JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.13212

Bigger testes increase paternity in a simultaneous hermaphrodite, independently of the sperm competition level

N. VELLNOW*, L. MARIE-ORLEACH†, K. S. ZADESENETS‡ & L. SCHÄRER*

*Zoological Institute, Evolutionary Biology, University of Basel, Basel, Switzerland
†School of Biology, University of St Andrews, Fife, UK
‡Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Keywords:

local sperm competition (LSC);
male fitness gain curve;
mating group size;
paternity success;
sex allocation;
simultaneous hermaphrodite;
sperm competition;
testis size.

Abstract

Hermaphroditic animals face the fundamental evolutionary optimization problem of allocating their resources to their male vs. female reproductive function (e.g. testes and sperm vs. ovaries and eggs), and this optimal sex allocation can be affected by both pre- and post-copulatory sexual selection. For example, local sperm competition (LSC) – the competition between related sperm for the fertilization of a partner's ova – occurs in small mating groups and can favour a female-biased sex allocation, because, under LSC, investment into sperm production is predicted to show diminishing fitness returns. Here, we test whether higher testis investment increases an individual's paternity success under sperm competition, and whether the strength of this effect diminishes when LSC is stronger, as predicted by sex allocation theory. We created two subsets of individuals of the simultaneously hermaphroditic flatworm *Macrostomum lignano* – by sampling worms from either the highest or lowest quartile of the testis investment distribution – and estimated their paternity success in group sizes of either three (strong LSC) or eight individuals (weak LSC). Specifically, using transgenic focal individuals expressing a dominant green-fluorescent protein marker, we showed that worms with high testis investment sired 22% more offspring relative to those with low investment, corroborating previous findings in *M. lignano* and other species. However, the strength of this effect was not significantly modulated by the experienced group size, contrasting theoretical expectations of more strongly diminishing fitness returns under strong LSC. We discuss the possible implications for the evolutionary maintenance of hermaphroditism in *M. lignano*.

Introduction

General aspects of sex allocation

All sexually reproducing organisms face the evolutionary optimization problem of how much of their limited resources they should invest into male vs. female reproduction, that is, the problem of sex allocation (Charnov, 1982). Consider, for example, producing the optimal offspring sex ratio in gonochoristic (separate-sexed) animals. As during sexual reproduction both parents contribute the same amount of nuclear genetic material to the zygote, the total fitness of males and females – on the population level – has to be equal (Düsing, 1884; Fisher, 1930; Houston & McNamara, 2006; Queller, 2006). Therefore, negative frequency-dependent selection will tend to select for the production of more daughters when females are rare in the population and of more sons when males are rare. Given certain, arguably strong, simplifying assumptions, including random mating (i.e. any male gamete in the population has the same probability of fusing with any female gamete) and large population size (i.e. mates and competitors are always unrelated) (Hamilton, 1967; Schärer & Ramm, 2016), this is expected to lead to equal investment into sons and daughters over evolutionary time (Düsing, 1884; Fisher, 1930; Charnov, 1982).

Although the principle of equal investment resulting from this so-called Fisher condition was initially

Correspondence: Nikolas Vellnow, Zoological Institute, Evolutionary Biology, University of Basel, Vesalgasse 1, CH-4051 Basel, Switzerland.
Tel.: +41 61 267 03 75; fax: +41 61 207 03 62;
e-mail: nikolas.vellnow@unibas.ch

formulated with gonochoristic animals in mind, it actually applies more generally. Under the same simplifying assumptions, an equal investment into male and female function could also be expected in (nonselfing) simultaneous hermaphrodites, because also here every zygote gets half of its nuclear genetic material from each parent (though, as we outline below, these assumptions are likely often broken in hermaphrodites). In these organisms, it is not the amount of investment into daughters and sons that is selected to be equal. Instead, the balance between investment into the male function (e.g. sperm, male-specific tissues and male-specific behaviours) on the one hand, and investment into the female function (e.g. eggs, female-specific tissues and female-specific behaviours) on the other hand, is the target of selection (Charnov, 1982, pp. 7–9; Schärer, 2009).

In spite of this general tendency towards equal investment, there are many conditions under which equal investment into the male and female function is not the evolutionary stable strategy. For instance, Hamilton (1967) drew attention to local mate competition (i.e. competition between related individuals for access to mates) in gonochoristic animals, in which – in the extreme case – the sons of a single female compete for the fertilization of their own sisters. Such local mate competition violates the random mating and large population size assumptions and selects for increased investment into daughters, because investment into more than just a few sons will not increase the number of grandchildren the mother produces. Investment into additional sons is therefore wasteful, favouring the production of more daughters and fewer sons (Charnov, 1982; West, 2009). Analogously, nonrandom mating can also affect the optimal sex allocation in simultaneous hermaphrodites, which we explore in the following section.

Sexual selection influences sex allocation

Although in anisogamous species, both parents contribute an equal amount of nuclear genetic material to the zygote, the mother contributes much more resources via her egg than the father via his sperm. In the absence of substantial post-zygotic paternal investment, this is expected to lead to 'classical' sex roles (Darwin, 1871; Bateman, 1948; Dewsbury, 2005; Parker & Birkhead, 2013; Lehtonen *et al.*, 2016), with males competing for access to females and the fertilization of their eggs, whereas females may choose which males to give access to their unfertilized eggs. In copulating organisms, sexual selection can be subdivided into precopulatory and post-copulatory episodes. During the precopulatory episode, males will tend to compete among each other for matings and females will tend to choose with whom to mate. When a female has mated with two or more males, they will often continue to compete for the fertilization of her eggs via their ejaculates – termed sperm competition (Parker, 1970) – and the female may bias which male's sperm to fertilize her eggs with – termed cryptic female choice (Thornhill, 1983). These episodes of pre- and post-copulatory sexual selection take place both in gonochorists and in simultaneous hermaphrodites, whereas in the latter, individuals may simultaneously compete and chose during the pre- and post-copulatory episodes in their respective roles as sperm donors and sperm recipients (Charnov, 1979).

In his influential work on sex allocation, Charnov proposed that the evolution of sex allocation is influenced by the shape of so-called fitness gain curves for the investment into the respective sex functions (Charnov, 1979, 1982), which in turn can be influenced by sexual selection (Charnov, 1979, 1980). One such case is local sperm competition (LSC), that is the competition between related sperm for the fertilization of ova of the mating partner (Greeff *et al.*, 2001; Schärer & Wedekind, 2001), which was named in analogy to local mate competition that we discussed above (Schärer, 2009; Schärer & Pen, 2013). Here, investment into the male function in the form of sperm production confers diminishing fitness gains when related sperm compete for fertilizations. A hermaphrodite that experiences LSC may therefore achieve higher fitness returns when it re-allocates resources to its own female function, in which the fitness returns are often expected to be more linear (Schärer, 2009). Alternatively, it may re-allocate those resources to other forms of male investment, such as seminal fluids, love darts or male-specific behaviours like mate searching or courtship (Michiels *et al.*, 2009; Schärer, 2009; Schärer & Pen, 2013), provided that these show higher returns than investment in its own female function.

When the sperm of every given sperm donor in a population of hermaphroditic animals is not equally likely to be represented in every recipient's sperm receiving organ, then the random mating assumption is broken. This occurs, for instance, when individuals only mate and compete within a small subset of the population (i.e. if they compete in small mating groups), leading to strong LSC, because (related) sperm from the same sperm donor will then compete for fertilizations. In the extreme case of a mating group of just two individuals, all sperm in a mating partner's female reproductive tract will be from the same sperm donor. Consequently, the production of more sperm than necessary to assure the fertilization of the partner's eggs can be considered wasteful. Instead, resources could be more profitably invested into an individual's own female function. Therefore, a decrease in mating group size is predicted to lead to stronger LSC and hence to a more female-biased sex allocation, provided that other forms of male investment are not available or also show diminishing returns (Charnov, 1980; Schärer, 2009; Schärer & Pen, 2013). If, in contrast, the sperm from a

given individual compete with unrelated sperm from other sperm competitors, producing additional sperm may indeed pay off and result in higher paternity success (Parker *et al.*, 1990). In this case, it may not be beneficial to re-allocate resources towards the female function and sex allocation is predicted to be less female-biased (Charnov, 1980; Schärer, 2009; Schärer & Pen, 2013).

Besides a small mating group size, other processes can also lead to an increase in LSC and favour a more female-biased sex allocation. In particular, processes leading to the different mating partners in a mating group having unequal chances to fertilize the eggs are likely to make the male fitness curve more saturating (Schärer, 2009; Schärer & Pen, 2013). For instance, sperm displacement can lead to higher LSC than expected by the number of sperm donors alone (Charnov, 1996). Similar effects can result from cryptic female choice, if a fixed proportion of sperm from a disfavoured sperm donor is removed and/or rejected by the sperm recipient (van Velzen *et al.*, 2009). And finally, random paternity skews resulting from stochastic effects can also lead to stronger LSC and therefore favour a more female-biased sex allocation (Greeff *et al.*, 2001; Schärer, 2009; Schärer & Pen, 2013).

Empirical evidence for LSC

Phenotypically plastic sex allocation adjustments that correspond to the predictions of the mating group size model have been observed in several hermaphroditic animals and corroborate the LSC perspective (reviewed in Schärer, 2009). For example, the trematode *Echinostoma caproni* increases its male allocation, measured as a composite of the size of the testes and the cirrus sac, when spending its metacercarial stage in larger groups compared to when it occurs in pairs or in isolation (Trouvé *et al.*, 1999). Another parasite, the intestinal trematode *Gyliauchen volubilis*, showed more female-biased sex allocation with decreasing mating group size in field-caught rabbitfish *Siganus rivulatus* (Al-Jahdali, 2012). And finally, the free-living simultaneously hermaphroditic flatworm *Macrostomum lignano* changes its phenotype according to different experimentally manipulated mating group sizes. For example, it increases testis size (Janicke *et al.*, 2013), testicular stem cell proliferation activity (Schärer *et al.*, 2004b) and sperm production rate (Schärer & Vizoso, 2007), presumably in part by increasing the speed of spermatogenesis (Giannakara *et al.*, 2016), when growing up in larger mating groups where sperm competition between unrelated sperm is stronger (and LSC thus weaker).

There is, however, only very limited empirical evidence that fitness gains for the male function indeed show more strongly diminishing returns under strong LSC. To our knowledge, the only study that has unequivocally documented a relationship between the level of LSC and the shape of the male fitness gain function was performed in the spermcast mating colonial ascidian *Botryllus schlosseri*. Here, the male fitness gain curve – measured as the relationship between the cross-sectional testis area of a focal colony and the percentage of available eggs fertilized by that colony in nearby partner colonies – did change according to the predictions of sex allocation theory. Namely, colonies placed under strong LSC had a gain curve with more strongly diminishing returns compared to those placed under weaker LSC (Yund, 1998). The scant empirical support for diminishing male fitness gains is problematic, because it is a key component of the theoretical foundation to explain the evolution of hermaphroditism and sex allocation (Charnov, 1979, 1980; Schärer, 2009; Schärer & Pen, 2013) and currently lacks empirical support in copulating hermaphrodites.

Objective

In this study, we use the copulating simultaneously hermaphroditic flatworm *M. lignano* to test (i) whether greater testis investment actually increases paternity success under sperm competition, and (ii) whether, as predicted by the LSC perspective, the gains in paternity success for greater investment into testes are more substantial in large compared to small mating groups (i.e. in a situation with presumed low or high LSC, respectively).

Materials and methods

Study organism and cultures

The experiment was conducted with *Macrostomum lignano* (Macrostomorpha, Platyhelminthes), a free-living flatworm that lives between sand grains in the intertidal zone of the Northern Adriatic sea and the Eastern Mediterranean basin (Ladurner *et al.*, 2005). It can be cultured in the laboratory (at 20 °C, 14 : 10 h light : dark and 60% humidity) in glass Petri dishes filled with artificial sea water or nutrient-enriched f/2 algal culture medium (Andersen *et al.*, 2005) and with the diatom *Nitzschia curvilineata* as the sole food source. It is small (adult length ~1.5 mm) and has a generation time of ~18 days, with eggs hatching ~5 days after laying and individuals reaching maturity in both sex functions ~13 days after hatching (Schärer & Ladurner, 2003). *Macrostomum lignano* is an obligatorily outcrossing simultaneous hermaphrodite with frequent and reciprocal copulation (Schärer *et al.*, 2004a) and, because of its highly transparent body, detailed measurements of internal reproductive structures are possible and noninvasive measures of testis and body size (among others) can be obtained (Schärer & Ladurner, 2003; Marie-Orleach *et al.*, 2016).

The worms used as mating partners in this experiment came from LS1, an outbred wild-type culture

(Marie-Orleach *et al.*, 2013). Focal worms came from the outbred transgenic BAS1 culture, which carries a green-fluorescent protein (GFP) marker and expresses GFP in all cell types (Marie-Orleach *et al.*, 2016). As some lines and cultures of *M. lignano* exhibit a karyotype polymorphism (Zadesenets *et al.*, 2016, 2017), we established a new BAS1 culture that exclusively included individuals whose karyotype was 2n=8 and that were homozygous for the GFP allele, increasing stable inheritance of this marker. More specifically, we performed metaphase chromosome preparation to count the number of chromosomes for 277 worms from the original BAS1 culture (Zadesenets *et al.*, 2016). Subsequently, we paired some of them with worms from the LS1 culture to assess the penetrance of the GFP allele (i.e. the proportion of GFP expressing offspring produced). To found the new BAS1 culture, we used only those 76 individuals (i) that showed the 'normal' 2n=8 karyotype (Zadesenets *et al.*, 2016), (ii) for which we could phenotype a progeny array of ≥ 17 offspring (mean: 48.7, range: 17–92) and (iii) for which all offspring expressed the GFP allele, thus indicating that their BAS1 parent was homozygous for the GFP allele. This culture is now being kept in a meta-population structure and at a total population size of 1200 individuals to maintain its genetic diversity.

As the GFP allele is dominant and as we expect it to be fixed within the BAS1 culture, offspring from BAS1 × LS1 crosses will always show GFP expression and can thereby be distinguished from the offspring of LS1 × LS1 crosses, which will lack GFP expression (Marie-Orleach *et al.*, 2014). Both LS1 and BAS1 are maintained at the Zoological Institute in Basel.

Experimental design

The rationale of the experimental design was to divide focal worms into two subsets with either a large or small testis size (while excluding individuals with intermediate trait values) and to then measure their resulting paternity success in two different mating group sizes, which each included a GFP-positive BAS1 individual as the focal worm and either two (hereafter called 'triplets') or seven (hereafter called 'octets') GFP-negative LS1 individuals as the partners (i.e. a full-factorial 2 × 2 design). Note that here 'mating group size' refers to the number of worms able to interact (sometimes also called the social group size); in this system, a social group size of three vs. eight is known to lead to a substantial difference in the mating group size (Janicke & Schärer, 2009a; Janicke *et al.*, 2013).

This experimental design permitted us to investigate the effect of testis size on paternity success and whether the magnitude of this effect was modulated by the mating group size in which a focal resided. Following the LSC rationale, we predicted that the same increase in testis size should confer a higher relative increase in paternity success in octets than in triplets, as we outline in more detail in the 'Statistics' section below. But before we do so, it is helpful to generate some theoretical expectations to which we could compare the observed paternity successes, which we do in the following section.

Theoretical expectations

We calculated the expected paternity successes of focal worms in the respective treatments, making the following, highly simplifying assumptions: (i) every individual in the mating group mates with every other individual, (ii) paternity success is the outcome of a fair raffle sperm competition, weighted by an individual's sperm production, (iii) sperm production is proportional to testis investment (measured as testis size), and (iv) focal worms from the BAS1 culture have, on average, a similar testis investment as the competitors from the LS1 culture.

We set the sperm production of a BAS1 focal with low testis investment (L) to, a, and with high testis investment (H) to, $2a$, because their testis size turned out to be approximately twice as high (Section 'Results' in Table 1). The sperm production of a LS1 competitor was set to, $1.5a$, because worms of this culture are expected to have, on average, an intermediate sperm production (as the BAS1 culture is derived from the LS1 culture; Marie-Orleach *et al.*, 2016). The expected paternity successes for the L and H focal worms in the triplets (3) and octets (8) are then as follows:

$$P_{L3} = \frac{a}{a+b} = \frac{a}{a+1.5a} = \frac{a}{2.5a} = 0.4$$

$$P_{H3} = \frac{2a}{2a+b} = \frac{2a}{2a+1.5a} = \frac{2a}{3.5a} \approx 0.5714$$

$$P_{L8} = \frac{a}{a+6b} = \frac{a}{10a} = 0.1$$

$$P_{H8} = \frac{2a}{2a+6b} = \frac{2a}{11a} \approx 0.1818$$

Thus, the paternity success of worms with high vs. low testis investment, respectively, is expected to be higher by 42.9% in the triplets (i.e. 0.5714 vs. 0.4) and by 81.8% in the octets (i.e. 0.1818 vs. 0.1). Note that the higher relative increase in the larger mating group size is consistent with the LSC scenario. Furthermore, it is important to note that we do not necessarily expect the outcomes to exactly match this simplified scenario, but these expectations serve as a useful comparison to our observed results and allow us to explore which of the assumptions may not have been met.

Experimental procedures

For logistic reasons, we divided the experiment into three blocks, each processed 1 day apart, but otherwise

Table 1 Shown are medians (and interquartile ranges) for measured morphological traits ($\times 10^3$ μm^2) in the different experimental groups.

	Triplets		Octets	
	Low testis investment	High testis investment	Low testis investment	High testis investment
Testis size	15.7 (12.6–22.3)	31.8 (19.6–40.2)	15.0 (10.4–18.5)	26.9 (19.4–40.2)
Ovary size	9.8 (7.1–12.7)	10.0 (8.2–12.1)	9.5 (7.6–12.9)	9.1 (6.8–11.4)
Seminal vesicle size	13.0 (9.1–20.5)	14.1 (10.7–18.9)	11.1 (8.0–14.9)	11.4 (8.2–19.3)
Body size	464.5 (344.1–496.6)	394.6 (336.4–465.9)	422.6 (323.3–491.7)	373.1 (315.6–483.7)

Note the approximately two-fold higher testis size in the subset with high testis investment compared to the one with low testis investment.

treated the worms in the same way. In the following, we explain the experimental procedures for one block only.

We kept all worms used here either in (i) glass Petri dishes in groups of one hundred during the growth phase or in (ii) triplets or octets in wells of 24-well tissue culture plates during isolation and mating trials (TPP AG, Switzerland), in 20 mL or 1.5 mL of f/2 medium (Andersen *et al.*, 2005), respectively. Except for a 24-h period of food deprivation to which the focal worms were submitted to facilitate the morphological measurements, all worms received *ad libitum* diatom algae during the whole experiment. Partner worms were raised under the same conditions as the focal individuals, except that they were not measured (and therefore also never food deprived).

On day 0, we put the parents (F0) of the focal worms into three Petri dishes (100 adult worms per dish) allowing them to lay eggs and then removed them on day 2, so that all focal worms (F1) used in this experiment were of similar age. On day 19, we transferred 100 focal worms to a new Petri dish, and on day 28, we transferred them into new Petri dishes *without* algae (in order to allow them to regurgitate the consumed algae in their gut, thus facilitating morphometry). On the following day, we isolated focal worms in 24-well plates, took their morphological measurements (block 1: $n = 81$; block 2: $n = 72$; block 3: $n = 85$) and then put them back into isolation in 24-well plates with algae until the next day.

To take morphological measurements, we performed a noninvasive squeeze preparation on the focal worms (Schärer & Ladurner, 2003; Vizoso & Schärer, 2007). Briefly, we anesthetized worms with a 2:1 mixture of 7.14% $MgCl_2$ and f/2 medium and squeezed them dorsoventrally between a glass slide and a haemocytometer cover glass separated by a 35-μm plastic spacer. Then, we captured digital micrographs at 40–400× magnification with a digital video camera (DFK 41BF02, The Imaging Source, Bremen, Germany) attached to a DM 2500 microscope (Leica Microsystems, Heerbrugg, Switzerland). We used BTV Pro 6.0b7 (http://www.bensoftware.com/) to acquire the images.

During the evening of the same day, we measured body and testis size (and later also ovary and seminal vesicle size) of all focal worms using ImageJ 1.47v (http://imagej.nih.gov/ij). For testis size and ovary size, we used the sum of the areas of both testes and ovaries, respectively (note that for 23 and 26 of the 103 focals, respectively, two testes and ovaries could be seen in the pictures, but only one could be properly imaged, in which case that gonad was measured twice to estimate total gonad size). Then, we performed a linear regression of testis on body size to determine residual testis size, which we used as a measure of relative testis investment that is uncorrelated to body size. Subsequently, we retained only worms belonging to the lowest and highest quartiles of the distribution of residual testis size for the further experiment (called the *low testis investment* and *high testis investment* subsets, respectively; see Figs 1 and 2).

The rationale for controlling for body size in this way is that bigger worms can be expected to have a higher resource budget available and could therefore have higher paternity success for that reason alone. As we were interested in the effect of testis size *per se*, we used residual testis size to assign each worm to their subset, which, as we show in the Results, nevertheless resulted in an approximately two-fold difference in absolute testis size between the subsets.

On day 30, we assigned GFP-positive focal worms of the low and high testis investment subsets to one of the two mating group size treatments (Fig. 1). For this, we grouped focal worms in wells with either two (triplets) or seven (octets) GFP-negative partner worms taken directly from a Petri dish with 100 worms, thus forming the first mating group. After 24 h, we then transferred the focal worm to a new mating group (with the same number of partner worms as in the first one) and isolated the partners from the first mating group in 24-well plates for 7 days to allow them to lay eggs and produce offspring. In total, we repeated this four times, so that every focal worm passed through four consecutive mating groups of the same size, thereby increasing the number of offspring based on which the paternity success of the focal could be estimated (Fig. 1). Note that, although *M. lignano* can phenotypically adjust its testis size and sperm production rate according to the mating group size, it was previously shown that it needs > 4 days to do so (Brauer

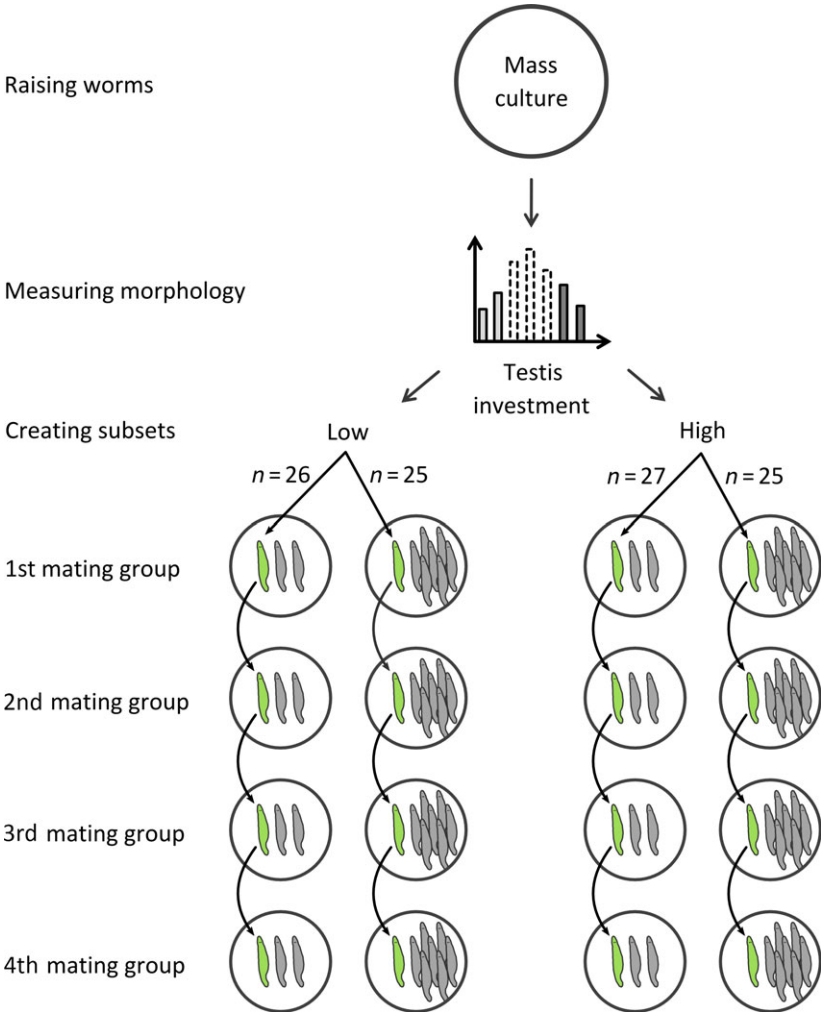

Fig. 1 Schematic illustration of the experimental design. Note how each GFP-positive focal worm (green) had to compete with wild-type worms (grey) in four consecutive mating groups. The sample sizes (n) refer to the final number of focal worms in the respective treatment combinations.

et al., 2007; Schärer & Vizoso, 2007). The isolated mating partners were removed from the wells after 7 and 10 days later the by now hatched offspring were genotyped by checking their GFP status under a MZ10 F stereo microscope with epifluorescence illumination (LEICA Microsystems).

Statistics

We first tested whether the assignment of worms to the different treatment groups resulted in the intended distributions of morphological trait values, by testing for effects of the subset, the mating group size and their interaction, as well as the block on the respective morphological traits using linear models in R (version 3.4.0; R Development Core Team, 2016; also used for all following statistical analyses). We log-transformed the values for the morphological traits to fulfil the normality assumption for residuals.

To assess the effect of the subset, the mating group size and their interaction (but see below) on paternity success, and to statistically control for block effects and account for the detected overdispersion, we fitted a quasi-binomial generalized linear model (GLM), in which the variance is given by the product of the mean and φ, the dispersion parameter. For this, we used the function 'glm' with family 'quasi-binomial' and a logit-link. Paternity was estimated as the number of offspring that expressed GFP among the total number of offspring over all partners from the four mating groups.

For all fitted models, the factors subset and mating group size were centred by encoding their levels as either −0.5 or 0.5 to facilitate hypothesis testing in the presence of interaction terms (Schielzeth, 2010). As the levels of both subset and mating group size were chosen as extreme values that presumably occur at both ends of the distribution in worm populations, it is preferable to estimate the interaction term at the intermediate levels (Schielzeth, 2010). We tested the significance of effects by removing single effects from the model and comparing the reduced to the complete model using an F-test. Model assumptions for all

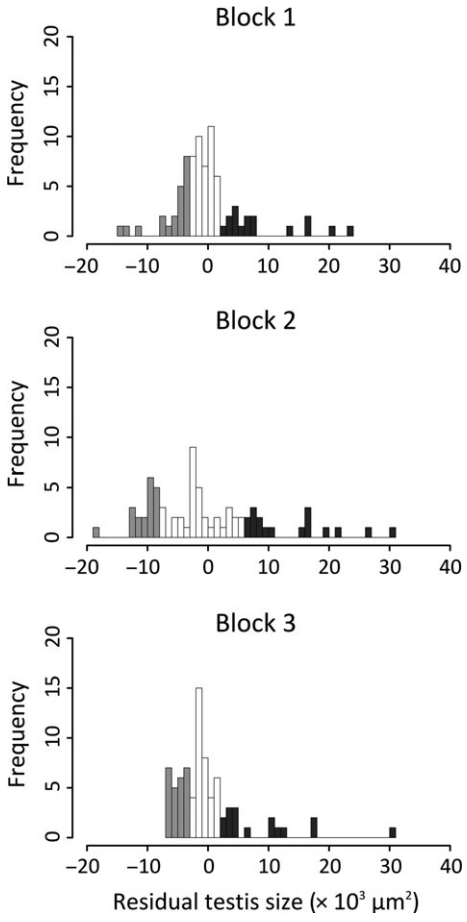

Fig. 2 Distributions of the residual testis size according to which we assigned the worms into subsets with low (light grey) or high (dark grey) testis investment for each of the three experimental blocks.

models were assessed by visually inspecting residuals vs. predicted values plots and normal quantile–quantile plots (Faraway, 2016).

Although the quasi-binomial GLM permits us to test for the effects of the subset and the mating group size, the biological interpretation of the interaction term is less clear in the light of our theoretically derived hypotheses (cf. 'Theoretical expectations' section). Specifically, we wanted to test whether the increase in paternity success in (i) the *triplets* and (ii) the *octets* was significantly different from the theoretical expectations, and whether (iii) the *relative* (rather than absolute) increase in the octets was significantly higher compared to that in the triplets, as predicted by the LSC perspective.

To test these hypotheses, we implemented permutation tests in R, by sampling n paternity success values for *each* mating group size treatment from our empirical data without replacement, and randomly assigned them to either the low or high subset (where n is the sample size for the mating group size treatments in our experiment). To test hypotheses (i) and (ii), we calculated, as our test statistic, the increase in paternity success for triplets (i.e. $\text{mean}_{H3}-\text{mean}_{L3}$) and octets (i.e. $\text{mean}_{H8}-\text{mean}_{L8}$), respectively. And to test hypothesis (iii), we first calculated the relative increase in paternity success due to larger testes in triplets $[(\text{mean}_{H3}-\text{mean}_{L3})/\text{mean}_{L3}\cdot 100]$ and in octets $[(\text{mean}_{H8}-\text{mean}_{L8})/\text{mean}_{L8}\cdot 100]$, and then the difference between them, as our test statistic. We repeated this sampling 10 000 times to generate null distributions of these test statistics. Then, we estimated the *P*-value as the proportion of those permutations where, for hypotheses (i) and (ii), the test statistic was lower (lower limit in triplets: 0.075; lower limit in octets: 0.023) or higher (upper limit in triplets: 0.268; lower limit in octets: 0.140) than expected under the theoretical calculations and where, for hypothesis (iii), the absolute value of the test statistic was greater than or equal to the absolute observed test statistic (note that using absolute values yields a two-sided test). Additionally, we used 10000 bootstrap iterations with replacement to estimate the 95% confidence intervals (as the 95 percentile confidence interval) for the relative increase in the respective mating groups.

To test whether the paternity success of a focal individual was repeatable across its four consecutive mating groups, we estimated repeatability with the R package 'rptR' (Nakagawa & Schielzeth, 2010; Schielzeth & Nakagawa, 2013). For this, we used the function 'rpt.remlLMM' to calculate repeatabilities from a linear mixed-effects model fitted with restricted maximum likelihood. Because the variance for binomial proportions is a quadratic function of the mean and this violates assumptions for linear models, the logit-transformation, which removes this mean-variance relationship, is often recommended for analysing binomial proportions (Warton & Hui, 2010; Engqvist, 2013). Therefore, we logit-transformed paternity values according to the formula $\log[(p + 0.01)/((1 - p) + 0.01)]$, where p is the paternity success value in each mating group (Warton & Hui, 2010). We added the constant of 0.01 to paternity values of 0 and −0.01 to paternity values of 1 to permit their inclusion, in the analysis, for which the logit would otherwise not be defined. Here, repeatability is the proportion of the total variance in paternity success explained by interindividual differences between paternity values in the four consecutive mating groups.

Finally, we tested whether the repeatability estimates between the two mating group size treatments differed statistically with a two-tailed permutation test. For this, we (i) randomly reassigned replicates to the two mating group sizes 10 000 times and calculated the differences between the repeatability estimates (repeatability for triplets minus repeatability for octets) for each

permutation and ii) estimated the *P*-value as the proportion of those permutations where the absolute difference was greater than or equal to the absolute observed difference.

Results

Morphology and assignment over treatments

As intended, worms from the high testis investment subset had testes that were almost twice as large in absolute terms compared to the worms from the low testis investment subset (high testis investment, median: 29 000 μm^2; low testis investment, median: 15 300 μm^2; Tables 1 and 2a), but there was no significant difference in testis size between the mating group size treatments, nor was there a significant interaction between the testis investment subset and mating group size treatments. The assignment of individuals over the treatment groups was also successful in so far as they did not differ in either body or ovary size (all $P > 0.3$; Table 2a). The only exception was seminal vesicle size, in that worms allocated to the octets had significantly smaller seminal vesicles (with the difference being approximately 0.17 ± 0.07 standard deviations), whereas worms with low and high testis investment did not differ in seminal vesicle size (cf. Tables 1 and 2a). We do not think, however, that this initial difference in seminal vesicle size should have had a strong effect on the amount of sperm transferred, because the focal worms were exposed to many mating partners over the four consecutive 24-h time periods. Therefore, not the initial amount of sperm in the seminal vesicle, but rather the sperm production rate during these periods will likely have determined the amount of sperm available for transfer (Schärer & Vizoso, 2007).

Testis investment and mating group size effects on paternity

Overall, worms with higher testis investment had a 22.3% (95% CI: −12.9% to 74.0%) higher mean paternity success relative to worms with low testis investment, but although worms in smaller groups (which of course had fewer competitors) had higher paternity success, the interaction between subset and mating

Table 2 Treatment effects of the (a) linear models for morphological traits and (b) quasi-binomial generalized linear model for paternity success.

(a) Response	Effect	$\beta \pm$ SE	d.f.	AIC	*F*-value	*P*-value
Testis size	None	–	–	−239.4	–	–
	Testis investment	0.65 ± 0.06	1	−159.0	118.88	**<0.0001**
	Mating group size	−0.05 ± 0.06	1	−240.5	0.82	0.37
	Testis investment × mating group size	0.01 ± 0.12	1	−241.4	0.01	0.94
	Block	–	2	−176.6	44.21	**<0.0001**
Ovary size	None	–	–	−195.1	–	–
	Testis investment	0.08 ± 0.07	1	−196.0	1.07	0.303
	Mating group size	−0.07 ± 0.07	1	−196.1	0.96	0.331
	Testis investment × mating group size	−0.06 ± 0.15	1	−197.0	0.14	0.706
	Block	–	2	−168.1	17.04	**<0.0001**
Sem. vesicle size	None	–	–	−194.4	–	–
	Testis investment	0.04 ± 0.07	1	−196.0	0.36	0.549
	Mating group size	−0.17 ± 0.07	1	−191.2	5.03	**0.027**
	Testis investment × mating group size	0.02 ± 0.15	1	−196.4	0.02	0.899
	Block	–	2	−150.3	28.89	**<0.0001**
Body size	None	–	–	−308.8	–	–
	Testis investment	−0.04 ± 0.04	1	−309.6	1.08	0.302
	Mating group size	−0.04 ± 0.04	1	−310.0	0.72	0.398
	Testis investment × mating group size	0.01 ± 0.09	1	−310.7	0.03	0.869
	Block	–	2	−259.3	33.01	**<0.0001**

(b) Response	Effect	$\beta \pm$ SE	d.f.	Deviance	*F*-value	*P*-value
Paternity success	None	–	–	616.1	–	–
	Testis investment	0.38 ± 0.16	1	648.2	5.06	**0.027**
	Mating group size	−2.17 ± 0.16	1	1653.4	163.30	**<0.0001**
	Testis investment × mating group size	−0.10 ± 0.32	1	616.5	0.08	0.781
	Block	–	2	659.7	3.43	**0.036**

F-tests were performed by removing single effects and comparing the full with the reduced model and significant *P*-values are written in bold face. Note that the parameter estimate β is on the log scale for the morphological traits and on the logit scale for paternity success.

group size was not significant (Fig. 3, Table 2b; but see the 'Statistics' section for caveats in interpreting the interaction term). Worms in triplets had a relative increase in paternity success of 18.7% (95% CI: −6.6% to 53.4%) when doubling their testis investment, a substantial and significant deviation from the theoretical expectation of 42.9% (two-tailed permutation test: $P = 0.013$; Fig. 3). In octets, the worms increased their paternity success by 33.1% (95% CI: −15.4% to 107.6%) with increased testis investment, which was also substantially lower than the expected 81.8%, although the difference between the observed and the expected increase was not statistically significant (two-tailed permutation test: $P = 0.132$). Although the relative increase in paternity was more pronounced in octets, as predicted by the LSC perspective, it was not significantly different between triplets and octets (two-tailed permutation test: $P = 0.453$; Fig. 3). In particular, worms with high testis investment tended to have lower paternity success than theoretically expected (Fig. 3).

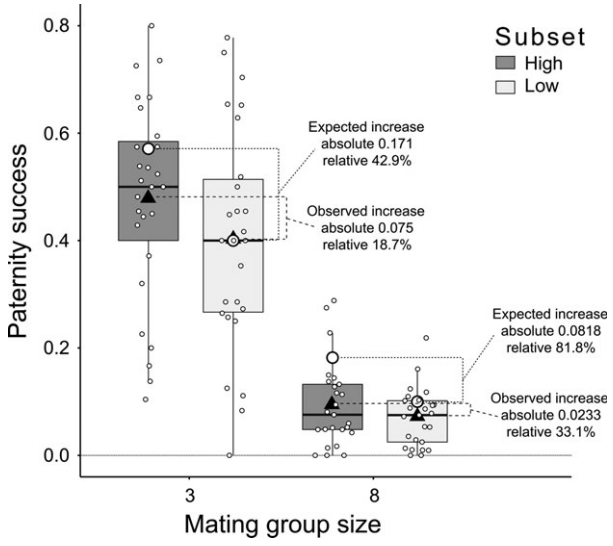

Fig. 3 Paternity success achieved across all four subsequent mating groups by focal worm of low testis investment (light grey) and high testis investment (dark grey) tested in two mating group sizes, either triplets or octets. Small circles represent the individual measurements (and are jittered along the *x*-axis for better visibility). The box plots show medians, and the 25th and 75th percentiles, respectively, and whiskers extend to 1.5 times the interquartile range. The large white circles represent theoretical expectations (see the 'Theoretical expectations' section), and the black triangles represent treatment means (see the 'Methods' section). Note that, although the focal worms with larger testes are expected to increase their paternity success more in triplets than in octets in absolute terms (0.171 vs. 0.0818), focal worms are expected to benefit more in octets in relative terms (42.9% vs. 81.8%).

Repeatability of paternity success

The repeatability of paternity success across the four mating groups was significant and moderate in octets (R = 0.360, CI = 0.197–0.499, $P = 0.004$) and substantially lower and not quite statistically significant in triplets (R = 0.093, CI = 0.000–0.227, $P = 0.061$). The difference in repeatability between the two mating group sizes was significant (two-tailed permutation test: $P = 0.040$).

Discussion

In the present experiment, we found that worms with higher testis investment sired a higher proportion of offspring in an environment where sperm competitors were present. We found the relative gain in paternity success resulting from higher testis investment to be lower than our theoretical expectations, and there was no strong evidence that the observed effect of testis investment was more pronounced under low LSC (i.e. in octets as opposed to triplets), a result that does not match what we predicted based on the LSC perspective. Moreover, the within-individual repeatability for paternity success was significantly higher in octets. In the following, we discuss these points in turn and explore implications for the evolutionary maintenance of hermaphroditism in *M. lignano*.

Testis investment effect

The positive effect of testis investment on paternity success found here confirms previous correlative findings that suggested an increased male reproductive success for individuals with larger testes in *M. lignano* (Marie-Orleach *et al.*, 2016), although worms tended not to benefit as much from higher testis investment as predicted under our simplifying theoretical assumptions. Moreover, this testis investment effect is in line with empirical results from other species and confirms predictions from sperm competition theory, as we discuss in the following.

In *M. lignano*, previous studies have shown that worms with larger testes also have more active testes, produce more sperm per unit time and also produce sperm more quickly (Schärer *et al.*, 2004b; Schärer & Vizoso, 2007; Giannakara *et al.*, 2016). This leads to an increase in sperm transfer success (Janicke & Schärer, 2009a; Marie-Orleach *et al.*, 2016), which in turn increases paternity success (Marie-Orleach *et al.*, 2016). Our results confirm this mechanism, because we can attribute the increased paternity success of the high subset to their larger testes (while they did not differ in body size). Interestingly, a recent and well-replicated experimental evolution study in *M. lignano* showed no evolutionary response for testis size in selection lines that were either kept in monogamous pairs or

polygamous octets for 20 generations (Janicke *et al.* 2016). But this lack of response to selection might potentially be linked to the pronounced phenotypic plasticity that *M. lignano* exhibits in many reproductive traits, including testis size (Schärer *et al.*, 2004b; Schärer & Vizoso, 2007; Janicke *et al.*, 2013; Giannakara *et al.*, 2016), which may have shielded the available genetic variation from selection (Price *et al.*, 2003; Kopp & Matuszewski, 2014; Ghalambor *et al.*, 2015). Overall, however, there is strong evidence that testis size influences male reproductive success and is therefore under selection in this species.

The observed effect of testis investment in our study is unlikely due to differences in the overall resource budget between the subsets, because individuals with low and high testis investment were deliberately chosen to have a similar body size. Similarly, an effect of ovary size is unlikely, as it did not differ between the testis investment subsets either. The significantly lower seminal vesicle size of worms from octets compared to triplets seems to be an unfortunate sampling effect, because this effect was present only in the first block of the experiment ($F_{1,33} = 5.017$, $P = 0.032$), but not in the other two blocks (second block: $F_{1,27} = 1.444$, $P = 0.240$; third block: $F_{1,31} = 0.100$, $P = 0.754$). Having said that, we do not think that it strongly influenced the results, because (i) as already mentioned above, we expect the initial fill grade of the seminal vesicle to be less important than the sperm production rate during the four consecutive 24 h mating trials and (ii) the worms with low and high testis investment did not differ in their seminal vesicle size, nor was there a subset by mating group size interaction effect on seminal vesicle size. This unwanted seminal vesicle size effect could, however, have somewhat reduced the difference between the octet and triplet treatment groups.

Given that our experiment did not directly manipulate testis size, any unmeasured traits that correlate with residual testis size and themselves affect paternity success could, at least in theory, have been responsible for, or contributed to, the observed relationship between residual testis size and paternity success. Sensible candidates for such traits might be seminal fluid proteins, which have been shown to interact with sperm and the female reproductive system and thereby influence male reproductive success in other species (Chapman, 2001; Arnqvist & Rowe, 2005; Wigby *et al.*, 2009). However, we currently have no evidence for seminal fluid effects in *M. lignano*, nor do we know whether there is a correlation between seminal fluid production and residual testis size. Mating rate may also correlate with testis size, as worms raised in octets with consequently larger testes mated more often and for longer than worms raised in pairs with smaller testes (Janicke & Schärer, 2009b). On the one hand, a higher mating rate can lead to increased paternity success simply because more sperm are transferred. On the other hand, if a sperm donor with a higher mating rate also displaces more sperm of his competitors, this could be an alternative mechanism for how testis investment led to increased paternity success in our study. However, we cannot distinguish between these two alternative hypotheses with the present data.

As we found lower paternity benefits due to higher testis investment than expected under our theoretical predictions, processes that deviate from a fair raffle sperm competition, such as sperm displacement (Charnov, 1996), cryptic female choice with a removal of a fixed proportion of the ejaculate (van Velzen *et al.*, 2009) or simply random paternity skews resulting from stochastic effects (Greeff *et al.*, 2001; Schärer, 2009; Schärer & Pen, 2013), may play a role in this system. However, the paternity success that our BAS1 focal did not achieve must have been achieved by its LS1 competitors. Therefore, these processes can only explain our results if they affected the focals in a different way than the competitors, for which we have no evidence (see also next section).

Looking at a broader range of taxa, three different types of studies lend support to the important role of relative testis size during sperm competition. First, there is correlational evidence for males with larger testes achieving higher paternity success in both laboratory studies and natural populations (e.g. Preston *et al.*, 2003; Schulte-Hostedde & Millar, 2003; Awata *et al.*, 2006; Holleley *et al.*, 2006). Second, there is evidence for the importance of testis size during sexual selection from experimental evolution studies under different enforced sexual selection regimes (e.g. Hosken & Ward, 2001; Pitnick *et al.*, 2001; Simmons & García-González, 2008), although some studies generated inconclusive results regarding the role of testis size (Crudgington *et al.*, 2009; Firman & Simmons, 2010). And third, many comparative studies also suggest this link between testis size and paternity success (e.g. Parker *et al.*, 1997; Parker & Pizzari, 2010; Simmons & Fitzpatrick, 2012). Altogether, these studies suggest that relative testis size may be a valid proxy for sperm production.

The testis investment effect on paternity success that we found for *M. lignano* in this study, and evidence from other taxa, confirms predictions from sperm competition theory. In particular, it is often assumed that paternity success of a sperm donor is to some degree proportional to the number of sperm it ejaculates, that is, that sperm production rate is in many cases the target of selection. This is especially true when sperm competition operates like a 'fair raffle', in which every transferred sperm cell has the same chance to fertilize an egg (Parker, 1990; Parker *et al.*, 1990). And empirical data from several taxa show that a higher number of sperm in the ejaculate have a positive effect on paternity success (e.g. Martin *et al.*, 1974; Gage & Morrow, 2003; Garcia-González & Simmons, 2005; Stoltz & Neff, 2006; Boschetto *et al.*, 2010). But although the

speed of spermatogenesis and the amount of spermatogenic tissue *per se* are likely the most important determinants of sperm production rate (and therefore the available sperm number in the ejaculate), the majority of studies use a measure of relative testis size as a proxy for sperm production rate (Parker *et al.*, 1997; Ramm & Schärer, 2014).

In summary, the positive effect of testis investment on paternity success found in our experiment shows, in combination with previous findings, that testis size, sperm production and increased paternity success during sperm competition are indeed linked in *M. lignano*. This confirms the theoretical prediction that increased investment into ejaculates is selected due to sperm competition, although the paternity benefits arising from increased investment were somewhat lower than expected under a model with pure fair raffle sperm competition.

Mating group size effects on relative paternity gains for testis investment

It has been suggested that LSC causes diminishing fitness gains for investment into the male sex function – and hence testis size and sperm production – which may be one reason why hermaphroditism is favoured over gonochorism in some animals (Charnov, 1980; Schärer, 2009; Schärer & Pen, 2013). Therefore, we hypothesized that the fitness gain curve for investment into the male function would show more sharply diminishing returns already for lower values of male allocation in smaller mating groups (with strong LSC) compared to larger mating groups (with weak LSC). Our results do not strongly support this prediction, because the worms in octets did not benefit significantly more from bigger testes than worms in triplets did. This was true both when we used the GLM approach, which tested for the difference between triplets and octets in the benefit for worms with high testis investment in *absolute* terms, and when we used a permutation test to detect differences between triplets and octets in the increase in high testis investment individuals *relative* to low testis investment individuals.

We see three possible explanations for the observed similarity between the paternity gains for increased testis investment in triplets vs. octets. First, the fitness gain curves do differ in shape, but our experimental design was not optimal for detecting this difference. Second, we did not detect a difference because of insufficient statistical power in our study. Third, the fitness gain curves do not differ in shape in the different group sizes, which would question the LSC perspective. We explore these explanations in turn.

Possible limitations of the experimental design

Concerning the first point, it is possible that we were not able to detect the difference between the paternity gains for increased testis investment in triplets vs. octets because our experimental design was not ideal to test for it. One possibility may be that within the range of LSC that we explored in our experiment (i.e. one and six unrelated competitors), the fitness gains resulting from higher testis investment are already quite high (but see 'Difference in strength of LSC between triplets and octets' section for a discussion of the possibility that LSC was *high* in both treatments). Then, the fitness that can be gained for the same testicular investment might not be so different in these two mating group sizes and we may therefore not have been able to detect a difference. We are aware of only one study that provides evidence for a changing male fitness gain curve in response to different levels of LSC in simultaneous hermaphrodites (Yund, 1998). In that study on the marine, spermcast mating ascidian *Botryllus schlosseri*, the author compared three treatment groups of experimentally assembled mating arrays. These involved a focal (male phase) sperm donor colony (of varying testis size) competing for fertilizations in two (female phase) sperm recipient colonies, in either a (i) 'high intensity sperm competition' group with two competing sperm donors having high testis investment, (ii) an 'intermediate intensity sperm competition' group with two competing sperm donors having low testis investment, and (iii) a 'competitor-free' group in which no competing sperm donors were placed in close proximity to the focal. However, as these mating arrays were placed in the field, there were also other nonexperimental competing sperm donors that provided low levels of 'exogenous sperm' and sometimes achieved a considerable paternity success, especially in the 'competitor-free' group. Yund (1998) found that male fitness gains diminished more strongly in the treatments with lower level of sperm competition and by far the most strongly saturating fitness gain curve was found for the 'competitor-free' treatment, whereas the other two treatments were more linear and differed much less. Therefore, one could expect the strongest effect on the shape of the male fitness gain curve when the conditions change from *very* strong LSC to intermediate levels of LSC. A situation with only one permanently present competitor, as in the triplets in our study, might already have substantially lower than maximal levels of LSC, so that the difference in the level of LSC experienced in octets might be fairly small. Instead, it might have been preferable to choose an experimental design that introduces a situation with even higher LSC than is possible with one permanently present competitor in a mating group, as that might have brought us into the range where the effects of LSC on the fitness gains become more easily detectable. One possible experimental design that could achieve this would be one with a 'part-time competitor' that can mate with the sperm recipient only a fraction of the time, whereas the focal worm is allowed to mate the rest of the time. An

experimental treatment like that might increase LSC to an amount where its effects, compared to the presence of one or several 'full-time competitors', is more easily detectable.

Moreover, the explored range of testis investment is crucial for detecting differences in fitness gain curves. In our experiment, we already used the most extreme quartiles of the phenotypic distribution of residual testis size, but maybe it would have been preferable to aim at creating an even broader range of testis investment values to estimate diminishing effects on male fitness. A possible way to generate a broader range of testis investment or rather sperm production rate might be dose-dependent RNA interference, which has been successfully used in *M. lignano* in an earlier study (Sekii *et al.*, 2013).

In addition, testis investment (as measured in our experiment) likely is an incomplete proxy for the sperm production rate. Although testis investment surely does not reflect sperm production perfectly, there is considerable empirical evidence that it does so reasonably well. Namely, the number of testicular stem cells in S-phase, a dynamic measure of testicular activity of a focal worm, is positively correlated with the mean testis investment of the worms in the mating group in which the focal worm was raised ($r^2 = 0.32$, $P < 0.001$; Schärer *et al.*, 2004b). Furthermore, as seminal vesicle area is strongly and positively correlated with number of sperm it contains ($r^2 = 0.77$, $t = 10.9$, $P < 0.001$; see Schärer & Vizoso, 2007) and as the increase in seminal vesicle size during isolation can be predicted by testis size (ANCOVA, including also the factor group size: $r^2 = 0.56$; see Schärer & Vizoso, 2007), we consider testis investment a valid proxy for sperm production (see also 'Testis investment' section). It is, however, possible that worms with high testis investment can plastically down-regulate the amount of sperm they transfer so that they do not profit as much from their bigger testes as one would expect from their testis size. This could also explain the fact that especially worms with high testis investment tended to have lower paternity success than expected (Fig. 3). Although we have no knowledge of such a phenomenon *per se*, we know that the relationship between testis size and sperm production rate can be influenced by the mating group size in which *M. lignano* is raised (Schärer & Vizoso, 2007).

Finally, if the competitors (LS1 culture) for some reason had a different testis size than the average focals (BAS1 culture), then the expected effect sizes could vary as a result. When we explored this possibility by simulating different testis sizes for the competitors (data not shown), we found the lowest deviations between our data and those expectations when the competitors had a testis size very similar to the focals with high testis investment (i.e. when competitors had a testis size of around 29×10^3 μm^2). On the one hand, we have little reason to expect them to differ in testis size because (i) the BAS1 culture was established by backcrossing the transgenic inbred HUB1 line onto the LS1 culture (Marie-Orleach *et al.*, 2016), so that they should be very similar genetically, (ii) a comparison between a transgenic and wild-type inbred line that should otherwise be genetically similar showed no differences in siring ability (Marie-Orleach *et al.*, 2014), and (iii) the focals and competitors were raised under similar conditions until the mating trials started (cf. 'Methods' section). On the other hand, as we raised competitors and focals in different petri dishes, we cannot exclude the possibility that these slightly different raising conditions led to different testis sizes in the competitors compared to the focals. This could have led to a decreased effect size and could therefore explain that we did not find the expected effect of mating group size on paternity gains. Furthermore, if the assumptions of random mating, fair raffle sperm competition and sperm production proportional to testis investment were broken in a way that was different between the focals and the competitors, this could have led to a smaller effect size.

Possible limitations of statistical power

Concerning the second point, it appears possible that there is actually a difference in the effect of testis investment on paternity between triplets and octets in *M. lignano*, but that we could not detect it because of insufficient statistical power. Although we cannot exclude this possibility, each treatment combination had at least 25 replicates, and paternity values were calculated on a fairly high number of total offspring that the partner worms produced (triplets: median = 32, range = 14–58; octets: median = 118.5, range = 78–194). However, the low and moderate within-individual repeatabilities for paternity (triplets: $R = 0.093$; octets: $R = 0.360$) suggest that, particularly in triplets, there may have been quite substantial unknown sources of paternity variation that changed from one mating group to the next. This variation could have affected the estimation of paternity success for the focal worm and therefore obscured the effects we tried to detect here. Possible examples of such unmeasured factors are the mating rate, mating order and/or effects of variation in seminal fluid composition. Also, effects of the genotype of the mating partners and/or competitors could potentially have strong effects on the paternity of the focal (Clark *et al.*, 1999; Evans *et al.*, 2013; Simmons *et al.*, 2014; Travers *et al.*, 2016), and in the triplets, such stochastic variation might have been more pronounced, while it averaged out in the larger groups.

Difference in strength of LSC between triplets and octets

Concerning the third point, it may be possible that the fitness gain curves are not different between triplets and octets. As we did not record mating behaviour directly, we do not know whether the focal worm

actually mated with all possible mating partners. We do know, however, that the realized mating group size is usually lower than the (possible) social mating group size, that is, the focal worm does not successfully store sperm in every member of its mating group (cf. Figs. 2A and 3B in Janicke & Schärer, 2009a and Janicke *et al.*, 2013; respectively). In the present study, focal worms in octets and triplets sired offspring in on average 4.9 (± 2.3 SD) and 0.8 (± 0.8 SD) partners, respectively (compared to worms that managed to store sperm in 2.8 and 1.5 partners, respectively, in a previous study that also joined focals and recipients for 24 h; Janicke & Schärer, 2009a), and the paternity representation in different partners was often highly skewed (i.e. very high in some but very low in other partners; Fig. 4). If this within-focal skew is somewhat representative of the actual among-competitor skew in paternity success, this might indicate a reduced effective number of mates and therefore a higher than otherwise expected level of LSC even in octets. Possible reasons for an elevated LSC in octets could be sperm displacement (Charnov, 1996), cryptic female choice during which a fixed proportion of the ejaculate is removed by the recipient (van Velzen *et al.*, 2009) or simply random paternity skews resulting from stochastic effects (Greeff *et al.*, 2001; Schärer, 2009; Schärer & Pen, 2013). Of these three we have only evidence for sperm displacement in *M. lignano* (Marie-Orleach *et al.*, 2014), although the relatively low repeatability for paternity success suggests substantial variation, be it due to stochasticity or cryptic female choice, as a contributing factor to the relatively high levels of LSC in octets. However, considering that focal worms in our experiments that were tested in octets had on average 4.9 (± 2.3) partners to whom they transferred sperm whereas focals in triplets transferred sperm to only 0.8 (± 0.8) partners, we think that there was still a substantial difference in the number of effective mating partners between triplets and octets and thus a difference in the level of LSC between the mating group sizes.

In summary, we think it would be premature to conclude that there is no difference in the level of LSC between triplets and octets, before we have more data about how the mating group size translates into the effective number of mates when paternities are skewed

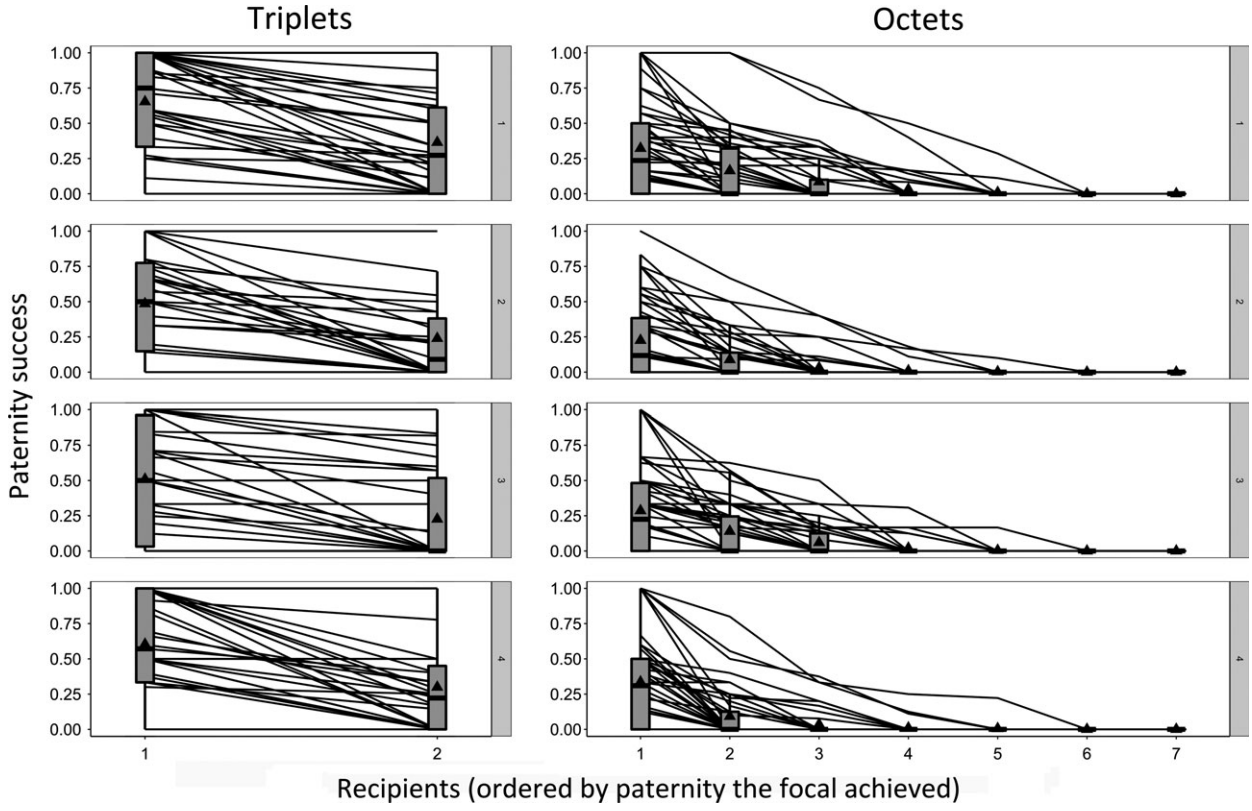

Fig. 4 Paternity success achieved by each focal worm in all of its possible mating partners, ordered by the paternity success that the focal achieved in a given partner. The black lines connect the values for individual focal worms. The box plots show medians, and the 25th and 75th percentiles, respectively, and whiskers extend to 1.5 times the interquartile range. The black triangles indicate mean paternity values, and the four different panels represent the four consecutive mating groups that the focals were tested in.

and how that in turn affects LSC. Furthermore, it seems unclear how, if not based on the LSC perspective, one would explain the well-documented phenotypic plasticity in sex allocation in response to mating group size in *M. lignano* (Schärer *et al.*, 2004b; Schärer & Vizoso, 2007; Janicke *et al.*, 2013; Giannakara *et al.*, 2016), as well as the evolutionary maintenance of hermaphroditism in this and other hermaphrodites, which we discuss next.

Local sperm competition and the evolutionary maintenance of simultaneous hermaphroditism

Although we measured surprisingly modest paternity gains in response to a two-fold difference in testis investment – itself suggesting diminishing fitness returns for testis investment – we were not able to confirm experimentally that LSC leads to diminishing fitness returns for male allocation in *M. lignano* and to show that LSC therefore can contribute to the evolutionary maintenance of simultaneous hermaphroditism in these and other copulating simultaneous hermaphrodites (Schärer & Pen, 2013). We thus briefly mention some other hypotheses besides LSC that have been considered to explain the evolution and maintenance of hermaphroditism, some of which are not mutually exclusive.

According to the fitness gain curve perspective, it is required that either the male and/or the female fitness gain curve shows diminishing returns for hermaphroditism to be the evolutionarily stable strategy. It is not necessary that it is the male fitness gain curve that shows diminishing returns, as long as the female fitness gain curve does so sufficiently (Charnov, 1979, 1982). Therefore, diminishing fitness gains for investment into the female function can also favour and maintain hermaphroditism. For example, in species that exhibit brooding and a limited brood space, the female fitness gain curve can saturate as soon as enough eggs have been produced to fill up the entire brood space (Heath, 1979; Charnov, 1982). Every additional egg produced will then result in much lower fitness returns, because there is no more room for it in the brood space. However, although there seems to be some evidence for a correlation between brooding and hermaphroditism in certain taxa (Ghiselin, 1969), evidence for brooding constraints on egg production is generally weak (Sewell, 1994) and we have no evidence for brooding behaviour, or any other forms of brood care, in *M. lignano*. Similarly, local resource competition, in which related individuals compete for resources in a local area, can also lead to diminishing female fitness gains, namely when offspring produced via the female function have a more clumped distribution in space than offspring produced via the male function, and when female function derived offspring therefore compete more strongly among each other (Charnov, 1982; Lloyd, 1982). Although this could possibly be an alternative explanation for why hermaphroditism is maintained in *M. lignano*, we lack any empirical evidence for sex-specific spatial clustering among offspring at present.

Finally, it has long been argued that hermaphroditism can be favoured over gonochorism when there is no strong trade-off between the male and the female function, for example due to a low overlap between resource requirements of the two sex functions (Charnov *et al.*, 1976), either because organs belonging to the two functions are built at different times or because they require different kinds of limiting resources and/or nutrients. But again, we have no empirical evidence that this is the case in *M. lignano*. On the contrary, we do have clear evidence for a trade-off between male and female allocation, at least under some conditions (Schärer *et al.* 2005; Janicke & Schärer 2009b). Furthermore, there seems to be no *a priori* reason to assume that testis tissue and ovary tissue should be built up by different resources. And while sex allocation is slightly more male-biased at early age in this species (Vizoso & Schärer, 2007), investment in male vs. female reproductive tissues cannot be considered separated in time either.

In summary, although there are several potentially plausible alternative hypotheses to explain the evolution and maintenance of hermaphroditism, none of them seems particularly likely in *M. lignano* given our current knowledge of the biology of this free-living flatworm. The LSC perspective therefore still seems a likely scenario and should be explored further.

Conclusions

Accumulating evidence from several studies, including the current one, suggests a clear, positive relationship between allocation into testes (be it in terms of phenotypic plasticity or standing variation), and the resulting sperm transfer and paternity success in *M. lignano*, which is in line with predictions from sperm competition and sex allocation theory. In contrast, we did not find strong evidence for more sharply saturating male fitness gains under LSC, another core prediction of sex allocation theory. Nevertheless, we think that it would be premature to reject the LSC hypothesis, because there are no plausible and empirically supported alternative hypotheses that could explain the evolutionary maintenance of hermaphroditism in this flatworm. Experiments with lower levels of sperm competition (e.g. with a competitor that is only present some of the time) and a broader range of variation in testis investment might be needed to get a more complete view of LSC and how it influences the male fitness gain curve.

Acknowledgments

We thank Christian Felber for helping with the morphological measurements and Dita Vizoso, Gudrun

Viktorin, Jürgen Hottinger and Urs Stiefel for technical support. Furthermore, we would like to thank two anonymous reviewers for their very helpful comments on an earlier version of this manuscript. This project was supported by grants from the Swiss National Science Foundation to Lukas Schärer (grants 31003A-143732 and 31003A-162543).

References

Al-Jahdali, M.O. 2012. Infrapopulations of *Gyliauchen volubilis* Nagaty, 1956 (Trematoda: Gyliauchenidae) in the rabbitfish *Siganus rivulatus* (Teleostei: Siganidae) from the Saudi coast of the Red Sea. *Parasite J. Société Fr. Parasitol.* **19**: 227–238.

Andersen, R.A., Berges, J.A., Harrison, P.J. & Watanabe, M.M. 2005. Recipes for freshwater and seawater media. In:*Algal Culturing Techniques* (R. A. Andersen, ed.), pp. 429–538. Elsevier, Amsterdam, The Netherlands.

Arnqvist, G. & Rowe, L. 2005. Sexual conflict after mating. In: *Sexual Conflict* (J.R. Krebs & T. Clutton-Brock, ed.), pp. 92–155. Princeton University Press, Princeton, NJ, USA.

Awata, S., Heg, D., Munehara, H. & Kohda, M. 2006. Testis size depends on social status and the presence of male helpers in the cooperatively breeding cichlid *Julidochromis ornatus*. *Behav. Ecol.* **17**: 372–379.

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

Boschetto, C., Gasparini, C. & Pilastro, A. 2010. Sperm number and velocity affect sperm competition success in the guppy (*Poecilia reticulata*). *Behav. Ecol. Sociobiol.* **65**: 813–821.

Brauer, V.S., Schärer, L. & Michiels, N.K. 2007. Phenotypically flexible sex allocation in a simultaneous hermaphrodite. *Evolution* **61**: 216–222.

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

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

Charnov, E.L. 1980. Sex allocation and local mate competition in barnacles. *Mar. Biol. Lett.* **1**: 269–272.

Charnov, E. 1982. *The Theory of Sex Allocation*. Princeton University Press, Princeton, NJ, USA.

Charnov, E.L. 1996. Sperm competition and sex allocation in simultaneous hermaphrodites. *Evol. Ecol.* **10**: 457–462.

Charnov, E., Smith, J. & Bull, J. 1976. Why be an hermaphrodite? *Nature* **263**: 125–126.

Clark, A.G., Begun, D.J. & Prout, T. 1999. Female × male interactions in *Drosophila* sperm competition. *Science* **283**: 217–220.

Crudgington, H.S., Fellows, S., Badcock, N.S. & Snook, R.R. 2009. Experimental manipulation of sexual selection promotes greater male mating capacity but does not alter sperm investment. *Evolution* **63**: 926–938.

Darwin, C. 1871. *The Descent of Man, and Selection in Relation to Sex*. John Murray, London, UK.

Dewsbury, D.A. 2005. The Darwin–Bateman paradigm in historical context. *Integr. Comp. Biol.* **45**: 831–837.

Düsing, C. 1884. Die Regulierung des Geschlechtsverhältnisses. *Jena. Z. Med. Naturwiss.* **17**: 593–940.

Engqvist, L. 2013. A general description of additive and nonadditive elements of sperm competitiveness and their relation to male fertilization success. *Evolution* **67**: 1396–1405.

Evans, J.P., Rosengrave, P., Gasparini, C. & Gemmell, N.J. 2013. Delineating the roles of males and females in sperm competition. *Proc. R. Soc. Lond. B Biol. Sci.* **280**: 20132047.

Faraway, J.J. 2016. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall, Boca Raton, FL, USA.

Firman, R.C. & Simmons, L.W. 2010. Experimental evolution of sperm quality via postcopulatory sexual selection in house mice. *Evolution* **64**: 1245–1256.

Fisher, R.A. 1930. *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, UK.

Gage, M.J.G. & Morrow, E.H. 2003. Experimental evidence for the evolution of numerous, tiny sperm via sperm competition. *Curr. Biol.* **13**: 754–757.

García-González, F. & Simmons, L.W 2005. Sperm viability matters in insect sperm competition. *Curr. Biol.* **15**: 271–275.

Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N. & Hughes, K.A. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **525**: 372–375.

Ghiselin, M.T. 1969. The evolution of hermaphroditism among animals. *Q. Rev. Biol.* **44**: 189–208.

Giannakara, A., Schärer, L. & Ramm, S.A. 2016. Sperm competition-induced plasticity in the speed of spermatogenesis. *BMC Evol. Biol.* **16**: 60.

Greeff, J.M., Nason, J.D. & Compton, S.G. 2001. Skewed paternity and sex allocation in hermaphroditic plants and animals. *Proc. R. Soc. Lond. B Biol. Sci.* **268**: 2143–2147.

Hamilton, W.D. 1967. Extraordinary sex ratios. *Science* **156**: 477–488.

Heath, D.J. 1979. Brooding and the evolution of hermaphroditism. *J. Theor. Biol.* **81**: 151–155.

Holleley, C.E., Dickman, C.R., Crowther, M.S. & Oldroyd, B.P. 2006. Size breeds success: multiple paternity, multivariate selection and male semelparity in a small marsupial, *Antechinus stuartii*. *Mol. Ecol.* **15**: 3439–3448.

Hosken, D.J. & Ward, P.I. 2001. Experimental evidence for testis size evolution via sperm competition. *Ecol. Lett.* **4**: 10–13.

Houston, A.I. & McNamara, J.M. 2006. John Maynard Smith and the importance of consistency in evolutionary game theory. *Biol. Philos.* **20**: 933–950.

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

Janicke, T. & Schärer, L. 2009b. Sex allocation predicts mating rate in a simultaneous hermaphrodite. *Proc. R. Soc. B Biol. Sci.* **276**: 4247–4253.

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

Janicke, T., Sandner, P., Ramm, S.A., Vizoso, D.B. & Schärer, L. 2016. Experimentally evolved and phenotypically plastic responses to enforced monogamy in a hermaphroditic flatworm. *J. Evol. Biol.* **29**: 1713–1727.

Kopp, M. & Matuszewski, S. 2014. Rapid evolution of quantitative traits: theoretical perspectives. *Evol. Appl.* **7**: 169–191.

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). *J. Zool. Syst. Evol. Res.* **43**: 114–126.

Lehtonen, J., Parker, G.A. & Schärer, L. 2016. Why anisogamy drives ancestral sex roles. *Evolution* **70**: 1129–1135.

Lloyd, D.G. 1982. Selection of combined versus separate sexes in seed plants. *Am. Nat.* **120**: 571–585.

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

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 Evol. Biol.* **14**: 148.

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**: 314–328.

Martin, P.A., Reimers, T.J., Lodge, J.R. & Dziuk, P.J. 1974. The effect of ratios and numbers of spermatozoa mixed from two males on proportions of offspring. *J. Reprod. Fertil.* **39**: 251–258.

Michiels, N.K., Crowley, P.H. & Anthes, N. 2009. Accessory male investment can undermine the evolutionary stability of simultaneous hermaphroditism. *Biol. Lett.* **5**: 709–712.

Nakagawa, S. & Schielzeth, H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol. Rev.* **85**: 935–956.

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

Parker, G.A. 1990. Sperm competition games: raffles and roles. *Proc. R. Soc. Lond. B Biol. Sci.* **242**: 120–126.

Parker, G.A. & Birkhead, T.R. 2013. Polyandry: the history of a revolution. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20120335.

Parker, G.A. & Pizzari, T. 2010. Sperm competition and ejaculate economics. *Biol. Rev.* **85**: 897–934.

Parker, G.A., Simmons, L.W. & Kirk, H. 1990. Analysing sperm competition data: simple models for predicting mechanisms. *Behav. Ecol. Sociobiol.* **27**: 55–65.

Parker, G.A., Ball, M.A., Stockley, P. & Gage, M.J.G. 1997. Sperm competition games: a prospective analysis of risk assessment. *Proc. R. Soc. Lond. B Biol. Sci.* **264**: 1793–1802.

Pitnick, S., Miller, G.T., Reagan, J. & Holland, B. 2001. Males' evolutionary responses to experimental removal of sexual selection. *Proc. R. Soc. Lond. B Biol. Sci.* **268**: 1071–1080.

Preston, B.T., Stevenson, I.R., Pemberton, J.M., Coltman, D.W. & Wilson, K. 2003. Overt and covert competition in a promiscuous mammal: the importance of weaponry and testes size to male reproductive success. *Proc. R. Soc. Lond. B Biol. Sci.* **270**: 633–640.

Price, T.D., Qvarnström, A. & Irwin, D.E. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **270**: 1433–1440.

Queller, D.C. 2006. Sex ratios and social evolution. *Curr. Biol.* **16**: R664–R668.

R Development Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramm, S.A. & Schärer, L. 2014. The evolutionary ecology of testicular function: size isn't everything. *Biol. Rev.* **89**: 874–888.

Schärer, L. 2009. Tests of sex allocation theory in simultaneously hermaphroditic animals. *Evolution* **63**: 1377–1405.

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

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. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20120052.

Schärer, L. & Ramm, S.A. 2016. Hermaphrodites. In: *Encyclopedia of Evolutionary Biology* (R. Kliman, ed.), pp. 212–224. Elsevier, Oxford, UK.

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

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

Schärer, L. & Wedekind, C. 2001. Social situation, sperm competition and sex allocation in a simultaneous hermaphrodite parasite, the cestode *Schistocephalus solidus*. *J. Evol. Biol.* **14**: 942–953.

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

Schärer, L., Ladurner, P. & Rieger, R.M. 2004b. Bigger testes do work more: experimental evidence that testis size reflects testicular cell proliferation activity in the marine invertebrate, the free-living flatworm *Macrostomum* sp. *Behav. Ecol. Sociobiol.* **56**: 420–425.

Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* **1**: 103–113.

Schielzeth, H. & Nakagawa, S. 2013. *Repeatability for Gaussian and non-Gaussian data*. R package version 0.6. 405/r52.

Schulte-Hostedde, A.I. & Millar, J.S. 2003. Intraspecific variation of testis size and sperm length in the yellow-pine chipmunk (*Tamias amoenus*): implications for sperm competition and reproductive success. *Behav. Ecol. Sociobiol.* **55**: 272–277.

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. *Proc. Biol. Sci.* **280**: 20122711.

Sewell, M.A. 1994. Small size, brooding, and protandry in the apodid sea cucumber *Leptosynapta clarki*. *Biol. Bull.* **187**: 112–123.

Simmons, L.W. & Fitzpatrick, J.L. 2012. Sperm wars and the evolution of male fertility. *Reproduction* **144**: 519–534.

Simmons, L.W. & García-González, F. 2008. Evolutionary reduction in testes size and competitive fertilization success in response to the experimental removal of sexual selection in dung beetles. *Evolution* **62**: 2580–2591.

Simmons, L.W., Lovegrove, M. & Almbro, M. 2014. Female effects, but no intrinsic male effects on paternity outcome in crickets. *J. Evol. Biol.* **27**: 1644–1649.

Stoltz, J.A. & Neff, B.D. 2006. Sperm competition in a fish with external fertilization: the contribution of sperm number, speed and length. *J. Evol. Biol.* **19**: 1873–1881.

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

Travers, L.M., Garcia-Gonzalez, F. & Simmons, L.W. 2016. Genetic variation but weak genetic covariation between pre- and post-copulatory episodes of sexual selection in *Drosophila melanogaster*. *J. Evol. Biol.* **29**: 1535–1552.

Trouvé, S., Jourdane, J., Renaud, F., Durand, P. & Morand, S. 1999. Adaptive sex allocation in a simultaneous hermaphrodite. *Evolution* **53**: 1599–1604.

van Velzen, E., Schärer, L. & Pen, I. 2009. The effect of cryptic female choice on sex allocation in simultaneous hermaphrodites. *Proc. R. Soc. Lond. B Biol. Sci.* rspb20090566.

Vizoso, D.B. & Schärer, L. 2007. Resource-dependent sex-allocation in a simultaneous hermaphrodite. *J. Evol. Biol.* **20**: 1046–1055.

Warton, D.I. & Hui, F.K.C. 2010. The arcsine is asinine: the analysis of proportions in ecology. *Ecology* **92**: 3–10.

West, S. 2009. *Sex Allocation*. Princeton University Press, Princeton, NJ, USA.

Wigby, S., Sirot, L.K., Linklater, J.R., Buehner, N., Calboli, F.C.F., Bretman, A. *et al.* 2009. Seminal fluid protein allocation and male reproductive success. *Curr. Biol.* **19**: 751–757.

Yund, P.O. 1998. The effect of sperm competition on male gain curves in a colonial marine invertebrate. *Ecology* **79**: 328–339.

Zadesenets, K.S., Vizoso, D.B., Schlatter, A., Konopatskaia, I.D., Berezikov, E., Schärer, L. et al. 2016. Evidence for karyotype polymorphism in the free-living flatworm, *Macrostomum lignano*, a model organism for evolutionary and developmental Biology. *PLoS ONE* **11**: e0164915.

Zadesenets, K.S., Schärer, L. & Rubtsov, N.B. 2017. New insights into the karyotype evolution of the free-living flatworm *Macrostomum lignano* (Platyhelminthes, Turbellaria). *Sci. Rep.* **7**: 6066.

Data deposited at Dryad: doi:10.5061/dryad.fd407

Received 23 February 2017; revised 31 October 2017; accepted 14 November 2017