Self-fertilization, but not mating strategy, predicts the evolution of sex allocation plasticity in a hermaphroditic flatworm genus

Pragya Singh[1*] and Lukas Schärer[1]

[1]University of Basel, Department of Environmental Sciences, Zoological Institute, Basel, Switzerland

[*]Corresponding Author:

Email: pragya.singh42019@gmail.com

ORCiD: https://orcid.org/0000-0002-7411-3206

Present address: Chemical Ecology, Bielefeld University, Universitätsstr. 25, 33615 Bielefeld, Germany

Running title: Evolution of sex allocation plasticity

Author Contributions: PS and LS designed the study, analysed the data and wrote the manuscript. PS performed the experiment. All authors read and approved the final manuscript.

Acknowledgements:

We would like to thank Christian Felber for help with the morphological measurements of *Macrostomum* sp. 22, Lucas Marie-Orleach, Steven Ramm and Tim Janicke for providing access to their datasets, Dita Vizoso and Kaja Wasik for help with collecting specimens of *Macrostomum* sp. 22, Jeremias Brand, Axel Wiberg and Nikolas Vellnow for helpful discussions, and Gudrun Viktorin, Daniel Lüscher, Jürgen Hottinger, Lukas Zimmerman and Urs Stiefel for technical support. This research was supported by grants 31003A_162543 and 310030_184916 of the Swiss National Science Foundation (SNSF) to LS.

Competing interests: The authors hereby declare no conflicts of interest.

Ethical note: All animal experimentation was carried out in accordance to Swiss legal and ethical standards.

Data accessibility: All data in this manuscript will be deposited online (xx).

Abstract

Sex allocation (SA) theory in simultaneous hermaphrodites predicts that optimal SA is influenced by local sperm competition (LSC), which occurs when related sperm compete to fertilize a given set of eggs. Different factors, including the mating strategy and the ability to self-fertilize, are predicted to affect LSC and hence the optimal SA. Moreover, since the LSC experienced by an individual can vary temporally and spatially, this can favour the evolution of SA plasticity. Here, using seven species of the free-living flatworm genus *Macrostomum,* we document sizable interspecific variation in SA, but neither their mating strategy nor their ability to self-fertilize significantly predicted SA among these species. Since we also found considerable interspecific variation in SA plasticity, we further estimated standardized effect sizes for plasticity in response to i) the presence of mating partners (i.e. in isolation vs. with partners) and ii) the strength of LSC (i.e. in small vs. large groups). We found that self-fertilization predicted SA plasticity with respect to the presence of mating partners, with plasticity being lower for self-fertilizing species. Finally, we showed that interspecific variation in SA is higher than intraspecific variation due to SA plasticity. Our study suggests that both SA and SA plasticity are evolutionarily labile, with self-fertilization predicting the latter in *Macrostomum*.

Keywords

phenotypic plasticity, sperm competition, self-fertilization, reciprocal mating, hypodermic insemination, local mate competition

Introduction

Sex allocation (SA) theory in simultaneous hermaphrodites predicts the optimal allocation of resources towards the male versus female function (Charnov 1982). Local sperm competition (LSC), which can be thought of as the inverse of sperm competition (Parker 1970, 1998), occurs when related sperm (usually from the same individual) compete for access to a given set of eggs, and it can favour a female-biased optimal SA in simultaneous hermaphrodites (Schärer 2009; Schärer and Pen 2013). LSC can be considered analogous to local mate competition (Hamilton 1967), i.e. the competition between related males for access to mates, which can lead to a female-biased sex ratio in species with separate sexes. However, LSC does not require mates to be related or a spatial population structure to be present, since the local competition occurs among competing sperm, not competing males. An example of a factor that can affect LSC, and hence the optimal SA, is the mating group size (Charnov 1980, 1982), denoted as K+1, where K is the number of sperm donors from which a sperm recipient receives sperm at the time its eggs are fertilized. Specifically, Charnov's mating group size model predicts that the optimal proportion of resources an individual allocates to the male function should increase with the number of sperm donors that are present in a local mating group.

The link between optimal SA and LSC can be visualised in terms of male fitness gain curves, which describe how much fitness is gained through the male function per unit resource investment (Charnov 1982; Lloyd 1984; Schärer 2009). Under monogamy, LSC is necessarily very high, since the competing sperm are maximally related to each other, resulting in the male fitness gain curve saturating quickly. Thus, any investment into the production of more sperm than are required to fertilize the partner's eggs is a waste of resources, as the related sperm simply compete amongst themselves for fertilizations. Instead, these resources could be more profitably invested into the individual's own female function (Charnov 1982), since the

85 female fitness gain curve is often assumed to be linear (Rademaker and de Jong 1999;
86 Campbell 2000; Schärer 2009), and may sometimes even be accelerating (Rosas and
87 Domínguez 2009). Consequently, under monogamy, a simultaneous hermaphrodite is
88 expected to have a highly female-biased SA (Charnov 1980, 1982). But as the mating group
89 size increases—i.e. with more competitors contributing sperm to a given recipient—an
90 individual's sperm are more and more competing with unrelated sperm. So, it now pays off to
91 invest into the male function to gain a greater share of the fertilizations, and the male fitness
92 gain curve is predicted to linearize, with a subsequent shift towards a more equal SA
93 (Charnov 1982; Schärer 2009; Schärer and Pen 2013).

94 An interesting question that arises then is whether and how interspecific variation in
95 reproductive biology, including different mating strategies or the ability to self-fertilize, could
96 influence the strength of LSC and hence the optimal SA. For example, hermaphrodites often
97 evolve different mating strategies in response to sexual conflict over mating roles, i.e. over
98 whether to mate as a sperm donor and/or as a sperm recipient (Charnov 1979; Michiels 1998;
99 Schärer et al. 2015). One such mating strategy is reciprocal mating (also called reciprocal
100 copulation), in which both partners mate in both mating roles, and thus donate and receive
101 sperm simultaneously. This strategy can lead to individuals being quite willing to engage in
102 mating in order to get an opportunity to donate sperm, and this general willingness likely
103 results in reduced precopulatory sexual selection. The ensuing higher mating rates could in
104 turn result in increased sperm competition (and thus a decrease in LSC) and more intense
105 postcopulatory sexual selection. Interestingly, however, the presence of certain postcopulatory
106 processes could also increase LSC (Schärer 2009). Indeed, theoretical studies have predicted
107 that different postcopulatory sexual selection processes, such as sperm displacement, sperm
108 digestion, and cryptic female choice, can lead to the removal of sperm of one or multiple
109 donors from competition, resulting in changes in LSC and hence the predicted optimal SA

110 (Charnov 1996; Greeff and Michiels 1999; Pen and Weissing 1999; van Velzen et al. 2009;
111 Schärer and Pen 2013).

112 In addition, it has been suggested that postcopulatory sexual selection and sexual conflict
113 among simultaneous hermaphrodites can also favour the evolution of mating strategies that
114 allow donors to bypass these postcopulatory processes and fertilize the eggs more directly
115 (Charnov 1979; Michiels 1998). One such mating strategy is forced unilateral insemination,
116 in which one partner mates only in the male role and donates sperm, e.g. via traumatic or
117 hypodermic insemination (Charnov 1979; Lange et al. 2013; Reinhardt et al. 2015), while the
118 other partner only mates in the female role, likely often against its interests. It is currently
119 difficult to predict the direction and magnitude of changes in LSC resulting from such a shift
120 in mating strategy—and the effects that such shifts may thus have on optimal SA—since we
121 still know relatively little about the particular postcopulatory processes involved in these
122 mating strategies. In species with reciprocal mating there may be some control over who or
123 how many partners an individual receives and stores sperm from, and there could be strategies
124 that lead to the (conceivably selective) removal of received sperm (van Velzen et al. 2009),
125 possibly leading to increased LSC. In species with hypodermic insemination, however, such
126 control might be more limited, since hypodermic sperm might function in a less easily
127 controlled environment, with sperm from multiple donors thus mixing more randomly in a
128 fair-raffle-type sperm competition. This would be expected to lower the strength of LSC
129 (Schärer and Janicke 2009).

130 In addition to the contrast between reciprocal copulation and hypodermic insemination,
131 species that employ self-fertilization might also differ in optimal SA from species that
132 primarily outcross, because self-fertilization greatly increases the strength of LSC. Thus, self-
133 fertilization leads to diminishing fitness returns from investing in the male function, favouring
134 the evolution of a more female-biased optimal SA compared to outcrossing species

135 (Charlesworth and Charlesworth 1981; Charnov 1982, 1987), which is supported by empirical
136 work in both plants and animals (Lemen 1980; Schoen 1982; McKone 1987; Johnston et al.
137 1998; Lemen 1980). In general, we would expect any process that has an effect on LSC to
138 affect the shape of the male fitness gain curve. And this in turn leads to different predictions
139 for the optimal SA (Schärer 2009), as shown in Figure 1.

140 While most SA models investigate the effect of LSC on SA over evolutionary timescales,
141 LSC can also vary temporally and spatially within an individual's lifetime. In hermaphrodites,
142 altering the SA in response to the current conditions can influence the immediate reproductive
143 success of an individual, favouring the evolution of SA plasticity. Indeed, plastic SA has been
144 suggested to be one of the advantages of hermaphroditism over separate-sex species (Charnov
145 1982; Michiels 1998, 1999). However, for species that do not often experience variation in
146 LSC during their lifetime, we may not expect high levels of plasticity, particularly if there are
147 costs to plasticity (DeWitt et al. 1998; Auld et al. 2010; Siljestam and Östman 2017).
148 Interestingly, while SA plasticity has now been documented in many hermaphroditic species
149 (Schärer 2009), its evolution is still comparatively poorly understood, with few studies having
150 investigated variation in SA plasticity across animal species in a controlled experimental
151 context (Schleicherová et al. 2014). And finally, as SA can vary both between-species (e.g.
152 due to differences in mating strategy over evolutionary timescales) and within-species (e.g.
153 due to phenotypic plasticity), it is interesting to examine the relative magnitude of the
154 interspecific versus intraspecific variation in SA.

155 An excellent model system for testing the effect of the mating strategy and the ability to self-
156 fertilize on SA and SA plasticity is the free-living flatworm genus *Macrostomum*
157 (Macrostomorpha, Platyhelminthes). This genus contains many species exhibiting one of at
158 least two different mating strategies; one involving reciprocal mating and the other
159 hypodermic insemination (Brand et al. 2021a; Schärer et al. 2011; Singh et al. in prep.). In

160 reciprocally mating species, a facultative postcopulatory suck behaviour has been observed, in
161 which an individual places its pharynx on top of its female antrum (the sperm-receiving and
162 egg-laying organ) and appears to suck, and this has been hypothesised to remove received
163 ejaculate components (Singh et al. in prep.; Schärer et al. 2004a; Vizoso et al. 2010). No such
164 postcopulatory suck behaviour has been documented in hypodermically inseminating species
165 (Singh et al. in prep.), which presumably exhibit forced unilateral mating, with sperm being
166 hypodermically injected into the partner via a needle-like male copulatory organ (Schärer et
167 al. 2011; Brand et al. 2021a). Moreover, both reciprocally mating and hypodermically
168 inseminating species possess a female antrum, but in the latter this organ has a simple
169 morphology and presumably serves only for egg-laying, while in the former it is usually more
170 complex and used both for egg-laying and for receiving sperm from the partner (Schärer et al.
171 2011; Brand et al. 2021a). Interestingly, hypodermic insemination is also associated with a
172 suite of morphological traits that potentially facilitate self-fertilization (Ramm et al. 2012,
173 2015; Giannakara and Ramm 2017), although self-fertilization has recently also been
174 documented in at least one reciprocally mating species, *M. mirumnovem* (Singh et al. 2020b).

175 Here, we collected literature data on SA estimates from a range of experiments performed in
176 six *Macrostomum* species, and we present additional data from a new and currently still
177 undescribed species, *Macrostomum* sp. 22 (see Supplementary Information S1). In all studies
178 the experimental design generally consisted in raising worms from juveniles in three different
179 group sizes (isolated, pairs or octets) and then obtaining estimates of their SA once the worms
180 had reached maturity. To facilitate SA plasticity comparisons between species, we calculated
181 standardized effect sizes for SA plasticity in response to i) the presence of mating partners
182 (i.e. comparing isolated worms vs. worms with partners) and ii) the strength of LSC (i.e.
183 comparing paired worms vs. octet worms). Using SA estimates and these SA plasticity effect
184 sizes, we then examined if the mating strategy and ability to self-fertilize predicted SA or SA

185 plasticity, while accounting for the phylogenetic interrelationships. Lastly, considering the
186 interspecific SA differences and the extensive SA plasticity shown by some of the species, we
187 examined how much of the variation in SA occurs between-species versus within-species, by
188 partitioning the SA variance into its interspecific and intraspecific components.

189

190 *Materials and Methods*

191 *Study species*

192 To examine how SA and SA plasticity evolves across the *Macrostomum* genus, we gathered
193 data on SA and SA plasticity in seven *Macrostomum* species (we use the "Genus species
194 Author, Year" format to refer to binomials and include citations to the relevant taxonomic
195 works). Three species are hypodermically inseminating, namely *M. pusillum* Ax, 1951 (Ax
196 1951), *M. hystrix* Örsted, 1843 sensu Luther, 1905 (Örsted 1843; Luther 1905; Schärer et al.
197 2011), and the currently undescribed species *Macrostomum* sp. 22, with the SA data reported
198 in, respectively, Giannakara and Ramm (2017), Winkler and Ramm (2018), and
199 Supplementary Information S1. An additional four species are reciprocally mating, namely
200 *M. janickei* Schärer, 2019, *M. cliftonense* Schärer and Brand, 2019, and *M. mirumnovem*
201 Schärer and Brand, 2019 (Schärer et al. 2020; Zhang et al. 2021), with the SA data reported in
202 Singh et al. (2020b), plus multiple studies in *M. lignano* Ladurner, Schärer, Salvenmoser, and
203 Rieger, 2005 (Ladurner et al. 2005), with the SA data reported in several studies (see below).
204 Details of all the SA studies are given in Supplementary Table S1.

205 The species *M. mirumnovem* (Singh et al. 2020b) and all three hypodermically inseminating
206 species (Ramm et al. 2012, 2015; Giannakara and Ramm 2017, this study) can self-fertilize,
207 and were classified accordingly. Moreover, we classified species as either reciprocally mating
208 or hypodermically inseminating using both behavioural and morphological data. Behavioural

209 data showed the presence of both reciprocal mating and the postcopulatory suck behaviour in
210 *M. lignano, M. janickei*, *M. cliftonense*, and *M. mirumnovem*, while neither of these
211 behaviours were observed in the hypodermically inseminating *M. pusillum*, *M. hystrix*, and
212 *Macrostomum* sp. 22 (Singh et al., in prep.; Schärer et al. 2004a; Schärer et al. 2020). In
213 addition, a classification of species using only morphological traits, which are known to be
214 correlated with the mating strategy in *Macrostomum* (Singh et al. in prep.; Schärer et al.
215 2011), corroborated our above classification based on behaviour. Thus, species could also be
216 classified morphologically as reciprocally mating or hypodermically inseminating depending
217 on their male and female genital morphology, the sperm morphology, and the location of
218 (received) allosperm (for details see Brand et al. 2021a).

219 ***SA estimates and SA plasticity effect sizes across species***

220 For most species the experimental procedure was similar to that described for
221 *Macrostomum* sp. 22 (Supplementary Information S1), i.e., worms were raised from juveniles
222 in three different group sizes (isolated, pairs or octets) until adulthood, when testis and ovary
223 size was measured in one randomly chosen worm per replicate as a proxy for an individual's
224 male and female allocation. The SA was then estimated as testis size/(testis size+ovary size).
225 In some experiments, the density of worms was varied independently of group size by using
226 two enclosure sizes (small and large) (for details and sample sizes see Supplementary Table
227 S1). We expect that the experiments are nevertheless comparable, since all experiments where
228 density was included as a factor revealed that it did not have a significant effect on the
229 estimate of SA.

230 For each experiment, we calculated the mean SA and respective standard deviation across the
231 group sizes, weighted by the sample size for each group size, using the R package 'Hmisc'
232 (Harrell 2020). To facilitate interspecies comparisons of the SA plasticity estimates, we
233 calculated standardized effect sizes and their standard deviation, using Cohen's d (Cohen

234 1988) with Hedges correction for small sample size (Hedges 1981), including confidence
235 intervals (Howell 2011) of the effect sizes using the R package 'effsize' (Torchiano 2017).
236 For each species, we calculated an effect size for SA plasticity to assess the effect of the
237 presence of mating partners (i.e. comparing isolated worms vs. worms with partners) and the
238 strength of LSC (i.e. comparing worms in pairs vs. worms in octets). Pairs represent a
239 condition with high LSC whereas octets represent a condition with low LSC. Studies in
240 *M. lignano* have shown that in larger groups individuals mate with many of the available
241 partners (Janicke and Schärer 2009a; Janicke et al. 2013), that there is sperm displacement
242 (Marie-Orleach et al. 2014) and that paternity is usually shared (Marie-Orleach et al. 2016;
243 Vellnow et al. 2018).

244 For *M. lignano*, we collected datasets from multiple published experiments (up to 25th
245 November 2019), and calculated SA and SA plasticity effect sizes for each of them
246 separately. In total, we found two and six studies, respectively, where we could extract the
247 effects of the presence of mating partners (Schärer and Janicke 2009; Ramm et al. 2019) and
248 the strength of LSC (Schärer and Ladurner 2003; Schärer et al. 2004b; Janicke and Schärer
249 2009b; Janicke et al. 2013; Marie-Orleach et al. 2014; Ramm et al. 2019). For Schärer and
250 Ladurner (2003), we calculated SA from the available data, since that study only reported
251 values for testis and ovary size. We excluded one previous study (Janicke and Schärer 2010)
252 from our analysis, since the worms used in that study came from the same experiment as
253 Janicke and Schärer (2009), and so these did not represent fully independent datasets (we
254 used the latter study since it included more replicates). In addition, in Marie-Orleach et al.
255 (2014) we combined the data from both inbred lines (HUB1 and DV1), since they did not
256 differ in any of the traits measured. Also, note that in Schärer and Janicke (2009), there are
257 only isolated and paired worms and hence the estimated effect size for the presence of mating
258 partners does not include the effect of octets. This should not have a large effect on the

259 calculated effect size, since in most cases the difference between the isolated and paired
260 treatments is much larger than the difference between paired and octet treatments (Figure 2).

261 For *M. lignano*, a weighted mean across experiments was used to estimate the SA, and SA
262 plasticity effect sizes (Turner and Bernard 2006), for both the presence of mating partners and
263 the strength of LSC. For SA, we took the mean of the SA across experiments, weighted by the
264 sample size of each experiment. To weight each effect size, it was multiplied by the inverse of
265 its variance (inverse variance weight, IVW), which allowed us to take into consideration the
266 differences in sample size across experiments, namely:

267 $$Hedges's\ g^* = \frac{\sum_{i=1}^{n} Hedges's\ g_i \times IVW_{Hedges's\ g_i}}{\sum_{i=1}^{n} IVW_{Hedges's\ g_i}}$$

268 And its standard error was calculated as:

269 $$SE_{g^*} = \sqrt{\frac{1}{\sum_{i=1}^{n} IVW_{Hedges's\ g_i}}}$$

270 ***Evolution of SA and SA plasticity***

271 For the phylogenetically corrected analyses of the SA and SA plasticity effect sizes, we used a
272 trimmed (ape package, Paradis et al. 2004) ultrametric phylogenomic tree constructed using
273 an alignment containing 385 genes (and 74,175 variable sites) from 98 transcriptome-
274 sequenced species, and calculated with a maximum likelihood approach (called H-IQ-TREE
275 in Brand et al. 2021b). All bipartitions in this trimmed tree had maximal support (Figure 3),
276 and trees constructed using a range of phylogenetic approaches had identical topologies with
277 respect to the seven species studied here (Brand et al. 2021b), thus obviating the need to
278 account for phylogenetic uncertainty in the following analyses.

279 Recent studies have shown that there are at least nine independent origins of hypodermic
280 insemination in *Macrostomum* (Brand et al. 2021a). Thus *M. hystrix* clearly represents an

281 independent origin of hypodermic insemination in an otherwise largely reciprocally mating
282 clade, when compared to the other two hypodermically inseminating species, *M. pusillum* and
283 *Macrostomum* sp. 22, which belong to a uniformly hypodermically mating clade (as indicated
284 by the inferred ancestral states of the internal nodes in the phylogeny in Figure 3). Although a
285 similarly unequivocal analysis for multiple independent origins of self-fertilization is
286 currently not yet available, self-fertilization was not readily observed in multiple reciprocally
287 mating species that we also cultured in the laboratory, including *M. spirale* Ax, 1956 (Ax
288 1956; Schärer et al. 2011), *M. axi* Papi, 1959 (Papi 1959), *M. clavituba* Ax, 2008 (Ax 2008),
289 and *M. poznaniense* Kolasa, 1973 (Kolasa 1973), which were held in isolation for extended
290 periods of time for behavioural experiments (P. Singh, personal observations). All of these
291 species are members of the reciprocally mating clade and they fall outside of the subclade
292 containing both *M. mirumnovem* and *M. hystrix* (Supplementary Figure S1). Thus, if we
293 assume that the absence of self-fertilization is the most likely ancestral state in the entire
294 reciprocally mating clade, then *M. hystrix* and *M. mirumnovem* represent at least one, or
295 possibly two, independent origins of self-fertilization relative to that observed in *M. pusillum*
296 and *Macrostomum* sp. 22 (Supplementary Figure S1).

297 As a preliminary analysis, we next examined if there was an association between self-
298 fertilization and the mating strategy, since such an association would render these non-
299 independent predictor variables. For this, we used the DISCRETE model in BayesTraits
300 V.3.0.1 to test for correlated evolution between the presence of self-fertilization and the
301 presence of hypodermic insemination using the Reversible Jump Markov Chain Monte Carlo
302 (RJ MCMC) approach (Pagel 1994; Pagel and Meade 2006; Meade and Pagel 2016). To test
303 support for correlated evolution, we compared the marginal likelihood of a dependent model,
304 in which the presence of self-fertilization depends on the presence of hypodermic
305 insemination, with an independent model, in which hypodermic insemination and self-

306 fertilization evolve independently. We ran each RJ MCMC chain for ten million iterations,
307 while discarding the first one million iterations as burn-in, after which the chain was sampled
308 every 1000th iteration. We placed 1000 stepping stones (iterating each 10000 times), and used
309 a gamma hyperprior (gamma 0 1 0 1) to obtain the marginal likelihood values for the models.
310 We performed three separate runs for both the independent and dependent models to check
311 for the stability of the likelihood values and convergence. We established that the chains had
312 converged using Gelman and Rubin's convergence diagnostic (Gelman and Rubin 1992) and
313 that the effective sample size was >200 for all parameters, using the R package coda
314 (Plummer et al. 2006). In addition, we also confirmed that the acceptance rate was usually
315 between 20-40% (ideal when the chain is at convergence and indicating good mixing; Pagel
316 and Meade 2006). We evaluated the alternative models using the Log Bayes Factor (BF) and
317 used the convention that a BF value greater than 2 is considered as positive support for the
318 best-fit model (Pagel and Meade 2006). We found only very weak support for the dependent
319 model of evolution over the independent model for the association between self-fertilization
320 and mating strategy, with all three runs for each model providing similar values (mean
321 marginal likelihood, independent = -9.59, dependent = -9.57, BF: 0.05; see also
322 Supplementary Table S2). This showed that the presence of self-fertilization appears
323 uncorrelated to the mating strategy in our dataset, and we could therefore use the mating
324 strategy and self-fertilization as independent predictors for our analysis of SA and SA
325 plasticity.

326 We estimated the phylogenetic signal using Blomberg's K (picante package version 1.8.2,
327 Blomberg et al. 2003; Kembel et al. 2010) and Pagel's λ (phytools package version 0.7-70,
328 Pagel 1999; Revell 2012), for both SA and the two SA plasticity effect sizes. A phylogenetic
329 signal value close to zero is suggestive of a trait evolving independently of the phylogeny,
330 while a value close to one suggests that the traits evolve under Brownian motion. In our case,

331 the estimates for SA showed a phylogenetic signal, while the two SA plasticity effect sizes
332 did not differ significantly from 0 (Supplementary Table S3). Since our sample sizes are
333 small, the likelihood ratio tests used to assess Pagel's λ can be unreliable (Boettiger et al.
334 2012), as the asymptotic properties of maximum-likelihood estimation may not hold for small
335 sample sizes. Thus, we present both phylogenetically corrected analyses, as well as analyses
336 without correcting for phylogeny.

337 For the phylogenetically corrected analyses, we compared the fit of different character
338 evolution models (i.e. Brownian motion, Ornstein–Uhlenbeck, and Early-burst) (geiger
339 package, Harmon et al. 2008) for SA and the two SA plasticity effect sizes. While it has been
340 suggested that datasets with small sample sizes can erroneously favour Ornstein–Uhlenbeck
341 models over simpler models like Brownian motion (Cooper et al. 2016b,a), our small dataset
342 was found to be more consistent with a Brownian motion model (see Supplementary Table
343 S4). Thus, we used Brownian motion as the preferred model for the subsequent analysis. This
344 does not necessarily imply, however, that the traits here actually evolve at random, but solely
345 that a Brownian motion model fits our data better than the alternative Ornstein–Uhlenbeck or
346 Early-burst models.

347 Using phylogenetic generalized least squares (PGLS) regressions and a Brownian motion
348 model of character evolution (nlme package version 3.1-152, Pinheiro et al. 2014), we tested
349 if the mating strategy (reciprocal mating vs. hypodermic insemination) or self-fertilization
350 (presence or absence) predicted SA and the two SA plasticity effect sizes (presence of mating
351 partners and strength of LSC). PGLS allows us to account for the phylogenetic non-
352 independence of observations resulting from common evolutionary history of species.
353 Furthermore, we also conducted an analysis without correcting for phylogeny, where we
354 calculated Wilcoxon rank-sum tests to assess if the mating strategy or self-fertilization
355 predicted the differences in SA or SA plasticity effect sizes. While comparative studies

356 usually consider only the mean value of a trait, this does not allow us to incorporate
357 intraspecific variability, which can be a possible source of error (Boucher et al. 2012;
358 Tonnabel et al. 2018). Here, we therefore incorporated intraspecific variation for both the
359 PGLS and Wilcoxon rank-sum tests using a resampling approach (with 10000 iterations).
360 Each time we generated a random value from a normal distribution with the observed means
361 and standard deviations for each species (Supplementary Table S1) and performed the
362 analysis on this value. We report the mean values of PGLS in the main text and the
363 distributions in Supplementary Figure S2.

364 Finally, to partition the observed variance in SA into its between-species and within-species
365 components, we used the mean SA (or weighted mean SA across all experiments for *M.*
366 *lignano*) for each group size and species (i.e. 3 observations per species), and fit a linear
367 mixed model (lme4 package, Bates et al. 2015) using species as a random effect and
368 calculated the percentage of variance explained by species. And additionally, to explore the
369 impact that the variable outcomes of the different experiments in *M. lignano* might have on
370 these estimates, we picked one mean SA value per group size and experiment at random, and
371 then redid the above analysis (resampling 10000 times).

372 All statistical analyses were carried out using R version 4.0.5 (R Core Team), unless stated
373 otherwise.

374 *Results*

375 *SA and effect sizes of SA plasticity for all species*

376 We found substantial variation in overall SA across the seven *Macrostomum* species (Figure
377 2, Supplementary Table S1), both with respect to the mating strategy (solid vs. stippled lines)
378 and self-fertilization (circles vs. triangles). Two of the three hypodermically inseminating
379 species (*Macrostomum* sp. 22 and *M. pusillum)*, exhibited a low SA. Similarly, and in line

380 with our predictions, three of the four self-fertilizing species (*Macrostomum* sp. 22,
381 *M. pusillum* and *M. mirumnovem*), exhibited a relatively low SA, likely indicating that SA in
382 these species was female-biased. Interestingly, the exception to these patterns was *M. hystrix*,
383 which had the highest SA of all the studied species. There was also interspecific variation in
384 the effect size estimates of SA plasticity, even among relatively closely related species
385 (Figure 3, Supplementary Table S1). For example, *M. cliftonense* and *M. lignano* exhibited
386 the highest SA plasticity in response to the presence of mating partners and the strength of
387 LSC, respectively. In addition, *M. lignano* exhibited SA plasticity across all experiments and
388 while the magnitude varied somewhat, the confidence intervals overlapped, indicating that the
389 estimates were fairly consistent across experiments (Figure 3).

390 ***Evolution of SA and SA plasticity***

391 The PGLS models showed that neither the mating strategy (mean: $t_5 = -0.56$, $P = 0.62$) nor the
392 ability to self-fertilize (mean: $t_5 = -0.54$, $P = 0.61$) significantly predicted SA (Supplementary
393 Figure S2a-d). And similar to SA, the mating strategy also did not significantly predict the SA
394 plasticity effect size, neither due to the presence of mating partners (mean: $t_5 = 0.97$, $P = 0.38$)
395 nor the strength of LSC (mean: $t_5 = 0.29$, $P = 0.78$) (Supplementary Figure S2e-f, S2i-j). In
396 contrast, self-fertilization significantly predicted plasticity of SA in response to the presence
397 of mating partners (mean: $t_5 = -2.98$, $P = 0.03$, Supplementary Figure S2g-h), with this kind of
398 plasticity being lower for selfing species (compare the black effect size estimates between
399 selfing vs. non-selfing species in Figure 3). No significant association between self-
400 fertilization and plasticity of SA was observed in response to the strength of LSC (mean: $t_5 = -$
401 0.72, $P = 0.52$, Supplementary Figure S2k-l). Qualitatively similar results were obtained from
402 the analyses that did not correct for phylogeny using Wilcoxon rank-sum tests
403 (Supplementary Table S5).

404 Lastly, the linear-mixed model, using the weighed mean SA for *M. lignano*, showed that
405 between-species variance explained 73.6% of the total SA variance (interspecific variance =
406 0.0177, and residual variance = 0.0063), suggesting that the observed SA variance between-
407 species is nearly three times larger than the observed variance within-species (and the
408 resampling approach showed very similar results, Supplementary Table S6).

409 ## *Discussion*

410 Our study showed that there was substantial interspecific variation in both SA and SA
411 plasticity in *Macrostomum*, also among closely related species, suggesting that both SA and
412 SA plasticity are evolutionarily labile. Furthermore, while the mating strategy predicted
413 neither SA nor the SA plasticity effect sizes, self-fertilizing species had a lower SA plasticity
414 in response to the presence of mating partners. In the following, we discuss these findings in
415 some detail.

416 In the context of phenotypic plasticity, very few studies have explored predictions of SA
417 theory across multiple species in hermaphroditic animals (Hoch and Levinton 2012;
418 Schleicherová et al. 2014), with most studies focussing solely on intraspecific comparisons
419 (Schärer 2009). Hoch and Levinton (2012) tested Charnov's (1980, 1982) mating group size
420 model in two species of acorn barnacles, *Semibalanus balanoides* and *Balanus glandula,* by
421 manipulating both the number and density of individuals in a natural setting. They showed
422 that, while both species exhibit an increased male allocation (estimated using the sum of the
423 mass of testes, sperm, seminal vesicle, and penis) at higher densities, there was no clear effect
424 on SA from either treatment. In part this was because the species also responded in terms of
425 their female allocation (estimated using egg mass), with both the number and density
426 manipulation having a positive effect in *S. balanoides*—i.e. an effect in the opposite direction
427 than predicted by the model—while there was no clear effect in *B. glandula.* The authors
428 argued that this might result from interspecific differences in life-history traits. In contrast to

429 the acorn barnacles, a study on resource allocation in response to different numbers of mating
430 partners in three related polychaete worms, *Ophryotrocha diadema, O. adherens* and
431 *O. gracilis*, showed that there was no significant plasticity in male allocation (estimated as the
432 number of sperm produced by an individual) across the species (Schleicherová et al. 2014),
433 while the species differed in the plasticity in female allocation (using multiple estimates, such
434 as resource investment in eggs, total number of egg cocoons, and time interval between egg
435 laying). The authors proposed that the magnitude of plasticity depended on the species-
436 specific costs of the sex function, and also on the mating system of the species, and that there
437 might be considerable resource allocation to behaviours linked to the male role in at least one
438 of these species (Lorenzi et al. 2006; Santi et al. 2018). So, while both of these studies
439 document interspecific variation in SA plasticity, they lacked the power to explore
440 statistically whether interspecific differences in reproductive traits may affect the level of SA
441 plasticity.

442 Our results showed that both SA and SA plasticity varied considerably across the studied
443 *Macrostomum* species. Both hypodermically inseminating and self-fertilizing species tended
444 to show a low (i.e. more female-biased) SA, with the notable exception of *M. hystrix*. This is
445 interesting, since *M. hystrix* has been shown to represent an independent origin of hypodermic
446 insemination in the reciprocally mating clade (Schärer et al. 2011). It has previously been
447 suggested that SA could become more male-biased when a species shifts to hypodermic
448 insemination, since hypodermic sperm might compete more in a fair-raffle-type sperm
449 competition (Schärer and Janicke 2009), unless selfing were to become the dominant mating
450 mode. Although capable of self-fertilization, *M. hystrix* is known to reach very high densities
451 in the field (L. Schärer, pers. obs.), which could potentially favour the evolution of a high SA.
452 In this context it is also important to note that our estimates of male and female allocation are
453 not absolute, such that when we obtain SA estimates >0.5 that probably does not suggest a

454 male-biased SA. This is because our SA estimate implicitly assumes that both testis and ovary
455 area are similarly suitable and complete proxies for male and female reproductive allocation,
456 respectively, across species. However, it is clear that these proxies do not provide absolute
457 estimates of SA, for multiple reasons (see also Singh et al. 2020b). For example, the energetic
458 expenditure per unit tissue could differ between the testes and ovaries (Schärer 2009), or there
459 could be other components of male and female allocation that are not quantified by assessing
460 gonad size, such as substantial provisioning of developing oocytes with yolk, sex-specific
461 behaviours (Picchi and Lorenzi 2019), or the production of seminal fluid (Patlar et al. 2019),
462 components that could themselves also be plastic.

463 We show that *M. lignano* and *M. janickei,* which are sibling species capable of hybridization
464 (Singh et al. 2020a), exhibit differences in SA plasticity in response to the strength of LSC
465 (although the 95% confidence intervals of two of the six *M. lignano* studies overlap somewhat
466 with that of *M. janickei*). Variation in SA plasticity in species with similar reproductive
467 biology could stem from different environmental conditions experienced by the species. For
468 example, evolution of SA plasticity might not be favoured in species that inhabit relatively
469 stable environments, especially if the maintenance of plasticity is costly (DeWitt et al. 1998;
470 Auld et al. 2010; Siljestam and Östman 2017). Even in the absence of maintenance costs,
471 there could still be production costs of plasticity, though if the benefits of adjusting SA
472 outweigh these costs, then phenotypic plasticity could still be maintained. This cost-benefit
473 ratio of SA plasticity can vary across species (Van Buskirk 2002; Steiner 2007) and
474 environments (Ratikainen and Kokko 2019), leading to the retention or loss of plasticity in a
475 given species. Moreover, such costs of SA plasticity might play an important role in both the
476 evolution and the maintenance of simultaneous hermaphroditism (St. Mary 1997), and could
477 also constrain plastic changes in SA, potentially leading to a suboptimal SA. A study in
478 *O. diadema* showed that there were no large short-term fitness costs of sex adjustment for the

479 species (Lorenzi et al. 2008), while in *M. lignano* production costs of SA plasticity have
480 previously been documented (Sandner 2013), with an alternating group size environment
481 leading to a lower hatchling production compared to a stable group size environment. We
482 currently do not have data on production costs of plasticity in *M. janickei* when exposed to a
483 stable vs. fluctuating environment.

484 Across all species, the estimates of plasticity effect sizes are larger for the presence of mating
485 partners than for the strength of LSC (number of mating partners), and for *M. cliftonensis* and
486 *M. janickei*, the confidence intervals of the two effect size estimates do not overlap. This
487 suggests that the increase in SA going from paired to octet groups is not as drastic as going
488 from isolated worms to worms in larger groups. A similar phenomenon has been observed in
489 the freshwater snail, *Lymnaea stagnalis*, where paired snails showed higher expression of six
490 seminal fluid protein (SFP) genes compared to isolated snails, while the SFP expression of
491 snails in larger (6 individuals) groups was similar to that of paired snails (Nakadera et al.
492 2019). In our study, this situation could potentially arise if in the presence of multiple partners
493 in octets, worms do not only increase their testis size, but also their rate of sperm production
494 per unit testis size (Schärer and Vizoso 2007), which in *M. lignano* has been shown to occur
495 due to an increase in the speed of spermatogenesis (Giannakara et al. 2016). If such effects
496 occur generally (but see Giannakara & Ramm 2017 for *M. pusillum*, a species that lacks SA
497 plasticity, see also below), we might underestimate the extent of plasticity in male allocation
498 when solely measuring testis size. Alternatively, there could also be increased investment in
499 other components of the male function (see above). A study in *O. diadema* showed that
500 increased mating opportunities are accompanied by higher investment into behavioural
501 components of male function, such as aggressive behaviour (Lorenzi et al. 2006; Santi et al.
502 2018).

503 The SA plasticity effect sizes did not differ significantly between species exhibiting the
504 reciprocal and hypodermic mating strategies, while self-fertilizing species had a lower SA
505 plasticity (in response to presence of mating partner, but not the strength of LSC).
506 Interestingly, to date *M. mirumnovem* is the only reciprocally mating *Macrostomum* species in
507 which self-fertilization has been recorded (Singh et al. 2020b), while studies have usually
508 found self-fertilization in hypodermic species. Including *M. mirumnovem* in our study
509 allowed us to disentangle the association between mating strategy and self-fertilization with
510 respect to SA plasticity effect sizes, clearly demonstrating how including additional species in
511 a phylogenetically-informed context leads to more informed interpretations.

512 Interesting questions that arise then concern the causes of the variation in SA and SA
513 plasticity between the self-fertilizing species. An effect of self-fertilization on SA has been
514 theoretically predicted, and while there has been both theoretical and empirical work
515 exploring the effect of self-fertilization on SA in plants (Lemen 1980; Charlesworth and
516 Charlesworth 1981; Schoen 1982; Charnov 1987; McKone 1987; Brunet 1992), there have
517 been few such studies in animals (Johnston et al. 1998). A comparative study on SA in plants
518 showed that resource allocation to flower parts can differ depending on the mating systems,
519 such as the pollen-ovule ratio increasing with the likelihood of cross-pollination (Cruden
520 1977; but see Cruden and Lyon 1985).

521 In our study, one possible explanation for variation between self-fertilizing species could be if
522 the species differ in their rate and pattern of self-fertilization, which could have an effect on
523 both optimal SA and optimal SA plasticity. For example, in obligatorily self-fertilizing
524 species, we would expect that they do not increase their SA with social group size, since LSC
525 would be expected to remain high irrespective of the number of individuals in the local group
526 (Figure 1). On the other hand, for species that exhibit facultative self-fertilization, we might
527 expect considerable SA plasticity despite a low overall SA. Our results do conform somewhat

528 to this pattern, with *M. pusillum*—hypothesised to be obligately self-fertilizing (Giannakara
529 and Ramm 2017)—showing both the lowest overall SA and essentially no SA plasticity. Its
530 closest congener, *Macrostomum* sp. 22, showed a similarly female-biased SA in isolated and
531 paired worms, but significant SA plasticity between worms in pairs and octets. If worms
532 shifted from self-fertilization when isolated, to exclusively outcrossing when in the presence
533 of a mating partner, then the strength of LSC would not be expected to change from when
534 they are alone to being in a pair. We would, however, expect the strength of LSC to drop from
535 worms in pairs to worms in octets, favouring an increased in male allocation and SA, a
536 scenario that would match the observed SA plasticity patterns in this species. This could
537 suggest that *Macrostomum* sp. 22 may only self in isolation. In contrast, *M. hystrix*—thought
538 to be a preferentially outcrossing species, with studies showing costs of self-fertilisation
539 (Ramm et al. 2012, Giannakara and Ramm 2020) and delayed selfing in isolation, at least in
540 some populations (Ramm et al. 2012)—was found to be more plastic than *M. pusillum.*

541 Finally, our results indicate that there appears to be considerably more variation in SA
542 between-species than within-species in the sample of species we studied here, which is
543 remarkable, since we saw extensive plasticity in SA in certain species in response to changes
544 in group size, including across the multiple experiments in *M. lignano*. This result suggests
545 that we can probably interpret variation in SA among field-collected worms of different
546 *Macrostomum* species as being at least partially due to interspecific variation, even if they
547 happen to have been sampled from different social environments, and may therefore also vary
548 in part due to SA plasticity. Thus, although we cannot currently explain the observed
549 interspecific variation in SA among the species studied here, our study will facilitate research
550 in understanding the evolution of SA patterns across *Macrostomum* species, by allowing
551 future studies to include SA estimates from field-collected worms.

552 Collectively, our results suggest that there is substantial interspecific variation in SA and SA
553 plasticity in *Macrostomum*, with self-fertilization being a significant predictor for the latter
554 (although the relatively small number of species we could include in this comparative study
555 of course needs to be taken into consideration). Future studies should explore how the extent
556 of self-fertilization, be it obligatorily or facultative, affects SA and SA plasticity using data
557 from a greater number of species.

558

559 ## *References*

560 Auld, J. R., A. A. Agrawal, and R. A. Relyea. 2010. Re-evaluating the costs and limits of
561 adaptive phenotypic plasticity. Proceedings of the Royal Society B: Biological Sciences
562 277:503–511.

563 Ax, P. 1951. Die Turbellarien des Eulitorals der Kieler Bucht. Zoologische Jahrbuecher
564 Abteilung fuer Systematik Oekologie und Geographie der Tiere 80:277–378.

565 Ax, P. 1956. Les Turbellariés des étangs côtiers du littoral méditerranéen de la France
566 méridionale... Hermann et Cie.

567 Ax, P. 2008. Plathelminthes aus Brackgewässern der Nordhalbkugel. Abhandlungen der
568 Mathematisch-naturwissenschaftligen Klasse 2008:1.

569 Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker. 2015. Fitting linear mixed-effects
570 models using lme4. Journal of Statistical Software, doi: 10.18637/jss.v067.i01.

571 Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in
572 comparative data: behavioral traits are more labile. Evolution 57:717–745.

573 Boettiger, C., G. Coop, and P. Ralph. 2012. Is your phylogeny informative? Measuring the
574 power of comparative methods. Evolution 66:2240–2251.

575 Boucher, F. C., W. Thuiller, C. Roquet, R. Douzet, S. Aubert, N. Alvarez, and S. Lavergne.
576 2012. Reconstructing the origins of high-alpine niches and cushion life form in the genus
577 Androsace S.L. (Primulaceae). Evolution 66:1255–1268.

578 Brand, J. N., L. J. Harmon, and L. Schärer. 2021a. Frequent origins of traumatic insemination
579 involve convergent shifts in sperm and genital morphology. BioRxiv doi:
580 10.1101/2021.02.16.431427.

581 Brand, J. N., G. Viktorin, R. A. W. Wiberg, C. Beisel, and L. Schärer. 2021b. Large-scale
582 phylogenomics of the genus *Macrostomum* (Platyhelminthes) reveals cryptic diversity
583 and novel sexual traits. In press. Molecular Phylogenetics and Evolution. BioRxiv doi:
584 2021.03.28.437366.

585 Brunet, J. 1992. Sex allocation in hermaphroditic plants. Trends in Ecology & Evolution
586 7:79–84.

587 Campbell, D. R. 2000. Experimental tests of sex-allocation theory in plants. Trends in
588 Ecology & Evolution 15:227–232.

589 Charlesworth, D., and B. Charlesworth. 1981. Allocation of resources to male and female
590 functions in hermaphrodites. Biological Journal of the Linnean Society 15:57–74.

591 Charnov, E. L. 1987. On sex allocation and selfing in higher plants. Evolutionary Ecology
592 1:30–36.

593 Charnov, E. L. 1980. Sex allocation and local mate competition in barnacles. Marine Biology
594 Letters 1:269–272.

595 Charnov, E. L. 1979. Simultaneous hermaphroditism and sexual selection. Proceedings of the
596 National Academy of Sciences 76:2480–2484.

Charnov, E. L. 1996. Sperm competition and sex allocation in simultaneous hermaphrodites. Evolutionary Ecology 10:457–462.

Charnov, E. L. 1982. The theory of sex allocation. Monographs in population biology 18:1–355. Princeton University Press, Cambridge.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences.

Cooper, N., G. H. Thomas, and R. G. FitzJohn. 2016a. Shedding light on the 'dark side' of phylogenetic comparative methods. Methods in Ecology and Evolution 7:693–699.

Cooper, N., G. H. Thomas, C. Venditti, A. Meade, and R. P. Freckleton. 2016b. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. Biological Journal of the Linnean Society 118:64–77.

Cruden, R. W. 1977. Pollen-Ovule Ratios: A Conservative Indicator of Breeding Systems in Flowering Plants. Evolution 31.

Cruden, R. W., and D. L. Lyon. 1985. Patterns of biomass allocation to male and female functions in plants with different mating systems. Oecologia 66.

DeWitt, T. J., A. Sih, and D. S. Wilson. 1998. Costs and limits of phenotypic plasticity. Trends in Ecology & Evolution 13:77–81.

Friedman, J., and S. C. H. Barrett. 2011. Genetic and environmental control of temporal and size-dependent sex allocation in a wind-pollinated plant. Evolution 65:2061–2074.

Gelman, A., and D. B. Rubin. 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science 7:457–472.

Giannakara, A., and S. A. Ramm. 2020. Evidence for inter-population variation in waiting times in a self-fertilizing flatworm. Invertebrate Reproduction & Development 64:158–168.

620 Giannakara, A., and S. A. Ramm. 2017. Self-fertilization, sex allocation and spermatogenesis
621 kinetics in the hypodermically inseminating flatworm *Macrostomum pusillum*. The
622 Journal of Experimental Biology 220:1568–1577.

623 Giannakara, A., L. Schärer, and S. A. Ramm. 2016. Sperm competition-induced plasticity in
624 the speed of spermatogenesis. BMC Evolutionary Biology 16:60.

625 Greeff, J. M., and N. K. Michiels. 1999. Sperm Digestion and Reciprocal Sperm Transfer Can
626 Drive Hermaphrodite Sex Allocation to Equality. The American Naturalist 153:421–430.

627 Guo, Q., D. G. Brockway, and X. Chen. 2017. Temperature-related sex allocation shifts in a
628 recovering keystone species, Pinus palustris. Plant Ecology & Diversity 10:303–310.

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

630 Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER:
631 investigating evolutionary radiations. Bioinformatics 24:129–131.

632 Harrell, F. E. 2020. Hmisc: Harrell miscellaneous. R package version 4.4-0.

633 Hedges, L. V. 1981. Distribution theory for glass's estimator of effect size and related
634 estimators. Journal of Educational Statistics 6:107–128.

635 Hoch, J. M., and J. S. Levinton. 2012. Experimental tests of sex allocation theory with two
636 species of simultaneously hermaphroditic acorn barnacles. Evolution 66:1332–1343.

637 Howell, D. C. 2011. Confidence intervals on effect size. Univresity of Vermont, doi:
638 10.1134/S0362119713030110.

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

642 Janicke, T., and L. Schärer. 2009a. Determinants of mating and sperm-transfer success in a
643 simultaneous hermaphrodite. Journal of Evolutionary Biology 22:405–415.

644 Janicke, T., and L. Schärer. 2009b. Sex allocation predicts mating rate in a simultaneous
645 hermaphrodite. Proceedings of the Royal Society B: Biological Sciences 276:4247–
646 4253.

647 Janicke, T., and L. Schärer. 2010. Sperm competition affects sex allocation but not sperm
648 morphology in a flatworm. Behavioral Ecology and Sociobiology 64:1367–1375.

649 Johnston, M. O., B. Das, and W. R. Hoeh. 1998. Negative correlation between male
650 allocation and rate of self-fertilization in a hermaphroditic animal. Proceedings of the
651 National Academy of Sciences 95:617–620.

652 Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S.
653 P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and
654 ecology. Bioinformatics 26:1463–1464.

655 Kolasa, J. 1973. Two New Species of Macrostomum (Turbellaria), A Redescription of an
656 Established Species and New Records from Poland. Bolletino di zoologia 40:181–200.

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

661 Lange, R., K. Reinhardt, N. K. Michiels, and N. Anthes. 2013. Functions, diversity, and
662 evolution of traumatic mating. Biological Reviews 88:585–601.

663 Lemen, C. 1980. Allocation of reproductive effort to the male and female strategies in wind-
664 pollinated plants. Oecologia 45:156–159.

Lloyd, D. G. 1984. Gender allocaations in outcrossing cosexual plants. Pp. 277–300 *in* R. Dirzo and J. Sarukhan, eds. Perspectives in Plant Population Ecology. Sinauer, Sunderland, MA.

Lopez, S., and C. A. Dominguez. 2003. Sex choice in plants: facultative adjustment of the sex ratio in the perennial herb Begonia gracilis. Journal of Evolutionary Biology 16:1177–1185.

Lorenzi, M. C., D. Schleicherová, and G. Sella. 2006. Life history and sex allocation in the simultaneously hermaphroditic polychaete worm Ophryotrocha diadema: The role of sperm competition. Integrative and Comparative Biology 46:381–389.

Lorenzi, M. C., D. Schleicherova, and G. Sella. 2008. Sex adjustments are not functionally costly in simultaneous hermaphrodites. Marine Biology 153:599–604.

Luther, A. 1905. Zur Kenntnis der Gattung *Macrostoma*. Festschrift für Palmén, Helsingfors 5:1-61, 64 plates.

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

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

McKone, M. J. 1987. Sex Allocation and Outcrossing Rate: A Test of Theoretical Predictions Using Bromegrasses (Bromus). Evolution 41:591.

Meade, A., and M. Pagel. 2016. Bayes Traits.

Michiels, N. K. 1998. Mating Conflicts and Sperm Competition in Simultaneous Hermaphrodites. Pp. 219–254 *in* Sperm Competition and Sexual Selection. Elsevier.

Michiels, N. K. 1999. Sexual adaptations to high density in hermaphrodites. Invertebrate Reproduction & Development 36:35–40.

Nakadera, Y., A. Giannakara, and S. A. Ramm. 2019. Plastic expression of seminal fluid protein genes in a simultaneously hermaphroditic snail. Behavioral Ecology 30:904–913.

Örsted, A. S. 1843. Forsog til en ny classification af Planarierne (Planariea Duges) grundet paa mikroskopisk-anatomiske Undersogelser. Naturhistorisk Tidsskrift 4:519–581.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society of London. Series B: Biological Sciences 255:37–45.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Pagel, M., and A. Meade. 2006. Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. The American Naturalist 167:808–825.

Papi, F. 1959. Specie nuove o poco note del gen. Macrostomum (Turbellaria: Macrostomida) rinvenute in Italia. Monitore Zoologico Italiano 66:84–102.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290.

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

Parker, G. A. 1998. Sperm Competition and the Evolution of Ejaculates: Towards a Theory Base. Pp. 3–54 *in* Sperm competition and Sexual selection.

710 Patlar, B., M. Weber, and S. A. Ramm. 2019. Genetic and environmental variation in
711 transcriptional expression of seminal fluid proteins. Heredity 122:595–611.

712 Pen, I., and F. J. Weissing. 1999. Sperm competition and sex allocation in simultaneous
713 hermaphrodites: A new look at Charnov's invariance principle. Evolutionary Ecology
714 Research 517–525.

715 Picchi, L., and M. C. Lorenzi. 2019. Gender-related behaviors: evidence for a trade-off
716 between sexual functions in a hermaphrodite. Behavioral Ecology 30:770–784.

717 Pinheiro, J., D. Bates, S. DebRoy, and D. Sarkar. 2014. R Core Team (2014). nlme: linear and
718 nonlinear mixed effects models. R package version 3.1–117. URL: http://cran. r-project.
719 org/web/packages/nlme/index. html.

720 Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. {CODA}: Convergence Diagnosis
721 and Output Analysis for {MCMC}. R News.

722 Rademaker, M. C. J., and T. J. Jong. 1999. The Shape of the Female Fitness Curve for
723 Cynoglossum officinale: Quantifying Seed Dispersal and Seedling Survival in the Field.
724 Plant Biology 1:351–356.

725 Ramm, S. A., B. Lengerer, R. Arbore, R. Pjeta, J. Wunderer, A. Giannakara, E. Berezikov, P.
726 Ladurner, and L. Schärer. 2019. Sex allocation plasticity on a transcriptome scale:
727 Socially sensitive gene expression in a simultaneous hermaphrodite. Molecular Ecology
728 28:2321–2341.

729 Ramm, S. a, A. Schlatter, M. Poirier, and L. Schärer. 2015. Hypodermic self-insemination as
730 a reproductive assurance strategy. Proceedings of the Royal Society B: Biological
731 Sciences 282:20150660.

732 Ramm, S. A., D. B. Vizoso, and L. Schärer. 2012. Occurrence, costs and heritability of
733 delayed selfing in a free-living flatworm. Journal of Evolutionary Biology 25:2559–
734 2568.

735 Ratikainen, I. I., and H. Kokko. 2019. The coevolution of lifespan and reversible plasticity.
736 Nature Communications 10:538.

737 Reinhardt, K., N. Anthes, and R. Lange. 2015. Copulatory wounding and traumatic
738 insemination. Cold Spring Harbor Perspectives in Biology 7:a017582.

739 Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other
740 things). Methods in Ecology and Evolution 3:217–223.

741 Rosas, F., and C. A. Domínguez. 2009. Male sterility, fitness gain curves and the evolution of
742 gender specialization from distyly in Erythroxylum havanense. Journal of Evolutionary
743 Biology 22:50–59.

744 Sandner, P. 2013. Impacts of sperm competition on mating behaviour and life history traits in
745 a simultaneous hermaphrodite. University of Basel.

746 Santi, M., L. Picchi, and M. C. Lorenzi. 2018. Dynamic modulation of reproductive strategies
747 in a simultaneous hermaphrodite and preference for the male role. Animal Behaviour
748 146:87–96.

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

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

755 Schärer, L., and T. Janicke. 2009. Sex allocation and sexual conflict in simultaneously
756 hermaphroditic animals. Biology Letters 5:705–708.

757 Schärer, L., T. Janicke, and S. A. Ramm. 2015. Sexual conflict in hermaphrodites. Cold
758 Spring Harbor Perspectives in Biology 7:a017673.

759 Schärer, L., G. Joss, and P. Sandner. 2004a. Mating behaviour of the marine turbellarian
760 Macrostomum sp.: these worms suck. Marine Biology 145:373–380.

761 Schärer, L., and P. Ladurner. 2003. Phenotypically plastic adjustment of sex allocation in a
762 simultaneous hermaphrodite. Proceedings of the Royal Society B: Biological Sciences
763 270:935–941.

764 Schärer, L., P. Ladurner, and R. M. Rieger. 2004b. Bigger testes do work more: experimental
765 evidence that testis size reflects testicular cell proliferation activity in the marine
766 invertebrate, the free-living flatworm Macrostomum sp. Behavioral Ecology and
767 Sociobiology 56:420–425.

768 Schärer, L., D. T. J. Littlewood, A. Waeschenbach, W. Yoshida, and D. B. Vizoso. 2011.
769 Mating behavior and the evolution of sperm design. Proceedings of the National
770 Academy of Sciences 108:1490–1495.

771 Schärer, L., and I. Pen. 2013. Sex allocation and investment into pre- and post-copulatory
772 traits in simultaneous hermaphrodites: the role of polyandry and local sperm
773 competition. Philosophical transactions of the Royal Society of London. Series B,
774 Biological sciences 368:20120052.

775 Schärer, L., and D. B. Vizoso. 2007. Phenotypic plasticity in sperm production rate: There's
776 more to it than testis size. Evolutionary Ecology 21:295–306.

777 Schleicherová, D., G. Sella, S. Meconcelli, R. Simonini, M. P. Martino, P. Cervella, and M.
778 C. Lorenzi. 2014. Does the cost of a function affect its degree of plasticity? A test on

plastic sex allocation in three closely related species of hermaphrodites. Journal of Experimental Marine Biology and Ecology 453:148–153.

Schoen, D. J. 1982. Male reproductive effort and breeding system in an hermaphroditic plant. Oecologia 53:255–257.

Siljestam, M., and Ö. Östman. 2017. The combined effects of temporal autocorrelation and the costs of plasticity on the evolution of plasticity. Journal of Evolutionary Biology 30:1361–1371.

Singh, P., D. N. Ballmer, M. Laubscher, and L. Schärer. 2020a. Successful mating and hybridisation in two closely related flatworm species despite significant differences in reproductive morphology and behaviour. Scientific Reports 10:12830.

Singh, P., J. N. Brand, and L. Schärer. n.d. Evolution of the suck behaviour, a postcopulatory female resistance trait in a hermaphroditic flatworm genus.

Singh, P., N. Vellnow, and L. Schärer. 2020b. Variation in sex allocation plasticity in three closely related flatworm species. Ecology and Evolution 10:26–37.

St. Mary, C. M. 1997. Sequential Patterns of Sex Allocation in Simultaneous Hermaphrodites: Do We Need Models That Specifically Incorporate This Complexity? The American Naturalist 150:73–97.

Steiner, U. K. 2007. Investment in defense and cost of predator-induced defense along a resource gradient. Oecologia 152:201–210.

Tonnabel, J., F. M. Schurr, F. Boucher, W. Thuiller, J. Renaud, E. J. P. Douzery, and O. Ronce. 2018. Life-History Traits Evolved Jointly with Climatic Niche and Disturbance Regime in the Genus Leucadendron (Proteaceae). The American Naturalist 191:220–234.

802 Torchiano, M. 2017. Package “effsize”: efficient effect size computation. CRAN Package,
803 doi: 10.5281/ZENODO.196082.

804 Turner, H. M., and R. M. Bernard. 2006. Calculating and Synthesizing Effect Sizes.
805 Contemporary Issues in Communication Science and Disorders 33:42–55.

806 Van Buskirk, J. 2002. A Comparative Test of the Adaptive Plasticity Hypothesis:
807 Relationships between Habitat and Phenotype in Anuran Larvae. The American
808 Naturalist 160:87–102.

809 van Velzen, E., L. Schärer, and I. Pen. 2009. The effect of cryptic female choice on sex
810 allocation in simultaneous hermaphrodites. Proceedings of the Royal Society B:
811 Biological Sciences 276:3123–3131.

812 Vellnow, N., L. Marie-Orleach, K. S. Zadesenets, and L. Schärer. 2018. Bigger testes increase
813 paternity in a simultaneous hermaphrodite, independently of the sperm competition
814 level. Journal of Evolutionary Biology 31:180–196.

815 Vizoso, D. B., G. Rieger, and L. Schärer. 2010. Goings-on inside a worm: Functional
816 hypotheses derived from sexual conflict thinking. Biological Journal of the Linnean
817 Society 99:370–383.

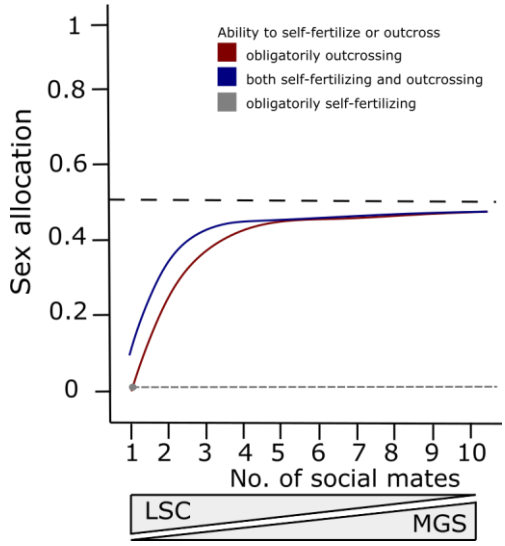
818 Winkler, L., and S. A. Ramm. 2018. Experimental evidence for reduced male allocation under
819 selfing in a simultaneously hermaphroditic animal. Biology Letters 14:20180570.

820 Zhang, S., Y. Shi, Z. Zeng, F. Xin, L. Deng, and A. Wang. 2021. Two New Brackish-Water
821 Species of Macrostomum (Platyhelminthes: Macrostomorpha) from China and Their
822 Phylogenetic Positions. Zoological Science 38.

823

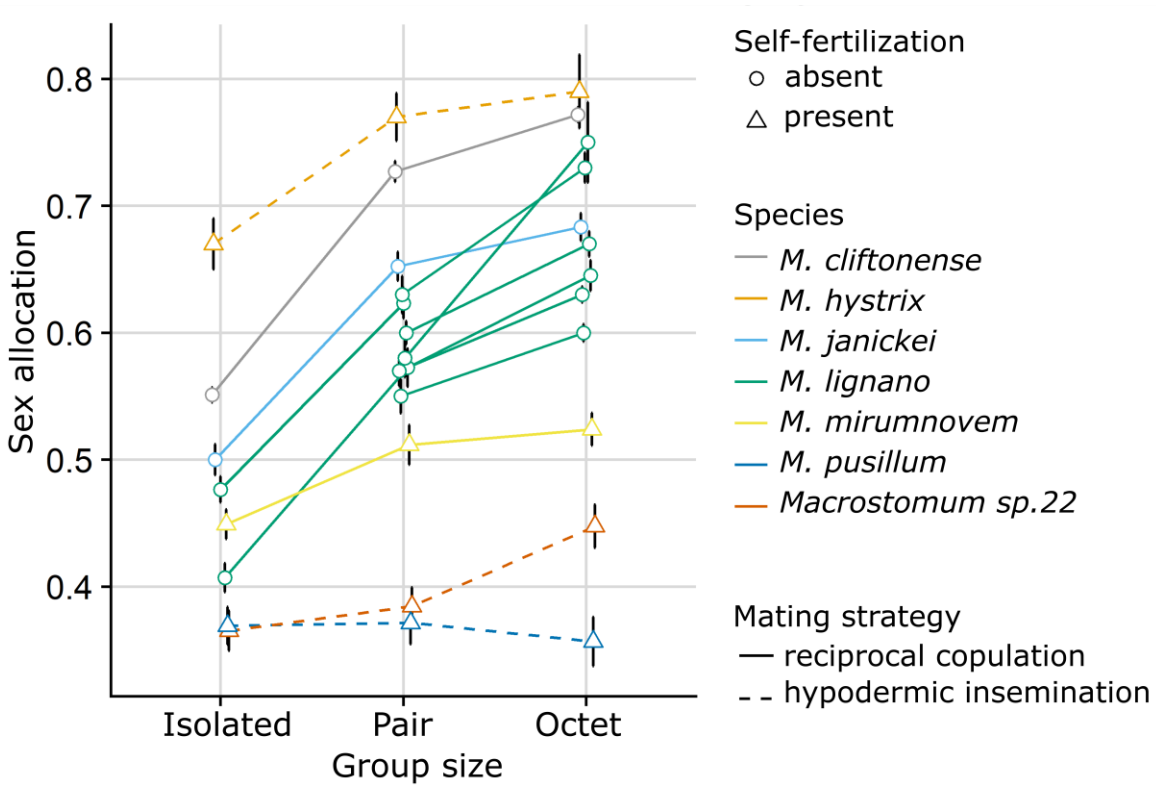
824

825 ***Figures***



826

827 Figure 1. A visualization of our hypotheses for the effect of the number of social mates on the
828 predicted sex allocation (i.e. the proportion of resources allocated to the male function) for
829 species that either obligatorily outcross (red), species that both self-fertilize and outcross
830 simultaneously (blue), or species that obligatorily self-fertilize (grey). As the number of social
831 mates increases, the so-called mating group size (MGS, i.e. the number of actual mates plus
832 one) is expected to increase, while the level of local sperm competition (LSC) is expected to
833 decrease. Note that for species that obligatorily self-fertilize, mating group size always
834 remains one and LSC is always maximal, leading to a highly female-biased sex allocation (i.e.
835 the minimal male allocation to allow for full self-fertility), independently of the number of
836 social mates (indicated by grey dot and stippled line). For species that both self-fertilize and
837 outcross simultaneously, the MGS is already increased when the number of mates is one,
838 since own sperm will compete with the partner's sperm. And in species that outcross only, the
839 prediction for when the number of mates is one is only the minimal male allocation to allow
840 for full outcross-fertility. Note that these SA predictions are only approximate, since the
841 degree to which MGS increases with the number of social mates will likely vary, and the
842 extent of self-fertilization and outcrossing is unclear in the species that show both.



843

844 Figure 2. Effect of group size on estimates of sex allocation in seven different *Macrostomum*
845 species (given by different colours). The line types represent the two mating strategies and the
846 symbols represent the ability to self-fertilize. The plots show means and standard errors of the
847 raw (untransformed) data. Note that for *M. lignano*, data from seven independent experiments
848 are shown.

849

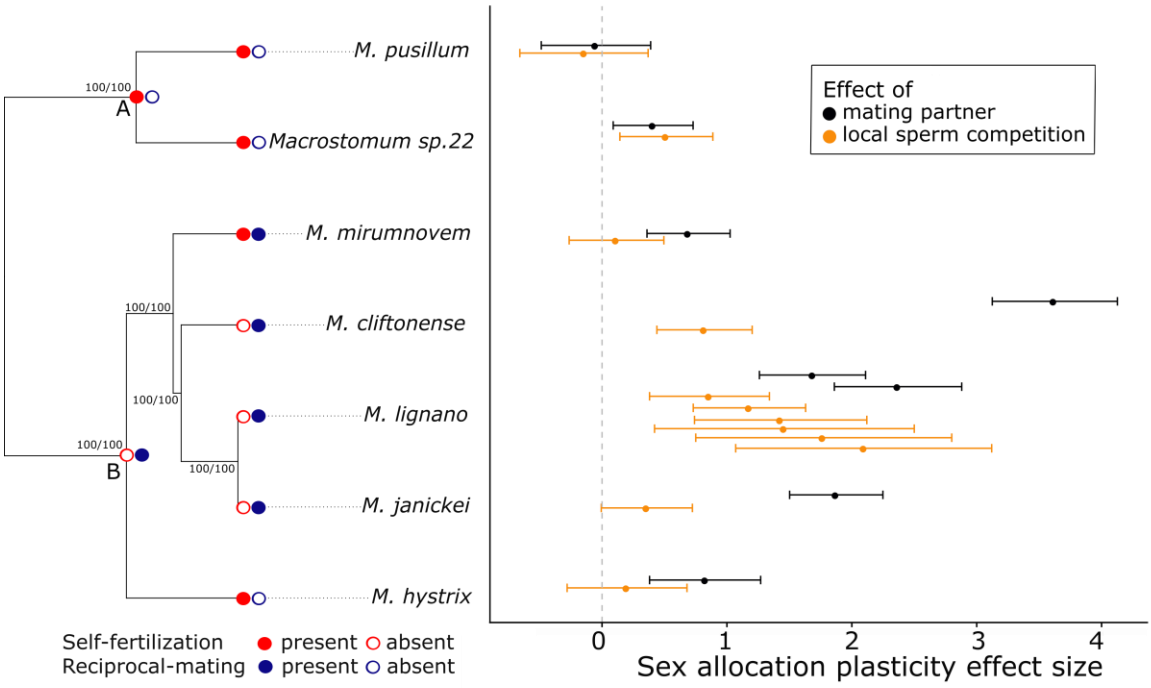


Figure 3. Standardized SA plasticity effect sizes for the effect of the presence of mating partners (i.e. comparing isolated worms vs. worms with partners) and the strength of local sperm competition (i.e. comparing worms in pairs vs. octets) among seven *Macrostomum* species (right side). The error bars represent 95% confidence intervals. For *M. lignano*, data from multiple experiments are shown. Also indicated is whether self-fertilization or/and reciprocal mating is present or absent in a species (left side), and these traits are mapped onto a trimmed maximum likelihood phylogeny of the genus (i.e. the H-IQ-TREE phylogeny from Brand et al. 2021b, which is based on 385 genes in 98 *Macrostomum* species), and all the shown bipartitions in this tree had maximal support, as indicated by ultrafast bootstrap support (first number) and approximate likelihood ratio tests (second number), respectively. A and B represent the inferred ancestral states at important internal nodes, suggesting that there are at least two independent origins of hypodermic insemination (Schärer et al. 2011; Brand et al. 2021a) and probably also self-fertilization in the genus (see also Methods and Supplementary Figure S1 for details).