Article

Faster rates of molecular sequence evolution in reproduction-related genes and in species with hypodermic sperm morphologies

R. Axel W. Wiberg[1*], Jeremias N. Brand[1,2], Lukas Schärer[1]

1. University of Basel, Department of Environmental Sciences, Zoological Institute, Vesalgasse 1, 4051 Basel, Switzerland

2. Current affiliation: Max Planck Institute for Biophysical Chemistry, Department of Tissue Dynamics and Regeneration, Am Fassberg 11, 37077 Göttingen, Germany

***** Correspondence: raw.wiberg@unibas.ch and raw.wiberg@protonmail.com

ORCID Numbers:

RAWW	0000-0002-8074-8670
JNB	0000-0001-6126-8279
LS	0000-0002-0209-9871

Keywords: Sexual conflict, sexual selection, comparative genomics, transcriptomics, *dN/dS*

Abstract

Sexual selection is expected to drive the evolution of many striking behaviours and morphologies, leaving signatures of selection at loci underlying these phenotypes. However, relatively few studies have contrasted molecular sequence evolution at such loci across lineages that differ in their sexual selection context. Our comparative genomics study of *Macrostomum,* a large genus of free-living simultaneously hermaphroditic flatworms, takes advantage of functional annotations from the model species, *M. lignano*, and transcriptome assemblies of 97 congeners. We compare molecular sequence evolution in species with contrasting sperm morphologies, which are strongly associated with multiple convergent shifts in the mating strategy and thus reflect the sexual selection context in *Macrostomum*. The sperm of most reciprocally mating species carry lateral bristles, likely functioning as anchoring mechanisms against post-copulatory sperm removal. Hypodermically mating species lack these bristles, potentially as adaptations to a different environment experienced by hypodermic sperm. We document faster molecular sequence evolution in reproduction-related, compared to ubiquitously-expressed, genes across all sperm morphologies, consistent with more intense selection acting on the former. Furthermore, we observed faster molecular sequence evolution in species with hypodermic sperm morphologies, in both reproduction-related and ubiquitously-expressed genes. These genome-wide patterns suggest that shifts to hypodermic mating reduce the efficiency of selection, possibly due to higher selfing rates in hypodermically mating species. Moreover, we find little evidence for convergent amino acid changes across species. We provide the first comprehensive comparative analysis of molecular sequence evolution in a group of simultaneously hermaphroditic animals, across well-replicated contrasts of lineages with divergent sperm morphologies.

Introduction

Sexual selection and the various associated processes (e.g. sperm competition, mate choice, and sexual conflict) are potent forces of diversifying selection, driving the evolution of behavioural, morphological, and biochemical traits in many organisms (Andersson 1994; Arnqvist and Rowe 2005). These forces can also be important drivers of population divergence and speciation, due to ensuing rapid evolution and coevolution resulting from intense competition over reproductive resources (Lande 1981; Gavrilets 2000; Hayashi et al. 2007). Understanding the genomic manifestations and consequences of sexual selection has been a long-standing goal in evolutionary biology. For example, sex chromosome evolution is expected to be driven, to a large extent, by sexual conflict (e.g. Frank and Patten 2019), and sex-biased gene expression is thought to result from sexual selection on the underlying sexually dimorphic traits (Garth and Parsch 2016). At the same time, sex-biased genes tend to show rapid rates of molecular sequence evolution (Ellegren and Parsch 2007). But the study of the genomics of sexual selection remains in its infancy, partly because it is difficult to distinguish between the processes of sexual selection, natural selection, and genetic drift based on patterns in sequence data alone (Wilkinson et al. 2015). To make progress, more detailed functional information on the loci in question is needed, as are replicated contrasts between species and populations that differ in how sexual selection operates (Wilkinson et al. 2015).

Rapid molecular sequence evolution of reproduction-related genes

Reproduction-related genes are defined in several different ways (e.g. Swanson and Vacquier 2002 *vs.* Dapper and Wade 2020). A broad definition can be given as genes linked to some aspect of reproduction. This definition includes those involved in conserved reproductive functions, such as meiosis and gamete production (which may or may not be subject to sexual selection), but also any genes that underlie variation in sexually-selected traits (e.g. ornaments, preferences or biases, and persistence/resistance traits). A common way to identify at least a subset of such genes is to focus on genes expressed in various reproductive tissues (e.g. testes, ovaries, accessory glands, and reproductive tracts). In principle, all reproduction-related genes, insofar as they are under sexual selection, could show evidence of positive selection (Swanson and Vacquier 2002; Dapper and Wade 2020).

As expected from models of sexual selection, large-scale studies often show elevated rates of molecular sequence evolution in reproduction-related genes when compared to the rest of the genome (for reviews see for example Swanson and Vacquier 2002; Vacquier and Swanson, 2011; Wilburn and Swanson, 2016). This is often assessed by analysis of the rates of molecular sequence

evolution, e.g. by computing the ratio of non-synonymous (*dN*) and synonymous (*dS*) substitution rates in protein-coding DNA sequences (Yang and Bielawski 2000; see below for more details). Among mammals, rapid evolution has been observed for a range of such genes. For example, genes governing sperm-egg interactions show strong evidence of positive selection at a few amino acid sites (Swanson et al. 2001). Additionally, in mice, ejaculate proteins and proteins isolated from the seminal vesicle have high rates of molecular sequence evolution that are indicative of positive selection (Ramm et al. 2008; Dean et al. 2009). Meanwhile, proteins isolated from other reproduction-related tissues, such as the prostate, show lower rates compared to the genome-wide background, suggesting sequence conservation for some subsets of reproduction-related genes (Dean et al. 2009). High rates of molecular sequence evolution have also been found for proteins involved in sperm-egg interactions in several marine invertebrates (Clark et al. 2009; Aagaard et al. 2013; Vacquier and Swanson 2011). In comparative studies of *Drosophila*, some subsets of reproduction-related genes (namely genes expressed in the testes and accessory glands), are among the fastest evolving genes in the genome, both in terms of sequence divergence and gene gain/loss (Haerty et al. 2007; Findlay et al. 2008).

To date, however, few studies have contrasted molecular sequence evolution in closely related lineages that differ in the way in which sexual selection manifests itself. Even fewer studies have taken a wide comparative genomics approach, be it by investigating many genes or a large sample of species. Instead, studies are generally limited either by their focus on only a few genes or by sampling only a few species (e.g. Dorus et al. 2004; Hurle et al. 2007; Ramm et al. 2008; Vicens et al. 2014; Harrison et al. 2015). Furthermore, almost no work has been done to generalise findings to different sexual systems (e.g. from the commonly studied separate-sexed organisms to simultaneous hermaphrodites), especially in animals. Some recent studies of hermaphrodites exist from plant systems, but these generally also focus on only a few species (Szövényi et al. 2013; Arunkumar et al. 2013; Gossmann et al. 2013; 2016; Harrison et al. 2019; Moyle et al. 2021), and comparisons between plants and animals are challenging for reasons which will be discussed below. It is clear that more work is sorely needed to probe the consequences of different sexual selection contexts and sexual systems on the genes underlying sexually-selected traits and on the genome as a whole.

Signatures of rapid molecular sequence evolution are typically taken as evidence for strong sexual selection acting on reproduction-related genes. However, relaxed selection on coding sequences, either due to reduced intensity or reduced efficiency of selection, can also result in accelerated molecular sequence evolution (Wertheim et al. 2015; Dapper and Wade 2016; 2020). Indeed, there are several reasons why selection may in fact be less efficient on reproduction-related genes than other genomic loci (Dapper and Wade 2016; 2020). For example, reproduction-related

genes are often sex-biased, or even sex-limited, in their expression (Dapper and Wade 2016; 2020). In the case of sex-limited expression, such genes are "visible" to selection only in one sex and, all
111 else being equal, the selection they experience is therefore expected to be only half that of genes expressed equally in both sexes (Dapper and Wade 2016; 2020). It may therefore not be appropriate to compare reproduction-related genes to a genomic background set of genes, and more nuanced
114 null models should be considered (Dapper and Wade 2016; 2020). However, such null models have, to date, only been developed for species with separate sexes, and theory about how they might apply to, for example, simultaneous hermaphrodites is currently lacking. Every simultaneous
117 hermaphrodite is both male and female at the same time. It would therefore seem that a gene cannot, on average, be sex-biased or sex-limited in expression in a sense that would produce the kind of consistent differences in the efficiency of selection highlighted by Dapper and Wade (2016;
120 2020). As a result, reproduction-related genes in simultaneous hermaphrodites may not experience lower efficiency of selection compared to other genes. Therefore, in hermaphrodites any differences in the rates of molecular sequence evolution between reproduction-related genes and other genes
123 may be more easily attributable to differences in the actual intensity of selection. This potentially makes simultaneous hermaphrodites powerful systems to study the evolution of reproduction-related genes and the effects of sexual selection on the genome.

126 Estimating the effects of evolutionary forces on molecular sequences is a challenging task and relying on a single approach is problematic. As noted above, the effect of selection on molecular sequence evolution is often assessed by computing the ratio of non-synonymous (*dN*) to
129 synonymous (*dS*) substitution rates in protein-coding DNA sequences ($dN/dS = \omega$; Yang and Bielawski 2000). If synonymous substitutions, at the coding sequence level, are assumed to be neutral, then a value of $\omega = 1$ suggests that a gene is evolving neutrally, since both types of
132 substitution occur at the same rate. If $\omega < 1$ or > 1 a gene is inferred to be evolving under purifying or positive selection, respectively, since non-synonymous substitutions occur less or more frequently than expected. Maximum-likelihood methods for the computation of ω using codon-
135 substitution models are readily available. "Branch models" estimate a gene-wide ω to contrast average ω values for genes across different lineages in a phylogeny (Yang and Bielawski 2000). While flexible and useful, these models have some limitations. For instance, even if positive
138 selection is acting on a gene, this is unlikely to raise $\omega > 1$ for the entire coding sequence. While changes may be favoured in a specific binding region, the overall protein backbone may still be conserved. Thus, most genes typically show values of $\omega < 1$. "Branch-site models", in turn, permit
141 more complex scenarios where most sites are under purifying selection while allowing a few sites to be under positive selection (Yang and Bielawski 2000). While there is ongoing work on how robust

branch-site tests are to violations of model assumptions (e.g. Fletcher and Yang 2010; Yang and dos
144 Reis 2011; dos Reis and Yang 2013; Gharib and Robinson-Rechavi 2013; Venkat et al. 2018), the consensus suggests that it is a fairly robust approach. Additionally, because a shift from purifying to relaxed selection can result in similar values of ω to those expected from positive selection,
147 especially in branch models, it is prudent to also explicitly test for evidence of relaxed selection. Specifically, branch-site models can be employed to test if the differences in the distributions of ω values across sites between a set of test and reference branches in a phylogeny are consistent with
150 expectations from a relaxation of selection in test branches (Wertheim et al. 2015). In summary, using multiple methods with different assumptions and approaches allows for a more complete picture of the role of selection in shaping patterns of sequence divergence.

153

Reproductive biology of the genus Macrostomum

Free-living flatworms of the species-rich genus *Macrostomum* are members of
156 Macrostomorpha, a large group of Platyhelminthes, representatives of which vary tremendously in reproductive traits (Janssen et al. 2015). *Macrostomum* are simultaneous hermaphrodites, producing sperm and eggs at the same time, and several species from this genus, particularly the main model
159 species *M. lignano* (Ladurner et al. 2005; Wudarski et al. 2020), are used to study mating behaviour, sexual selection, and sexual conflict (e.g. Schärer et al. 2005; Ramm et al. 2015; Marie-Orleach et al. 2017; Giannakara et al. 2016; Giannakara and Ramm 2017; Patlar et al. 2020; Singh et al. 2020a,
162 2020b). In hermaphrodites, sexual conflict is expected to occur not just over the mating and re-mating rates, but also over the sex roles that the partners assume during copulation, with the male (or "sperm donor") role generally thought to be preferred over the female (or "sperm recipient")
165 role (Charnov 1979; Michiels 1998; Schärer et al. 2015). These conflicts have likely been important in driving the evolution of the extraordinary diversity in reproductive morphologies and behaviours seen in this genus (Schärer et al. 2011).

168 Many *Macrostomum* species employ a strategy of reciprocal copulation (figure 1A), where both partners donate and receive sperm at the same time, by reciprocally inserting their male intromittent organ, the stylet, into the partner's sperm receiving organ, the antrum (Schärer et al.
171 2004; Vizoso et al. 2010). In contrast, other species employ a strategy of hypodermic insemination (figure 1A), where individuals inject sperm into the body of the partner through the epidermis using a needle-like stylet (Schärer et al. 2011). These two mating strategies are associated with
174 characteristic differences in stylet and sperm morphology (Schärer et al. 2011). Specifically, reciprocal copulation is associated with sperm having stiff lateral bristles (figure 1B), which are thought to serve as anchoring mechanisms preventing sperm removal by the recipient (Vizoso et al.

177 2010; Schärer et al. 2011). Indeed, in several species it is observed that, after copulation, the recipient will place its mouth opening over its own female genital opening in a stereotypical *sucking* behaviour in an apparent attempt to remove sperm and/or other ejaculate components (Schärer et al. 180 2004; Vizoso et al. 2010; Schärer et al. 2020). Meanwhile, hypodermic insemination is associated with a simpler sperm morphology lacking these bristles (Schärer et al. 2011; figure 1B). Additionally, some species display an intermediate sperm morphology with reduced bristles 183 (Schärer et al. 2011; figure 1B).

Recent comparative work on 145 *Macrostomum* species has now greatly expanded the number of known species (Brand et al. *unpublished data a*) and investigated the evolution of 186 morphological traits across the genus *Macrostomum* (Brand et al. *unpublished data b*). These studies provide evidence for striking convergent evolution, with at least 9 independent losses and 7 independent reductions of sperm bristles across the genus (figure 1C). They also confirm that the 189 contrasting bristle and sperm morphologies are strongly associated with differences in the mating strategies exhibited by the species, by showing that the evolution of hypodermic insemination predicts variation in major axes of morphospace (Brand et al. *unpublished data b*). The reproductive 192 morphology can therefore be considered a proxy for the contrasting mating strategies, which represent different outcomes of sexual conflict over sex roles during mating, and probably strongly affect the sexual selection context experienced by sperm and other reproductive traits. For example, 195 a shift to a hypodermic mating strategy might result in a more fair-raffle like sperm competition between the ejaculates of rival individuals that compete within a recipient (Schärer and Janicke 2009; Schärer et al. 2011). A shift in the sperm competition arena, namely from the female 198 reproductive tract to different tissues in the body of the recipient, may also drastically alter the types of sperm traits selected that are favoured. This should also be reflected in patterns of molecular evolution in genes underlying these traits.

201

Study overview

Here we combine several sources of information to investigate patterns of molecular sequence 204 evolution of reproduction-related genes in the genus *Macrostomum*. We produce alignments of orthologous genes using *de novo* transcriptome assemblies of ~100 species of *Macrostomum* available from a recent large-scale phylogenomic study (Brand et al. *unpublished data a*; figure 207 1C). Moreover, annotations for a large number of genes are available from gene expression and functional studies in the model species *M. lignano*. For example, using a positional RNA-Seq approach Arbore et al. (2015) identified transcripts that were primarily expressed in body regions 210 containing specific reproductive tissues, i.e. the testes, the ovaries, and the tail (containing male and

female genitalia, as well as prostate gland cells). Further work using *in situ* hybridisation narrowed down transcripts expressed in these regions to those exclusively localised in, for example, the testis, the ovary, the female reproductive tract (the antrum), or the prostate gland cells, thus largely validating these positional annotations (Arbore et al. 2015; Lengerer et al. 2018; Weber et al. 2018; Ramm et al. 2019). Evidence has also been presented that these reproduction-related genes evolve rapidly (Brand et al. 2020), but a crucial limitation of that study is that it investigated only four species and thus had little power to compare lineages with different sexual selection contexts.

In this study, we take advantage of a greatly expanded dataset to investigate the evolution of reproduction-related genes throughout the genus *Macrostomum*. We fit codon models of molecular sequence evolution to alignments of genes expressed in different body regions that contain important reproductive tissues in *M. lignano* (reproduction-related genes). Our study tests the hypotheses that molecular divergence in these genes are due to positive or relaxed selection. Specifically, we compare reproduction-related genes to genes that are likely ubiquitously expressed throughout the worm. Moreover, we contrast patterns of molecular sequence evolution between species that differ in whether their sperm morphology includes sperm bristles or not, in order to assess the effect of a changing sexual selection context on the evolution of reproduction-related genes.

Results

Branch models

We identified a total of 1,149 orthogroup alignments (OGs) that passed all successive steps of the bioinformatic pipeline (see Materials and Methods; figure S1) and thus could be used in analyses of molecular sequence evolution with PAML v. 4.9 (table S1; Yang 2007). Based on RNA-Seq data from the model organism *M. lignano* (Arbore et al. 2015), we annotated 892 OGs as likely reproduction-related (yielding 643, 135, and 114 testis-, ovary-, and tail-region OGs, respectively), and 257 as likely showing ubiquitous expression (ubiquitously-expressed OGs). To identify broad differences in the patterns of molecular sequence evolution among these annotation groups, we first fit a codon model of molecular sequence evolution that assumes a single common ω value for all lineages. On the assumption that OGs can be treated as statistically independent, we found a significant effect of annotation group on average ω (Kruskal-Wallis rank-sum test, $\chi^2 = 34.93$, d.f. = 3, $p < 0.001$; figure 2A). In *post-hoc* pairwise comparisons we found that this was driven by ubiquitously-expressed OGs having lower average ω, while the other annotation groups showed similar values (figure 2A). Furthermore, we found a negative correlation between the expression level of the *M. lignano* gene in each OG (as measured from the whole body) and the estimated ω

value for that OG (Spearman rank correlation, rho = -0.72, p = 0.015). Thus, more highly expressed
246 OGs show slower rates of molecular sequence evolution.

We next asked if ω differed across the sperm bristle states, both between and within annotation groups. Specifically, for 828 OGs that contained representative species of all three bristle
249 states, we compared the above single ω models to models with three separate ω values, one for each bristle state. Overall, we found that 514 (~62%) OGs were better explained by the model with three states (table S1), thus clearly documenting variation in ω across sperm bristle states. Ideally, we
252 would test the two factors bristle state and annotation group simultaneously, but given the distributional properties of the ω values (i.e. skewed distributions and unequal variances between the bristle states), we in the main text report only separate non-parametric statistics for the
255 differences (i) between annotation groups within each bristle state and (ii) across bristle states within each annotation group (but see also Supplementary Text T1 in the Supplementary Materials for parametric statistics on $\log_{10}()$ transformed ω values, including an interaction term, which give
258 qualitatively similar results).

For species with bristles present, we found significant differences in ω between annotation groups (Kruskal-Wallis rank-sum test, $\chi^2 = 25.2$, d.f. = 3, $p < 0.001$), in that testis-, ovary-, and tail-
261 region OGs showed higher average ω values than ubiquitously-expressed OGs (figure 2B), and similar patterns were observed for species with reduced bristles ($\chi^2 = 24.9$, d.f. = 3, $p < 0.001$) or absent bristles ($\chi^2 = 32.3$, d.f. = 3, $p < 0.001$). In summary, we found that the pattern of higher ω
264 values for reproduction-related OGs, compared to the ubiquitously-expressed OGs (figure 2A), was recapitulated within the lineages with different bristle states (figure 2B).

Because estimates of ω are arguably more informative when more evolutionary lineages are
267 included, we further checked if these results were robust when we only considered OGs containing sequences from species covering ≥3 phylogenetically independent losses of bristles. Here the only qualitative change was that there was no longer a significant difference between the tail-region OGs
270 and the ubiquitously-expressed OGs for any bristle state (figure S2). Removing OGs with few independent origins greatly reduced the sample sizes, also for the OGs with the already sparse tail-region annotations (from N = 114 to N = 48). Identifying orthologs for tail-region genes might be
273 particularly difficult due to their rapid evolution. Brand et al. (2020) also found that testis-, ovary-, and tail-region OGs were under-represented among OGs that contained representative sequences from all four species in their study, indicating that homologs were more difficult to detect for these
276 genes. The above differences among annotation groups were also qualitatively the same if we removed low-expressed OGs (figure S3) or if we removed some OGs containing residual cDNA synthesis primer sequences (see Supplementary Text T2 in the Supplementary Materials).

Next, we compared ω across the different bristle states within annotation groups. Although this effect is already evident in figure 2B, for visualisation it is helpful to split the data according to different patterns observed among the bristle states. When we split the data by whether ω decreases (–) or increases (+) as one moves from species with bristles to those with absent bristles (figure 3), we observed that ω values are more often higher in species with absent bristles (i.e. the right panel contains more OGs). Among the testis- and ovary-region OGs we found a significant overall difference among the three bristle states (Kruskal-Wallis rank-sum tests, testis-region, $\chi^2 = 42.2$, d.f. = 2, $p < 0.001$; ovary-region, $\chi^2 = 7.3$, d.f. = 2, $p = 0.03$), while tail-region OGs showed a non-significant pattern in the same direction ($\chi^2 = 3.5$, d.f. = 2, $p = 0.18$), and also the highest percentage of OGs with the opposite pattern (32.6%, figure 3, left panel).

Intriguingly, we found that this overall pattern was true also for ubiquitously-expressed OGs ($\chi^2 = 19.1$, d.f. = 2, $p < 0.001$), indicating that the higher ω for species with absent sperm bristles appears to be a genome-wide characteristic, rather than one that is specific for reproduction-related OGs. Pooling data across all annotation groups, we found an even stronger effect of bristle state ($\chi^2 = 59.1$, d.f. = 2, $p < 0.001$). Across all annotation groups, the majority (67.4-87.5%) of OGs had higher ω values in species with absent bristles (figure 3). Moreover, testis-region OGs had the highest percentage of OGs with higher ω among species with absent bristles (87.5%; figure 3, right panel). However, there was no significant overall deviation from the proportions expected based on the distribution of all tested OGs in these annotation groups (χ^2-test, $\chi^2 = 4.6$, d.f. = 3, $p = 0.2$).

We also observed that the variance in the estimates of ω across OGs was highest for the lineages with reduced bristles (figure 3). This was even more apparent when we visualized the data split into four different patterns of the ω values across the three bristle states (see the four panels labelled –/–, –/+, +/–, and +/+ in figure S4). The patterns –/– and +/+ are, respectively, interpretable as ω always decreasing (–) or increasing (+) as one moves from species with bristles present, *via* those with reduced bristles, to those with absent bristles. The other two patterns (–/+ and +/–) reveal that species with reduced bristles sometimes do not show intermediate ω values. Deviations from the –/– and +/+ pattern could occur if species with reduced bristles do not represent a straightforward transitional state between present and absent bristles. In this case, the peculiarities of mating strategies within this reduced bristle state may impose different selection pressures on reproduction-related genes for some species, leading to increased variation within this state. Alternatively, a higher variance in estimates of ω for the reduced bristles category could arise due to the generally lower number of representatives for this bristle state in our data, both in terms of their absolute numbers (figure S5A), and in terms of the represented number of phylogenetically

312 independent losses compared to reductions of the sperm bristles (figure S5B and C), thus potentially making estimation less reliable.

We also confirmed that these results were robust to various potentially confounding factors. 315 Visual inspection of the data suggested that the OGs that were better explained by a model with three ω values compared to a single ω value did not show strong biases in terms of average pairwise *dS* (figure S6A) or overall tree length (figure S6B). Thus, we concluded that *dS* saturation was not 318 driving these results. Moreover, systematic differences in codon usage bias between lineages with contrasting sperm morphologies could also give spurious signals of positive selection. However, OGs did not show evidently different levels of codon usage bias (figure S6C). Nor were OG 321 alignment lengths different (figure S6D), suggesting that a simple difference of power stemming from longer sequences (and the correspondingly greater opportunity for mutations to occur and fix) was not driving these results. Finally, across the annotation groups, 42.1-54.1% of OGs used in this 324 analysis contained representatives covering ≥3 independent losses of sperm bristles (figure S5B) and 24.6-45.2% of OGs contained representatives covering ≥3 independent reductions of sperm bristles (figure S5C). Thus, in many cases our estimates were derived from several independent 327 transitions, giving us a greater confidence in them.

Branch-site models

330 While branch models give an informative measure of the relative rates of molecular sequence evolution, they are not ideal for detecting strong positive selection acting on only a few sites within a landscape of overall purifying selection (Zhang et al. 2005). We therefore also applied branch-site 333 models to a subset of the OGs (see also Materials and Methods for details). Briefly, we subset the data to perform contrasts for species with bristles present compared to species with absent bristles (present-absent) and for species with bristles present compared to species with reduced bristles 336 (present-reduced). In both cases we considered species with bristles present as the background branch. We were mainly interested in these two contrasts because present-absent represents a comparison between the hypothesised start and endpoints of the transition in morphology (and in 339 mating strategy), while the present-reduced contrast could represent early stages of transition (Brand et al. *unpublished data b*). We fit the branch-site model A (Zhang et al. 2005), which allows for positive selection ($\omega > 1$) among branches labelled as foreground, and compare this to the 342 corresponding null model (A_{null}) that is otherwise identical, but does not allow for positive selection. Thus, for a given OG a significantly better fit of model A and a ω estimates > 1 can be taken as evidence of positive selection acting on some subset of sites within the gene (see the Materials and 345 Methods for more details).

We found that relatively few OGs showed strong evidence for having sites evolving under positive selection in either of these contrasts (table 1, table S2), compared to the proportion of OGs with evidence for different ω values depending on the bristle state (see *Branch models* above). In the present-absent and present-reduced contrast, we found that only a total of 64 (6.8%) and 73 (7.6%) OGs showed evidence of positive selection (i.e. LRT p-values < 0.05), respectively. However, in neither contrast was there a clear difference in the ω_2 parameter (see Materials and Methods) across the annotation groups (Kruskal-Wallis rank-sum tests, present-absent, $\chi^2 = 6.6$, d.f. = 3, p = 0.09; figure S7A; present-reduced, $\chi^2 = 1.0$, d.f. = 3, p = 0.81; figure S7B). Furthermore, in neither contrast did we find a clear deviation of the distribution of significant OGs from that expected based on the distribution of all OGs across annotation groups (χ^2-tests, present-absent, $\chi^2 = 3.3$, d.f. = 3, p = 0.35; present-reduced, χ^2-test, $\chi^2 = 3.3$, d.f. = 3, p =0.34; table 1, table S2). These results suggest that in both contrasts all annotation groups are equally likely to contain genes inferred to contain sites under positive selection and there is no clear evidence for more intense selection in any one annotation group. Finally, a large proportion of the alignments with evidence for positive selection represent alignments that contain only a single loss or reduction of bristles (present-absent, 64%; present-reduced, 47.9%), although in both contrasts some alignments contained up to 7 and 4 independent shifts, respectively (table S2).

Based on posterior probabilities that ω > 1 for each site (Yang et al. 2005; see Materials and Methods for more details), relatively few genes showed evidence for positive selection in branch-site tests. Full details of which genes/sites are inferred to be under positive selection in the present-absent and present-reduced contrasts are given in table S2, and a selection of alignments is highlighted in figure S8. In those alignments with evidence for positive selection it was often the case that only a single foreground branch was driving this result in an otherwise good alignment (e.g. figure S8A). One should, therefore, arguably place more weight on signals from well-resolved alignments, containing multiple losses/reductions of sperm bristles, where amino acid changes have occurred in several foreground species within long stretches of well-aligned sequence (e.g. figure S8B and C), and not necessarily on those with the lowest p-values or those that pass corrections for multiple testing. We partially avoided this by stratifying our analysis approach by the number of independent losses/reductions (see Materials and Methods). In other alignments, probably the majority, some parts of the alignment are fragmented, but the sites inferred to be under selection lie within relatively well-aligned regions (e.g. figure S8D). Finally, in some cases the sites inferred to be under positive selection lay in regions of fragmented sequences from species with either absent or reduced bristles (e.g. figure S8E, F, G). Robust inferences are difficult in the last category of alignments, and the estimated values should be taken with some scepticism. However, although

such fragmented regions may, in part, be due to incomplete transcripts in the *de novo* transcriptome
381 assemblies we have used here, it is also possible that they represent biological reality. For example, occasionally the foreground species have a very different sequence structure from other species (e.g. figure S8F). Such cases may represent structural evolution, be it adaptive or neutral, of molecular
384 sequences, which is not easily captured by the codon substitution models we employed here. Identification of sites under selection therefore often requires careful examination of alignments (which we therefore provide in full in Supplementary Archive 1, available at:
387 https://doi.org/10.5281/zenodo.4972188), although this naturally becomes difficult in large-scale studies involving hundreds or thousands of alignments.

390 *RELAX models*

The genome-wide elevated ω values in lineages with absent or reduced bristles implied a strong overall difference in selective regime. It seemed unlikely that this could entirely be attributed
393 to positive selection. Instead, the increase in ω could be due to a more general process such as relaxed selection. We, therefore, used RELAX (Wertheim et al. 2015) to test whether some of the elevated ω values were consistent with relaxed selection in foreground branch species. RELAX
396 compares the distributions of ω values in test *vs.* reference branches, by fitting a model where test branches have a distribution of ω values different from that of reference branches according to an exponential function with a parameter k, which is interpreted as a "selection intensity parameter"
399 (Wertheim et al. 2015). A value of $k > 1$ indicates intensified selection, and a value of $k < 1$ indicates relaxed selection. This alternative model is compared to a null model that is otherwise identical, but where k is fixed to 1 (see Materials and Methods for details). In the analyses
402 performed here the species with bristles present were always used as the reference branch. For visualization, we also fit the less constrained Partitioned Descriptive Model (PDM) to the data (see Materials and Methods for details).

405 In the present-absent contrast, a larger proportion of OGs showed evidence for relaxed ($k<1$) selection (32.1%), than for intensified ($k > 1$) selection (6.8%; table 2). Overall, the proportion of OGs with a significantly different distribution of ω values varied across the annotation groups,
408 deviating significantly from those expected from the distribution of all OGs tested across the annotation groups, in that there were more observed OGs with the testis-region annotation (239) than expected (195; χ^2-test, $\chi^2 = 21.6$, d.f. = 3, $p < 0.001$; table 2). However, when we split the data
411 into OGs with $k > 1$ and $k < 1$, we only found a significant deviation from expected distributions for the latter ($k > 1$, $\chi^2 = 2.7$, d.f. = 3, $p = 0.43$; $k < 1$, $\chi^2 = 31.2$, d.f. = 3, $p < 0.001$; table 2). We also found that both for OGs with $k > 1$ and $k < 1$ there was no significant difference in the k parameter

414 across the annotation groups (Kruskal-Wallis rank-sum tests, $k > 1$, $\chi^2 = 0.8$, d.f. = 3, p = 0.85; $k <$
1, $\chi^2 = 2.4$, d.f. = 3, p = 0.49; figure 4A). Thus, we found that relaxed selection was relatively
common and that OGs annotated as testis-region specific were more likely to be evolving by
417 relaxed selection in the species with absent bristles than expected. However, we found that there
was no clear difference in the degree to which selection is relaxed or intensified across the
annotation groups (see figure S9 for examples of how distributions of ω covary with *k* estimates).
420 To better understand the changes in the distributions of ω, we further chose to visualize the OGs
with the 5 highest and 5 lowest values of *k,* as well as 20 random OGs for closer examination
(figure S10). From figure S10 it is evident that where $k > 1$, most sites were actually under more
423 intense purifying selection in the test branches rather than under increased positive selection.

In the present-reduced contrast, the proportion of OGs with evidence for a different ω
distribution among target species was generally lower and showed less variation across annotation
426 groups (table 2). There was a non-significant trend for a deviation from the expected distribution for
OGs with $k > 1$ (χ^2-tests, $\chi^2 = 6.7$, d.f. = 3, p = 0.08), with an excess of ovary-region (observed, 10;
expected, 6) and ubiquitously-expressed (observed, 12; expected, 9) OGs. For OGs with $k < 1$ there
429 was a significant deviation from the expected distribution (χ^2-tests, $\chi^2 = 14.1$, d.f. = 3, $p < 0.001$),
with an excess of testis-region OGs (observed, 61; expected, 45), as observed for the present-absent
contrast. Moreover, also here there was no significant difference in *k* across annotation groups
432 (Kruskal-Wallis rank-sum tests, $k > 1$, $\chi^2 = 1.01$, d.f. = 3, p = 0.8; $k < 1$, $\chi^2 = 6.1$, d.f. = 3, p = 0.11;
figure 4B). Figure S11 (with OGs chosen as for figure S10) shows that, just as in figure S9 and S10,
in OGs with $k > 1$ most sites were under more intense purifying selection in species with reduced
435 bristles rather than elevated positive selection.

Counter-intuitively, in neither contrast did we find much overlap between OGs that we
identified as showing evidence for positive selection in the branch-site analyses and those with a
438 different distribution of ω between test and reference branches in the RELAX analyses (figure 4 and
figure S11). Unexpectedly, among the OGs that do overlap, a larger number were identified as being
under relaxed rather than intensified selection (present-absent, $k < 1$; N = 19, $k > 1$; N = 9; present-
441 reduced, $k < 1$; N = 7, $k > 1$; N = 5).

Discussion

444 We present a comparative study of molecular sequence evolution of reproduction-related and
ubiquitously-expressed genes across the genus *Macrostomum,* a group of simultaneously
hermaphroditic flatworms that show tremendous diversity in sperm design and other reproductive
447 morphologies (Schärer et al. 2011; Brand et al. *unpublished data b*). We characterised the nature of

selection operating at reproduction-related and ubiquitously-expressed genes in species with contrasting and convergent sperm morphologies. These sperm morphologies have evolved
450 convergently multiple times and are strongly associated with different mating strategies involving hypodermic insemination or reciprocal copulation that characterise a large part of the sexual selection context of these species (Schärer et al. 2011; Brand et al. *unpublished b*).

453 Branch models clearly indicate an overall pattern of higher ω values in testis-, ovary-, and tail-region annotated OGs compared to ubiquitously-expressed OGs (figure 2A). This pattern was also largely recovered within lineages with contrasting sperm morphologies, indicating that OGs
456 containing reproduction-related genes show elevated ω; regardless of sperm morphology (figure 2B, figure 3, figure S2). These patterns are consistent with expectations of rapid evolution due to intense selection on reproduction-related genes seen also in other studies (Vacquier and Swanson
459 2011; Wilburn and Swanson 2016; but see Dapper and Wade 2016, 2020), including a recent small-scale study investigating protein distance and transcript presence/absence in four *Macrostomum* species (Brand et al. 2020). In addition, we found that species lacking sperm bristles, indicative of
462 hypodermic insemination, generally had higher values of ω. Crucially, even the ubiquitously-expressed OGs, with no obvious reproductive function showed elevated w. This pattern (figure 2, figure 3, and figure S2) strongly suggested that general differences in the evolutionary history of
465 species with contrasting sperm morphologies have influenced molecular sequence evolution throughout the genome, and not just at specific subsets of genes. We discuss possible reasons for this observation in more detail below.

468 In the branch-site models we found relatively little evidence for convergent positive selection at specific amino acid sites in any annotation group. Additionally, although many OGs with evidence for positive selection had sequences from species representing multiple independent losses
471 or reductions of sperm bristles, many only had a single "foreground" lineage driving that result. Such cases do not allow for associating patterns of positive selection with changes in sperm morphology. Evidence of widespread convergent molecular sequence evolution in other systems is
474 mixed. A few examples of convergent amino acid changes in individual genes or groups of genes have been identified in, for example, the Na^+ channel proteins of electric fishes (Zakon et al. 2006), or certain mitochondrial genes in snakes and agamid lizards (Castoe et al. 2009). However, when
477 investigating genome-wide patterns for convergent evolution to marine environments, Foote et al. (2015) did not find a strong overall signal. Similarly, few of the convergent increases in haemoglobin-oxygen affinity among birds living at high altitudes can be attributed to convergent
480 amino acid changes (Natarjan et al. 2016). Such variation is perhaps expected given that selection on complex traits with multiple genetic underpinnings need not necessarily target variants in the

same genes across lineages (Foote et al. 2015; Chikina et al. 2016). Additionally, selection could act on the same genes, but on different sites within those genes, which is a model of molecular sequence evolution not covered by the framework of branch-site models. Consistent with such a model of less strict convergent selection, recent analyses of convergent molecular sequence evolution, using methods that estimate whole-gene evolutionary rates along phylogenetic branches and identifying genes that are strong outliers, found that many genes show a signal of convergent shifts in selective pressure in response to a shift to a marine environment (Chikina et al. 2016), and in parallel adaptation to subterranean lifestyles in mammals (Partha et al. 2019). Understanding the genetic basis for convergent phenotypic evolution and the extent of convergent evolution at the molecular level remains an active field of research (Sackton and Clark 2019).

The finding that ω values in branch tests were higher at reproduction-related genes, even when accounting for the differences between sperm bristle categories, indicated that selection is indeed more intense at these loci. However, the paucity of evidence for convergent selection at specific amino acid sites in the branch-site models suggested that selection is likely acting in different regions of the genes involved. Additionally, some of the alignments shown in figure S8 revealed that uncertain alignments in fragmented regions could potentially confound the signal of positive selection. Alignment uncertainty could reflect biological reality that is not captured by codon models or it could be due to the incomplete nature of the available *de novo* transcriptome assemblies. Genome-guided transcriptome assembly methods could potentially improve transcriptome contiguity and provide for more robust analyses, but currently there is only a single published genome in the genus *Macrostomum* (Wasik et al. 2015; Wudarski et al. 2017).

Because it is not expected that positive selection should be acting across all genes in the genome, the observed genome-wide signal of elevated ω in species with absent bristles spoke to a general feature of the evolutionary history of these species driving patterns of molecular divergence. This raised the possibility that a process such as relaxed selection due to reduced efficiency of purifying selection, rather than positive selection, could be responsible for our observations of genome-wide elevated ω in those species. While relaxed selection could also result from reduced intensity of selection, this would not be expected to affect genome-wide patterns by increasing mean ω throughout, but rather only target specific genes and thus increase the variance in ω. We found that evidence for relaxed selection was common in the present-absent contrast (overall ~32% of OGs), and that it was more common than intensified purifying or positive selection (overall ~5% of OGs). Additionally, there was little overlap between OGs identified as being under positive selection in branch-site tests and OGs identified as being under intensified selection in the RELAX analyses (figure 4, figure S11), which could be partly explained by the different assumptions made

516 in PAML and RELAX. For example, branch-site models in PAML do not allow for positive selection in background branches, while the models in RELAX do (Kosakovsky Pond et al. 2011; Wertheim et al. 2015). Furthermore, the assumption, in RELAX, that intensified selection should 519 simultaneously affect sites under positive selection and sites under purifying selection may not be realistic (i.e. the effects may not necessarily be symmetric).

One possible explanation for genome-wide elevated ω in species with absent bristles is a 522 reduced efficiency of natural selection due to smaller effective population size (N_e) in self-fertilising species. A few species of hypodermically inseminating *Macrostomum*, and at least one reciprocally copulating species (namely *M. mirumnovem*), are known to be capable of self-fertilisation (Ramm 525 et al. 2012; Ramm et al. 2015; Giannakara and Ramm 2017; Giannakara and Ramm 2020, Singh et al. 2020). Some hypodermically inseminating species also show no evidence for inbreeding depression, suggesting they could be preferential self-fertilisers (Giannakara and Ramm 2017), 528 while other species seem to show facultative selfing (Ramm et al. 2012; Ramm et al. 2015). Furthermore, heterozygosity is lower among hypodermically inseminating species (Brand et al. *unpublished data c*), which is consistent with higher rates of selfing in these species driving down 531 genetic diversity and N_e. Relaxed selection, or more specifically reduced efficiency of selection, due to self-fertilisation is in fact predicted from theoretical models (Glémin 2007; Glémin et al. 2019). Support for this is found in a number of plant systems, but evidence from self-fertilising animals is 534 much rarer (reviewed in Hartfield 2016). Sexual conflict over sex roles could have repeatedly driven the evolution of alternative mating strategies (i.e. hypodermic insemination), and the resulting adaptations (injecting stylets), in turn, may facilitate the evolution of selfing (Ramm et al. 537 2012) allowing higher rates of selfing that drive down N_e (Brand et al. *unpublished data c*). This increases ω genome-wide *via* the fixation of deleterious mutations as a result of the reduced efficiency of natural selection (Glémin 2007). We find some evidence for this in that a substantial 540 proportion of OGs show evidence of relaxed selection in species with absent bristles.

To further explore the relationship between a hypodermic mating strategy and N_e, which may contain clues about selfing rates and the efficiency of selection in these species, population genomic 543 data will be necessary. Additionally, more information about the taxonomic distribution of selfing in *Macrostomum* will be invaluable. For example, it is currently unknown whether any of the intermediate species are capable of selfing. Thus, differences in the mating behaviour among 546 species of *Macrostomum* may be shaping genome-wide patterns of molecular sequence evolution, seemingly to a large extent *via* inefficient selection due to lower N_e, likely resulting from higher selfing rates in hypodermically inseminating species. Meanwhile, sexual selection may still be 549 driving the more rapid evolution observed at reproduction-related genes compared to ubiquitously-

expressed genes. This combination of selection and drift is an underappreciated point in the genomic consequences of sexual selection. Different evolutionary forces act together to produce the patterns seen in genomic variation. Such combinations of effects make observed patterns difficult to interpret, especially without functional information with which to classify the genes of interest. Functional information for as many genes as possible is invaluable, and models of molecular sequence evolution that explicitly test for the effects of different evolutionary forces have to be deployed in combination to obtain a comprehensive picture.

In this study we re-capitulate the common finding of rapid evolution at reproduction-related genes and show that this is the case also in hermaphroditic animals. This study therefore highlights the importance of considering diverse sexual systems, including hermaphroditism, as much of the theory and empirical work currently remains focused primarily on separate-sexed organisms. Evidence from hermaphroditic organisms, especially animals, is much rarer. In particular, the effects of sex-biased or sex-limited expression of genes on the efficacy of selection, especially with respect to reproduction-related genes (Dapper and Wade 2016; 2020), can likely be avoided, since in simultaneously hermaphroditic animals a gene cannot be strictly sex-limited in expression. Some studies have investigated rates of sequence evolution at reproduction-related genes in self-compatible, hermaphroditic plants where reproduction-related genes are also often found to evolve rapidly, though the mechanism is still debated (Szövényi et al. 2013; Arunkumar et al. 2013; Gossmann et al. 2013; 2016; Harrison et al. 2019; Moyle et al. 2021). Additionally, the opportunity for haploid selection in plants (e.g. in transcriptionally active pollen) further complicates expectations and comparisons to animal systems, while the evidence for haploid selection in animals is limited, given that it is thought that few genes are expressed in the haploid stages of animals (Joseph and Kirkpatrick 2004, but see Alavioon et al. 2017).

The conditions under which genes could experience differences in the efficiency of selection due to the fact that hermaphroditic individuals have dual sex roles remain largely unexplored. For example, it is plausible that genetic variation for sex allocation (i.e. the degree to which a hermaphroditic individual allocates resources to its male or female function) could become linked to sexually antagonistic genetic variation ("intra-individual sexual antagonism"; Abbott 2011), and that this process might precede the evolution of what could be considered proto-sex chromosomes. Such linkage effects might result in genes that are differentially expressed between more "male-biased" *vs.* more "female-biased" individuals, i.e. genotypes that consistently differ in the relative amount of fitness they obtain through one or the other sex function ("effective gender"; *sensu* Lloyd 1980, sometimes also "functional gender"; Charnov 1982; Verdú et al. 2004). In addition, genes involved in sperm competition traits experience less efficient selection, because the genetic

variation in these traits that is present within a multiply mated female only represents a fraction of the total genetic variation in the population, especially when female re-mating rates are low (Dapper
585 and Wade 2016; 2020). Intuition would suggest that a reduced efficiency of selection at genes related to sperm competition should be expected also in hermaphrodites. Thus, some caution in interpretation remains warranted when comparing separate-sexed organisms and hermaphrodites,
588 and relaxed selection should at least be considered as a possible explanation for elevated rates of sequence evolution in (some) reproduction-related genes also for simultaneous hermaphrodites. Exploring these ideas will require further development of theory to clarify expectations.

591 Finally, our results also allow generating candidates for functional investigations on genes that may underlie some of the interesting divergence in morphology observed throughout the genus *Macrostomum*. In the Supplementary Materials (Supplementary Text T3) we highlight a few OGs
594 that we opportunistically selected based on a combination of i) an examination of individual alignments (including alignment quality and the number of independent losses/reductions of sperm bristles represented), ii) evidence of rapid evolution (for example, values of ω above the median for
597 ubiquitously-expressed OGs from branch models), iii) evidence for positive selection (from branch-site models), and/or iv) more detailed expression information, where available. Several of these represent intriguing candidates for the evolution of reproductive traits, including the sperm bristles.
600 Indeed, one already has an established knockdown phenotype, producing aberrant sperm morphology in *M. lignano* (Grudniewska et al. 2018). Another candidate is known to be expressed in the developing stylet (Weber et al. 2018), which is another aspect of the extreme diversity in
603 reproductive morphology in this genus. Thus, our findings, combined with an examination of the alignments, allow the generation of promising individual candidate genes for further investigation. Some of these genes might be good candidates to explore in the context of differences among
606 species in sperm morphology and other sperm traits, in the stylet morphology, in the seminal fluid, and in other reproductive functions. Combined with more detailed information about gene expression patterns, these candidates can provide great insights and guide the exploration of gene-
609 phenotype associations across the genus.

Conclusions

612 We have conducted a comparative study of molecular sequence evolution in a genus of simultaneously hermaphroditic flatworms. To our knowledge this is the first robust study of molecular sequence evolution at reproduction-related genes in hermaphroditic animals. This study
615 highlights the importance of studies that contrast lineages with different mating strategies and associated morphologies and behaviours, which presumably also differ in the sexual selection

context they experience. Importantly, the extent of taxon sampling allows the comparison of multiple independent changes in sperm morphology (and the associated mating strategies), a feature largely lacking from previous studies. Furthermore, we expand the scope of work investigating the effects of sexual selection on genomic patterns of variation to include hermaphroditic animals. We show that changes in sperm morphology are associated with elevated ω values throughout the genome, which suggests a general feature of the evolutionary history of species lacking bristles driving patterns of sequence divergence. A possible explanation lies in the mating strategy associated with the bristle-less sperm morphology, namely hypodermic insemination. Hypodermically inseminating species may engage in high rates of self-fertilisation, which should drive down effective population sizes and increase the effect of genetic drift throughout the genome. On top of these genome-wide patterns we also find higher ω values at reproduction-related genes consistent with a strong effect of sexual selection acting on such genes.

Materials and Methods

Candidate reproduction-related genes

Previous work on the model species *M. lignano* has characterised gene expression in different body regions of adult worms using a "positional RNA-Seq" approach (Arbore et al. 2015). Briefly, worms were cut once just anterior to the testes, between testes and ovaries, just posterior to the ovaries, or not at all, yielding RNA samples that represented pools of progressively larger fragments of the worm. To make use of more recent and improved genome and transcriptome assemblies (Wudarski et al. 2017; Grudniewska et al. 2018), Brand et al. (2020) repeated the analyses of that RNA-Seq data using transcriptome assembly Mlig_RNA_3_7_DV1.v3 (Grudniewska et al. 2018). Transcripts could then be identified with specific expression in the testis-, ovary-, or tail-region of the worm, while other transcripts could be identified as being likely ubiquitously-expressed across the regions. Additionally, some transcripts were annotated as being "low-expressed" (i.e. with <50 reads mapping from the uncut worm sample) for downstream filtering (see Additional file 14 in Brand et al. 2020).

Ortholog inference and codon-alignment preparation

We made use of recently generated whole transcriptome data for 97 species of *Macrostomum* and three outgroups (i.e. three undescribed species of the sister genus *Psammomacrostomum*; Janssen et al. 2015), including clusters of homologous sequences from across all species (for details see Brand et al. *unpublished data a*). Briefly, homolog groups for all identified protein-coding genes in the assembled transcriptomes were obtained by first conducting an all-vs-all BLASTp search and

651 then clustering these results using OrthoFinder (v. 1.1.10; Emms and Kelly 2015), keeping only homolog groups containing at least 10 species. These homolog groups were aligned with MAFFT (v. 7.3; Katoh and Standley 2013), the alignments were trimmed with ZORRO (v. 1.0; Wu et al. 654 2012), and finally gene trees were built with IQ-TREE (v. 1.5.6; Nguyen et al. 2015) using the best fitting model of molecular sequence evolution for each alignment (Brand et al. *unpublished data a*).

For our study, we identified those homolog groups, and their associated gene trees, which 657 contained a *M. lignano* gene, either from one of the above-mentioned reproduction-related tissue regions or from a randomly chosen set of 1,000 ubiquitously-expressed genes, and annotated them according to the inferred function of the *M. lignano* gene (yielding four functional annotation 660 groups). We then pruned gene trees in three steps according to the procedure outlined by Yang and Smith (2014). First, we pruned tips from the tree if they were longer than an absolute length of 2 substitutions per site or if the length of a tip was > 10x the length of its sister tip. Next, if a clade 663 was composed of multiple sequences from just a single species, we retained only the sequence with the most aligned columns. Finally, we split trees into subtrees at long internal branches, and only retained the resulting subtrees if the number of species in these trees was ≥ 5. Finally, we inferred 666 ortholog groups (OGs) from the resulting subtrees using both the "rooted ingroups" (RT) and "maximum inclusion" (MI) algorithms (Yang and Smith 2014). We used the union of OGs inferred by both of these methods in downstream analyses, always keeping the OG produced by RT where 669 there was overlap, because they were typically larger.

We then extracted the amino acid and nucleotide sequences corresponding to the coding region of these OGs from each species' transcriptome assembly, aligned them using MAFFT, and 672 converted the amino acid alignments to codon alignments with pal2nal (v. 14; Suyama et al. 2006). We subsequently trimmed these alignments by first computing column-wise alignment scores for the amino acid alignments with ZORRO and then removing the codon columns from the codon 675 alignments that corresponded to low-scoring (score ≤ 3) amino acid columns. If trimmed, gap-less sequences were shorter than 50% of the alignment length, we removed these sequences from the alignment. Finally, we checked alignments for internal stop codons and removed alignments < 80 678 nucleotides long. Note that these last alignment trimming steps in some cases led to the removal of the original *M. lignano* sequence from the alignment. Where this occurred (N = 19, 1.7%), we nevertheless retained these OGs, preserving the annotation from *M. lignano*, for further analysis. 681 We used a recently produced maximum likelihood species tree based on 385 genes with an occupancy rate of >80% across the genus *Macrostomum*, representing 94,625 aligned amino acids (and referred to as "H-IQ-TREE" in Brand et al. *unpublished data a*; figure 1C). Since our 684 alignments contained different sets and numbers of species, we pruned the species tree to contain

only the species actually present in each alignment. Only alignments with at least four species were retained.

687 The pipeline outlined above involves many steps, each of which has the potential to discard genes from the analysis. To evaluate the success of recovering OGs from across the species, we quantified the proportion of OGs remaining that contained an annotated gene from *M. lignano* after 690 each step (figure S1). We found that there is a substantial loss during orthology identification regardless of the functional annotation of the *M. lignano* gene (table S3; figure S1). However, the trend suggests that recovery is somewhat lower (13.6-16.4%) for reproduction-related genes than 693 for ubiquitously expressed genes (25.7%). These results suggest that ortholog identification, in general, has a low recovery rate, at least on the phylogenetic scales investigated here, but also that recovery is worse for groups of reproduction-related genes that are expected to be evolving more 696 rapidly (see also Brand et al. 2020). For all analyses below, the final trimmed alignments used are made available in a supplementary archive (Supplementary Archive 1, available from: https://doi.org/10.5281/zenodo.4972188).

699

Branch models

In order to test whether species with different sperm morphologies show evidence for 702 different patterns of molecular sequence evolution we first classified species into three categories of sperm morphology according to Brand et al. (*unpublished data b*; figure 1B), namely species with bristles present, reduced, or absent. We then identified orthologs that showed variation in *dN/dS* 705 ratios across these three bristle states by fitting a series of codon substitution branch models using the codeml software in the PAML package (v.4.9; Yang et al. 2000; Yang 2007). As outlined in the Introduction, *dN/dS* (ω) is the ratio of non-synonymous (*dN*) to synonymous (*dS*) substitution rates. 708 This value allows inference of selection on a protein-coding gene, with $\omega = 1$ implying that the gene is evolving neutrally, while $\omega > 1$ or $\omega < 1$ are generally considered to imply positive or purifying selection, respectively (but see also below). First, we fit a model that assumes a single ω across the 711 entire tree for each OG alignment. Second, we fit a model where terminal branches leading to species with reduced bristles and species with absent bristles were labelled separately, effectively assuming a different ω for each of the three bristle states (note that the alternative model does not 714 require that all branches have different estimates for ω). Both models have the parameters "fix_omega = 0", "omega = 0.001, "fix_kappa = 0", "kappa = 2", differing only in the input tree with either a single branch label (null model) or three branch labels (alternative model). These 717 models are then compared using a likelihood-ratio test (LRT) with 2 degrees of freedom for each model comparison. FDR was controlled at a rate of 0.05 by converting p-values from the LRT to q-

values (Storey and Tibshirani 2003). Note that, by definition, q-values can never be larger than the
720 estimated π_0 parameter, and therefore with a low estimate for π_0, q-values can be smaller than the corresponding p-value. We then compare ω estimates from the model with a single ω between annotation groups using a Kruskal-Wallis rank-sum test, followed by a Dwass-Steele-Critchlow-
723 Fligner all-pairs post-hoc test. We also tested whether the proportion of OGs for which the alternative model was a better fit, was larger than expected in some annotation groups using χ^2-tests. The expected proportions were based on the distribution of all OGs across annotation groups.

726 Finally, we evaluated the mean level of codon usage bias for species without bristles using the method of Karlin et al. (2001), always setting species with bristles in each alignment as the reference set. This approach determines whether there are codon-usage biases among species with
729 absent bristles compared to species with bristles present. Such a systematic difference could potentially be a confounding variable in a comparison of molecular sequence evolution between two groups.

732

Branch-site models

The branch models we described above fit average estimates of ω across the entire alignment,
735 but the entire coding sequence of a gene is often not expected to be under the same kind of selection. Rather, in many cases only a few amino acid sites are likely the targets of intense selection (Yang and Nielsen 2002; Zhang et al. 2005). Therefore, we also fit branch-site models, as
738 implemented in PAML. Since branch-site models only allow for two branch classes ("foreground" and "background"), we performed two contrasts. We set species with absent bristles or species with reduced bristles as the foreground branch in the present-absent and the present-reduced contrasts,
741 respectively, and species with bristles present were always set as the background branches. We chose these contrasts because reciprocal copulation is thought to be the ancestral state for a large part of the *Macrostomum* phylogeny (Brand et al. *unpublished data b*).

744 We fit the branch-site model A (Zhang et al. 2005), which has four site classes (0, 1, 2a, and 2b). This model allows two of these classes to have $\omega > 1$ among branches labelled as foreground (namely 2a and 2b, with parameter ω_2). Model A was specified with the following PAML
747 parameters "model = 2", "NSsites = 2", "fix_omega = 0", "omega = 0.001". The other site classes are constrained to $0 < \omega < 1$ (site class 0, with parameter ω_0), or $\omega = 1$ (site class 1, with parameter ω_1) in both background and foreground branches. We also fit the corresponding null model (A_{null},
750 specified with PAML parameters "model = 2", "NSsites = 2", "fix_omega = 1", "omega = 1"). A_{null} is otherwise identical, but the parameter ω_2 is fixed to 1 (Zhang et al. 2005). Thus, the null model allows for sites that were under purifying selection in the background branches, but are now

evolving neutrally in the foreground branches (site class 2a) as well as sites that are evolving neutrally in both the background and foreground branches (site class 2b; see Zhang et al. 2005 for full details). The null and alternative models were then compared using an LRT with 2 degrees of freedom (as above). Branch-site tests, as implemented in PAML, also allow the assignment to each site in the alignment of a posterior probability that it belongs to a particular site class (Yang et al. 2005). The probability that a site belongs to a site class under positive selection in the foreground branch is of particular interest. Thus, for those OGs with evidence of positive selection we identified sites with a high posterior probability (>0.95) of being under positive selection.

These branch-site tests amount to a stringent test for a pattern of positive selection, in common to the lineages either lacking or having reduced sperm bristles, on a few sites of a protein-coding gene. Where alignments include multiple independent origins (see also figure S5), this can be taken as a test of convergent molecular sequence evolution associated with shifts in sperm morphology. Though many OGs contain multiple species with absent bristles, these may often only represent one or two phylogenetically independent losses (e.g. multiple representatives of the morphologically uniform hypodermically mating clade, top of figure 1C). OGs with more independent changes are arguably more informative concerning convergent molecular sequence evolution, and therefore more informative about patterns of molecular sequence evolution associated with differences or changes in bristle morphology. Simply correcting for multiple testing would not account for the variation across OGs in the number of phylogenetically independent losses or reductions of bristles that are represented within each OG. Therefore, in order to keep as much information from independent changes as possible, we chose to retain all OGs with an uncorrected LRT p-value < 0.05. This permitted a more nuanced exploration of results stratified by the number of independent losses within OGs. However, we nevertheless also provide overall q-values (as for the branch tests above). Among genes inferred to show evidence for positive selection (LRT p-values < 0.05) we tested whether there were differences in the ω_2 parameter estimates from foreground branches across annotation groups using Kruskal-Wallis rank-sum tests. We further tested whether some annotation groups had a larger proportion of OGs for which the alternative model was a better fit, than expected based on the distribution of OGs tested as a proportion of all OGs using χ^2-tests.

RELAX models

Finally, we also performed analyses with RELAX (Wertheim et al. 2015) in the HyPhy suite of analysis tools (v. 2.5.29; Kosakovsky Pond et al. 2019). In the hypothesis-testing framework we use here, RELAX compares a set of test branches to a set of reference branches in a phylogeny, in

order to test if the distribution of ω values across sites is different between them. Specifically, it fits a discrete distribution of ω values with $n = 3$ site categories to model the distribution of ω (such that
789 the three categories have $\omega_0 \leq \omega_1 \leq 1 \leq \omega_2$), and the proportion of sites in each category is also estimated from the data. It further assumes that the ω values of the distribution for the test branches is related to the distribution in reference branches by a value, *k,* that describes either relaxation or
792 intensification of selection. It does so by comparing two models by an LRT, namely a null model where $k = 1$ and an alternative model where k can take any value > 0. If the LRT rejects the null model it indicates that selection is either intensified (when $k > 1$) or that selection is relaxed (when k
795 < 1). Note that in the hypothesis-testing framework, the proportion of sites in each category is constrained to be the same across test and reference branches. RELAX can also fit a less constrained Partitioned Descriptive Model (PDM) where the ω values as well as the proportion of
798 sites in each category may vary between test and reference branches. The PDM estimates can then be used to make more refined interpretations of differences between test and reference branches, while hypothesis testing is performed with the null and alternative models described above
801 (Wertheim et al. 2015).

In the RELAX framework, relaxation of selection should result in the distribution of ω values becoming more narrowly distributed around a value of $\omega = 1$ (Wertheim et al. 2015), since *dN* and
804 *dS* become more similar. In contrast, intensified selection results in the distribution of ω becoming wider. In other words, RELAX assumes that intensified selection simultaneously intensifies both positive and purifying selection throughout the gene. As above, we ran RELAX for data sets that
807 only include two sperm bristle categories (present-absent and present-reduced), and we adjusted LRT p-values with the q-value method to control the FDR. We then split those data where the null model was rejected into those where $k < 1$ and $k > 1$, respectively. Finally, we compared the
810 distribution of k values across the reproduction-related annotation groups within these two groups using Kruskal-Wallis rank-sum tests.

813 *Data analysis and processing*

Unless otherwise stated, all statistical tests have been performed in R (v. 3.6.3; R Development Core Team 2020) and figures are plotted with the "ggplot2" package (v. 3.3.1;
816 Wickham 2016) with the aid of various helper packages: "reshape" (v. 0.8.8; Wickham 2007), "plyr" (v. 1.8.6; Wickham 2011), and "gridExtra" (v. 2.3; Auguie 2017). The Dwass-Steele-Critchlow-Fligner post-hoc all-pairs tests were performed with tools from the "PCMRplus" package
819 (v. 1.4.4; Pohlert 2020), MANOVA and mixed model results given in the Supplementary Materials were obtained with the "MANOVA.RM" (v. 0.4.3; Friedrich et al. 2021) and "nlme" (v. 3.1-149;

Pinheiro et al. 2020) R packages respectively. Q-values were computed using the "qvalue" package (v. 2.16; Storey et al. 2019). The phylogenetic tree was processed and plotted with tools from the "ape" (v. 5.4; Paradis and Schliep 2019), "phytools" (v. 0.7-47; Revell 2012), and "ggtree" (v. 1.17.4; Yu et al. 2017) packages. Codon usage bias analyses were performed with the "coRdon" package (v. 1.2.0; Elek et al. 2019). See the R scripts in the supplementary materials for full details.

Acknowledgments

This work was supported by grants from the Swiss National Science Foundation (grant numbers 31003A_162543 and 310030_184916 to LS). We also thank Lukas Zimmermann for IT support. We would also like to thank Mike Wade and Lynda Delph for valuable discussion. Computations were performed, in part, at sciCORE (http://scicore.unibas.ch/) scientific computing center at the University of Basel.

Data availability

Transcriptome assemblies forming the basis of these analyses are available from https://doi.org/10.5281/zenodo.4543289. The final codon alignments, in addition to analysis scripts and datasets used in the above analyses are provided as supplementary files to this manuscript and are available from: https://doi.org/10.5281/zenodo.4972188.

References

Aagaard JE, Springer SA, Soelberg SD, Swanson WJ. 2013. Duplicate abalone egg coat proteins bind sperm lysin similarly, but evolve oppositely, consistent with molecular mimicry at fertilisation. *PLoS Genet.* 9: e1003287.

Abbott J. 2011. Intra-locus sexual conflict and sexually antagonistic genetic variation in hermaphroditic animals. *Proc. R. Soc. B Biol. Sci.* 278: 161-169.

Alavioon G, Hotzy C, Nakhro K, Rudolf R, Scofield DG, Zajitschek S, Maklakov AA, Immler S. 2017. Haploid selection within a single ejaculate increases offspring fitness. *Proc. Natl. Acad. Sci. USA.* 114:8053-8058.

Andersson M. 1994 *Sexual Selection*. Eds. J. Krebs and T. Clutton-Brock. Princeton University Press. Princeton, New Jersey, USA.

Arbore R, Sekii K, Beisel C, Ladurner P, Berezikov E, Schärer L. 2015. Positional RNA-Seq identifies candidate genes for phenotypic engineering of sexual traits. *Front. Zool.* 12: 14.

Arnqvist G, Rowe L. 2005. *Sexual Conflict.* Princeton University Press. Princeton, New Jersey, USA.

Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Mol. Biol. Evol*. 30: 2475-2486.

Auguie B. 2017. gridExtra: Miscellaneous Functions for "Grid" Graphics. Available from: https://CRAN.R-project.org/package=gridExtra

Brand JN, Wiberg RAW, Pjeta R, Bertemes P, Beisel C, Ladurner P, Schärer L. 2020. RNA-Seq of three free-living flatworm species suggests rapid evolution of reproduction-related genes. *BMC Genomics* 21: 462.

Brand JN, Viktorin G, Wiberg RAW, Beisel C, Schärer L. Large-scale phylogenomics of the genus *Macrostomum* (Platyhelminthes) reveals cryptic diversity and novel sexual traits. *unpublished data a*. preprint at *bioRxiv*: https://doi.org/10.1101/2021.03.28.437366

Brand JN, Harmon LJ, Schärer L. Frequent origins of traumatic insemination involve convergent shifts in sperm and genital morphology. *unpublished data b*. preprint at *bioRxiv*: https://doi.org/10.1101/2021.02.16.431427

Brand JN, Harmon LJ, Schärer L. The evolution of sex allocation is linked to hypodermic insemination. *unpublished data c*

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.

Castoe TA, de Koning APJ, Kim H-M, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106: 8986-8991.

Charnov EL. 1976. Simultaneous hermaphroditism and sexual selection. *Proc. Natl. Acad. Sci. U.S.A.* 76: 2480-2484.

Charnov EL. 1982. *The theory of sex allocation.* Princeton University Press. Princeton, New Jersey, USA.

Chikina M, Robinson JD, Clark NL. 2016. Hundreds of genes experienced convergent shifts in selective pressure in mammals. *Mol. Biol. Evol.* 33: 2182-2192.

882 Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. 2009. Coevolution of interacting fertilisation proteins. *PLoS Genet.* 5:e1000570.

Dapper A, Wade M. 2020. Relaxed selection and the rapid evolution of reproductive genes. *Trends*
885 *in Genetics*. 36: 640-649.

Dapper A, Wade M. 2016. The evolution of sperm competition genes: The effect of mating system on levels of genetic variation within and between species. *Evolution*. **70**: 502-511.

888 Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, MacCoss MJ, Nachman MW. 2009. Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Mol. Biol. Evol.* 26: 1733-1743.

891 Dorus S, Evans PD, Wyckoff GJ, Sun SC, Lahn BT. 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat. Genet.* 36: 1326-1329.

894 Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* 8: 689-698.

Elek A, Kuzman M, Vlahovicek K. 2019. coRdon: Codon usage analysis and prediction of gene
897 expressivity. Available from: https://github.com/BioinfoHR/coRdon

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16: 157.

900 Findlay GD, Yi X, MacCoss MJ, Swanson WJ. 2008. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol.* 6: e178.

Fletcher W, Yang Z. 2010. The effect of insertions, deletions, and alignment errors on the branch-
903 site test of positive selection. *Mol. Biol. Evol.* 27: 2257-2267.

Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, Dugan S, Elk CE Van, Hunter ME, Joshi V, et al. 2015. Convergent evolution of the genomes of marine mammals. *Nat. Genet.* 47:
906 272-274.

Frank SA, Patten MM. 2019. Sexual antagonism leads to a mosaic of X–autosome conflict. *Evolution* 74: 495-498.

909 Friedrich S, Konietschke F, Pauly M. 2021. MANOVA.RM: Resampling-based analysis of multivariate data and repeated measures designs. Available from: https://CRAN.R-project.org/package=MANOVA.RM

912 Grath S, Parsch J. 2016. Sex-biased gene expression. *Annu. Rev. Genet.* 50: 29-44.

Gavrilets S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* 403: 886-889.

915 Gharib WH, Robinson-Rechavi M. 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol. Biol. Evol.* 30: 1675-1686.

918 Giannakara A, Ramm SA. 2017. Self-fertilisation, sex allocation and spermatogenesis kinetics in the hypodermically inseminating flatworm *Macrostomum pusillum*. *J. Exp. Biol.* 220: 1568-1577.

921 Giannakara A, Ramm SA. 2020. Evidence for inter-population variation in waiting times in a self-self-fertilising flatworm. *Invertebr. Reprod. Dev.* 64: 158-168.

Glémin S. 2007. Mating systems and the efficacy of selection at the molecular level. *Genetics* 916:
924 905-916.

Glémin S, François CM, Galtier N. 2019. Genome evolution in outcrossing vs. selfing vs. asexual species. Pp. 331-369. in *Evolutionary Genomics: Statistical and Computational Methods,*
927 *Methods in Molecular Biology.* Ed. M. Anisimova. Humana, New York

Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2013. Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 31:
930 574-583.

Gossmann TI, Saleh D, Schmid MW, Spence MA, Schmid KJ. 2016. Transcriptomes of plant gametophytes have a higher proportion of rapidly evolving and young genes than sporophytes.
933 *Mol. Biol. Evol.* 33: 1669-1678.

Grudniewska M, Mouton S, Grelling M, Wolters AHG, Kuipers J, Giepmans BNG, Berezikov E. 2018. A novel flatworm-specific gene implicated in reproduction in *Macrostomum lignano*.
936 *Sci. Rep.* 8: 3192.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related
939 genes in *Drosophila*. *Genetics.* 177: 1321-1335.

Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, Mank JE. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 112: 4393-4398.

Harrison MC, Mallon EB, Twell D. Hammond RL. 2019. Deleterious mutation accumulation in *Arabidopsis thaliana* pollen genes: A role for a recent relaxation of selection. *Genome Biol. Evol.* 11: 1939-1951.

Hartfield M. 2016. Evolutionary genetic consequences of facultative sex and outcrossing. *J. Evol. Biol.* 29: 5-22.

Hayashi TI, Vose M, Gavrilets S. 2007. Genetic differentiation by sexual conflict. *Evolution* 61: 516-529.

Hurle B, Swanson W, Green ED. 2007. Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. *Genome Res*. 17: 276-286.

Janssen T, Vizoso DB, Schulte G, Littlewood DTJ, Waeschenbach A, Schärer L. 2015. The first multi-gene phylogeny of the Macrostomorpha sheds light on the evolution of sexual and asexual reproduction in basal Platyhelminthes. *Mol. Phylogenet. Evol.* 92: 82-107.

Joseph SB, Kirkpatrick M. 2004. Haploid selection in animals. *Trends Ecol. Evol.* 19: 592-597.

Karlin S, Mrázek J, Campbell A, Kaiser D. 2001. Characterisations of highly expressed genes of four fast-growing bacteria. *J. Bacteriol.* 183: 5025-5040.

Kasimatis KR, Nelson TC, Phillips PC. 2017. Genomic signatures of sexual conflict. *J. Hered.* 108: 780-790.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30: 772-780.

Kosakovsky Pond SL, Murrel B, Fourment M, Frost SDW, Delport W, Scheffler K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* 28: 3033-3043.

Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, et al. 2019. HyPhy 2.5 — A customisable platform for evolutionary hypothesis testing using phylogenies. *Mol. Biol. Evol.* 37: 295-299.

Lande R. 1981. Models of speciation by sexual selection on polygenic traits. *Proc. Natl. Acad. Sci. U.S.A.* 78: 3721-3725.

Ladurner P, Schärer L, Salvenmoser W, Rieger RM. 2005. A new model organism among the lower Bilateria and the use of digital microscopy in taxonomy of meiobenthic Platyhelminthes: *Macrostomum lignano*, n. sp. (Rhabditophora, Macrostomorpha). *J. Zool. Syst. Evol. Res*. 43: 114-126.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357-359.

Lengerer B, Wunderer J, Pjeta R, Carta G, Kao D, Aboobaker A, Beisel C, Berezikov E, Willi S, Ladurner P. 2018. Organ specific gene expression in the regenerating tail of *Macrostomum lignano*. *Dev. Biol.* 433: 448-460.

Lloyd, D. G. 1980. Sexual strategies in plants III. A quantitative method for describing the gender of plants. *N.Z. J. Zool.* 18:103-108.

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

Marie-Orleach L, Vogt-Burri N, Mouginot P, Schlatter A, Vizoso DB, Bailey NW, Schärer L. 2017. Indirect genetic effects and sexual conflicts: Partner genotype influences multiple morphological and behavioral reproductive traits in a flatworm. *Evolution.* 71: 1232-1245.

Michiels NK. 1998. Mating conflicts and sperm competition in simultaneous hermaphrodites. Pp 219-254. in *Sperm Competition and Sexual Selection.* Eds. Birkhead TR and Møller AP. Academic Press. London

Moyle LC, Wu M, Gibson MJS. 2021. Reproductive proteins evolve faster than non-reproductive proteins among *Solanum* species. *Front. Plant Sci.* 12: 635990.

Muyle A. 2019. How different is the evolution of sex-biased gene expression between plants and animals? A commentary on: 'Sexual dimorphism and rapid turnover in gene expression in pre-reproductive seedlings of a dioecious herb' *Ann. Bot.* 123: iv-v.

Natarjan C. Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. 2016. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science.* 354: 336-339.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32: 268-274.

Paradis E. Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 35: 526-528.

Partha R, Chauhan BK, Ferreira Z, Robinson JD, Lathrop K, Nischal KK, Chikina M, Clark NL. 2019. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife.* 6: e25884.

Patlar B, Weber M, Ramm SA. 2019. Genetic and environmental variation in transcriptional expression of seminal fluid proteins. *Heredity.* 122: 595-611.

Patlar B, Weber M, Temizyürek T, Ramm SA. 2020. Seminal fluid-mediated manipulation of post-mating behavior in a simultaneous hermaphrodite. *Curr. Biol.* 30: 143-149

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods.* 14: 417-419.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2020. nlme: linear and nonlinear mixed effects models. Available from: https://CRAN.R-project.org/package=nlme.

Pohlert T. 2020. PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended. Available from: https://CRAN.R-project.org/package=PMCMRplus

R Development Core Team. 2019. R: A language and environment for statistical computing. Available from: http://www.r-project.org.

Ramm SA, Lengerer B, Arbore R, Pjeta R, Wunderer J, Giannakara A, Berezikov E, Ladurner P, Schärer L. 2019. Sex allocation plasticity on a transcriptome scale: Socially-sensitive gene expression in the hermaphroditic flatworm *Macrostomum lignano*. *Mol. Ecol.* 28: 2321-2341.

Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD. 2008. Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Mol. Biol. Evol.* 25: 207-219.

Ramm SA, Schlatter A, Poirier M, Schärer L. 2015. Hypodermic self-insemination as a reproductive assurance strategy. *Proc. R. Soc. B Biol. Sci.* 282: 20150660.

Ramm SA, Vizoso DB, Schärer L. 2012. Occurrence, costs and heritability of delayed selfing in a free-living flatworm. *J. Evol. Biol.* 25: 2559-2568.

dos Reis M, Yang Z. 2013. Why do more divergent sequences produce smaller nonsynonymous/synonymous rate ratios in pairwise sequence comparisons. *Genetics*. 195: 195-204.

Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3: 217-223.

Rowe L, Chenoweth SF, Agrawal AF. 2018. The genomics of sexual conflict. *Am. Nat.* 192: 274-286.

Sackton TB and Clark N. 2019. Convergent evolution in the genomics era: New insights and directions. *Phil. Trans. R. Soc. B.* 374: 20190102.

Schärer L, Janicke T. 2009. Sex allocation and sexual conflict in simultaneously hermaphroditic animals. *Biol. Lett.* **5**: 705-708.

Schärer L, Janicke T, Ramm SA. 2015. Sexual conflict in hermaphrodites. *Cold Spring Harb. Perspect. Biol.* 7: a017673.

Schärer L, Littlewood DTJ, Waeschenbach A, Yoshida W, Vizoso DB. 2011. Mating behavior and the evolution of sperm design. *Proc. Natl. Acad. Sci. U.S.A.* 108: 1490-1495.

Singh P, Vellnow N, Schärer L. 2020a. Variation in sex allocation plasticity in three closely related flatworm species. *Ecol. Evol.* 10: 26-37.

Singh P, Ballmer DN, Laubscher M, Schärer L. 2020b. Successful mating and hybridisation in two closely related flatworm species despite significant differences in reproductive morphology and behaviour. *Sci. Rep.* 10: 12830.

Storey JD, Tibshirani R. 2003. Statistical significance for genome-wide studies. *Proc. Natl. Acad. Sci. U.S.A.* 100: 9440-9445.

Storey JD, Bass AJ, Dabney A, Robinson D. 2019. qvalue: Q-value estimation for false discovery rate control. Available from: http://github.com/jdstorey/qvalue

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34: 609-612.

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. U.S.A.* 98: 2509-2514.

Swanson WJ. Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3: 137-144.

1056 Szövényi P, Ricca M, Hock Z, Shaw JA, Shimizu KK, Wagner A. 2013. Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. *Mol. Biol. Evol.* 30:1929-1939.

1059 Vacquier VD, Swanson WJ. 2011. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. *Cold Spring Harb. Perspect. Biol.* 3: e002931.

Venkat A, Hahn MW, Thornton JW. 2018. Multinucleotide mutations cause false inferences of
1062 lineage-specific positive selection. *Nat. Ecol. Evol.* 2: 1280-1288.

Verdú M, Montilla AI, Pannell J. 2004. Paternal effects on functional gender account for cryptic dioecy in a perennial plant. *Proc. R. Soc. B Biol. Sci.* 271: 2017-2023.

1065 Vicens A, Gómez Montoto L, Couso-Ferrer F, Sutton KA, Roldan ERS. 2014. Sexual selection and the adaptive evolution of PKDREJ protein in primates and rodents. *Mol. Hum. Reprod.* 21: 146-156.

1068 Vizoso DB, Rieger G, Schärer L. 2010. Goings-on inside a worm: Functional hypotheses derived from sexual conflict thinking. *Biol. J. Linn. Soc.* 99: 370-383.

Wasik K, Gurtowski J, Zhou X, Ramos OM, Delás MJ, Battistoni G, El Demerdash O, Falciatori I,
1071 Vizoso DB, Smith AD, Ladurner P. 2015. Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano*. *Proc. Natl. Acad. Sci. U.S.A.* 112: 12462-12467.

Weber M, Wunderer J, Lengerer B, Pjeta R, Rodrigues M, Schärer L, Ladurner P, Ramm SA. 2018.
1074 A targeted in situ shybridisation screen identifies putative seminal fluid proteins in a simultaneously hermaphroditic flatworm. *BMC Evol. Biol.* 18: 81.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX : Detecting
1077 relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32: 820-832.

Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Wickham H. 2011. The split-apply-combine strategy for data analysis. *J. Stat. Softw.* 40: 1-29.

1080 Wickham H. 2007. Reshaping data with the reshape package. *J. Stat. Softw.* 21.

Wilburn DB, Swanson WJ. 2016. From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *J. Proteomics.* 135: 12-25.

1083 Wilkinson GS, Breden F, Mank JE, Ritchie MG, Higginson AD, Radwan J, Jaquiery J, Salzburger W, Arriero E, Barribeau SM, et al. 2015. The locus of sexual selection: Moving sexual selection studies into the post-genomics era. *J. Evol. Biol.* 28: 739-755.

1086 Wu M, Chatterji S, Eisen JA. 2012. Accounting for alignment uncertainty in phylogenomics. *PLoS ONE* 7: e30288.

Wudarski J, Egger B, Ramm SA, Schärer L, Ladurner P, Zadesenets KS, Rubtsov NB, Mouton S,
1089 Berezikov E. 2020. The free-living flatworm *Macrostomum lignano*. *EvoDevo* 11: 1-8.

Wudarski J, Simanov D, Ustyantsev K, De Mulder K, Grelling M, Grudniewska M, Beltman F, Glazenburg L, Demircan T, Wunderer J, et al. 2017. Efficient transgenesis and annotated
1092 genome sequence of the regenerative flatworm model *Macrostomum lignano*. *Nat. Commun*. 8: 2120.

Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and
1095 low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* 31:3081-3092.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24: 1586-
1098 1591.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15: 496-503.

1101 Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19: 908-917.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous
1104 selection pressure at amino acid sites. *Genetics.* 155: 431-449.

Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22: 1107-1117.

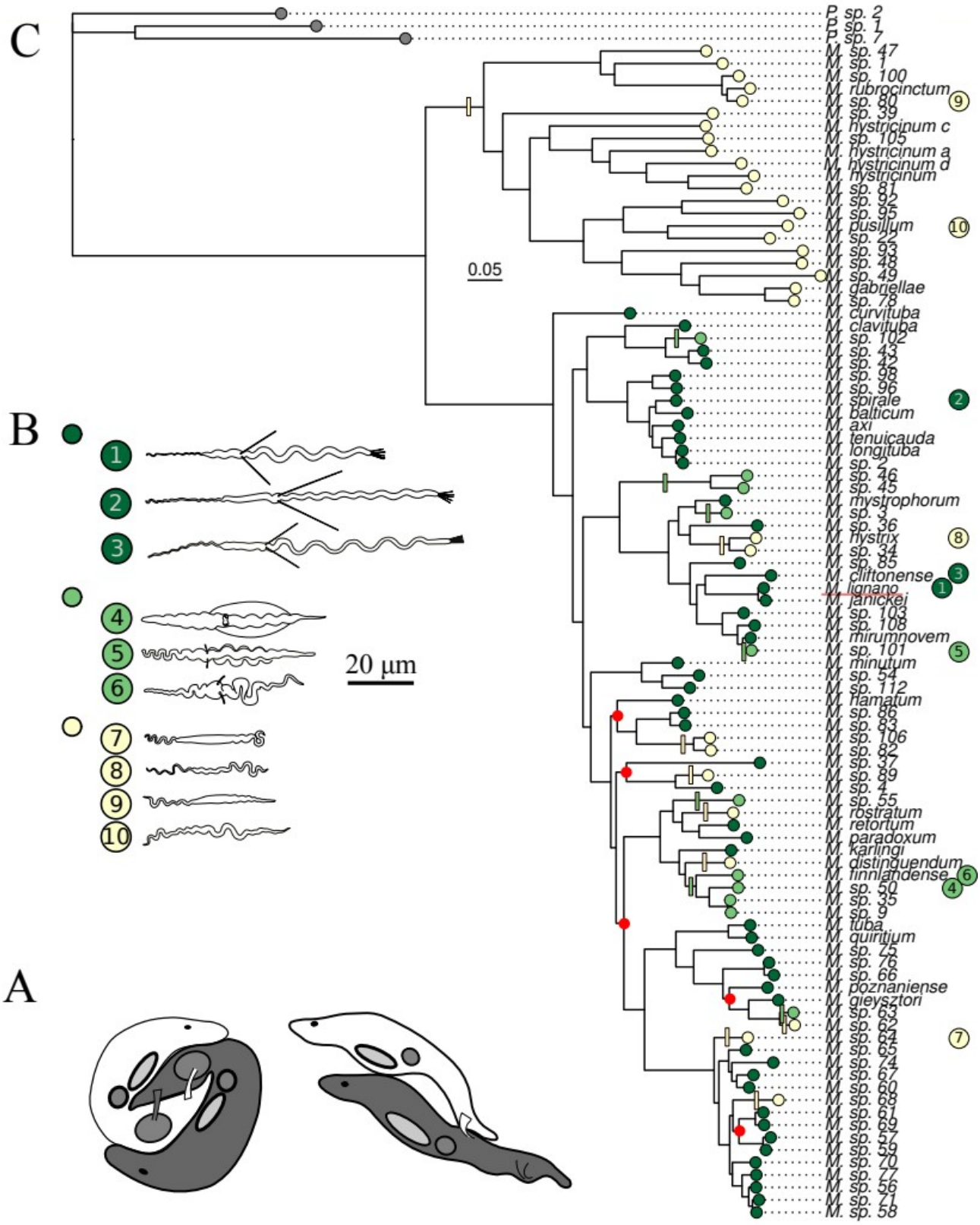
1107 Yang Z, dos Reis M. 2011. Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* 28: 1217-1228.

Yu G, Smith D, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an R package for visualization and
1110 annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8: 28-36.

Zakon HH, Lu Y, Zwicki DJ, Hillis DM. 2006. Sodium channel genes and the evolution of diversity
1113 in communication signals of electric fishes: Convergent molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 103: 3675-3680.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for
1116 detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22: 2472-2479.

Figures and Tables



1119 **Figure 1 (on previous page). A** Stylised drawings of *Macrostomum* species with reciprocal copulation (left) and hypodermic insemination (right), respectively, adapted from drawings by Dita Vizoso. **B** Examples of sperm from each bristle state category, from top to bottom: dark green -
1122 present, shown are *M. lignano* (1), *M. spirale* (2), and *M. cliftonense* (3); light green - reduced, shown are *M.* sp. 50 (4), *M.* sp. 101 (5), and *M. finnlandense* (6), and yellow - absent, shown are *M.* sp. 64 (7), *M. hystrix* (8), *M.* sp. 80 (9), and *M. pusillum* (10). **C** Phylogeny of the genus
1125 *Macrostomum* based on *de novo* transcriptomes (referred to as H-IQ-TREE in Brand et al,. *unpublished data a*), pruned to contain only the 97 species used in this study, as well as three outgroup species of the sister genus *Psammomacrostomum*. The scale bar depicts the number of
1128 substitutions per site. Tips are coloured according to the character state of the sperm bristles (see **B**). Coloured bars along branches indicate the phylogenetically independent losses (yellow) and reductions (light green) inferred from ancestral state reconstructions (from Brand et al. *unpublished*
1131 *data b*). The outgroup species (grey labels) do not have sperm bristles. Red points indicate nodes with < 90% ultrafast bootstrap support (from Brand et al. *unpublished data a*). *M. lignano,* an established model species within this group, is underlined in red.

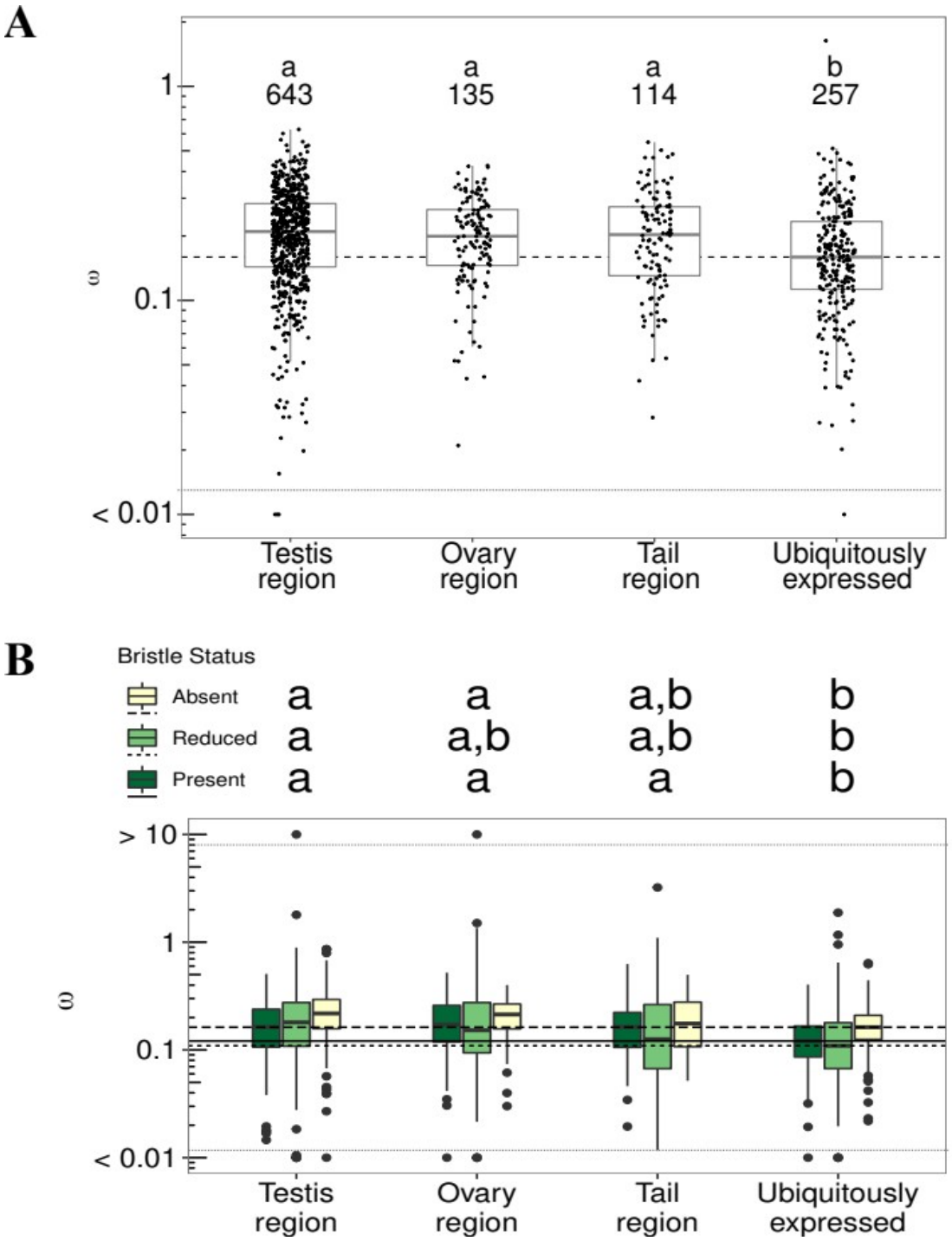


Figure 2. A *dN/dS* (ω) values, from a model assuming a single ω value for the whole tree, for orthogroups (OGs) assigned to different (functional) annotation groups, as well as a set of OGs annotated as ubiquitously expressed (out of 1,000 such ubiquitously-expressed OGs initially selected at random). The horizontal dashed line gives the median ω value for the ubiquitously expressed OGs. Sample sizes for each annotation group are given as inset text. **B** *dN/dS* (ω) values across OGs with in different annotation groups for species with different sperm morphologies (bristle state). The solid, short-dashed, and long-dashed horizontal lines give the median ω values across ubiquitously-expressed OGs for species with different sperm morphologies (see inset

legend). For both **A** and **B** the letters show nonparametric all-pairs post-hoc tests (groups with different letters are significantly different at an adjusted p-value < 0.05). Note that for ease of plotting, some OGs with ω estimates > 10 or < 0.01 (N = 15) have been fixed to ω values of 10 or 0.01, respectively, and separated from the remaining points with stippled horizontal lines.

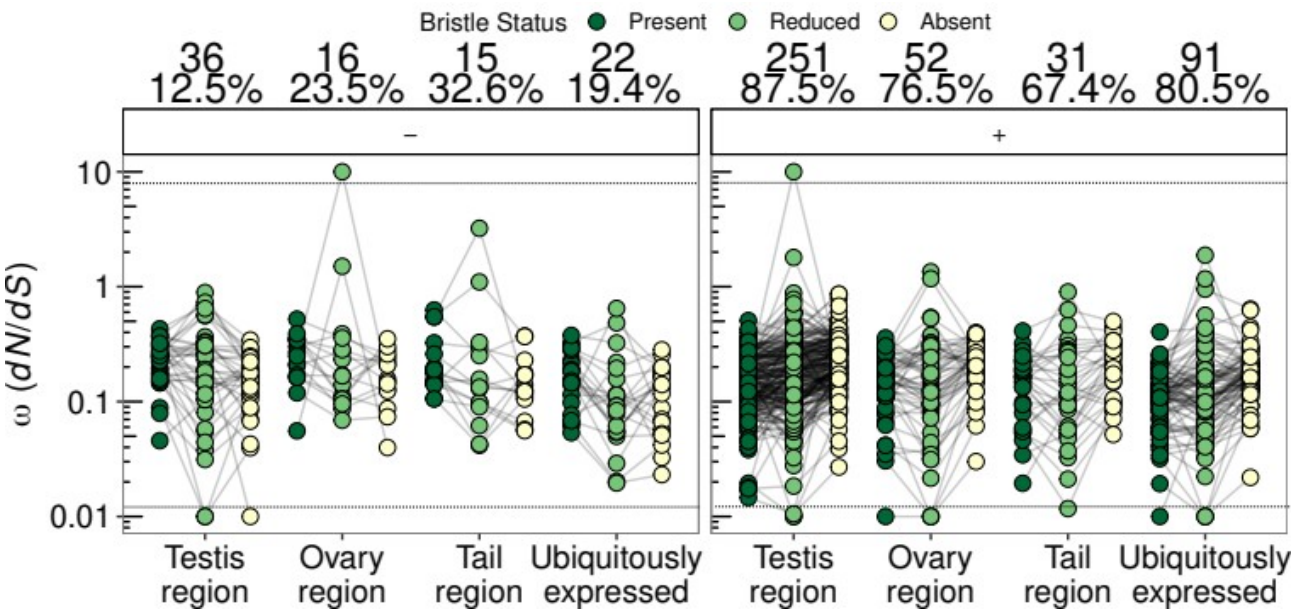
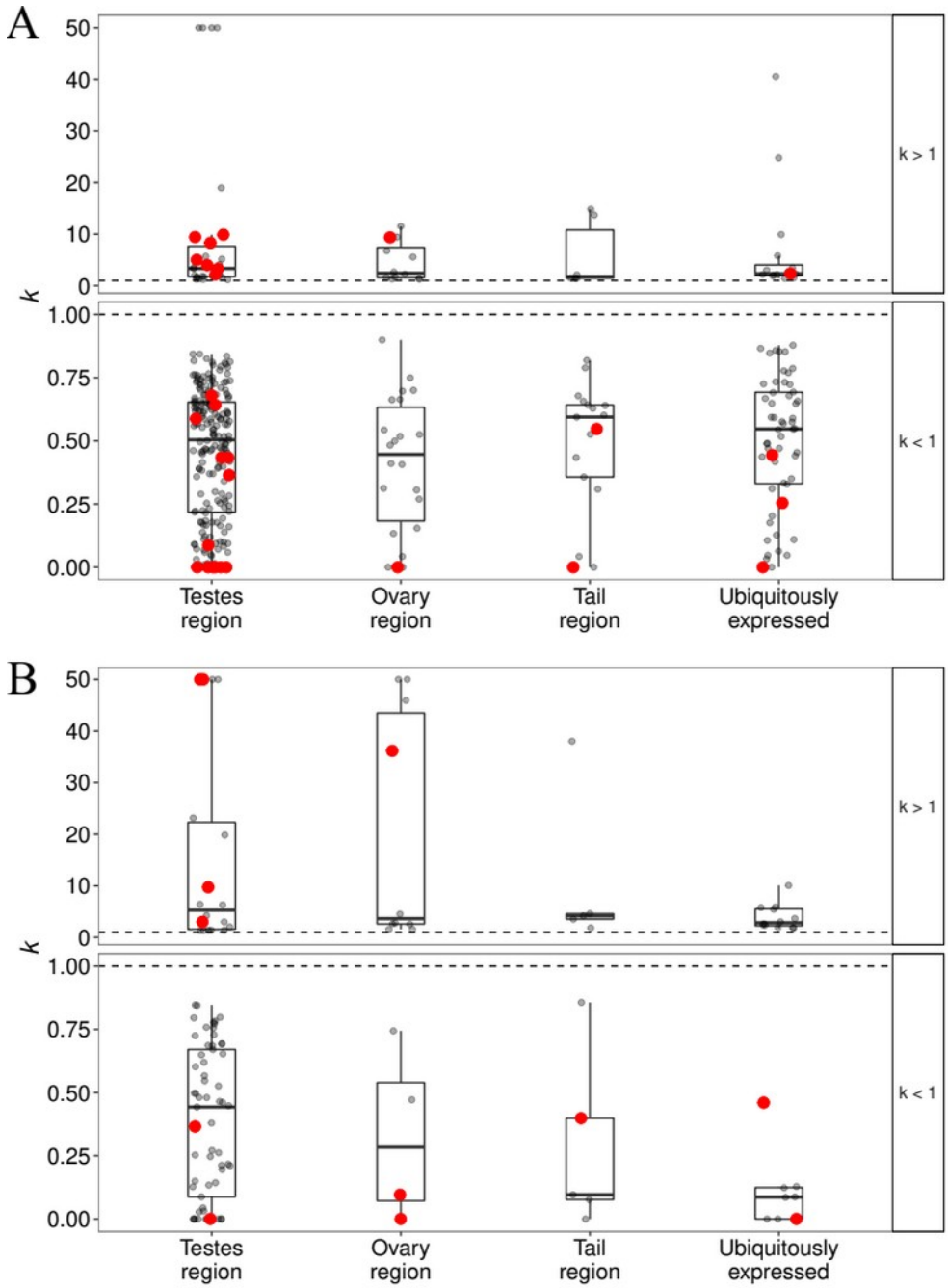


Figure 3. *dN/dS* (ω) values for orthogroups (OGs) that were better explained by a model assuming different ω values for species with the three contrasting sperm morphologies (bristle states). Values for the same OG in species with different bristle states are connected by a line. Panels show the data split into different patterns, namely with ω decreasing (–, left panel) or increasing (+, right panel) as one moves from species with bristles to those lacking bristles. The text above each panel gives the number and percentage of OGs across annotation groups and patterns of ω. Note that for ease of plotting, some OGs with ω estimates > 10 or < 0.01 have been fixed to ω values of 10 or 0.01, respectively, and separated from the remaining points with stippled horizontal lines.



1155 **Figure 4.** *k* values for orthogroups (OGs) in different (functional) annotation groups from the contrast of species with bristles present and species with absent bristles (**A**), and the contrast of species with bristles present and species with reduced bristles (**B**). In both **A** and **B**, plotted values 1158 are restricted to those that show a significant difference in the distribution of ω values between test and reference branches. Additionally, data are split by whether the test showed evidence for intensified ($k > 1$) and relaxed ($k < 1$) selection, respectively. The horizontal black dashed line gives 1161 as a reference the $k = 1$ line, where the distribution of ω values would be considered the same between test and reference branches. Red points highlight OGs that also showed evidence for positive selection among species with absent bristles in a branch-site models.

1164 **Table 1.** Analyses of molecular sequence evolution using branch-site models. Total number of orthogroups (OGs) tested (in square brackets), and the number and percentage (in brackets) of OGs in each annotation group for which a branch-site model—allowing for specific sites to be under 1167 positive selection in foreground compared to background branches—represents a better fit (LRT p-value < 0.05). Results are shown for both contrasts tested, namely species with bristles present *vs.* species with bristles absent (present-absent), and species with bristles present *vs*. species with 1170 reduced bristles (present-reduced).

Annotation group	**present-absent**	**present-reduced**
	Significant (%) [total]	**Significant (%) [total]**
Testis region	38 (7.6) [503]	41 (7.4) [555]
Ovary region	6 (5.2) [116]	7 (5.8) [121]
Tail region	9 (9.5) [95]	11 (12.3) [89]
Ubiquitously expressed	11 (4.7) [231]	14 (7) [201]
Total	64 (6.8) [945]	73 (7.6) [966]

Table 2. Analyses of molecular sequence evolution using RELAX. The number and percentage of orthogroups (OGs) in each annotation group with a significantly different distribution of ω across sites in test branches (either species with bristles absent or reduced) compared to reference branches (species with bristles present). The percentage of OGs in each annotation group is given in brackets. The total number of OGs tested for each annotation group, from which the expected distribution of counts is computed, is given in square brackets. Results are shown for both contrasts tested: species with bristles present *vs.* species with bristles absent (present-absent), and species with bristles present *vs*. species with reduced bristles (present-reduced). The data are split according to whether the *k* parameter is inferred to be > 1 or < 1.

	present-absent			**present-reduced**		
	intensified *k* > 1	**relaxed *k* < 1**		**intensified, *k* > 1**	**relaxed, *k* < 1**	
Annotation group	**Significant (%) [total]**	**Significant (%) [total]**	**Total**	**Significant (%) [total]**	**Significant (%) [total]**	**Total**
Testis region	30 (6) [503]	209 (41.6) [503]	239	18 (3.2) [555]	61 (11) [555]	79
Ovary region	12 (10.3) [116]	22 (19) [116]	34	10 (8.3) [121]	4 (3.3) [121]	14
Tail region	6 (6.3) [95]	17 (17.9) [95]	23	5 (5.6) [89]	5 (5.6) [89]	10
Ubiquitously expressed	16 (7) [231]	55 (23.8) [231]	71	12 (6) [201]	8 (4) [201]	20
Total	64 (6.8) [945]	303 (32.1) [945]	367	45 (4.7) [966]	78 (8.1) [966]	123