Sperm competition favours intermediate sperm size in a hermaphrodite

Santhosh Santhosh[1], Dieter Ebert[1], and Tim Janicke[2]

1 Department of Environmental Sciences, Zoology, University of Basel, CH-4051 Basel, Switzerland.
2 CEFE, Univ. Montpellier, CNRS, EPHE, IRD, 34293 Montpellier Cedex 05, France.

* Corresponding author. E-mail: santhosh.s@unibas.ch

Running head: Sperm competition and sperm morphology

Keywords: sexual selection, sperm competition, sperm size, hermaphrodites, stabilising selection

Author contributions: SS conceived the study, conducted the experiment and collected the data. SS performed the statistical analysis and wrote the manuscript with the help of DE and TJ.

Acknowledgements: We thank Lukas Schärer for significant contributions to conceiving the study and designing the experiment. We also thank Gudrun Viktorin, Jurgen Hottinger and Daniel Lüscher for technical support, and Yasmin Picton for administrative support.

Funding: The study was supported by a Swiss national Science Foundation grant to Lukas Schärer (Grant no: 310030_184916).

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

Data Availability: All relevant dataset and R scripts will be made publicly available in an online data repository at the time of publication.

Summary Statement: This study is one of the first to present direct empirical evidence suggesting that sperm size within a species is under stabilising selection.

Abstract

Sperm competition is a potent mechanism of post-copulatory sexual selection that has been found to shape reproductive morphologies and behaviours in promiscuous animals. Especially sperm size has been argued to evolve in response to sperm competition through its effect on sperm longevity, sperm motility, the ability to displace competing sperm and ultimately fertilization success. Theoretical work predicts that sperm competition may favour longer sperm but may also select for shorter sperm if sperm size trades off with number. In this study, we studied the relationship between sperm size and post-mating success in the free-living flatworm, *Macrostomum lignano*. Specifically, we used inbred isolines of *M. lignano* that varied in sperm size to investigate how sperm size translated into the ability of worms to transfer and deposit sperm in a mating partner. Our results revealed a hump-shaped relationship with individuals producing sperm of intermediate size having highest sperm competitiveness. This finding broadens our understanding of the evolution of sperm morphology by providing evidence for stabilizing selection on sperm size under sperm competition.

Introduction

Ever since Parker's pioneering work on sperm competition, the study of post-copulatory sexual selection has drastically improved our understanding of male and female reproductive biology (Birkhead and Moller, 1999; Simmons, 2001). In species where individuals mate multiply with different partners, post-mating fertilisation success determined by the efficiency to outcompete rival ejaculates can become a major predictor of male reproductive success (Collet et al., 2012; Pélissié et al., 2014; Marie-Orleach et al., 2021). Consequently, sperm competition is expected to promote post-copulatory intrasexual selection on various traits that influence fertilization success. Over the last decades, great efforts have been made to gain a better understanding of such phenotypic and genetic relationships between ejaculate traits and the outcome of sperm competition. Especially, sperm size became an ejaculate trait that has received substantial theoretical and empirical attention (reviewed in Pizzari and Parker, 2009). This is because sperm size has been found to vary tremendously not only within ejaculates but also among males and among species, which is often considered to result from selection arising from sperm competition and/or cryptic female choice (Pitnick et al. 2009). Specifically, sperm size is expected to covary positively with sperm motility, sperm longevity and the defensive behaviour of an ejaculate (Snook 2005; Fitzpatrick and Lüpold 2014), such that most theoretical models assume a positive, saturating relationship between sperm size and its fertilization success under sperm competition (Pizzari and Parker, 2009). However, sperm size is often considered to trade-off with sperm number, which complicates the prediction of how sperm size is selected under varying sperm competition risks and intensities. Theoretical work suggests that if sperm size trades-off against number of sperm produced or resources required for pre-copulatory mate acquisition, selection may favour ejaculates with sperm of intermediate size (Parker, 1993; Parker et al., 2010; Parker and Begon, 1993).

Although variation in ejaculate performance traits, such as sperm size, is generally intrinsic to males, it has also been shown that females play an active and crucial role in post-copulatory sexual selection and that competitive fertilization success of males is influenced by interactions between sperm and the female reproductive tract (Lüpold et al., 2013; Miller and Pitnick, 2002; Pitnick et al., 2009). Notably, sperm

size has been found to covary with the length of female reproductive organs, such as seminal receptacle or spermatheca (Minder et al., 2005; Morrow and Gage, 2000; Pitnick et al., 1999, 2003). Moreover, using an experimental evolution study, Miller and Pitnick (2002) showed a correlated evolution in sperm length as a result of an evolutionary increase in seminal receptacle length in *Drosophila melanogaster*. Thus, selection on sperm size may also arise from selection imposed by properties of the female reproductive tract.

Empirical tests on the relationship between sperm size and sperm competitiveness provided mixed results within and across species (Pizzari and Parker, 2009). Comparative studies indicate that higher levels of sperm competition are associated with larger and/or more elaborate sperm in fish (Balshine et al., 2001), birds (Briskie et al., 1997) and mammals (Tourmente et al., 2011) suggesting that sperm competition promotes directional selection on sperm size. By contrast, negative or no relationship between sperm competition and sperm size has been reported for Sylvian warblers (Immler and Birkhead, 2007) and mammals (Gage and Freckleton, 2003). Another line of across-species comparisons suggests that higher levels of sperm competition are associated with reduced intra-specific and intra-male variation in sperm size across birds, which has been speculated to arise from stabilising selection for an optimum sperm size (Lüpold et al., 2009).

The most direct evidence for selection acting on sperm size comes from studies quantifying how phenotypic variation in sperm size translates into sperm competitive success within species. Empirical studies attempting to establish such associations within species have found positive (Radwan, 1996), negative (Gage and Morrow, 2003; Garcìa-González and Simmons, 2007) or no clear relationship (Cascio Sætre et al., 2018; Kahrl et al., 2021) between sperm size and male reproductive success (or paternity share). In addition, studies comparing the relative fertilization success between extreme sperm sizes (e.g. large vs small) also provided mixed results (Bennison et al., 2015; Morrow and Gage, 2001). These contrasting results might, at least partly, result from substantial within-individual variation in male post-mating success and the associated uncertainty in estimating individual differences in sperm competitiveness. Marie-Orleach et al. (2021) took measures of reproductive performance of the same individual over multiple mating events and detected low repeatabilities for different proxies of post-mating success, indicating the significance

107 of stochastic variation in measures of post-mating success. Importantly, the vast
108 majority of empirical studies assume a linear relationship between sperm size and
109 reproductive (or post-mating) success and do not test for non-linear associations,
110 which limits our understanding on how selection operates on sperm size.

111 Here, we report an experimental study exploring the relationship between sperm size
112 and post-mating success within a species. Specifically, we tested the hypothesis that
113 sperm size predicts sperm transfer success in the simultaneously hermaphroditic
114 flatworm, *Macrostomum lignano,* a species in which sperm competition has been
115 shown to be intense (Janicke et al., 2013; Marie-Orleach et al. 2021). In *M. lignano*,
116 a recent study explored how sperm morphology predicts male reproductive success
117 (and other fitness components) (Marie-Orleach et al., 2023), focussing on purely
118 phenotypic relationships between sperm traits and male fitness components. By
119 using inbred lines, we aimed to extend previous findings by exploring how sperm
120 competition generates selection on standing genetic variance in sperm length.

121 An outbred lab culture and a set of 12 inbred lines of *M. lignano* expressing a green
122 fluorescent protein (GFP) in all cell types, including sperm cells, allowed us to
123 distinguish between sperm from a focal GFP-positive (hereafter denoted as GFP+)
124 donor from those of wildtype competitors with no GFP expression (hereafter GFP-).
125 We used these inbred lines to conduct mating trials testing how variation in sperm
126 size among the lines predicted sperm transfer success (i.e., the number of sperm
127 stored in the partner's sperm storage) in the presence of sperm competition. While
128 we hypothesised an association between sperm size and sperm competitiveness, we
129 did not have a clear prediction for the shape of this relationship as the factors
130 determining its exact form (e.g., resource allocation among ejaculate components,
131 sperm displacement propensity and spatial constraints for sperm storage) are largely
132 unknown for *M. lignano*.

133 Methods

134 Model organism

135 The free-living flatworm *Macrostomum lignano* (Macrostomorpha, Platyhelminthes) is
136 an obligatorily outcrossing simultaneous hermaphrodite found in the intertidal zones

137 of the Northern Adriatic sea and the Aegean sea (Ladurner et al., 2005; Schärer et
138 al., 2020). Laboratory cultures of *M. lignano* are kept at 20°C in glass petri dishes
139 with artificial sea water (ASW) at a salinity of 32 ‰ and fed with the diatom *Nitzschia*
140 *curvilineata*. Under laboratory conditions, *M. lignano* are found to be highly
141 promiscuous and mate frequently when kept in groups. Copulation in this species
142 involves reciprocal intromission of the male copulatory organ into the female sperm-
143 receiving and -storage organ, the antrum (Schärer et al., 2004; Vizoso et al., 2010).
144 *M. lignano* exhibits a post-copulatory suck behaviour, in which worms bend down
145 and place their pharynx on top of their own genital opening and appear to suck,
146 which may have the function to remove sperm out of the antrum (Vizoso et al.,
147 2010).

148 The morphology of the sperm in *M. lignano* is complex with multiple specialized
149 structures, including two lateral bristles and an anterior feeler (Fig. 1) (Ladurner et
150 al., 2005; Vizoso et al., 2010; Willems et al., 2009). The lateral bristles are
151 hypothesized to reduce the likelihood of sperm being removed from the antrum
152 during the suck behaviour (Vizoso et al., 2010). The feeler has been observed to
153 anchor sperm cells to the epithelium at the anterior region of the antrum (called
154 cellular valve), where fertilization is hypothesized to occur. The transparent body of
155 these worms enables the observation of numerous internal structures and processes
156 *in vivo* such as testis and ovaries, stylet morphology and sperm (both GFP+ and
157 GFP-) stored in the female antrum (Janicke et al., 2011; Marie-Orleach et al., 2016,
158 2021).

159 Experimental setup

160 In this study, we aimed to (i) quantify inter-individual variation in sperm size and (ii)
161 assess the relationship between sperm size and sperm transfer success in *M.*
162 *lignano*. To this end, we measured the sperm size of adult worms (N = 39) from the
163 outbreeding BAS1 culture to quantify the natural phenotypic variation in sperm
164 morphology in *M. lignano*. The BAS1 culture is an outbred GFP+ culture derived
165 from the outbred GFP- LS1 culture by backcrossing the GFP marker from the GFP+
166 HUB1 inbred line onto LS1 for 9 generations (Marie-Orleach et al., 2016). BAS1
167 worms express GFP ubiquitously in all cells, including sperm cells. For our second
168 aim, we conducted mating experiments and measured relative sperm

169 competitiveness of worms known to vary in sperm size. Sperm size of worms cannot
170 be measured before mating experiments, as it is a destructive process harming the
171 worms (see next section). Hence, we used inbred lines of *M. lignano*, which are
172 expected to show low within-, but large among-line variation in sperm size. Using
173 inbred lines also allowed us to estimate the amount of variation in sperm size
174 attributable to genetic variance across lines.

175 We investigated variation in sperm size among LM lines, a set of 12 GFP+ inbred
176 lines of *M. lignano* derived from the inbred GFP- DV lines by backcrossing the GFP
177 marker from the HUB1 inbred line onto themselves for several generations (Vellnow
178 et al., 2017). For the sperm competition experiment, we chose 8 of the 12 LM lines
179 with relatively narrow variation in sperm size, competing against common
180 competitors from a single GFP- inbred line called DV1. Using worms from the GFP-
181 DV1 line as common competitors to all the LM lines allowed us to distinguish GFP+
182 sperm of LM sperm donors from the GFP- sperm of their DV1 competitors. Sperm
183 competitiveness was estimated by calculating sperm transfer success as the
184 proportion of GFP+ focal sperm relative to total number of sperm stored in the
185 antrum. Sperm transfer success has been shown to positively covary with mating
186 success and paternity share in *M. lignano* (Marie-Orleach et al., 2016), thus
187 representing an important fitness component of the male sex function. As mating
188 rates have been shown to be high (Schärer et al., 2004), using unmated (virgin)
189 worms would not represent a realistic scenario to study reproductive traits in this
190 species. Therefore, worms were raised in small groups of 3 worms until the mating
191 trials so that they were sexually experienced and had reached a biologically realistic
192 steady state of sperm and egg reserves ready to be donated and fertilized,
193 respectively.

194 Sperm size measurements

195 Measurements taken from each sperm cell included the total sperm length,
196 measured from anterior tip of feeler to the posterior tip of brush, and the length of
197 sperm feeler, body, shaft, brush and lateral bristles (mean length of the two bristles).
198 Measurements were done using an established technique previously used in *M.*
199 *lignano* (Janicke and Schärer, 2010). Adult worms were isolated for a day in wells of
200 a 24-well plate to allow them to restore sperm reserves in their seminal vesicles,

201 where they are stored until the sperm is transferred to a partner during copulation.
202 The tail plates of the worms consisting the seminal vesicle were amputated and
203 transferred onto a microscopic slide with 1 µl of ASW medium, rupturing the seminal
204 vesicle and causing sperm to spill out into the surrounding medium. This restricted
205 the movement of sperm cells greatly, facilitating the imaging of sperm morphology
206 with a digital camera (The Imaging Source, DFK 41BF02) connected to a light
207 microscope (Leica DM2500, Leica Micro-systems, Germany) and imaging software
208 BTV Pro 6.0b7 (http://www.bensoftware.com/).

209 We took digital micrographs of sperm cells at 1000x magnification and the
210 micrographs were analyzed in ImageJ 1.53c using the ObjectJ plugin
211 (https://sils.fnwi.uva.nl/bcb/objectj/). Sperm body length was not measured in the
212 BAS1 culture worms, but was measured in the LM line worms. About 15 sperm cells
213 were measured from all individuals that were measured as part of this study. This
214 number has been shown to be adequate to provide an accurate estimate of the
215 individual mean values of sperm length (Janicke and Schärer, 2010). Additionally,
216 we estimated the measurement error in this method by calculating the repeatability
217 when the same sperm cell was measured thrice for 10 sperm cells.

218 Rearing conditions

219 On day 1, 200 GFP(+) adults from each of the 12 LM lines and 2400 GFP- adults of
220 the DV1 line were distributed into petri dishes, 100 adults in each, for egg laying with
221 20 ml ASW and ad libitum algae. On day 3, we removed all adults from the dishes
222 controlling the age differences in the resulting juveniles to less than 48 hours. On
223 day 9, we sampled the hatchlings to form rearing groups of three hatchlings, all from
224 the same inbred line. In total, we formed 30 such groups for each LM line and 360
225 groups from the DV1 line. The groups were placed in wells of 24-well plates with 1.5
226 ml of 32‰ ASW and ad libitum algae. To control for any environmental effects during
227 development, every 24-well plate contained 2 groups from each of the LM lines in
228 randomly assigned wells. All groups were transferred to fresh medium and algae
229 every 6-10 days until adulthood. During this period, sperm size of the 12 LM lines
230 were measured from a different set of adult worms (56 - 63 days old) and 8 lines with
231 relatively narrow within-line variation in sperm morphologies were chosen for the
232 sperm competition experiment.

233 Sperm competition experiment

234 On days 35 through 55, we assessed sperm transfer success of worms from the 8
235 LM lines when competing against the common competitor DV1 line. On every
236 experiment day, we formed 'mating groups' by moving one randomly picked adult
237 from a LM rearing group, verified for GFP expression, to a fresh well along with three
238 DV1 worms. Thus, a mating group consisted of one GFP+ focal worm from one of
239 the 8 LM lines and three GFP- worms of a DV1 rearing group, and each focal LM
240 worm competed against two DV1 competitors for access to eggs of a third DV1
241 worm. On the following day, the focal LM worm was removed from the experimental
242 group.

243 Immediately after the 24 hour grouping phase, the number of GFP- and GFP+ sperm
244 in the female antrum of the three DV1 partners was assessed, as previously reported
245 (Janicke et al., 2013). We used a digital video camera (ORCA-Flash 4.0 V2:
246 C11440-22CU, Hamamatsu corporation) connected to a Leica DM2500 microscope
247 (Leica Micro-systems, Germany) and the software Leica Applications Suite X (LAS
248 X, Leica Microsystems) to make movies of the female antrum. For each DV1 partner,
249 we recorded a movie of the female antrum by focusing through the antrum under
250 differential interference contrast illumination to be able to see all the sperm in the
251 antrum (movie 1). Immediately after, we recorded a second movie of the female
252 antrum under epifluorescence illumination to visualize the GFP+ sperm that were
253 transferred and stored by the focal worm (movie 2). We analyzed the movies using
254 VLC media player and counted, for each focal donor, the total number of sperm
255 (from movie 1) and the number of GFP+ sperm (from movie 2) from the focal donor
256 in the female antrum of all its potential partners.

257 Similar to previous studies with *M. lignano*, we encountered worms with an egg (or
258 eggs) in the antrum (i.e., 263 of the total 720 DV1 worms used in the experiment),
259 which meant we could only count the GFP+ sperm in these partners. Previous
260 studies extrapolated the total number of sperm in such worms by using the average
261 value of the total number of sperm in all the other worms (Marie-Orleach at al., 2016,
262 2021). In this study, the number of GFP+ sperm was sometimes greater than the
263 total number of sperm resulting in proportion values higher than 1. Therefore, we

264 performed the analysis excluding the worms that had an egg (or eggs) in their female
265 antrum.

266 Statistical analysis

267 All statistical analysis for this study were performed in R (Version 4.2.3).

268 Variation in sperm size

269 The inter-male and the intra-male variation in sperm length measurements of the
270 outbred BAS1 culture were assessed by calculating the coefficient of variation for
271 each sperm morphological trait. We tested for between-individual differences on all
272 sperm traits using Kruskal Wallis rank sum tests with individual ID as the predictor
273 variable. We then assessed the relationships between the lengths of the different
274 sperm traits by calculating the Pearson correlation coefficients. Repeatabilities for
275 the measurement method were calculated based on three measurements of each
276 sperm cell using the rptR package (V0.9.22, Stoffel et al., 2017).

277 All further analysis were done using the arithmetic mean of sperm measurements
278 obtained from each individual. Differences in sperm traits among the 12 LM lines
279 were tested by performing post-hoc pairwise comparisons of a linear model using the
280 *emmeans* package (V1.8.6, Lenth, 2023) in R, with trait length as dependent variable
281 and the LM line ID as the independent group variable. We estimated the proportion
282 of phenotypic variation in sperm size that is explained by the LM line identity, i.e.,
283 variance that can be attributed to the genotype of the male using a linear mixed
284 models with LM line ID and individual ID as random effects and trait length as
285 dependent variable. Given the complex morphology of the sperm in this species and
286 that most sperm traits are highly correlated to each other, we performed a Principal
287 Components Analysis (PCA) using the *stats* package (V4.2.3) with the length data of
288 the five sperm traits to obtain a composite measure describing variation in overall
289 sperm size as a whole in further analysis. For simplicity, we present the analysis of
290 the sperm competition experiment using PC1 values as the proxy for overall sperm
291 size in the main text and report the analysis of original traits in the Supplementary
292 information (Fig. S1; Table S1).

293 Relationship between sperm length and sperm transfer success

294 Initially, we intended to run 240 sperm competition trials (30 per line) but we lost 4
295 replicates due to handling errors. Moreover, 263 (36.5%) of the 720 DV1 partners
296 had one or more eggs in their antrum, resulting in the loss of 24 replicates where all
297 partners had eggs in their antrum at the time of observation. Previous studies have
298 reported comparable frequencies for the occurrence of eggs in the female antrum in
299 *M. lignano* (Marie-Orleach at al., 2016, 2021). Thus, the final analysis was performed
300 with a total of 212 replicates (24 to 29 replicates per line, mean = 26.5). For each
301 replicate, we summed the total number of sperm and the number of focal GFP+
302 sperm found in the three partners (only 2 or 1 in cases of eggs/worms lost). Sperm
303 transfer success of a GFP+ focal worm was calculated as the proportion of GFP+
304 sperm among the total number of sperm stored in the antrum of all the partners.

305 Firstly, we assessed the between-line differences in sperm transfer success among
306 the LM lines. We fitted binomial Generalized linear-mixed models (GLMM) using the
307 lme4 package (V1.1-32, Bates et al., 2015), with line identity as a categorical fixed
308 effect, date of the mating group formation as a categorical random effect and the
309 relative number of GFP+ and GFP- sperm with the 'cbind' function as the response
310 variable representing sperm transfer success. The random factor only explained a
311 very small portion of the total variance in sperm transfer success (3.8%) and was
312 therefore excluded from all the models and results presented here. We observed
313 over-dispersion in binomial Generalized Linear Models (GLM) excluding the random
314 effect term. Therefore, we refitted the models assuming a quasibinomial data
315 distribution (Dispersion parameter = 9.343). Post-hoc tests were conducted and
316 Tukey adjusted *P*-values for pairwise comparisons were obtained using the
317 *emmeans* package (V1.8.6, Lenth, 2023).

318 Secondly, we assessed the relationship between sperm transfer successes and
319 sperm length (captured by PC1). As mentioned earlier, sperm length measurements
320 are invasive requiring the amputation of the tail plate so that we could not measure
321 sperm length of focal worms used in the mating experiments. Therefore, we used the
322 mean of the PC1 of each line (hereafter, called "line means") as our measure of
323 sperm size. This meant that all focal worms from the same LM line were assumed to
324 express similar sperm length. However, to account for the variation within lines such

that mean values of lines with higher precision are given relatively higher weight compared to those with lower precision, we used the inverse of the standard deviation in PC1 values within lines as a weighing factor in our models.

We built two quasibinomial regression models, one first-order linear and one including, in addition, a second-order quadratic term, with individual sperm transfer success as the response variable and line means of PC1 values as a continuous predictor variable. In order to assess the potential for a hump-shaped relationship between sperm length and sperm transfer success, we used a quadratic polynomial of line means as the independent predictor variable. We performed a likelihood ratio test to infer whether adding the quadratic term provides a better model fit. In addition, we compared the goodness of fit of the two models by using the Akaike information criterion, accounting for the over-dispersion in the data (Quasi-AIC), as implemented in the *bbmle* package (V1.0.25, Bolker, 2022) in R.

Results

The measurement repeatabilities were high for all the sperm traits measured (Table1). The total, bristles and shaft lengths were highly repeatable whereas repeatabilities were lower for the feeler lengths and brush lengths (Table 1). This was expected as undulations of the feelers and the very small size of the brush, caused difficulties in accurate measurement of the length s.

Variation in sperm length: BAS1 line

In the outbred BAS1 culture, we observed significant between-individual variation in all sperm traits measured (Table 1, also see Fig. 2A for the variation in total sperm length). The coefficients of variation (CV) of all sperm traits were comparable (4.3 to 6.1 %), except for the brush which had a much higher coefficient of variation (34.38 %, Table 1), which could partly be due to high measurement error in brush lengths. All sperm traits covaried positively with each other ($r > 0.5$), except brush length which hardly covaried with any of the other sperm traits (Table 2).

Variation in sperm length: LM lines

As expected, the overall variation in sperm length found in all the LM lines fell within the range of the variation in the outbred BAS1 culture (Fig. 3) and the combined

355 variation in the LM lines approximately covered the overall variation in the outbred
356 BAS1 culture. We found a significant between-line variation in all sperm traits
357 measured (total sperm length: Kruskal-wallis test, χ^2 = 65.341, df = 11, *P* = 9.3e^{-10}),
358 which is also reflected in pairwise differences between some of the lines (Table S2).
359 The linear mixed models revealed that a substantial portion of phenotypic variation in
360 all sperm traits could be explained by LM line identity (Table S3), explaining 45 % of
361 variation in total sperm length. The coefficients of variation within lines varied, with
362 some LM lines showing a wider range of sperm traits than others (results not
363 shown). Principal component analysis revealed that only PC1 had an eigenvalue of
364 >1.0, and was positively and almost equally loaded by all the sperm traits used,
365 except the brush length, which had a lower weight (Table 3). PC2 was mainly
366 negatively loaded by the brush length, and to a lesser extent positively by shaft
367 length. Therefore, PC1 was used as the only proxy for sperm length in the sperm
368 competition experiment analysis presented below.

369 Sperm competition experiment

370 The null expectation for sperm transfer success in all the lines was one third (3
371 individuals were competing for the eggs of the fourth member of each group), which
372 was close to the observed mean sperm transfer success of 0.29 (one-sample t-test, t
373 = -.1.719, df = 207, *P* = 0.09). We observed a significant between-line variation in
374 sperm transfer success of LM focal worms (line ID: χ^2 = 52.07, df = 7, *P* = 5.651e^{-9},
375 Fig. 4; pairwise comparison results in Table S4). To assess the form of the
376 relationship between sperm size and sperm transfer success, two binomial models
377 were fitted between PC1 (mean values of lines) and sperm transfer success. The
378 first-order liner model showed a significantly negative relationship between PC1 and
379 sperm transfer success (Linear: $\chi^2_{1,206}$ = 9.234, *P* = 0.002, R^2 = 0.038, Slope = -
380 0.118, SE = 0.039). The second-order quadratic model revealed a significant non-
381 linear relationship with intermediate PC1 values having higher sperm transfer
382 success (Quadratic term: $\chi^2_{1,205}$ = 12.693, *P* = 0.0004, R^2 = 0.089, Fig. 4). The
383 quadratic regression model provided a significantly better fit to the data compared to
384 the linear regression model (likelihood ratio test: χ^2_1 = 143.62.94, *P* = 0.0003; AIC_{linear} =
385 309.44, $AIC_{quadratic}$ = 298.38).

Linear regressions on original sperm traits revealed that all sperm traits, except sperm body length, exhibit a significant negative-linear relationship with sperm transfer success (Table S1). Quadratic regression on sperm traits were significant only for bristle and feeler lengths, also significantly improving the model fit to our sperm transfer success data (Table S1).

Discussion

Sperm size has repeatedly been found to evolve in response to sperm competition across species (Balshine et al., 2001; Godwin et al., 2017; LaMunyon and Ward, 2002; Tourmente et al., 2011). However, our understanding of how sperm size affects sperm competitiveness and hence male reproductive performance within species is limited. In this study, we observed a hump-shaped relationship between post-mating success and sperm size with sperm of intermediate size having on average the highest competitiveness in the simultaneously hermaphroditic flatworm, *M. lignano*. Our results provide support for the hypothesis that sperm competition may, at least in part, favor the evolution of intermediate sperm size (Calhim et al., 2007; Parker, 1993). This finding contributes to our knowledge about how selection shapes sperm and ejaculate traits. In the following, we discuss our findings and point to the strengths and limitations of our study.

Sperm transfer success and sperm size

Non-linear selection towards an optimal sperm size has been suggested mainly by comparative studies (Kleven et al., 2008; Lüpold et al., 2009), inferred from a reduction in sperm size variation in species with higher levels of sperm competition. Lymbery *et al.* (2018) observed a non-linear selection with an intermediate peak on canonical axes of combinations of sperm traits including sperm length, swimming velocity and swimming path linearity in the mussel, *Mytilus galloprovincialis*. Moreover, in *M. lignano*, a recent study explored selection on traits of sperm morphology arising from male reproductive success (and other fitness components) (Marie-Orleach et al., 2023), revealing both hump-shaped and bowl-shaped relationships. Contrary to our results, Marie-Orleach et al. (2023) found significant non-linear relationships with male reproductive success only for sperm body and shaft lengths. No significant relationships were found with sperm transfer efficiency –

417 a fitness component comparable to sperm transfer success assessed in this study.
418 The composite traits that showed non-linear relationships with fitness components
419 were also mainly loaded by either sperm body, shaft or brush length. Marie-Orleach
420 et al. (2023) explored purely phenotypic relationships between sperm traits and male
421 fitness components, which does not allow inference of the evolutionary trajectories.
422 In the present study, we focused on mean estimates of sperm length obtained from
423 inbred lines, which correspond to breeding values of worms within the lines. Thus,
424 this study extends our understanding by demonstrating how sperm competition
425 selects on standing genetic variance on sperm length suggesting that selection on
426 sperm transfer success is favouring the evolution of intermediate sperm length.

427 In our study, we focused on sperm transfer success, aiming to quantify potential
428 selection on sperm size arising from competition to get sperm deposited in the
429 female reproductive tract, which is a crucial step in order to obtain reproductive
430 success through male function. In *M. lignano*, sperm transfer efficiency (i.e., sperm
431 transfer success per successful mating) has been estimated to explain only 27% of
432 the variation in male reproductive success (Marie-Orleach et al., 2021). The fitness
433 component that explained most variation in male reproductive success (37%) in *M.*
434 *lignano* was sperm fertilizing efficiency (i.e., the proportion of sired offspring relative
435 to the number of sperm stored in the female sperm storage organ). However, sperm
436 transfer success has been shown to strongly predict the proportion of sired offspring
437 in *M. lignano* (Marie-Orleach et al., 2016), indicating sperm transfer success is a
438 good estimate for post-mating performance in the species. Nevertheless, we
439 acknowledge that other aspects, such as sperm fertilizing efficiency, may also affect
440 selection on sperm traits and remain unexplored in our study.

441 We used PC1 as a proxy for sperm size in our analysis, revealing a hump-shaped
442 relationship with sperm transfer success. The same analysis using original sperm
443 traits yielded additional insights. All sperm traits, including PC1, except sperm body
444 length exhibited significant negative linear relationships with sperm transfer success.
445 Such negative associations have been suggested to support the theory that sperm
446 competition generates selection for increased number of sperm, which trades-off
447 with sperm size resulting in smaller sperm (Gage and Morrow, 2003; Garcìa-
448 González and Simmons, 2007). However, hump-shaped relationships provided the
449 best fit to sperm transfer success for PC1, total length, bristle length and feeler

length, suggesting selection for intermediate sperm sizes. Selection for increased sperm numbers alone cannot explain this result satisfactorily. Selection for intermediate sperm size, however, was predicted by theoretical models (Parker 1993, Parker and Begon, 1993, Parker et al., 2010). These models assumed that larger sperm have a fertilization advantage over smaller sperm, but increase in the mass of each sperm traded off against other male reproductive expenditures, such as sperm number and mate acquisition. We suspect that the lower success of individuals with particularly long sperm observed in our study may be explained by this trade-off between sperm size and number and that selection for intermediate sperm size may provide a balance between the selective forces for more competitive, but fewer, large sperm and numerous, but less competitive, small sperm.

Different components of sperm morphology may have very different consequences for sperm competitiveness, such that different traits show divergent associations to male reproductive success within species. For example, in the red deer (*Cervus elaphus*), sperm head and tail length were positively associated with sperm swimming velocity and male fertility, but mid-piece length showed a negative correlation with male fertility (Malo et al., 2006). In *M. lignano*, bristle and feeler length have been hypothesized to play active roles in post-copulatory stages, preventing removal of sperm from the female antrum and enabling anchoring of sperm to the cellular valve, respectively (Vizoso et al., 2010). Selection arising from sperm competition can be expected to be stronger on these two traits compared to others, and our results showed a non-linear selection for intermediate sizes of these traits. This may be, at least partly, due to the fact that all sperm traits (except brush length) strongly contribute to PC1 and show strong positive correlations.

Using inbred worms provided us with groups of individuals that are expected to be genetically similar and we expected them to have relatively narrow range of sperm size within lines and significant differences among lines. While we found clear differences among lines, we also observed substantial variation in sperm length within lines. This may in part explain the high variation in sperm transfer success within lines, reiterating the significance of stochastic and non-genetic sources of variation in the measurements of post-mating success (Marie-Orleach et al., 2021). The inbred lines used in the study were derived from four geographically distinct

483 populations of *M. lignano* (Vellnow et al., 2017), and are expected to differ
484 genetically. Therefore, phenotypic differences other than sperm size, arising from
485 focal LM genotype and/or its interaction with the partner DV1 genotype, could have
486 contributed to between-line variation in sperm transfer success in the LM lines.
487 Partner genotypes have been shown to affect reproductive morphology and
488 behaviour in *M. lignano* (Marie-Orleach et al., 2017), but evidence for the influence of
489 partner genotype on fitness components, like mating success and sperm transfer
490 success, is still lacking. Thus, as true for most studies on selection gradients, our
491 results do not allow to establish a causal link between sperm size and sperm transfer
492 success.

493 A shortcoming of our study was a discontinuous gradient in sperm size among LM
494 lines with a particularly big difference between the smallest (LM47) and the second
495 smallest line (LM67). This resulted in the LM 47 being the only representative in the
496 lower end of the spectrum in sperm size. However, we decided not to remove LM47
497 from the analysis because our aim was to cover the largest possible variation in
498 sperm size that could be observed in the outbred BAS1 culture of *M. lignano*, which
499 itself is likely to harbor only a fraction of the variation in sperm size compared to
500 natural populations. Another potential shortcoming of our study is the assumption
501 that sperm length varies little within inbred lines. We did not measure the actual
502 sperm size of the focal LM worms used in the mating experiments, and used instead
503 mean values of a different set of worms measured before the mating experiments to
504 assess differences in sperm size among LM lines. Given that focal worms were
505 taken from inbred lines with relatively narrow range of variation, we expect the actual
506 sperm size of the focal worms would not differ substantially from the values used in
507 our analysis. We also accounted for among-line differences in the variance of sperm
508 length by adding the inverse of standard deviations in sperm size as a weighting
509 factor in our model. Our results remained qualitatively unaffected without the
510 weighting factor in the model.

511 Intra-specific variation in sperm size

512 Intra-specific variation in sperm size is a frequently observed phenomenon across
513 animal taxa (reviewed in Pitnick et al., 2009). We also observed substantial variation
514 in sperm size in both BAS1 culture and among the LM lines. The inbred line identity

515 explained a substantial fraction of sperm trait variation (45% variation in total sperm
516 length). This portion of variation is expected to be attributable to the genetic variance
517 between the LM lines and is an estimate of the total genetic variation (equivalent to
518 the broad-sense heritability) in *M. lignano*. This result is in line with the commonly
519 observed high heritability in sperm size across animals (Simmons and Moore, 2009).
520 In addition, also variation among worms within lines varied, with some lines
521 expressing a much wider range of sperm size than others (Fig. 3). The unusual
522 genetic system of *M. lignano* might contribute to this, with karyotype polymorphisms
523 frequently observed. Worms of this species have been reported to carry 2n = 8, 2n =
524 9 or 2n = 10 chromosomes, with rare cases of other karyotypes (Zadesenets et al.,
525 2016, 2020). Although the karyotypes of the LM lines have not yet been explored,
526 ancestral DV lines were derived from wild populations and karyotype polymorphisms
527 may still persist in these lines and thus contribute to phenotypic variation. Karyotype
528 polymorphism could result in increased variation in mRNA dosages of genes
529 influencing sperm morphology and additional copies of chromosomes may account
530 for more genetic material that has to be physically packed within the sperm shaft.

531 Conclusion

532 We used inbred lines and performed mating trials and *in vivo* sperm counting to
533 assess how genetic variance in sperm size translates into post-mating success in an
534 outcrossing simultaneous hermaphrodite. The hump-shaped association between
535 sperm size and post-mating success suggests that sperm competition may favour an
536 intermediate sperm size. We speculate that such stabilizing selection most likely
537 results from a trade-off between sperm size and sperm number in *M. lignano*, as
538 predicted by theory. Our study is one of the first to explore and present evidence for
539 non-linear selection arising from post-copulatory sexual selection on sperm size and
540 calls for future studies to consider divergent forms of selection on ejaculate traits.

541

542

References

Balshine, S., Leach, B. J., Neat, F., Werner, N. Y., and Montgomerie, R. (2001). Sperm size of African cichlids in relation to sperm competition. *Behavioral Ecology*, *12*(6), 726–731.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.

Bennison, C., Hemmings, N., Slate, J., and Birkhead, T. (2015). Long sperm fertilize more eggs in a bird. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1799), 20141897.

Birkhead, T., and Moller, A. P. (1999). Sperm Competition and Sexual Selection. *Academic Press*, London.

Bolker, B, R Development Core Team (2022). bbmle: Tools for General Maximum Likelihood Estimation_. R package version 1.0.25, <https://CRAN.R-project.org/package=bbmle>.

Briskie, J. V, Montgomerie, R., and Birkhead, T. (1997). The Evolution of Sperm Size in Birds. *Evolution*, *51*(3), 937–945.

Calhim, S., Immler, S., and Birkhead, T. (2007). Postcopulatory Sexual Selection Is Associated with Reduced Variation in Sperm Morphology. *PLoS ONE*, *2*(5), e413.

Cascio Sætre, C. Lo, Johnsen, A., Stensrud, E., and Cramer, E. R. A. (2018). Sperm morphology, sperm motility and paternity success in the bluethroat (Luscinia svecica). *PLoS ONE*, *13*(3), e0192644.

Collet, J., Richardson, D. S., Worley, K., and Pizzari, T. (2012). Sexual selection

566 and the differential effect of polyandry. *Proceedings of the National Academy of*
567 *Sciences*, *109*(22), 8641-8645.

568 **Fitzpatrick, John L., and Stefan Lüpold**. (2014) Sexual selection and the evolution
569 of sperm quality. *Molecular Human Reproduction,* 20(12), 1180-1189.

570 **Gage, M. J. G., and Freckleton, R. P.** (2003). Relative testis size and sperm
571 morphometry across mammals: No evidence for an association between sperm
572 competition and sperm length. *Proceedings of the Royal Society B: Biological*
573 *Sciences*, *270*(1515), 625–632.

574 **Gage, M. J., and Morrow, E. H.** (2003). Experimental evidence for the evolution of
575 numerous, tiny sperm via sperm competition. *Current Biology*, 13(9), 754-757.

576 **García-González, F., and Simmons, L. W.** (2007). Shorter sperm confer higher
577 competitive fertilization success. *Evolution*, 61(4), 816-824.

578 **Godwin, J. L., Vasudeva, R., Michalczyk, Ł., Martin, O. Y., Lumley, A. J.,**
579 **Chapman, T., and Gage, M. J.** (2017). Experimental evolution reveals that
580 sperm competition intensity selects for longer, more costly sperm. *Evolution*
581 *Letters*, *1*(2), 102-113.

582 **Immler, S., and Birkhead, T. R.** (2007). Sperm competition and sperm midpiece
583 size: no consistent pattern in passerine birds. *Proceedings of the Royal Society*
584 *B: Biological Sciences*, 274(1609), 561-568.

585 **Immler, S., Calhim, S., and Birkhead, T. R.** (2008). Increased postcopulatory
586 sexual selection reduces the intramale variation in sperm
587 design. *Evolution*, *62*(6), 1538-1543.

588 **Janicke, T., and Schärer, L.** (2010). Sperm competition affects sex allocation but
589 not sperm morphology in a flatworm. *Behavioral Ecology and Sociobiology*, 64,

1367-1375.

Janicke, T., Sandner, P., and Schärer, L. (2011). Determinants of female fecundity in a simultaneous hermaphrodite: the role of polyandry and food availability. *Evolutionary ecology*, *25*, 203-218.

Janicke, T., Marie-Orleach, L., De Mulder, K., Berezikov, E., Ladurner, P., Vizoso, D. B., and Schärer, L. (2013). Sex allocation adjustment to mating group size in a simultaneous hermaphrodite. *Evolution*, 67(11), 3233-3242.

Kahrl, A. F., Kustra, M. C., Reedy, A. M., Bhave, R. S., Seears, H. A., Warner, D. A., and Cox, R. M. (2021). Selection on sperm count, but not on sperm morphology or velocity, in a wild population of anolis lizards. *Cells*, 10(9), 2369.

Kleven, O., Laskemoen, T., Fossøy, F., Robertson, R. J., and Lifjeld, J. T. (2008). Intraspecific variation in sperm length is negatively related to sperm competition in passerine birds. *Evolution*, *62*(2), 494-499.

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

LaMunyon, C. W., and Ward, S. (2002). Evolution of larger sperm in response to experimentally increased sperm competition in Caenorhabditis elegans. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1496), 1125-1128.

Lenth R. (2023). _emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package version 1.8.6, <https://CRAN.R-project.org/package=emmeans>.

615 **Lüpold, S., Linz, G. M., and Birkhead, T. R.** (2009). Sperm design and variation in
616 the New World blackbirds (Icteridae). *Behavioral Ecology and Sociobiology*, 63,
617 899-909.

618 **Lüpold, S., Pitnick, S., Berben, K. S., Blengini, C. S., Belote, J. M., and Manier,**
619 **M. K.** (2013). Female mediation of competitive fertilization success in Drosophila
620 melanogaster. *Proceedings of the National Academy of Sciences*, 110(26),
621 10693-10698.

622 **Lymbery, R. A., Kennington, W. J., & Evans, J. P.** (2018). Multivariate sexual
623 selection on ejaculate traits under sperm competition. *The American*
624 *Naturalist*, *192*(1), 94-104.

625 **Malo, A. F., Gomendio, M., Garde, J., Lang-lenton, B., Soler, A. J., and Roldan,**
626 **E. R. S.** (2006). Sperm design and sperm function. *Biology Letters*, 2(2), 246–
627 249.

628 **Marie-Orleach, L., Janicke, T., Vizoso, D. B., David, P., and Schärer, L.** (2016).
629 Quantifying episodes of sexual selection: Insights from a transparent worm with
630 fluorescent sperm. *Evolution*, *70*(2), 314–328.

631 **Marie-Orleach, L., Vogt-Burri, N., Mouginot, P., Schlatter, A., Vizoso, D. B.,**
632 **Bailey, N. W., and Schärer, L.** (2017). Indirect genetic effects and sexual
633 conflicts: Partner genotype influences multiple morphological and behavioral
634 reproductive traits in a flatworm. *Evolution*, *71*(5), 1232–1245.

635 **Marie-Orleach, L., Vellnow, N., and Schärer, L.** (2021). The repeatable opportunity
636 for selection differs between pre- and postcopulatory fitness components.
637 *Evolution Letters*, *5*(1), 101-114.

638 **Marie-Orleach, L., Hall, M. D., and Schärer, L.** (2023). Contrasting the form and
639 strength of pre- and postcopulatory sexual selection in a transparent worm with

fluorescent sperm. *Biorxiv*, 2023-03.

Miller, G. T., and Pitnick, S. (2002). Sperm-Female Coevolution in Drosophila. *Science*, *298*(5596), 1230–1233.

Minder, A. M., Hosken, D. J., and Ward, P. I. (2005). Co-evolution of male and female reproductive characters across the Scathophagidae (Diptera). *Journal of Evolutionary Biology*, *18*(1), 60–69.

Morrow, E. H., and Gage, M. J. G. (2000). The evolution of sperm length in moths. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1440), 307–313.

Morrow, E. H., and Gage, M. J. G. (2001). Sperm competition experiments between lines of crickets producing different sperm lengths. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1482), 2281–2286.

Parker, G. A. (1993). Sperm competition games: Sperm size and sperm number under adult control. *Proceedings of the Royal Society B: Biological Sciences*, *253*(1338), 245–254.

Parker, G. A., and Begon, M. E. (1993). Sperm competition games: sperm size and number under gametic control. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 253(1338), 255-262.

Parker, G. A., Immler, S., Pitnick, S., and Birkhead, T. (2010). Sperm competition games: Sperm size (mass) and number under raffle and displacement, and the evolution of P2. *Journal of Theoretical Biology*, *264*(3), 1003–1023.

Pélissié, B., Jarne, P., Sarda, V., and David, P. (2014). Disentangling precopulatory and postcopulatory sexual selection in polyandrous species. *Evolution*, *68*(5), 1320–1331.

Pitnick, S., Hosken, D. J., and Birkhead, T. (2009). Sperm morphological diversity, In *Sperm biology* (pp. 69-149). Academic Press, London.

Pitnick, S., Markow, T., and Spicer, G. S. (1999). Evolution of Multiple Kinds of Female Sperm-Storage Organs in Drosophila. *Evolution*, *53*(6), 1804–1822.

Pitnick, S., Miller, G. T., Schneider, K., and Markow, T. A. (2003). Ejaculate-female coevolution in *Drosophila mojavensis*. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1523), 1507–1512.

Pitnick, S., Wolfner, M. F., and Suarez, S. S. (2009). Ejaculate–female and sperm–female interactions. In *Sperm biology* (pp. 247-304). Academic Press, London.

Pizzari, T., and Parker, G. A. (2009). Sperm competition and sperm phenotype. In *Sperm biology* (pp. 207–245). Academic Press, London.

Radwan, J. (1996). Intraspecific variation in sperm competition success in the bulb mite: a role for sperm size. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1372), 855-859.

Schärer, L., Brand, J. N., Singh, P., Zadesenets, K. S., Stelzer, C. P., and Viktorin, G. (2020). A phylogenetically informed search for an alternative Macrostomum model species, with notes on taxonomy, mating behavior, karyology, and genome size. *Journal of Zoological Systematics and Evolutionary Research*, *58*(1), 41–65.

Schärer, L., Joss, G., and Sandner, P. (2004). Mating behaviour of the marine turbellarian Macrostomum sp.: These worms suck. *Marine Biology*, *145*(2), 373–380.

Simmons, L. W. (2001). Sperm competition and its evolutionary consequences in

686 the insects. *Princeton University Press.*

687 **Simmons, L. W., and Moore, A. J.** (2009). Evolutionary quantitative genetics of
688 sperm. In *Sperm biology* (pp. 405-434). Academic Press, London.

689 **Snook, R.R.** (2005). Sperm in competition: not playing by the numbers. *Trends in*
690 *ecology & evolution*, 20(1), 46-53.

691 **Stoffel, M. A., Nakagawa, S., and Schielzeth, H.** (2017). rptR: repeatability
692 estimation and variance decomposition by generalized linear mixed-effects
693 models. *Methods in Ecology and Evolution*, *8*(11), 1639–1644.

694 **Tourmente, M., Gomendio, M., and Roldan, E. R.** (2011). Sperm competition and
695 the evolution of sperm design in mammals. *BMC evolutionary biology*, *11*(1), 1-
696 10.

697 **Vellnow, N., Vizoso, D. B., Viktorin, G., and Schärer, L.** (2017). No evidence for
698 strong cytonuclear conflict over sex allocation in a simultaneously
699 hermaphroditic flatworm. *BMC Evolutionary Biology, 17, 1-14.*

700 **Vizoso, D. B., Rieger, G., and Schärer, L.** (2010). Goings-on inside a worm:
701 Functional hypotheses derived from sexual conflict thinking. *Biological Journal*
702 *of the Linnean Society*, *99*(2), 370–383.

703 **Willems, M., Leroux, F., Claeys, M., Boone, M., Mouton, S., Artois, T., and**
704 **Borgonie, G.** (2009). Ontogeny of the complex sperm in the macrostomid
705 flatworm Macrostomum lignano (Macrostomorpha, Rhabditophora). *Journal of*
706 *Morphology*, 270(2), 162-174.

707 **Zadesenets, K. S., Vizoso, D. B., Schlatter, A., Konopatskaia, I. D., Berezikov,**
708 **E., Schärer, L., and Rubtsov, N. B.** (2016). Evidence for karyotype
709 polymorphism in the free-living flatworm, macrostomum lignano, a model

organism for evolutionary and developmental biology. *PLoS ONE*, *11*(10), 1–24.

Zadesenets, K. S., Jetybayev, I. Y., Schärer, L., and Rubtsov, N. B. (2020). Genome and karyotype reorganization after whole genome duplication in free-living flatworms of the genus macrostomum. *International Journal of Molecular Sciences*, *21*(2).

717 # Tables and Figures:

718 **Table 1.** Descriptive statistics and variation in sperm traits in the BAS1 culture
719 (based on all the sperm cells measured, *n* = 591) and Kruskal-Wallis analysis of
720 variance testing for variation in sperm morphology between individuals (n = 39). (SD
721 - standard deviation; CV - Coefficient of Variation)

Sperm trait	Mean **±** SD (μm)	Repeatability	CV (%)	X^2	df	*P*
Total length	79.4 ± 3.51	0.962	4.31	347.5	38	**$<2.2e^{-16}$**
Feeler length	16.92 ± 1.03	0.671	6.10	202.54	38	**$<2.2e^{-16}$**
Bristle length	10.91 ± 0.6	0.911	5.52	231.42	38	**$<2.2e^{-16}$**
Shaft length	44.61 ± 2.59	0.941	5.79	313.22	38	**$<2.2e^{-16}$**
Brush length	3.31 ± 1.15	0.398	34.38	188.46	38	**$<2.2e^{-16}$**

722

723 **Table 2.** Pearson correlation matrix showing correlation coefficients between the
724 individual means of all sperm morphological traits (*n* = 39). Bold values indicate
725 significant correlations. Note that sperm body length was not measured in BAS1
726 culture worms.

727

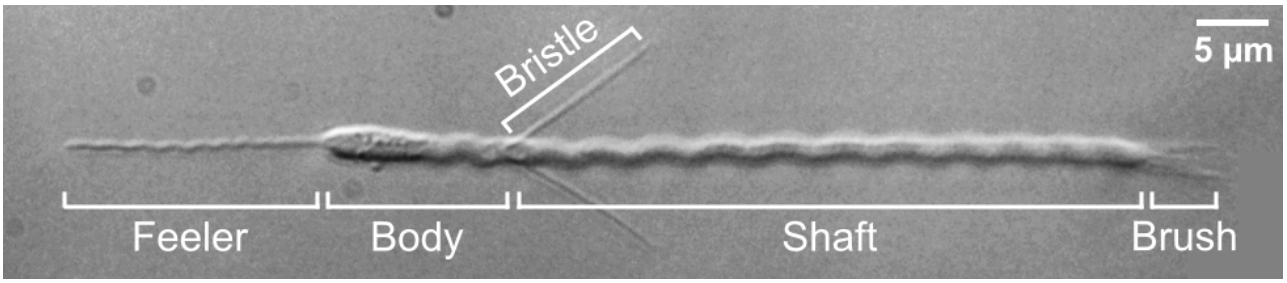
Sperm trait (μm)	Total	Feeler	Bristle	Shaft	Brush
Total length	1				
Feeler length	**0.717**	1			
Bristle length	**0.666**	**0.516**	1		
Shaft length	**0.853**	**0.500**	**0.655**	1	
Brush length	0.066	0.026	-0.111	-0.272	1

728
729
731
732

Table 3. Results from the principal component analysis of the 5 sperm morphological traits. Variance explained and loadings in PC1 and PC2 are shown.

	PC1	PC2
Eigenvalue	3.415	0.932
% variance explained	68.31	18.59
Pearson's ***r*** for correlation with total sperm length	0.988	0.105
Sperm trait loadings		
Feeler length (μm)	0.496	0.12
Body length (μm)	0.495	0.037
Bristle length (μm)	0.494	0.004
Shaft length (μm)	0.479	0.219
Brush length (μm)	0.191	-0.968

745
746
747
748



749 **Figure 1**. Morphology of a mature sperm cell in *Macrostomum lignano*.

750

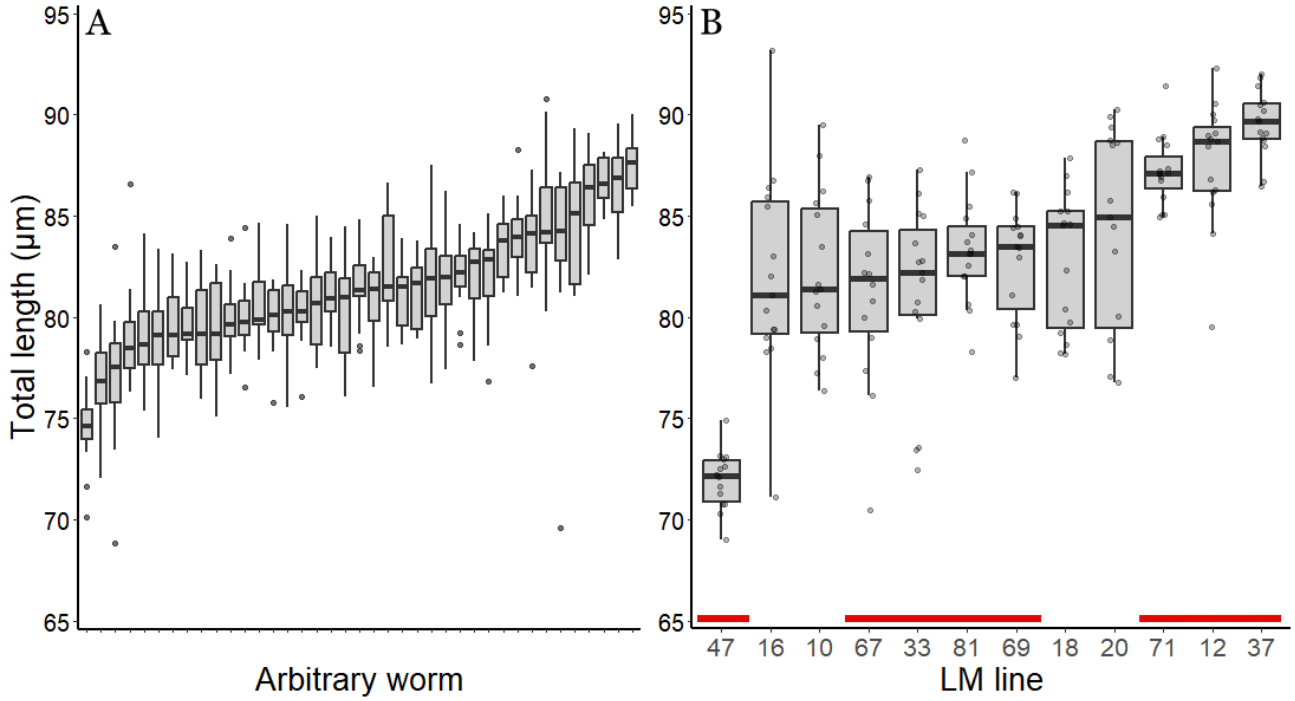


Figure 2. Between-individual and among-line variation in total sperm length in *Macrostomum lignano*. (A) Between-individual variation in total sperm length in the outbred BAS1 population. (N = 39). Individuals are arranged in increasing order of their median total sperm length (see Table 1 for statistical tests). (B) Variation in individual means of total sperm length among the 12 LM lines. LM lines are arranged in increasing order of medians. Jittered dots in B show individuals mean values of 14-15 worms measured for each line. About 15 sperm cells were measured for each worm in both BAS1 and LM lines. Red lines indicate the 8 LM lines chosen for the sperm competition experiment.

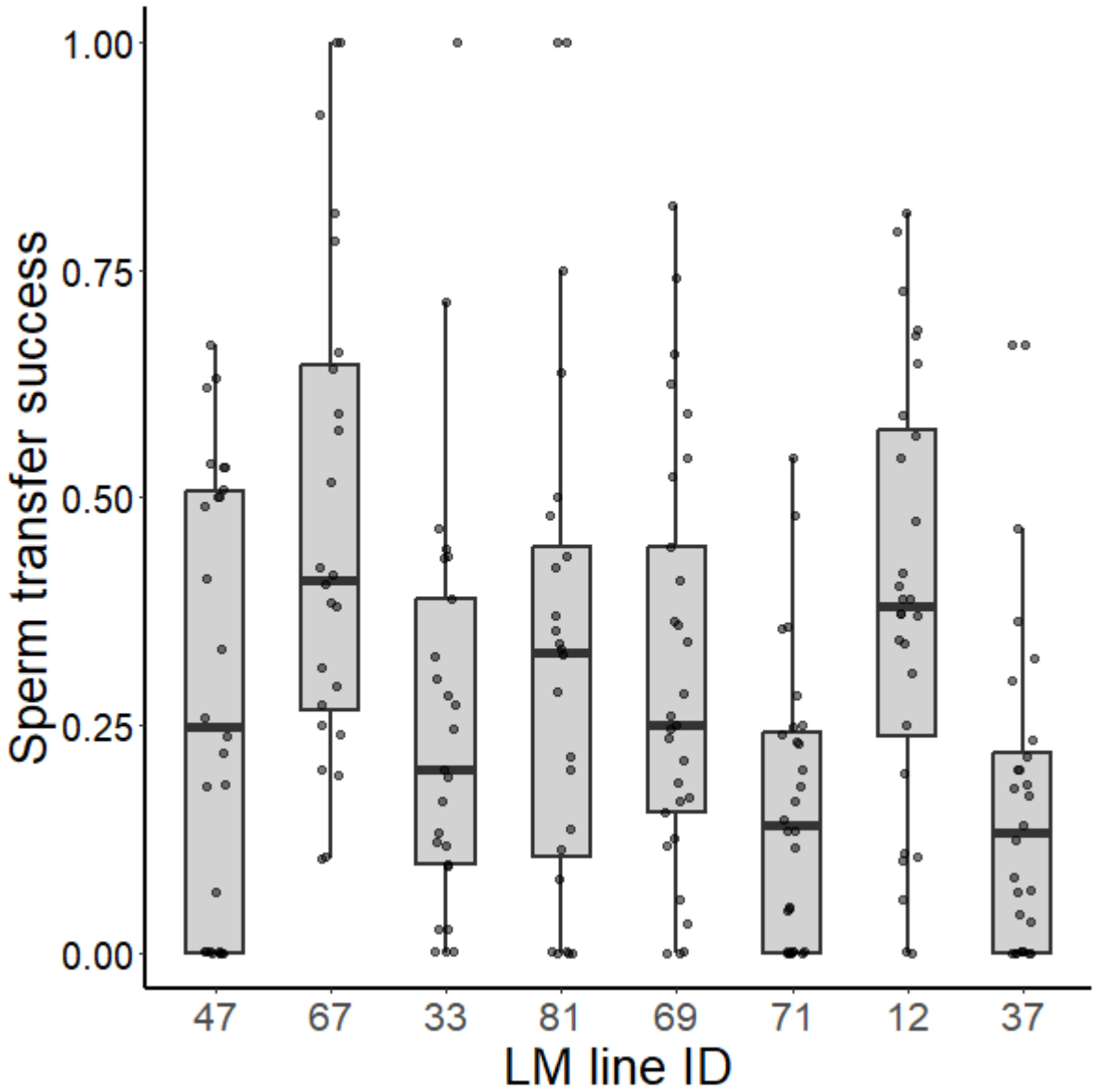


Figure 3. Comparison of sperm transfer success between LM lines inferred from competition experiments in which one focal LM worms competed against DV1 worms. The boxplots for the LM lines are arranged in increasing order of median sperm lengths.

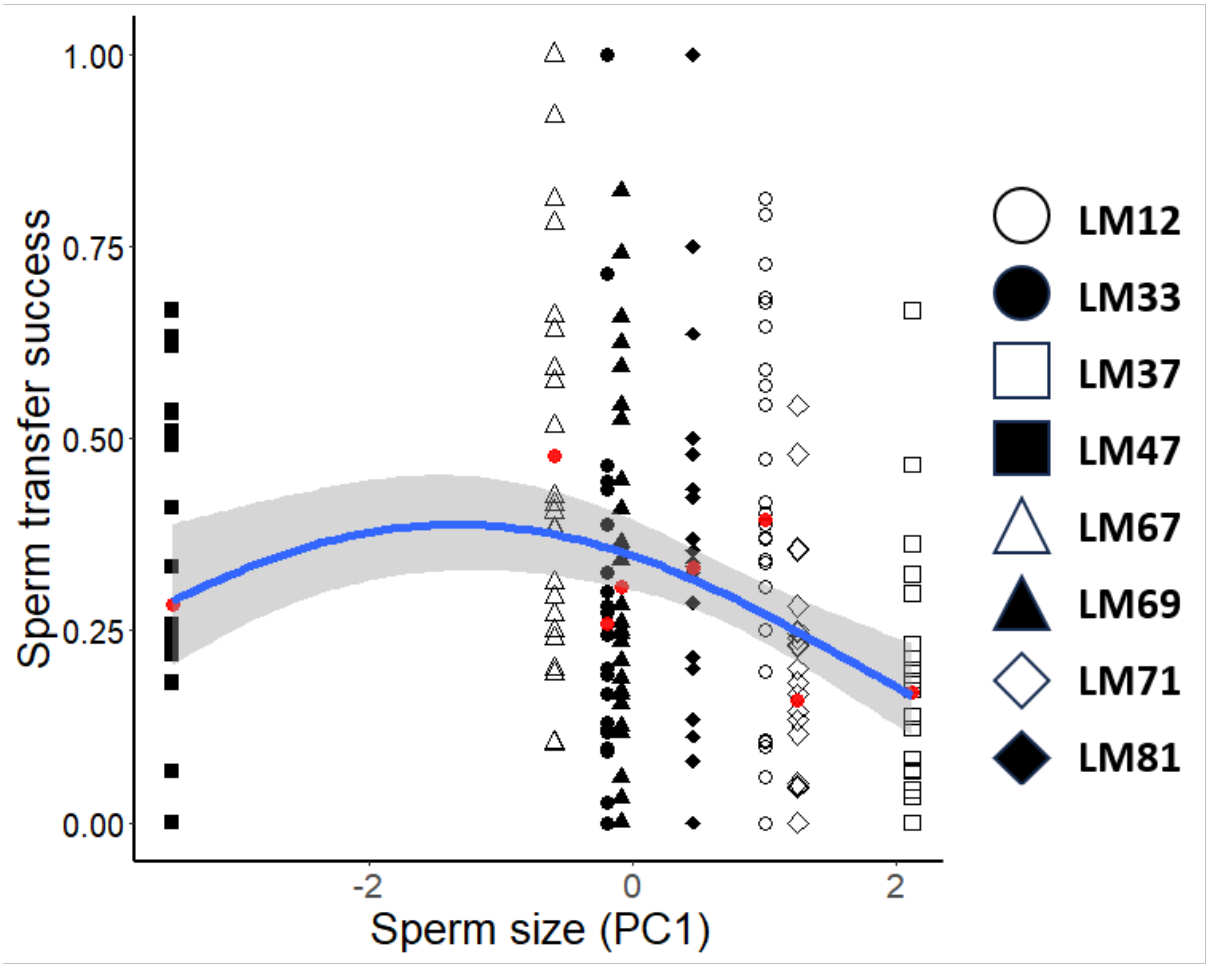


Figure 4. Relationship between sperm size (PC1) and sperm transfer success. Symbols indicate different LM lines (see legend). Red dots indicate mean sperm transfer success in each line. Solid blue line shows the fit of the quadratic model, and grey regions indicate the 95% confidence intervals.