extent be driven by stochastic events: if the same group of individuals were to play it out again, somewhat different mating interactions would likely result. Second, the variance assigned to a given episode depends on the specific fitness component(s) chosen to characterize that episode. For example, both the number of mates and matings are fitness components that can be used to characterize mating success, but they measure different facets of precopulatory sexual selection, and thus the variance ascribed to them may differ substantially. This can be particularly important if detailed observations of precopulatory interactions are considered. For example, Pischedda and Rice (2012) and Pélissié et al. (2014) found that a considerable portion of the variance initially ascribed to the postcopulatory fitness component (i.e., paternity success) could actually be ascribed to the mating order, and they suggested that this variance was therefore due to precopulatory selection. However, some sperm donors might end up in the last-male mating position because they successfully prevented the sperm recipient from remating as a result of their transfer of seminal fluids (e.g., Chen et al. 1988; Peng et al. 2005; Kimura et al. 2013), arguably a postcopulatory trait. It may therefore be misleading to consider the whole variance due to mating order as being linked to precopulatory traits.

In addition to the approach used here of decomposing variance in reproductive success into fitness components, one also needs to understand the traits that determine the outcome of the different episodes of selection to reach conclusions about the operation of sexual selection in a study system. We discuss this next.

SELECTION GRADIENTS AND MORPHOLOGICAL TRAITS

Our results indicate that individuals that mated more obtained a higher paternity share. Moreover, we could further decompose this relationship by assessing an intermediate step, sperm-transfer success, which seems to be important in mediating the correlation between mating success and paternity share in *M. lignano*. Paternity share may be influenced by many factors. In particular, in internally fertilizing animals, the sperm recipient may have some control over the fate of the partner's sperm and thus influence the fertilization success of some sperm donors by preferentially using their sperm. Importantly, in our experimental setup, such cryptic female choice might have occurred both before and after our measure of sperm-transfer success, for example, by preferentially storing sperm of some donors or by preferentially recruiting stored sperm from some donors for fertilization, respectively. Therefore, our results do not permit conclusions about the prevalence of cryptic female choice in *M. lignano*, but highlight that the number of sperm successfully transferred in partners is an important determinant of male reproductive success, and that any trait that positively influences sperm-transfer success is expected to confer a selective advantage. The positive relationship between mating success and sperm-transfer success is of particular interest in *M. lignano* as individuals mate more frequently in response to an increased level of sperm competition (Janicke and Schärer 2009b). Hence, together with our findings, this suggests that individuals may increase their mating rates to achieve numerical dominance against competitors in their mates' stored sperm (Bretman et al. 2011).

Our data also revealed strong selection differentials on testis size and seminal vesicle size, and showed that they both appear to affect sperm-transfer efficiency. This is in accordance with previous studies that showed a positive correlation between testis size and sperm-transfer success (Janicke and Schärer 2009a), and a positive effect of seminal vesicle size on paternity share (Sekii et al. 2013). Although testis size has been shown to be a reliable predictor of sperm production rate (Schärer and Vizoso 2007), seminal vesicle size reflects the amount of sperm available during mating. In our data, testis size and seminal vesicle size were highly correlated (Pearson's correlation, $r = 0.54$, $N = 52$, $P < 0.001$). Thus, this suggests that individuals with a bigger testis and seminal vesicle size have a higher sperm production rate and more sperm available, which allows them to transfer more sperm per copulation and thus to reach a higher siring success.

Moreover, in the current study, neither testis nor seminal vesicle size was correlated with mating success, which seems to contrast with the results of previous studies (Janicke and Schärer 2009b; Sekii et al. 2013). We think that these contrasting results are most probably due to differences in the chosen experimental paradigms. Namely, in the current study individuals were kept in a stable group size throughout the experiment, to assay the performance of focal worms that have experienced a steady-state situation, whereas in the other studies the focal worms were deliberately raised in different social environments compared to the conditions in which they were later assayed, which led to predictable differences in testis and/or seminal vesicle size that we intended to study there. We here decided to use a steady-state experimental approach, because we think that it is more appropriate to measure selection on natural variation and covariation in traits.

Contrary to our initial expectations (Janicke and Schärer 2009a; Vizoso et al. 2010; Schärer et al. 2011), the stylet or sperm morphology traits predicted none of our fitness components. These negative results should, however, be considered with some caution, as the analysis of selection differentials may require larger sample sizes than we were able to achieve here (Hersch and Phillips 2004). Moreover, larger sample sizes would also have allowed us to test for stabilizing or disruptive selection, or for more complex selection on combinations of traits. Overall, our data indicate that directional selection (if any) on the stylet and sperm morphology seems weaker than on testis size and seminal vesicle size in *M. lignano*. Importantly, the genetic architecture of these traits is being studied in an on-going experiment (S. A. Ramm et al., unpubl. data), which will provide critical information to predict their evolutionary trajectories.

COMPARING DIFFERENT ESTIMATORS OF MALE MATING SUCCESS

The rationale behind Bateman gradient analyses is to quantify the strength of precopulatory sexual selection, and so the way in which mating success is estimated is of crucial importance. Measures of mating success are usually inferred either from observations of mating interactions ("copulatory mating success"; e.g., Collet et al. 2012; Pélissié et al. 2012; Fritzsche and Arnqvist 2013) or from parentage analysis ("genetic mating success"; e.g., Bateman 1948; Jones et al. 2000; Gopurenko et al. 2007; Pischedda and Rice 2012). In the present study, the opportunity to assess the sperm-transfer success allowed us to infer an additional measure of mating success based on the presence of successfully transferred sperm in the partners, which we suggest to call "sperm mating success." These three measures of mating success have different meanings because they are inferred from different fitness components and thus capture the result of sexual selection up to different episodes of selection (Anthes et al. 2010; Collet et al. 2014). Namely, copulatory mating success encompasses exclusively the selection on achieving copulations; sperm mating success encompasses the selection on achieving copulations and successfully transferring sperm in the partners; and genetic mating success encompasses the selection on achieving copulations, successfully transferring sperm, fertilizing the partners' ova and, depending on when reproductive success is measured, possibly

also on developing viable zygotes, embryos, and/or juveniles. Consequently, as the measure of mating success gets inherently closer to the estimate of reproductive success, more components of sexual selection will merge, gradually leading to an increasingly strong autocorrelation with reproductive success (Anthes et al. 2010; Collet et al. 2014). This is likely the reason why genetic mating success has often been found to be a strong predictor for male reproductive success (Arnqvist 2013). We illustrate this critical point in Figure 5 and, in line with Anthes et al. (2010) and Collet et al. (2014), we advocate using several measures of mating success whenever possible, and decomposing the relationships between mating success and reproductive success, to provide a more complete understanding of the operation of sexual selection.

Conclusions

In his seminal contribution, Bateman (1948) concluded that "In the male [. . .] fertility is seldom likely to be limited by sperm production but rather by the number of inseminations or the number of females available to him." Our data on the simultaneously hermaphroditic free-living flatworm *M. lignano* contradict this statement, as we found that although male fitness of course depended to some extent on copulation activity, selection appeared to be stronger on the postcopulatory episodes of selection, in which the amount of sperm produced and transferred appeared to be a crucial determinant. Therefore, our findings support the hypothesis that postcopulatory selection is a potent evolutionary force (Parker 1970, 1998; Charnov 1979; Eberhard 2009; Birkhead 2010) selecting on traits that affect sperm-transfer efficiency and sperm-fertilizing efficiency.

ACKNOWLEDGMENTS

We are grateful to G. Arnqvist, A. Chippindale, and two anonymous reviewers for providing thoughtful comments on an earlier version of the manuscript, and to J. Lehtonen for helpful discussion. We thank M. Poirier for assistance during the paternity analysis, and J. Hottinger, V. Mislin, and U. Stiefel for technical support. LMO was partially supported by the *Freiwillige Akademische Gesellschaft* Basel and by the *Nikolaus und Bertha Burckhardt-Bürgin-Stiftung*. This project was funded by grants from the Swiss National Science Foundation to LS (SNF grants 31003A-127503 and 31003A-143732) and TJ (SNF grant PBBSP3-135985). All authors declare that they have no competing interests.

DATA ARCHIVING

The doi for our data is 10.5061/dryad.p093c.

LITERATURE CITED

Andersen, R. A., J. A. Berges, P. J. Harrison, and M. M. Watanabe. 2005. Recipes for freshwater and seawater media. Pp. 429-538 *in* R. A. Andersen, ed. Algal cultural techniques. Elsevier, Amsterdam.

Andersson, M. 1994. Sexual selection. Princeton Univ. Press, Princeton, NJ.

Anthes, N., P. David, J. R. Auld, J. N. A. Hoffer, P. Jarne, J. M. Koene, H. Kokko, M. C. Lorenzi, B. Pélissié, D. Sprenger, et al. 2010. Bateman gradients in hermaphrodites: an extended approach to quantify sexual selection. Am. Nat. 176:249–263.

Arnold, S. J., and D. Duvall. 1994. Animal mating systems—a synthesis based on selection theory. Am. Nat. 143:317–348.

Arnold, S. J., and M. J. Wade. 1984. On the measurement of natural and sexual selection: theory. Evolution 38:709–719.

Arnqvist, G. 2013. Comment on "Bateman in nature: predation on offspring reduces the potential for sexual selection". Science 340:549.

Arnqvist, G., and L. Rowe. 2005. Sexual conflict after mating. Pp. 92–155 *in* J. R. Krebs and T. Clutton-Brock, eds. Sexual conflict. Princeton Univ. Press, Princeton, NJ.

Bateman, A. J. 1948. Intra-sexual selection in *Drosophila*. Heredity 2:349–368.

Birkhead, T. R. 2010. How stupid not to have thought of that: post-copulatory sexual selection. J. Zool. 281:78–93.

Birkhead, T. R., D. J. Hosken, and S. Pitnick. 2009. Sperm biology: An evolutionary perspective. Elsevier Academic Press, Inc., San Diego, CA.

Bretman, A., M. J. G. Gage, and T. Chapman. 2011. Quick-change artists: male plastic behavioural responses to rivals. Trends Ecol. Evol. 26:467–473.

Chapman, T. 2001. Seminal fluid-mediated fitness traits in *Drosophila*. Heredity 87:511–521.

Charnov, E. L. 1979. Simultaneous hermaphroditism and sexual selection. Proc. Natl. Acad. Sci. USA 76:2480–2484.

Chase, R., and K. C. Blanchard. 2006. The snail's love-dart delivers mucus to increase paternity. Proc. R. Soc. B Biol. Sci. 273:1471–1475.

Chen, P. S., E. Stummzollinger, T. Aigaki, J. Balmer, M. Bienz, and P. Böhlen. 1988. A male accessory-gland peptide that regulates reproductive behaviour of female *Drosophila melanogaster*. Cell 54:291–298.

Collet, J., D. S. Richardson, K. Worley, and T. Pizzari. 2012. Sexual selection and the differential effect of polyandry. Proc. Natl. Acad. Sci. USA 109:8641–8645.

Collet, J. M., R. F. Dean, K. Worley, D. S. Richardson, and T. Pizzari. 2014. The measure and significance of Bateman's principles. Proc. R. Soc. B Biol. Sci. 281:32973.

Darwin, C. 1871. The descent of man, and selection in relation to sex. John Murray, London.

Demircan, T. 2013. Advancing the flatworm *Macrostomum lignano* as a versatile model organism for stem cell research. Hubrecht Institute of the Royal Netherlands Academy of Arts and Sciences, Utrecht.

Devigili, A., J. P. Evans, A. Di Nisio, and A. Pilastro. 2015. Multivariate selection drives concordant patterns of pre- and postcopulatory sexual selection in a livebearing fish. Nat. Commun. 6:8291.

Eberhard, W. G. 1996. Female control: Sexual selection by cryptic female choice. Princeton Univ. Press, Princeton, NJ.

———. 2009. Postcopulatory sexual selection: Darwin's omission and its consequences. Proc. Natl. Acad. Sci. USA 106:10025–10032.

Egger, B., P. Ladurner, K. Nimeth, R. Gschwentner, and R. Rieger. 2006. The regeneration capacity of the flatworm *Macrostomum lignano*—on repeated regeneration, rejuvenation, and the minimal size needed for regeneration. Dev. Genes Evol. 216:565–577.

Fritzsche, K., and G. Arnqvist. 2013. Homage to Bateman: sex roles predict sex differences in sexual selection. Evolution 67:1926–1936.

Gopurenko, D., R. N. Williams, and J. A. DeWoody. 2007. Reproductive and mating success in the small-mouthed salamander (*Ambystoma texanum*) estimated via microsatellite parentage analysis. Evol. Biol. 34:130–139.

Hersch, E. I., and P. C. Phillips. 2004. Power and potential bias in field studies of natural selection. Evolution 58:479–485.

Janicke, T., and L. Schärer. 2009a. Determinants of mating and sperm-transfer success in a simultaneous hermaphrodite. J. Evol. Biol. 22:405–415.

———. 2009b. Sex allocation predicts mating rate in a simultaneous hermaphrodite. Proc. R. Soc. B Biol. Sci. 276:4247–4253.

———. 2010. Sperm competition affects sex allocation but not sperm morphology in a flatworm. Behav. Ecol. Sociobiol. 64:1367–1375.

Janicke, T., P. Sandner, and L. Schärer. 2011. Determinants of female fecundity in a simultaneous hermaphrodite: the role of polyandry and food availability. Evol. Ecol. 25:203–218.

Janicke, T., L. Marie-Orleach, K. De Mulder, E. Berezikov, P. Ladurner, D. B. Vizoso, and L. Schärer. 2013. Sex allocation adjustment to local sperm competition in a simultaneous hermaphrodite. Evolution 67:3233–3242.

Janicke, T., P. David, and E. Chapuis. 2015. Environment-dependent sexual selection: Bateman's parameters under varying levels of food availability. Am. Nat. 185:756–768.

Jennions, M. D., and H. Kokko. 2010. Sexual selection. Pp. 343–364 *in* D. F. Westneat and C. W. Fox, eds. Evolutionary behavioral ecology. Oxford Univ. Press, Oxford, U. K.

Jennions, M. D., H. Kokko, and H. Klug. 2012. The opportunity to be misled in studies of sexual selection. J. Evol. Biol. 25:591–598.

Jones, A. G. 2009. On the opportunity for sexual selection, the Bateman gradient and the maximum intensity of sexual selection. Evolution 63:1673–1684.

Jones, A. G., G. Rosenqvist, A. Berglund, S. J. Arnold, and J. C. Avise. 2000. The Bateman gradient and the cause of sexual selection in a sex-role-reversed pipefish. Proc. R. Soc. B Biol. Sci. 267:677–680.

Kimura, K., K. Shibuya, and S. Chiba. 2013. The mucus of a land snail love-dart suppresses subsequent matings in darted individuals. Anim. Behav. 85:631–635.

Koene, J. M., and H. Schulenburg. 2005. Shooting darts: co-evolution and counter-adaptation in hermaphroditic snails. BMC Evol. Biol. 5:13.

Ladurner, P., L. Schärer, W. Salvenmoser, and R. M. Rieger. 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). J. Zool. Syst. Evol. Res. 43:114–126.

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

Lange, R., K. Reinhardt, N. K. Michiels, and N. Anthes. 2013. Functions, diversity, and evolution of traumatic mating. Biol. Rev. 88:585–601.

Marie-Orleach, L., T. Janicke, and L. Schärer. 2013. Effects of mating status on copulatory and postcopulatory behaviour in a simultaneous hermaphrodite. Anim. Behav. 85:453–461.

Marie-Orleach, L., T. Janicke, D. B. Vizoso, M. Eichmann, and L. Schärer. 2014. Fluorescent sperm in a transparent worm: validation of a GFP marker to study sexual selection. BMC Evol. Biol. 14:148.

Michiels, N. K. 1998. Mating conflicts and sperm competition in simultaneous hermaphrodites. Pp. 219–254 *in* T. R. Birkhead and A. P. Møller, eds. Sperm competition and sexual selection. Academic Press, San Diego, CA.

Michiels, N. K., and L. J. Newman. 1998. Sex and violence in hermaphrodites. Nature 391:647–647.

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

———. 1998. Sperm competition and the evolution of the ejaculates: towards a theory base. Pp. 3–54 *in* T. R. Birkhead and A. P. Møller, eds. Sperm competition and sexual selection. Academic Press, Cambridge, U. K.

Pélissié, B., P. Jarne, and P. David. 2012. Sexual selection without sexual dimorphism: Bateman gradients in a simultaneous hermaphrodite. Evolution 66:66–81.

Pélissié, B., P. Jarne, V. Sarda, and P. David. 2014. Disentangling precopulatory and postcopulatory sexual selection in polyandrous species. Evolution 68:1320–1331.

Peng, J., S. Chen, S. Büsser, H. F. Liu, T. Honegger, and E. Kubli. 2005. Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. Curr. Biol. 15:207–213.

Pischedda, A., and W. R. Rice. 2012. Partitioning sexual selection into its mating success and fertilization success components. Proc. Natl. Acad. Sci. USA 109:2049–2053.

Pizzari, T., and G. A. Parker. 2009. Sperm competition and sperm phenotype. Pp. 207–245 *in* T. R. Birkhead, D. J. Hosken, and S. Pitnick, eds. Sperm biology: An evolutionary perspective. Academic Press, San Diego, CA.

Pizzari, T., D. P. Froman, and T. R. Birkhead. 2002. Pre- and post-insemination episodes of sexual selection in the fowl, *Gallus g. domesticus*. Heredity 88:112–116.

Rose, E., K. A. Paczolt, and A. G. Jones. 2013. The contributions of premating and postmating selection episodes to total selection in sex-role-reversed Gulf pipefish. Am. Nat. 182:410–420.

Schärer, L., and P. Ladurner. 2003. Phenotypically plastic adjustment of sex allocation in a simultaneous hermaphrodite. Proc. R. Soc. Lond. Ser. B Biol. Sci. 270:935–941.

Schärer, L., and I. Pen. 2013. Sex allocation and investment into pre- and post-copulatory traits in simultaneous hermaphrodites: the role of polyandry and local sperm competition. Philos. Trans. R. Soc. B 368:20120052.

Schärer, L., and D. B. Vizoso. 2007. Phenotypic plasticity in sperm production rate: there's more to it than testis size. Evol. Ecol. 21:295–306.

Schärer, L., G. Joss, and P. Sandner. 2004. Mating behaviour of the marine turbellarian *Macrostomum* sp.: these worms suck. Mar. Biol. 145:373–380.

Schärer, L., P. Sandner, and N. K. Michiels. 2005. Trade-off between male and female allocation in the simultaneously hermaphroditic flatworm *Macrostomum sp*. J. Evol. Biol. 18:396–404.

Schärer, L., D. T. J. Littlewood, A. Waeschenbach, W. Yoshida, and D. B. Vizoso. 2011. Mating behavior and the evolution of sperm design. Proc. Natl. Acad. Sci. USA 108:1490–1495.

Schärer, L., T. Janicke, and S. A. Ramm. 2014. Sexual conflict in hermaphrodites. Pp. 265-290 *In* W. R. Rice and S. Gravilets, eds. The genetics and biology of sexual conflict. Cold Spring Harbor Press, New York.

Sekii, K., D. B. Vizoso, G. Kuales, K. De Mulder, P. Ladurner, and L. Schärer. 2013. Phenotypic engineering of sperm-production rate confirms evolutionary predictions of sperm competition theory. Proc. R. Soc. B Biol. Sci. 280:20122711.

Snook, R. R. 2005. Sperm in competition: not playing by the numbers. Trends Biochem. Sci. 20:46–53.

Thornhill, R. 1983. Cryptic female choice and its implications in the scorpionfly *Harpobittacus nigriceps*. Am. Nat. 122:765–788.

van Noordwijk, A. J., and G. De Jong. 1986. Acquisition and allocation of resources: their influence on variation in life-history tactics. Am. Nat. 128:137–142.

Vizoso, D. B., G. Rieger, and L. Schärer. 2010. Goings-on inside a worm: functional hypotheses derived from sexual conflict thinking. Biol. J. Linn. Soc. 99:370–383.

Webster, M. S., S. PruettJones, D. F. Westneat, and S. J. Arnold. 1995. Measuring the effects of pairing success, extra-pair copulations and mate quality on the opportunity for sexual selection. Evolution 49:1147–1157.

Zelditch, M. L., D. L. Swiderski, H. D. Sheets, and W. L. Fink. 2004. Geometric morphometrics for biologists: a primer. Elsevier Academic Press, London.

Associate Editor: A. Chippindale
Handling Editor: M. Servedio

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Supporting Information A. Experimental cultures.
Supporting Information B. Estimating penetrance of the GFP marker.
Supporting Information C. Experimental phases 1 and 2.
Supporting Information D. Geometric morphometrics of the copulatory stylet.
Supporting Information E. Estimation of the variance due to binomial sampling error.
Movie Clip S1. Time-lapse mating movie of five replicate groups.
Movie Clips S2, S3, S4 and S5. Movies of the female sperm-storage organ of two GFP(-) sperm recipients, recorded under both differential interference contrast illumination (Movie clips S2 and S4) and epifluorescence illumination (Movie clips S3 and S5).